Evaluation of quality level and selection of relevant parameters' values in manufacturing, business and education

Ahmet Demir
Ishik University
Erbil, Iraq
ahmet.demir@ishik.edu.iq

Irakli Rodonaia
International Black Sea University
Tbilisi, Georgia,
irodonaia@yahoo.com

Irina Milnikova,
Head Teacher of Manhattan Montessori school
New-York, USA
ira0226@yahoo.com

Abstract- The requirements for production and learning process quality are different in various manufacturing, business and educational organizations. A new approach to fit these requirements and evaluate the closeness of realistic (actual) quality of production or learning processes (based on quality indicators of output or scores of examination tests) is proposed in the paper. The technique uses the strictly defined approximation procedures and allows users automatically evaluate of closeness of actual quality level when changing quality requirements. In case of significant difference between actual and pattern distributions a new approach (using neural network of 'Generalized Regression Neural Network' type) of determining the minimum values of the factors that will bring the actual distribution to the pattern one is proposed in the paper.

Keywords: manufacturing processes quality, business processes quality, learning process quality, percentile function, non-parametric approximation, generalized lambda distribution, Generalized Regression Neural Networks (GRNN).

I. Introduction

Let's assume that we are given the next requirement for the learning process quality: "weak" (failed) students can be thought those ones whose grades are less than 60 and the percentage of them should be 30%; "ordinary" (of acceptable level) students are those ones whose grades are between 61 and 95 grades, the percentage of them should be 65%; in latter range so called "middle" level students are those whose percentage is no more than 50% of total number of students (including failed ones) and 20% of "ordinary" student; say, the grade of these "middle" students turns out to be 80 (or any other value), so the grade 80 can be considered as a median of grades distribution; "excellent" students are those ones whose grades are above 95% and the percentage of them is 5%. The corresponding cumulative distribution function (CDF) is shown in Figure 1. Obviously, the pattern distribution cannot be approximated by the normal distribution.

However, in all known for us papers such distributions were approximated by either the normal distribution or by some another well-known distributions (beta distribution, gamma distribution, Weibull distribution, etc.) ([1]-[3]. But in case of applying normal distribution the adequacy and precision of results strongly depends on the degree of "skewness" and often may not be acceptable. In case of applying other distributions (beta distribution, gamma distribution, Weibull distribution, etc.) the problem of estimating adequate distribution parameters arises. In many cases analytical expression cannot be obtained in close form. Besides, when requirements for quality changes, the corresponding shapes of PDF and CDF functions also change. As a result, it is necessary to use frequently complicated procedures of distribution parameters estimation.

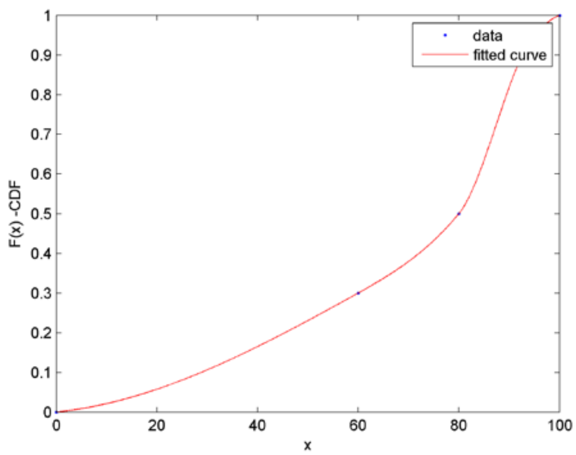

Figure 1. CDF of pattern distribution

The similar task is commonly met in the area of product quality control ([4]. Suppose that the quality requirement to the product quality is as follows. The percentage of deviation from required level of some quality parameter must be no more than ±5% in 95 % of the output of the product; in this case the quality of the product is regarded as "excellent". To be regarded as "acceptable" the product quality must be as follows: deviation from required level of the quality parameter is ±6%-20% in 3% of the output of the product. The product quality is regarded as "unacceptable" (or defective) if there is the deviation of more than 20% (so the percentage of defective production must be no more than 2%).The CDF is shown in Figure 2. As one can see, the distribution is also skewed and the problem of choosing the right type of distribution occurs here.

