Res Intell Manuf Assem, 2023, 2(1): 41-51
DOI: 10.25082/RIMA.2023.01.001

RESEARCH ARTICLE

Detection of abnormal situations in the operation of communication channels

Oleg V. Chikalo[1*] **Ilya A. Obukhov**[1]

[1] Research and production enterprise "Radiotekhnika", Moscow 115419, Russia

Correspondence to: Oleg V.Chikalo, Research and production enterprise "Radiotekhnika", Moscow 115419, Russia; Email: tchikalo@mail.ru

Received: January 1, 2023;
Accepted: March 4, 2023;
Published: March 9, 2023.

Citation: Chikalo OV and Obukhov IA. Detection of abnormal situations in the operation of communication channels. *Res Intell Manuf Assem*, 2023, **2**(1): 41-51. https://doi.org/10.25082/RIMA.2023.01.001

Abstract: Currently, many countries have high expectations for the digitalization of economies, meaning various elements of automation. One of the most effective tools in achieving a new level of digitalization can be the Internet of Things (IoT). The development of IoT provokes the fourth industrial revolution (Industry 4.0), which will be marked by the transition to fully automated digital production, the use of cyber-physical systems and cloud computing. Processes will be controlled by "smart" devices online. An example of such smart devices is modern telecommunications equipment, the operation of which accumulates large amounts of data – telemetry of various kinds. This "big data" can be used to predict possible future failures and other faults (abnormal situations) in the equipment itself. This article is devoted to the issue of creating models of normal behavior of various characteristics of communication channels, which is central in creating predictive diagnostics systems. Examples of such models are given.

Keywords: creating models, IP Quality Monitor (IQM), Model of Normal Behavior (MNB)

1 Introduction

In the course of communication channel operation, a large amount of statistical information (telemetry) about its operation is usually recorded. Each captured characteristic is a time series, *e.g.*, a sequence of measurements (counts) every n minutes. One of the problems of "big data" accumulated during the operation of modern communication equipment is the problem of rational use of this data.

One of the ideas of such rational use is that on the basis of these data it is possible to assess whether the channel worked normally, or in its work there were deviations from normal operation - anomalies. "Normality" or the benchmark of normal behavior must first be expressed in the form of a model, which is then used to assess the degree of normality.

The term "model" is used conventionally, usually a model is simply a set of points in a multidimensional space. One point is the value of a characteristic, for example, for a day. Thus, the model for a month will consist of a maximum of 31 points. Some number of characteristics is collected for those days when the channel worked normally. This is preliminarily evaluated by the experts.

Then we "build a model of normal behavior," *i.e.* outliers are searched for and discarded among the data obtained. The remaining points form the desired model. The key question is to determine the space in which to build the model. This space can be the original values of the time series or some other features. Usually the criterion is the density of the cluster of points that make up the model. The more dense the cluster, the better the model, the easier it is able to detect outliers.

During operation, the next portion of telemetry is compared to the points of the normal behavior model. The new portion (point) may be close to or among the existing points, which will indicate its normality, or at some distance from other points. In this case it may turn out to be an outlier. There are a large number of algorithms for determining outliers in multidimensional space [1]. It is even possible to estimate the degree of "outlierness" or normality of a new point.

Channel performance is a multivariate time series if recorded synchronously and characterizes different sides of channel performance. Therefore, the normality of the whole such time series should also be evaluated on the basis of its model of normal behavior, besides the evaluation of its individual components. Individual characteristics may be in certain relationships (linear and nonlinear) with each other, *e.g.*, be correlated. This fact should also be used to build appropriate models reflecting the normal joint behavior of these characteristics.

When a new portion of the counted (measured) characteristics is obtained, normal behavior models can be used to assess whether the entire set of characteristics is normal or abnormal. Each controlled characteristic and the behavior of controlled pairs or groups of characteristics is also evaluated. The assessment can be given in three ways: normal, weak outlier, strong outlier. The article uses the LDA_med method [2] to detect outliers, but others can be used as well. There can be a situation when each characteristic behaves normally, but the behavior of a pair (group) or the whole totality can be abnormal.

The deviation from normal indicates a possible abnormal situation that took place during the time corresponding to the received portion of telemetry.

2 Methods and materials

2.1 Materials

As an example, we will consider standard channel characteristics collected with IP Quality Monitor (IQM) hardware and software package [3]. For example, SDBytes – number of bytes transmitted from Source to Destination, DSBytes – number of bytes transmitted from Destination to Source, SDLostPercent - percentage of lost packets during transmission from Source to Destination, DSLostPercent – the same, but in reverse direction, *etc*. A complete list of measurable characteristics is given in Table 1.

