Analysis of integral characteristics in the IoT system

I Ya Lvovich¹, Ya I Lvovich², A P Preobrazhenskiy¹, Yu P Preobrazhenskiy¹ and O N Choporov²

¹Information Systems And Technologies Department, Voronezh institute of high technologies, 73a, Lenina st, Voronezh, 394043, Russia
²Information Security Department, Voronezh state technical university, 14, Moscow dist., Voronezh, 394026, Russia

E-mail: komkovvivt@yandex.ru

Abstract. The paper discusses the features of solving problems related to the modelling of information in the Internet of things in the organizations of the agricultural sector. The quality conversion method in numerical estimates is proposed. As a result of ranking indicators according to the degree of decrease or increase in their significance, each linguistic value can be assigned a certain rank. The integrated characteristics based on normalized indicators and scores in the organisations are shown. The features of using a combination of several methods for processing and modeling of reduced monitoring information in the Internet of things system are given.

1. Introduction

Internet of things is currently being used in a wide variety of agricultural enterprises. It is of interest to evaluate the effectiveness of its use for different groups of employees. For this, methods of statistical data processing can be applied. Among the methods used in statistical data processing and modeling, there are algorithms for processing both numerical and qualitative data. However, the arsenal of methods for processing numerical data is much richer, they allow you to get a more accurate result. In addition, most of the methods for constructing multidimensional models cannot be used to process qualitative indicators. In this connection, it becomes necessary to convert information containing fixed semantic meanings to a numerical form.

2. Quality conversion method in numerical estimates

It is proposed to use the following algorithm for analysis. The Internet of Things system [1-3] has many indicators. At the first stage, for each indicator, all possible values are ranked by degree of significance. For the rank assessment of each value, the a priori ranking method is used, which allows you to objectively evaluate the subjective opinion of experts, since with a large number of gradations, the opinions of experts can diverge. When collecting a priori information, experts (m> 7) are invited to fill out questionnaires in which it is necessary to evaluate n values of the indicator depending on the degree of their significance. As a result of ranking indicators [4] according to the degree of decrease or increase in their significance, each linguistic value is assigned a certain rank. If experts find it difficult to assign different grades to all the gradations, they can assign the same ranks to two or more values. In the case of matching ranks, the ranking matrix is reduced to normal so that the sum of the ranks in each row of the ranking matrix [5], where the opinion of the j-th researcher (j) is written, is equal to n(n + 1) / 2. For
this purpose, grades of the evaluated indicator, having the same ranks, are assigned a rank equal to the average value of the places that the variables divided among themselves. According to the ranking matrix, an assessment is made of the consistency of experts using the concordance coefficient:

\[ W = \frac{S(d^2)}{12m^2(n^3 - 1) - m \sum_{j=1}^{m} T_j} \]  

(1)

where \( S(d^2) \) — sum of squared differences; \( a_{ji} \) — generalized sum of ranks of the \( j \)-th variable, \( t_j \) — the number of repetitions of the \( i \)-th rank in the \( j \)-th row of the matrix.

\[ d = \left( \sum_{j=1}^{m} a_{ji} \right) - \frac{1}{2} mn(n + 1); \]

\[ T_j = \frac{1}{12} \sum_{i=1}^{n} (t_j^3 - t_j); \]

(2)

If the ranking matrix does not contain matching ranks, then

\[ W = \frac{12S(d^2)}{m^2(n^3 - 1)}. \]

(3)

The value of the coefficient of concordance lies in the range \((0 \ldots 1)\). With \( W = 1 \), experts are unanimous in assessing the significance of each linguistic value of the indicator; with \( W = 0 \), agreement is completely absent. The significance of the concordance coefficient \( W \) is estimated by \( \chi^2 \) — Pearson’s criterion [6]:

\[ \chi^2_{\text{calc}} = m(n - 1)W. \]

(4)

If with the number of degrees of freedom \( f = n - 1 \) critical value \( \chi^2_{\text{cr}} \) will be less than estimated \( \chi^2_{\text{calc}} \), then the hypothesis of expert consent is accepted. The resulting ranks can be used as a numerical estimate of the indicator values. If the differences between two adjacent pairs of indicator values are unequal, the second stage of converting ordered linguistic evaluations to numerical ones is carried out. Observations that have two possible options (such as “Yes”, “No”) are converted to 1 and 0, respectively. If an observation can take more than two different linguistic meanings \( L_i \) \((i = 1, \ldots, l; \ l \geq 3)\), the method of expert evaluations is again used. Before N experts \((N \geq 2)\), the question is posed: “How important is \( L_i \) more significant than \( L_{i-1} \)?” The answers for each pair are formed in the form of a linguistic variable \( \gamma_i < \text{observation} L_i < \text{more important than} \ L_{i-1} > (i = \overline{2,l}) \). The following terms are defined as terms of this variable:

\[ T_\gamma = \begin{cases} 
\text{strong}, & \\
\text{significantly}, & \\
\text{some}, & \\
\text{little}, & \\
\text{few}, & 
\end{cases} \]

(5)

The number of matches is from 1 to 5. As a result, for each pair \( L_i, L_{i-1} \) \((i = \overline{2,l}) \) N variable values are formed \( \gamma_{j,i}, i = \overline{2,l}, j = \overline{1,N} \). To illustrate the foregoing, we give an example of determining a numerical assessment for an indicator characterizing a person’s belonging to one of the social and professional groups. Suppose for our example that this indicator can take values: «worker», «employee», «pensioner», «leader». The generalized value and numerical estimate of each initial value are calculated by the formula
3. The ranking and sorting of the values