It is not clear in advance which type of distribution should be used in this case [5]. The above distributions (reflecting quality requirements) are called hereinafter "pattern" distributions (functions). It is desirable that distribution of grades of actual exams would be as close to the pattern distribution as possible. The question of closeness degree is a problem (and is considered further in the paper.

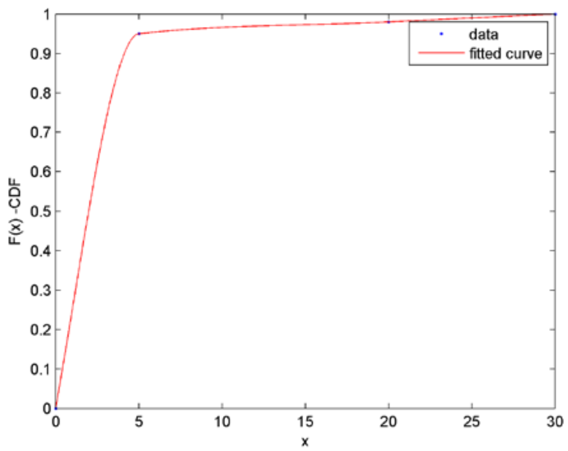

Figure 2. CDF of pattern distribution in product quality control

Moreover, the pattern distribution presents quality requirement for total learning process (which must take into account results of all relevant tests).

That is, grades of many subjects (obtained by a group of students in tests held during one of more courses) must match the pattern distribution in order that the group would be regarded as successful and meeting the requirements of learning quality. Of course, it is possible to compare grades of each actual test with the pattern distribution and then summarize the results. But this approach is associated with a large amount of additional and repeated calculations.

Taking into account all the above-mentioned, a new general method of using a unified non-parametric [6] estimation of relevant grades distributions and further application of its results to the evaluation process of learning quality is developed in this paper. It is important to point out that the method does not require the execution of rather complicated procedures of estimating distribution parameters (mean, standard deviation, third and fourth moments)). The method can be applied to fit grades of various multiple tests and compare them with pattern distribution by using the same unified techniques and algorithms. The approach provides forming of overall quality criterion for all test scores and method of comparing it with pattern quality requirement.

II. General Part

To provide fitting the wide variety of distribution shapes and to describe data by using a single functional form the approach used in the paper implements the Generalized Lambda Distribution (GLD)[7].The method specifies four parameter values for each case, instead of giving the basic data (which is what the empirical distribution essentially does) for each case. The one functional form allows us to group cases that are similar, as opposed to being overburdened with a mass of numbers or graphs.

The generalized lambda distribution family with parameters λ_1, λ_2, λ_3, λ_4, GLD (λ_1, λ_2, λ_3, λ_4), is most easily specified in

$$Q(y) = Q(y;\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2} \qquad (1)$$

where $0 \le y \le 1$. The parameters λ_1 and λ_2 are, respectively, location and scale parameters, while λ_3 and λ_4 determine the skewness and kurtosis of the GLD (λ_1, λ_2, λ_3, λ_4). Recall that percentile function (PF) of the stochastic variable X is the function Q(y) which, for each y between 0 and 1, tells us the value of x such that $F(x) = y$: $Q(y)$ = (The value of x such that $F(x) = y$), $0 \le y \le 1$

Here $F(x)$ is the cumulative distribution function (CDF) of the variable X:

$$F(x) = P(X \le x), \ -\infty < x < +\infty.$$

The restrictions on λ_1, λ_2, λ_3, λ_4 that yield a valid GLD(λ_1, λ_2, λ_3, λ_4) distribution and the impact of λ_3 and λ_4 on the shape of the GLD(λ_1, λ_2, λ_3, λ_4) PDF (Probability Density Function) will be considered later.

It is relatively easy to find the probability density function from the percentile function of the GLD ([7]. For the GLD (λ_1, λ_2, λ_3, λ_4), the probability density function is:

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}} \qquad (2)$$

at $x = Q(y)$.

As we have seen above, very often the quality requirement are given in the form of required percentiles (percent of failed, ordinary, middle and excellent students, percent of deviation of some product's quality parameters from their nominal values and so on). The percentile-based approach [7] fits a GLD(λ_1, λ_2, λ_3, λ_4) distribution to a given dataset by specifying four percentile-based sample statistics and equating them to their corresponding GLD (λ_1, λ_2, λ_3, λ_4) statistics. The resulting equations are then solved for λ_1, λ_2, λ_3, λ_4, with the constraint that the resulting GLD be a valid distribution.