Table 1 List of measurable characteristics

SDLost SDLostPercent DSLost DSLostPercent	Packet loss (for URL - sessions) in both directions in absolute numbers and as a percentage of the total.
SDBW SDBWPercent DSBW DSBWPercent	The resulting bandwidth of the network in both directions in kilobits per second and as a percentage of the expected bandwidth. The expected throughput is the one transmitted on the command line.
SDLossBW SDLossBWPercent DSLossBW DSLossBWPercent	"Lost" throughput in both directions. Represents the difference between the expected throughput and the throughput received as a result of testing.
SDRemarked SDRemarkedPercent DSRemarked DSRemarkedPercent	The number of packets delivered with a change of service class, in absolute numbers and as a percentage of the total received.
SDOOS SDOOSPercent DSOOS DSOOSPercent	The number of packets delivered with reordering, in absolute numbers and as a percentage of the total received.
MinRTT AvgRTT RMSRTT MaxRTT	Circumferential delay in packet delivery from initiator to initiator through the conjugate (minimum, average, quadratic, maximum)
SDMinDelay SDAvgDelay SDRMSDelay SDMaxDelay DSMinDelay DSAvgDelay DSRMSDelay DSMaxDelay	One-way packet delivery delay (minimum, average, quadratic, maximum)
SDJitter DSJitter	Packet delivery delay jitter in both directions, calculated according to RFC 3550
SDMinIPDV SDAvgIPDV SDRMSIPDV SDMaxIPDV DSMinIPDV DSAvgIPDV DSRMSIPDV DSMaxIPDV	Packet delivery delay variation, calculated by Y.1540 (basis - minimum delay) for test session time (minimum, average, quadratic, maximum)
SDMinMAPDV2 SDAvgMAPDV2 SDRMSMAPDV2 SDMaxMAPDV2 DSMinMAPDV2 DSAvgMAPDV2 DSRMSMAPDV2 DSMaxMAPDV2	Packet delivery delay jitter, calculated according to G.1020 during the test session (minimum, average, quadratic, maximum)
SDBytes DSBytes	Number of bytes transferred in the test session in both directions

Suppose we have data for each day of channel operation, broken down by five minutes. That is, each time series has 288 values (12 values for each hour, a total of 24 hours).

2.2 Models of normal behavior for individual characteristics

There can be many models of normal behavior of an individual characteristic, so we must be able to choose the best one. As a criterion, it is proposed to choose the number of outliers detected with its help. The model that is sensitive to the presence of outliers will be of higher quality. The simplest model is to use as coordinates in space the values of time series characteristics, we get 31 points (counting 31 days in a month) in 288-dimensional space.

For clarity, we will display the multidimensional space on the plane with the help of Sammon's projection [4].

When working with time series, a well-known technique is to use its characteristic values - features - instead of the original values of the series. There is also a great number of them - minimum and maximum values, average, dispersion, etc., etc. A very large list of such features is given, for example, in work [5]. Any combination of these features can also be used as a model of normal behavior (MNB).

In the course of our experiments it turned out that the desired quality of outlier detection is demonstrated by a model of mutual similarity, which is formed by points in 4-dimensional space (coordinates - three correlation coefficients - Kendall, Pearson, Spearman, and Euclidean distance between points). The point is the time series for the day.

A matrix of similarity M of points to each other is determined. For this purpose m_{ij} - correlation coefficients between points are calculated. Then we take the median by rows. The first three

coordinates differ only in the way the correlation is calculated. Then we calculate the fourth coordinate - determine the Euclidean distance between the source points, and then take the median by rows.

2.3 Model based on numerical pattern

Each characteristic of communication channel takes values from a certain numerical range.

The numerical pattern of the characteristic can be represented by a histogram. The number of bins of the histogram is the number of coordinates. The height of each bin (column of the histogram) is the value of the coordinate. Such model may be useful, when there is a gradual degradation of equipment, and you need to detect it in time.

2.4 Model based on strings patterns

One of the ways of working with time series is to transform them into

character strings. One possible transformation technique is presented in the table taken from [6] (Table 2).

Table 2 Fornell-Lecker criterion

Symbol	Meaning	Definition
a	Highly increasing transition	$\frac{d}{dt} > 5$
b	Slightly increasing transition	$5 \geq \frac{d}{dt} > 2$
c	Stable transition	$2 \geq \frac{d}{dt} \geq -2$
d	Slightly decreasing transition	$-2 > \frac{d}{dt} \geq -5$
e	Highly decreasing transition	$\frac{d}{dt} < -5$

If the difference between adjacent row values falls within the appropriate range, the numeric value is replaced by a symbol. This representation of the time series provides additional possibilities compared to a simple numerical pattern. You can count, for example, the number of different pairs of symbols or triples, fours, etc., to move to the MNB. Each combination of symbols can serve as a coordinate, and the number of occurrence of a particular combination can serve as a coordinate value.