The method, described above, requires usage of the complex tables of various values of parameters λ_3 and λ_4 . To automate the fitting process the algorithm P-KS ([8] is used in the paper. The strategy is to find the set of parameters (λ_1, λ_2, λ_3, λ_4) that give the lowest value of the Kolmogorov-Smirnov estimator EKS :

$$E_{ks} = \max\left|\hat{F}_n - F(x)\right| \qquad (3)$$

where $\hat{F}_n$ is the empirical cumulative distribution function (ECDF).

As it was stated above, the pattern distribution is given in the form of some percent. For the example of the section we have the following data (expressed in the form of Matlab statements):

x= [0, 60, 80, 95,100];
y= [0, 0.30, 0.50, 0.95, 1];
In order to form the pattern distribution (with which the actual tests grades should be compared) we need to fit a curve to the given data. The fitted curve will be used to generate data values in intermediate points (other than the original data points) -interpolation points. To provide the smoothness and maximum accuracy of generated data in interpolation points the technique of the shape-preserving cubic splines is used. The plot of the ECDF for pattern distribution looks like (Fig.3). The corresponding PDF function can be obtained similarly and is shown in Figure 4.
As one can see, the shape of the PDF is non-standard and it is difficult to guess which theoretical distribution can successfully fit it.

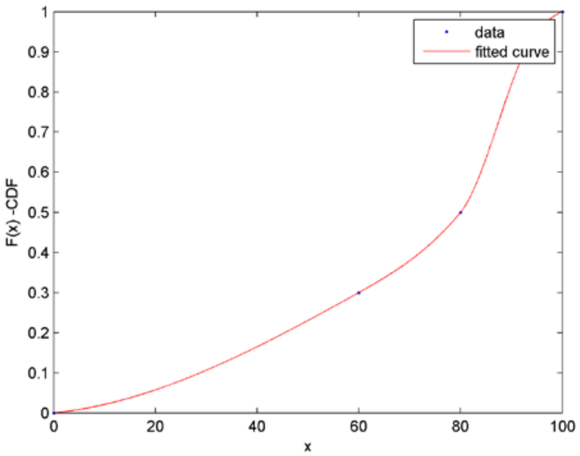

Figure 3. ECDF for pattern distribution

Now we can estimate (using relevant Matlab statements) values of the pattern distribution in interpolation points, that is, we can estimate the values of various percentiles (namely, 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th percentiles) of the pattern distribution to be compared with actual

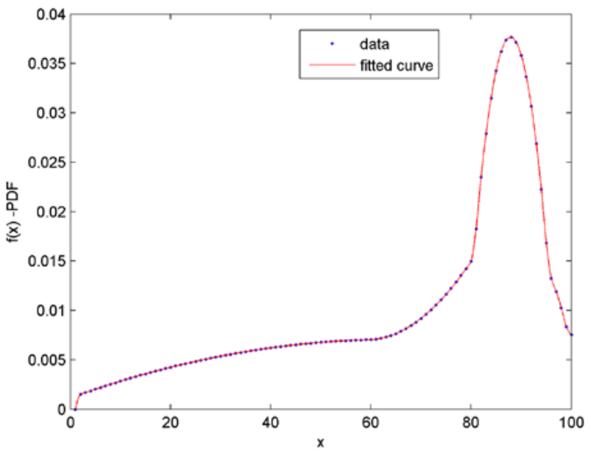

Figure 4. PDF function for pattern distribution

tests grades' percentiles. As we stated above, the GLD Percentile-Based Approach to Fitting Distributions intensively uses operations with percentile functions PF (inverse cumulative distribution functions ICDF). We can compute a nonparametric estimate of the inverse CDF. In fact, the inverse CDF estimate is just the CDF estimate with the axes swapped. Here we again use the Piecewise Cubic Hermite Interpolant Polinomial (PCHIP) to estimate values of ICDF (Fig.5).
Having values of PF we can compute now the values of $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$. Having computed these values, we now run the procedure P-KS. The solution with the best KS criteria for all possible combinations of pairs (λ_3, λ_4) and associated with them pairs of (λ_1, λ_2) is selected. As it was explained above, knowing $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ and using formulas (1) and (2) , we can

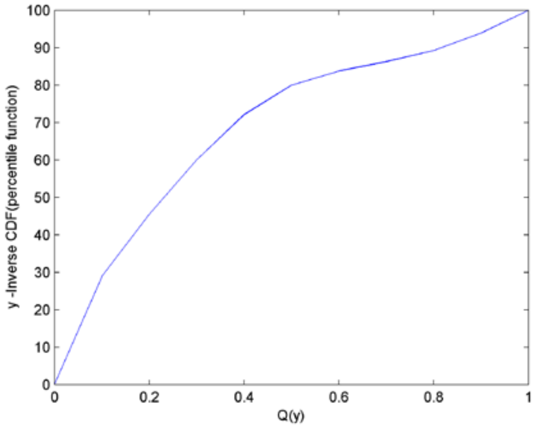

Figure 5. PCHIP to estimate values of ICDF

build the PDF curve: we take a grid of y values (such as .01, .02, .03, . . ., .99, that give us the 1%, 2%, 3%, . . ., 99% points), find x at each of those points from (1), and find f(x) at that x from (2). Then, we plot the pairs (x, f(x)) and link them with a smooth curve.
Now, by using a modification of the *desirability* [9] function, we have to create single integrated PDF curve (which represent PDF curves of all actual tests). For our goals it is enough just to create a single integrated PDF curve by using the arithmetical mean. Suppose that there are PDF curves of R actual tests (given in interpolation points i, namely, i mean points of 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th percentiles, see explanation above), denoted $F_r(x_i)$, ($r = 1 \ldots\ldots R$). They are combined to achieve an overall PDF curve D:

$$D(i) = \frac{\sum_{1}^{R}(F_r(x_i)}{R}, \qquad (4)$$

The integrated PDF curve should be compared with the pattern PDF curve obtained above. To determine the closeness (or distinction) of distribution functions (and, thereby, determine the quality of learning process) we'll use Kullback–Leibler Divergence [10]. Let D and P be two PDFs, defined on $\Re^n$, where n is the dimension of the observed vectors x. The Kullback-Leibler divergence (KL divergence) between D and P is defined as:

$$KL(D \parallel P) = \int_{\Re^n} D(x) \log \frac{D(x)}{P(x)} dx \qquad (5)$$

Here D(x) is an integrated PDF, obtained in (4), and P(x) is a pattern PDF.

The problem of obtaining good upper and lower bounds for the relative entropy attracts considerable interest in information theory .We use the following estimation of upper bounds[11]:

$$KL(D \| P) \le \min\left[\sum_{1}^{n}\frac{D(x_i)^2}{P(x_i)} - 1, \sum_{1}^{n}\sqrt{\frac{D(x_i)}{P(x_i)}}\left|D(x_i) - P(x_i)\right|\right] \quad (6)$$

we assume that the quality of educational or manufacturing processes does not match the required standards . In this case, relevant actions to improve quality must be undertaken. Concrete actions, based on the methodology, described in the paper, are considered below in the text.

Let us assume that comparison of integrated pattern and actual distribution gave us unsatisfactory result: the value (5) is more than the value defined in (6). This means that the quality of learning process is poor and we have to reveal courses and groups that caused this undesired result. Hence, we have to develop a method which can determine courses (or course) whose quality (performance) does not match requirement of the pattern distribution. Besides, we'll examine ways of improving learning quality in these courses.