2.5 Models of joint normal behavior for pairs of characteristics: Based on correlation

Correlation is the best-known form of linear relationship between characteristics. Correlation between two characteristics measures the similarity in form between those characteristics. Among channel characteristics, we can find strongly correlated and uncorrelated (weakly correlated).

2.6 Model for strongly correlated characteristics

To obtain an MNB, it is sufficient to measure the correlation between the points, as is done in the mutual similarity model. MNB consists of M values of the selected correlation coefficient (medians of matrix rows).

2.7 Models for uncorrelated characteristics: Two-cluster model

Initial feature values can form explicit clusters, such as, in the case of DSOOSPercent (blue) and DSRMSMAPDV2 (red). (Figure 1)

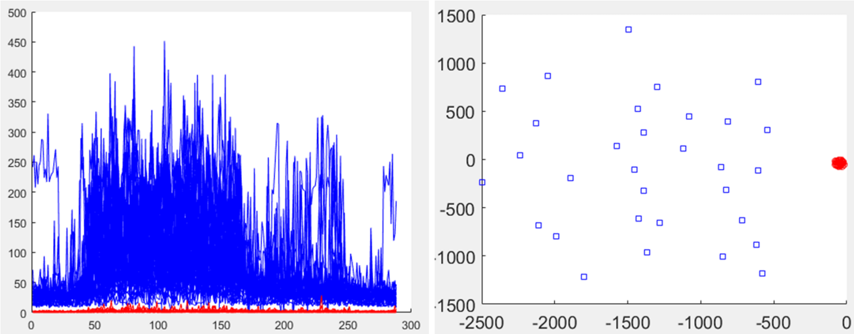

Figure 1 DSOOSPercent (blue) and DSRMSMAPDV2 (red) characteristics (left) and their mapping to plane (Sammon projection)

The characteristics have 288 values, and 288-dimensional space is used. On the left in Figure 1 are the time series of these characteristics, and on the right is a mapping of the 288-dimensional space to the plane. Clear clusters can be seen.

The presence of such clusters indicates that there are complex nonlinear relations between the values of the two characteristics. These relations should be preserved during the channel operation.

To verify the preservation of these relations, let us construct the MNB. For this purpose each pair of points (one from the first characteristic for day i, the second one from another characteristic for the same day) let us represent by the following four coordinates:

distance to the medoid of one's cluster, distance to the medoid of another's cluster - for the first point of the pair;

distance to own cluster's medoid, distance to foreign cluster's medoid - for the second point of the pair.

This is what the model looks like for the pair DSOOSPercent and DSRMSMAPDV2 in 4-dimensional space. (Figure 2)

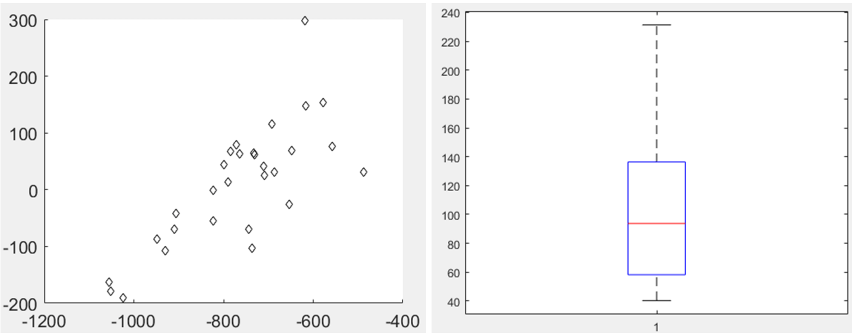

Figure 2 Resulting MNB and boxplot

Naturally, the points are checked for outliers. If there are no outliers, we save the obtained points as MPBs.

2.8 Models for uncorrelated features without clusters: Based on distances between pairs

The initial values of not all pairs of characteristics form clusters.