First of all, we'll consider actual exams. For the simplicity, we consider 5 groups, each containing 20 students (totally 100 students). So, we consider 100 points (grades) obtained in exams for 2 different courses. Moreover, for each exam we consider several factors which can have affect on the quality (that is , on grades obtained). Of course, we assume that such factors are available and can be determined on the basis of interviews of students (filling corresponding questionnaires). Again, for the simplicity we consider the following four factors (in general, number of factors is not crucial for the method developed and any number of factors can be considered):

1. Total midterm evaluation of the student, that is, grades obtained by a student for laboratory works, practical works, quizzes, midterm exam(s) during the current semester; the possible values of this parameter are in the range is: 20÷60; the values of the parameter are filled in the questionnaire by a teacher.
2. Average number of hours (per week) that each student has spent on home assignments or home work during the current semester; the possible values are in the range 0.1÷5 hours; the values of the parameter are filled in the questionnaire by a student.
3. Average grades that each student has obtained for all prerequisites of the current subject (the vector 'aver_prerequizites'); the possible values are in the range 51÷100; the values of the parameter are filled in the questionnaire by a teacher.
4. The difficulty level of the exam (the vector 'exam_difficulty'):
   1 = No study required
   2 = Light revision required
   3 = A reasonable effort required
   4 = Some real study required
   5 = A significant effort requires

The values of the parameter are filled in the questionnaire by a teacher.

So, we obtained training set for four parameters.

We have also a set (sorted) of grades obtained by students for one of the actual exams:

12 12 17 18 21 22 23 24 24 24 26 27 28
28 28 30 30 31 31 33 33 34 35 36 37 39
40 40 41 41 43 45 46 46 46 46 46 48 50
50 51 53 54 55 55 55 55 56 56 57 57 57
58 62 63 64 64 65 66 66 66 67 67 67 70
71 72 73 73 74 75 77 77 77 78 78 79 80
80 80 81 81 82 82 82 83 83 84 86 87 88
88 89 89 90 90 92 97 99 99

As one can see, the distribution of grades is as follows: about 50% of students have grades less or equal 60, about 30% of students have grades between 61 and 80, about 18% of students gave grades between 81 and 95, and 2% of students have grades between 96 and 100. This distribution of grades, of course, does not match the pattern (required) distribution.

Now we have to perform the following task: to find the dependence of the grades on these factors (parameters) and then try to determine the minimum values of the factors that will bring the actual distribution to the pattern one. That is, percents of students received corresponding grades must match the values required by the pattern distribution. For example, percent of students who received grades less or equal 60 must be 40%, percent of students who received grades between 61 and 80 must be 20%, percent of students who received grades between 81 and 95 must be 30%, and percent of students who received grades between 96 and 100 must be 10%. The percentage of actual grades (see above) is quite different.

To perform this task there are many difficulties. The character of the dependence of the percent distribution of students received certain marks on these parameters is absolutely unclear. Moreover, the dependencies in our case are likely *non-linear*. Consequently, it is impossible to determine in advance the type of regression dependence, which is necessary to carry out the regression analysis [12]. Based on the above, the most adequate approach is the use of the neural networks. Using this approach it is possible theoretically reasonable and objective research and identification of the *hidden* nature of the above dependence. Neural networks - a powerful modeling tool, allowing to reproduce extremely complex dependencies. Neural networks are non-linear in nature. In addition, neural networks can cope with the "curse of dimensionality", which does not allow to simulate non-linear dependencies in the case of a large number of variables. Then, after the determination of this relationship, one can use it to determine the needed values of the parameter. This is the purpose of the proposed approach

To build neural network model for our task we'll use the *Generalized Regression Neural Network (GRNN*). It is known the GRNN is a much efficient method for fitting or approximating the complex dependencies. Generalized Regression Neural Networks (GRNN) is a special case of

Radial Basis Networks (*RBN*) [13]. Here a *radial basis function* (RBF) (also called a *kernel* function) is used to predict value of the dependent variable in some point by taking into account the values of dependent variable in neighbor points. The RBF is applied to the distance to compute the weight (influence) for each point. The radial basis function is so named because the radius distance is the argument to the function.

Weight = RBF(distance)

Different types of radial basis functions could be used, but the most common is the Gaussian function. If there is more than one predictor variable, then the RBF function has as many dimensions as there are variables. The best predicted value for the current point (for which the prediction is being performed) is found by summing the values of the other points weighted by the RBF function. Unlike standard feed-forward networks, GRNN estimation is always able to converge to a global solution and won't be trapped by a local minimum.

In the paper we have used programmatic statements and the graphical user interface "nntool" of the MATLAB's toolbox "Neural Networks". To use the toolbox we have created the independent training set (the 4x100 array 'independent_training_set') on the basis of these factors (parameters): total midterm evaluation, average number of hours (per week) , average grades, difficulty level of the exam. We also have created the dependent training set (1x100 vector 'dependent_training_set') on the basis of the grades obtained by students for one of the actual exams.