For example, here is what the picture looks like for DSLossBW (red) and DSMaxMAPDV2 (blue) - the original time series values are shown (left) and the 288-dimensional space mapping to the plane on the right. (Figure 3)

Characteristics are not correlated (the values of three correlation coefficients for this pair, 5 points - days, are given below for example):

-0.107	0.038	0.056
0.221	0.057	0.089
0.236	0.040	0.057
0.221	-0.013	-0.015
0.234	0.094	0.140

Let's construct the MNB: represent each pair by a point in the 3-dimensional space of correlation coefficients between pairs. (Figure 4)

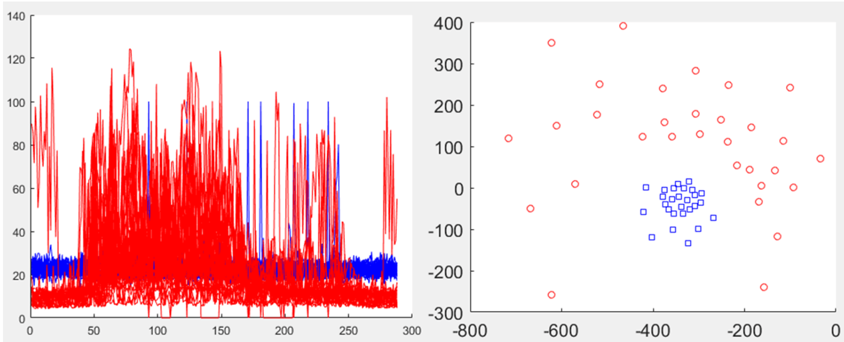

Figure 3 DSLossBW (red) and DSMaxMAPDV2 (blue) - the original time series values are shown (left) and the 288-dimensional space mapping to the plane on the right. There are no linearly separable clusters.

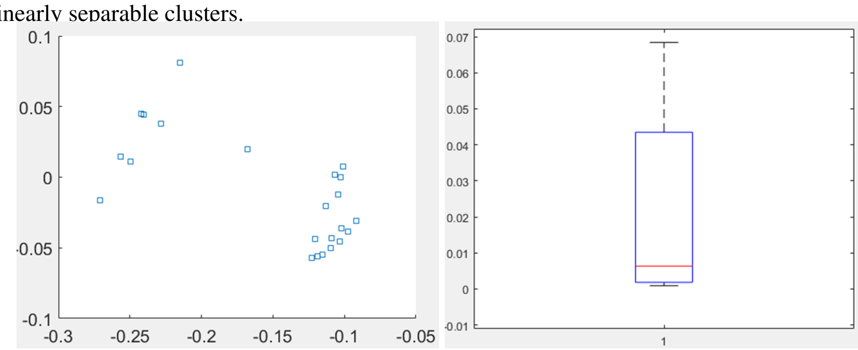

Figure 4 Resulting MNB(left) and boxplot (right)

2.9 Models of normal behavior of multivariate time series of characteristics

Theoretically, it is possible that all characteristics will show normal behavior, and their aggregate - multivariate time series - abnormal behavior.

In order to detect such a situation, we need special MNBs for the whole multivariate series. Such a solution is possible.

A model is constructed based on models of the internal similarity of individual characteristics.

It is suggested to use the space:

coordinate1 - outlarity of characteristic1;

coordinate2 - outlarity of characteristic2;

...

coordinateN - outlarity of characteristicN.

3 Experimental results

3.1 Models of normal behavior for individual characteristics

Let us consider the derivation of the MPB for the DSMinMAPDV2 characteristic. Figure 5 shows the initial values on the graph. (Figure 5)

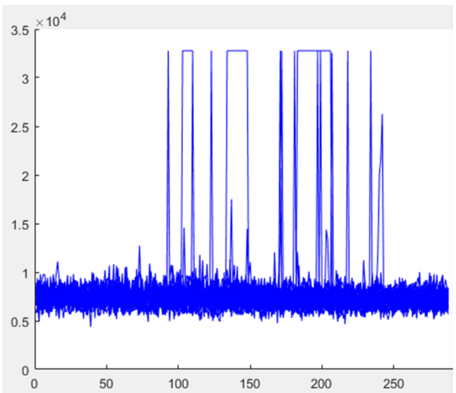

Figure 5 DSMinMAPDV2 characteristic

We can assume that all visual ejections will be outliers.

To map the multidimensional space to the plane, we use the Sammon projection. To detect outliers we use the LDA_med method.

The outliers (blue squares) are removed after detection. Then the process is repeated for the remaining points. The process ends when the outliers are no longer detected. (Figure 6)

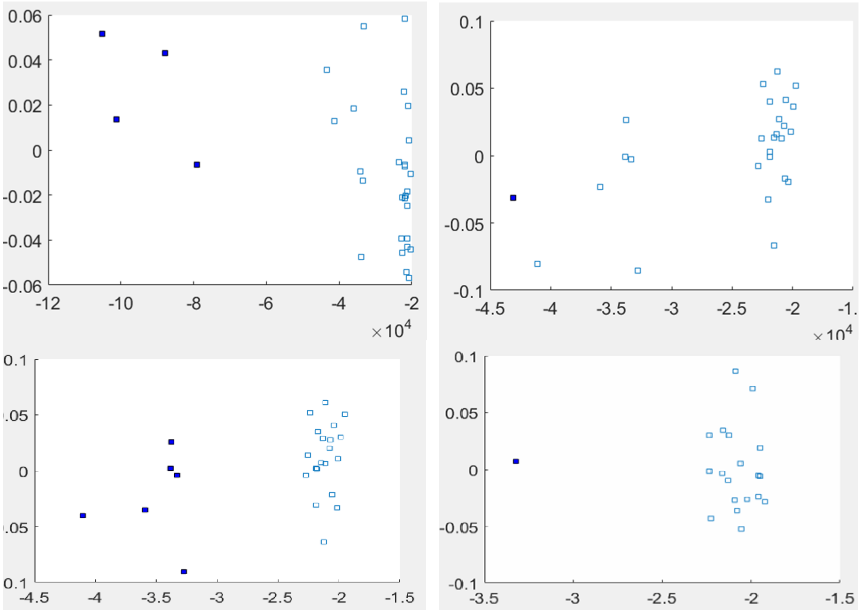

Figure 6 Outlier detection and removal

The pictures show how the outliers are searched for. First, four outliers are found (marked as blue squares), after removing them one new one is found, after removing it six new ones at once, and finally one outlier is found last.

The remaining points represent the model of normal behavior in 4-dimensional space. (Figure 7)

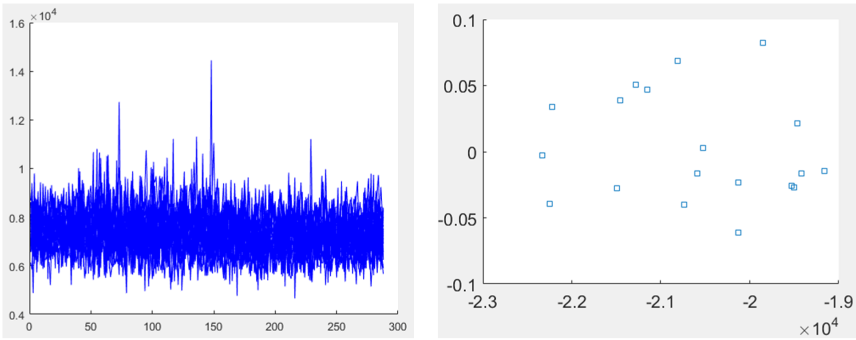

Figure 7 DSMinMAPDV2 characteristic after removing outliers (left) and resulting MNB (right). A model based on a numerical pattern.

We will take a histogram of one hundred bins as a numerical pattern. (Figure 8)

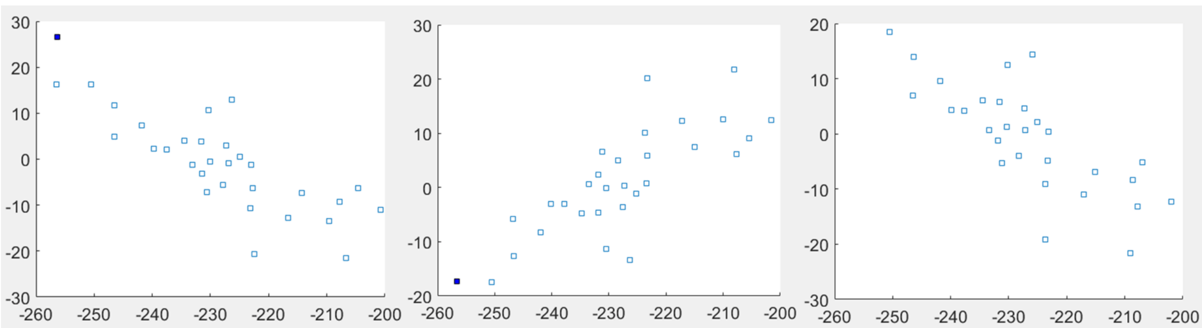

Figure 8 Outlier detection and removal

Each point has one hundred coordinates, the value of each is the height of the corresponding bin. When constructing the MNB we first remove one outlier, then the second, the remaining points make a model of normal behavior in the space of 100 bins.

If we take the artificially obtained characteristic of 288 ones, we get the result (red square) using this model. (Figure 9)

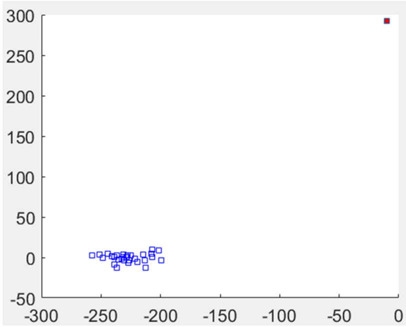

Figure 9 The entered point is a strong outlier and is very far from the model points

3.2 A model based on string patterns

As an example, let's look at the DSAvgMAPDV2 characteristic for one day. The time series turns into a string pattern:

cdccccbdccbdcccbdccccbdcccaecccbcbcecccccccccdaebdcbdcccccbbecc ccdcccbcdccccae-bcccadbecbdcbdbddcbdbdcccccaebbeccccccccccccbcdc cccccccccccbaecccccccccbaecdbcdccc-ccccbcdbbbecccccccccccccccccccd ccccccccbdcccccccccccccaeccccccbcccccccccdccbdccc-ccccccccccbd cccccccccccccccbdcccccccccccdccccbb.