We started by calling the command "nntool" of the MATLAB toolbox "Neural Networks". Next we import two datasets: 'independent_training_set' and 'dependent_training_set'. Then we created the neural network of the 'generalized regression neural network' type. Here we assign the spread constant the value 0.7. We use a spread slightly lower than 1, the distance between input values, in order, to get a function that fits individual data points fairly closely. A smaller spread would fit data better but be less smooth. The network looks like

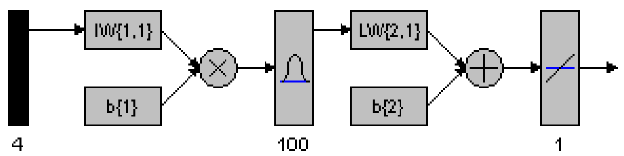

Fig. 6. The network of 'generalized regression neural network' type

The advantage of the GGRN networks is that the training process is carried out in parallel with creation of the network. So, we can immediately use (simulate) the network for the new data.

To fulfill our goal (to determine the minimum values of the factors that will bring the actual distribution to the pattern one) we can try to change the values of one of the factors (or all factors). The change may consist in increasing or decreasing of the factor (depending on whether the percentage of actual grades is more or less than the percentage of the pattern one). For the percentage of failed students (those who obtained less than 60 grades) the changes are as follows: if the actual percentage is more than pattern one, the algorithm has to increase the values of three first factors (total midterm evaluation of the student, average number of hours that each student has spent on home assignments, average grades that each student has obtained for all prerequisites of the current subject) and, maybe, to reduce the difficulty level of the exam (in case if changes of first three factors did not help). The order of factors (priorities) that must be changed is determined by the administration. One option of the priority is as follows:

1. total midterm evaluation of the student
2. average number of hours that each student has spent on home assignments
3. average grades that each student has obtained for all prerequisites of the current subject
4. difficulty level of the exam

Another option is when all factors have the same priorities.

For other percentages (percentages of students have grades between 61 and 80, percentages of students gave grades between 81 and 95, and percentages of students have grades between 96 and 100) the rule is as follows: if the actual percentage is less than pattern one, the algorithm has to increase the values of three first factors maybe, to reduce the difficulty level of the exam (in case if changes of first three factors did not help). If the actual percentage is more than pattern one, then maybe it is necessary to increase the difficulty level (since the recommendation of decrease values of first three factors is not acceptable from a pedagogical point of view).

Here it is necessary to emphasize the following point: changing the values of factors is intended to determine the values which may be useful in future, that is, the updated values of factors can be taken into account and recommended for preparation to future exams. For example, if the average prerequisite grades for failed students, found by the proposed procedure, is, say, 68, then the administration may issue the decree that students who have the average prerequisite less than 68, cannot be admitted to the exam, otherwise the probability of pattern (required) requirements' violations increases and, thereby, the quality of the educational process deteriorates. Besides, it is assumed that ability of students to learn (and which are fixed by grades obtained in the exams) will be unchanged in future. The main goal of the proposed approach is to meet the requirements of the quality of learning process developed by the university's administration.

In accordance with the above mentioned we can continue by creation a new training set, for example, for total midterm evaluation factor. The step of change is: (maximum value – minimum value)/10, or (59-20)/10=3.9. The rounded value is 4. The updated values of the factors are submitted (as training set) to the GRNN. The result (updated values of grades) are obtained, the percentage of failed is reduced and now is 46%.

The further action depends on the distribution of priorities among the factors. If all factors have the same priorities then the next action is changing the values of the next factors (here this is average number of hours that each student has spent on home assignments). If the current factor has higher priority, then the its values is increased by the step (equal to 4), the updated independent training set again is submitted to the network, the updated grades are again analyzed and so on. Only if the end of the range of factor's possible values is reached and the desired result is not obtained (that is, no reduction of the percentage of failed students to 40% is obtained), we continue with updating values of the next factor.