If we count, for example, all combinations of substrings of two characters and of three characters in this string, we can represent the time series by a point in 150-dimensional space. (150 is the number of different twos and threes of the alphabetic characters). The coordinate is the combination of symbols, the value of the coordinate is the number of this combination in the row of the time series.

In this space we can also construct a model of normal behavior. (Figure 10)

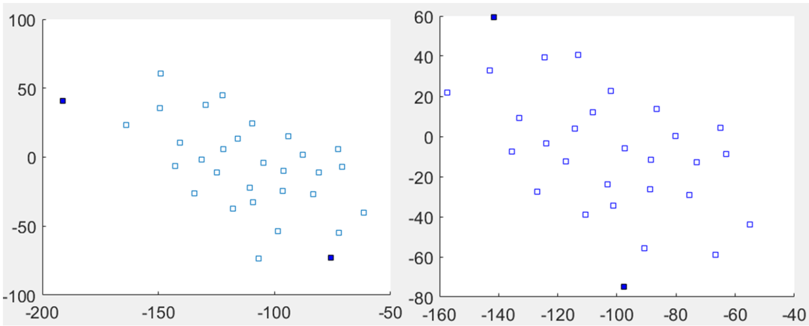

Figure 10 Outlier detection and removal

Two outliers are consecutively removed first, and then two more.

The result is a MNB. (Figure 11)

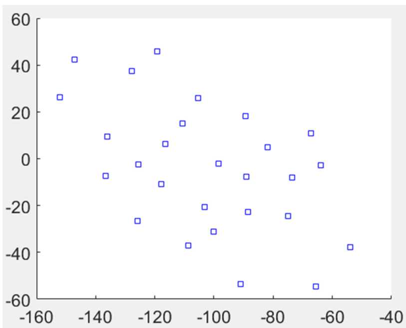

Figure 11 Resulting MNB

To check the strength of the model, let us estimate the place of the artificial point:

Let's create an artificial time series: 24 units, 26 threes, 25 threes, 25 fours, 49 units, 51 fives, the rest are units. This time series corresponds to a line:

ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccbccccccccc ccccccccccccccccdccccccccccccccccccccccccccccccccccccccccccccbccccccccccccccccccc cccccccccccccccccccccccccccccdcccccccccccccccccccccccccccccccccccccccccccccccccccc cccccccccccccccccccccccccccccccc.

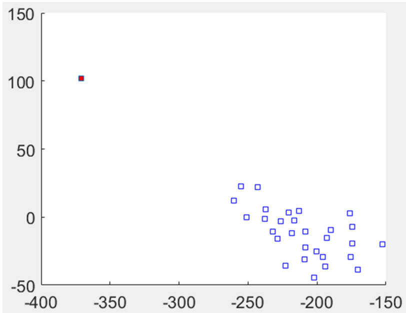

Figure 12 The new point (red square) is a strong outlier. The new point (red square) is a strong outlier and is very far from the model points.

3.3 Models of joint normal behavior for pairs of characteristics: For strongly correlated characteristics

Consider the DSBW and DSBWPercent characteristics. (Figure 13)

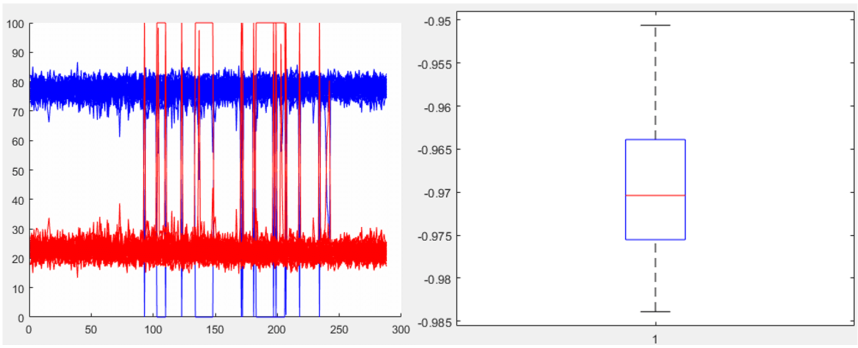

Figure 13 DSBW and DSBWPercent characteristics (left) and boxplot for Kendall's coefficient (right)

The values of the correlation coefficients of these characteristics are close to 1 (Kendall, Pearson, Spearman): -0,975641458348255 -0,999783256671490 –0,997684955502735.

As a model of normal behavior, we store the values of any of the 3 coefficients for each pair - a total of M values (it is preferable to use Kendall or Spearman coefficients because of their nonparametricity). Fig. 13 - the boxplot for Kendall's coefficient is shown on the right.

When new values of a pair of characteristics are obtained, we calculate a new value of the correlation coefficient and compare it to the model values.

3.4 Models for uncorrelated characteristics: Two-Cluster model

We have previously constructed such a MNP for the DSOOSPercent and DSRMSMAPDV2 pair. An artificial point, which will be depicted in the original data in Figure 14.

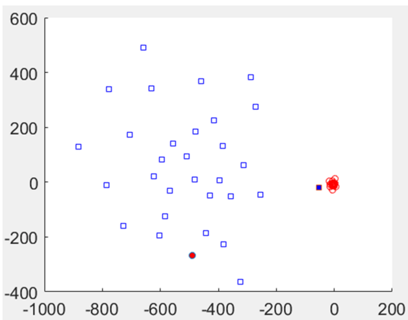

Figure 14 The new pair is a blue square and a red circle

Let's apply the model. (Figure 15)

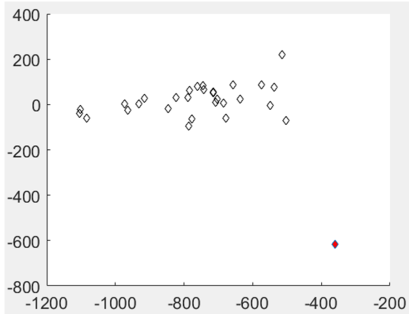

Figure 15 The new pair, the red diamond, is a strong outlier.

3.5 Models for uncorrelated characteristics without clusters: Based on distances between pairs

Previously, we built the MPP for DSLossBW and DSMaxMAPDV2.

Now we will test the model by introducing an artificial point (correlated pair values). (Figure 16)

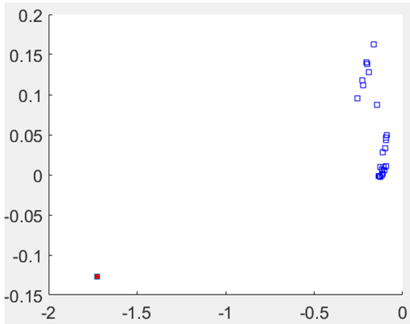

Figure 16 The new point is a strong outlier

4 Models of normal behavior of multivariate time series of characteristics

Consider an example for 5 characteristics:
DSMinIPDV, DSMinMAPDV2, DSOOSPercent, DSRemarkedPercent, DSRMSMAPDV2
The removal of outliers is done gradually. (Figure 17)

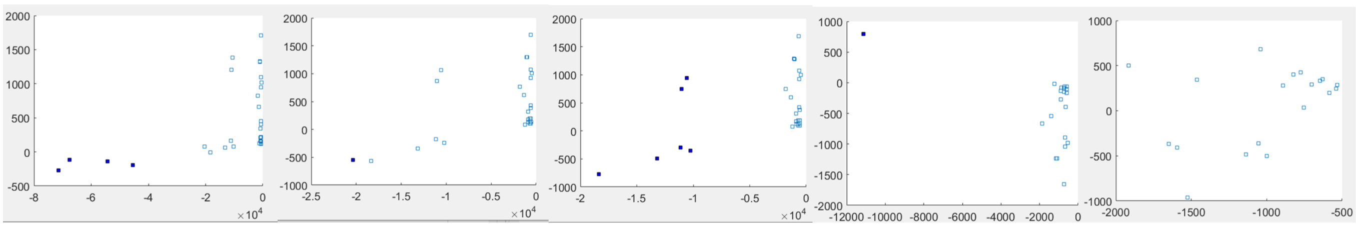

Figure 17 An example of outlier detection and removal for a multivariate time series

Total 12 points were deleted (data for 12 days). The remaining points in the selected space form a model of normal behavior of multivariate time series of characteristics.

5 Discussion

Above we considered examples of building normal behavior models for assessing the quality of communication channel performance. Of course, their list is far from being complete, but the presented MNBs can serve as a basis for building an automated system of predictive diagnostics. Let's imagine that characteristics are processed in portions, *i.e.* not in real time, but after accumulation of a certain amount of data during the set time period, in the considered example this period was equal to days.