Let us assume that all factors have the same priority. In this case we proceed with the next factor. So, we have to change (increase) the values of the second factor - average number of hours that each student has spent on home assignments. The step of change is: (maximum value –minimum value)/10=0.4441. The updated independent training set is again submitted to the network, it is simulated, the obtained results show that there was no reduction of failed students' percentage: this value remains 46%. .

Hence , we continue with the next factor - average grades that each student has obtained for all prerequisites of the current subject. We change values of this factor by the appropriate step, again submit updated set to the network, simulate the network, obtain the grades. Now we obtained the reduced percentage of failed students: 42%.

As the required (pattern) value 40% is not reached, we return to the first factor, update it, submit to the network, simulate it and obtain the new result: percentage of failed student is 39%. Hence, we obtain the desired result and it corresponds to the following minimum values of the factors:

Minimum value of the total midterm evaluation =28

Mimimum average number of hours that each student has spent on home assignments=0.93 hours

Minimum average grades that each student has obtained for all prerequisites = 60

Difficulty level of the exam = 3

The combination of values of these factors provides required quality of the learning process in the part of percentage of failed students. The values of the factor can be taken into account when preparing future exams. As for the other percentages, the search of appropriate values is been performed (with some difference that are describe above).

However, as one can see, this process (using the panel of "nntool" manually) is quite tedious. Therefore, the fully automated module has been developed for the paper. The MATLAB program constructions were used.

III. Conclusions.

The problem of evaluation of manufacturing, business and learning processes is defined in the paper. The need to use non-parametrical approximation methods is demonstrated. A new approach to evaluate the closeness of realistic (actual) quality of production or learning processes to the pattern requirements is proposed in the paper. In case of significant difference between actual and pattern distributions a new approach of determining the minimum values of the factors that will bring the actual distribution to the pattern one. The Generalized Regression Neural Network (GRNN) is used in the proposed approach. This approach might be used in manufacturing, Business and Educational fields.

References

[1] Rohwer G. Statistical Methods in Sociological Research of Education, Ruhr University Publish,, Germany, 2012.

[2] Schwarz J..Sampling, Regression, Experimental Design and Analysis for Environmental Scientists,Biologists, and Resource Managers, Simon Fraser University, 2011

[3] Milnikova I. Elaboration of statistical quality control models in education process, PhD thesis, Georgian technical University, Tbilisi, 2012

[4] Manzini R. A. Regattieri H.,Pham E., Ferrari. Maintenance for Industrial , © Springer 2010

[5] Lule Ahmedi, Eliot Bytyçi, Blerim Rexha, and Valon Raça Improving students learning process by analyzing patterns produced with data mining. International journal of education and information technologies, Issue 3, Volume 7, 2013, pp.90-97

[6] Erceg-Hurn D. M., Mirosevich V. M. Modern Robust Statistical Methods, University of Western Australia, American Psychologist, 2008, Vol. 63, No. 7, 591–601

[7] Karian Z.A.,. Dudewicz E.J. Handbook of Fitting Statistical Distributions with R", CRC Press, 2011.

[8] Fournier B. , Rupin N., Bigerelle M., Najjar D., and Iost A.. "Fitting a Generalized Lambda Distribution uing a percentile-KS (P-KS) Adequcy Criterion", in (1) Pages 279-309, ISBN: 78-1-58488-711-9

[9] Max Kuhn . The desirability Package., max.kuhn@p_zer.com., January 7, 2012

[10] Perez-Cruz F. Kullback-Leibler Divergence Estimation of Continuous Distributions,Department of Electrical Engineering, Princeton University, New Jersey, 2008

[11] Gofman, M. Kelbert. Un upper bound for Kullback-Leibler divergence with a small number of outliers. Moscow Economics National Research University,2012

[12] S. Norin Rahayu, M. R. Norazan, A. H. Az'lina, I. Adriana, M. R. Nornadiah and A. Nur Niswah Naslina. Determining the Significant Factors Affecting the Physical and Mental Components of Academicians Using Robust Linear Regression Models. International journal of mathematical models and methods in applied sciences, Issue 1, Volume 7, 2013, pp.30-37

[13] I. El-Feghi , Z. Zubia and S. Abozgaya. Efficient Weather Forecasting using Artificial Neural Network as Function Approximator. International journal of neural networks and advanced applications, Volume 1, 2014, pp.49-55