After that the forecast of the controlled equipment state for the next telemetry collection period is made. If the conclusion about danger of further operation is made, the decision to carry out repair or at least inspection is made.

It is envisaged to get a summary assessment of anomalousness of the next portion of telemetry for an instance of the device. Figure 18)

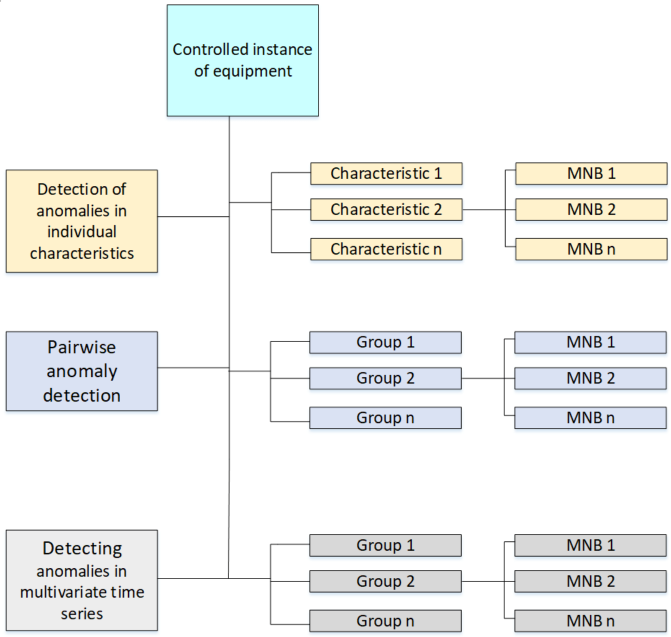

Figure 18 Anomaly detection procedure for data batch processing

Each characteristic is evaluated with the help of several MNBs. If the MNB shows that the characteristic is normal, this corresponds to, for example, 0 points, a weak outlier - 1 point, a strong one - 2 points. The sum of all the MNBs, taking into account the weighting coefficients of the models, gives an estimate of the normality of the characteristic, ideally 0 points.

Then you can get in a similar way the total score of all characteristics, groups of characteristics and the entire multivariate series as a whole.

The resulting estimates are stored and can be displayed graphically to visualize the dynamics of their change. The decision about the real danger of the current situation is made by a person. To help in decision-making, numerical thresholds of anomalousness for each device are established by experience.

The total score of abnormality of a device can serve as an index of the health of the system.

The disadvantage of this predictive diagnostics system is that it only detects anomalies, not specific faults (abnormal situations). But it can warn you about equipment problems long before the equipment fails [7].

The description of the detected anomalies (failures) in the operation of the devices, together with the formal description of the anomalies of specific characteristics, can be stored in the knowledge base. In the future, they are used both in the manual search for analogues and automatically in the detection of anomalies that match the previously saved descriptions.

6 Conclusion

One of the problems of "big data" accumulated during the operation of modern communication equipment is the problem of rational use of this data.

The modern trend is to introduce predictive maintenance into the enterprise practice, which provides timely detection of anomalous behavior of equipment. The most preferred approach to detecting anomalies in equipment behavior is a data-driven approach. In this approach, models are built for anomaly detection and fault diagnosis purposes on the basis of the vast amount of different types of telemetry data available. The advantage of such models is their independence from the knowledge of subject matter experts.

The construction of such models of normal behavior for characteristics of communication channels is considered in this paper. Examples, confirming the performance of the proposed models, are given.

The proposed models and approach to building a predictive diagnostics system on their basis can be useful for developers of such systems.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Aggarwal CC. Outlier Analysis (2nd. ed.). Springer Publishing Company, Incorporated, 2016.

[2] Angiulli F and Pizzuti, C Outlier mining in large high-dimensional data sets. IEEE Transactions on Knowledge and Data Engineering, 2005, **17**(2): 203-215.
https://doi.org/10.1109/TKDE.2005.31

[3] IP Quality Monitor (IQM), viewed February 27,2023,
http://www.net-probe.ru

[4] Sammon JW. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers, 1969, **18**(5): 401-409.
https://doi.org/10.1109/T-C.1969.222678

[5] Fulcher BD. Feature-based time-series analysis. arXiv preprint arXiv:1709.08055, 2017.

[6] Daw CS, Finney CEA and Tracy ER. A review of symbolic analysis of experimental data. Review of Scientific instruments, 2003, **74**(2): 915-930.
https://doi.org/10.1063/1.1531823

[7] Chikalo OV, Alexandrov VR and Obukhov IA. System of early detection of problematic telecommunication equipment. Microwave Engineering and Telecommunication Technologies, 2020, **1-2**: 128-129.