The assessment was carried out on a 4-point scale. Based on the totality of expert opinions, a matrix of weighting coefficients [7] was compiled (table 1).

| Indicator value | Evaluations of eight experts |
|-----------------|-------------------------------|
|                 | I | II | III | IV | V | VI | VII | VIII |
| 1. worker       | 2 | 3  | 2   | 1  | 2 | 1  | 1   | 2    |
| 2. projector    | 1 | 2  | 1   | 1  | 2 | 1  | 1   | 1    |
| 3. designer     | 4 | 1  | 3   | 2  | 3 | 3  | 2   | 4    |
| 4. leader       | 3 | 4  | 4   | 3  | 3 | 4  | 2   | 3    |

Since the same expert assigned the same weight to some indicator values, the ranking matrix had to be brought to its normal form (table 2), so that the sum of the ranks in each column was equal to $K * (K + 1) / 2$, where $K$ is the number of different values of the indicator "social and professional group" ($K = 4$). As a result of the calculation according to formulas (1) - (5), it was obtained $W = 0.75$, $\chi^2_{calc} = 24.00$.

| Indicator values | Evaluations of eight experts | Rang sum |
|------------------|-----------------------------|----------|
|                  | I | II | III | IV | V | VI | VII | VIII |
| 1. worker        | 2 | 3  | 2   | 1  | 1 | 1.5| 2   | 15   |
| 2. projector     | 1 | 2  | 1   | 1.5| 1 | 1  | 2   | 11   |
| 3. designer      | 4 | 1  | 3   | 3  | 3 | 3.5| 4   | 25   |
| 4. leader        | 3 | 4  | 4   | 3  | 3 | 3.5| 3   | 29   |

Since the calculated value turned out to be more critical for the number of degrees of freedom $v = n-1 = 3$ and significance level $q = 5\%$ ($\chi^2_{tabl} = 11.07$), then the hypothesis of expert agreement was accepted. Thus, the following ordered (in terms of the level of social status) sequence of values of the indicator “social-professional group” was obtained: 1) «designer»; 2) «worker»; 3) «designer»; 4) «leader». To determine the numerical evaluation of each indicator value, the linguistic values of indicators were compared in pairs based on expert estimates. The results are shown in table 3. As a result of calculations carried out according to formula (6), we obtain numerical estimates for each linguistic value [8] of the “social-professional group” indicator (table 4).

| The numbers of comparing values | The evaluations of eight experts | Average value |
|--------------------------------|---------------------------------|---------------|
| 1-2                            | 3 2 2 3 3 3 2 2               | 2,500         |
| 2-3                            | 3 1 2 1 2 3 2 2               | 2,000         |
| 3-4                            | 2 1 1 2 2 1 2 1               | 1,500         |
Table 4. Numerical estimates of the values of the indicator “employees of the organization”.

| Linguistic value of the indicator | Indicator score | Normalized rating |
|----------------------------------|-----------------|-------------------|
| «projector»                      | 6,000           | 1,0000            |
| «worker»                         | 3,500           | 0,5833            |
| «designer»                       | 1,500           | 0,2500            |
| «leader»                         | 0               | 0                 |

Thus obtained numerical estimates of the values of a qualitative indicator can be used when applying the methods of mathematical statistics and modeling.

4. Integrated characteristics based on normalized indicators and scores

Evaluation of the current state of the modeling object or forecast of its change over time is possible based on the study of various characteristics, while comparison is possible for individual indicators, and a generalized assessment is very difficult. Such an assessment is necessary when choosing one of the alternative managerial influences, when there is a forecast of changes in individual indicators, but none of them gives a complete description of the state of the system [9-11]. To compare the state of individual elements of the studied system and the forecast obtained, an indicator is needed that allows a comprehensive assessment of the level of the control object, taking into account the individual components and their significance. To build an integral indicator, a list of characteristics reflecting the state of the system under study is initially formed. Then, based on the a priori ranking method and “discrete correlation constellations”, the minimum number of the most significant and not interconnected indicators characterizing the system is selected. For each indicator, a unification system is developed for standardization or their normalization is carried out.

The integral indicator was determined on the basis of the following convolution:

$$IP = \sum_{i=1}^{N} w_i X_i^N$$

(7)

where $N$ – the number of factors included in the integral indicator; $w_i$ – weight (significance) of the $i$-th factor, $X_i^n$ – normalized (ball) assessment of the $i$-th factor.

To assess the significance of each component, the method of a priori ranking is used, which makes it possible to objectively evaluate the subjective opinion of experts.

When collecting a priori information based on the experience, intuition and knowledge of experts, $m$ experts are invited to fill out questionnaires that evaluate all $n$ indicators by their significance. The assessment is made on an $n$-point scale. Based on the set of expert opinions, a ranking matrix is compiled. If the same expert assigns the same rank to some factors, the ranking matrix must be brought to its normal form, so that the sum of the ranks in each column is equal to $n(n+1)/2$.

Consistency of opinions of the examination participants is determined by calculating the coefficient of concordance (consistency) with the subsequent determination of the assessment of the significance of the results. Weight values $w_i$ calculated by the formula

$$w_i = \frac{m \cdot n - \sum_{j=1}^{m} r_{ij}}{\sum_{i=1}^{m} \left( m \cdot n - \sum_{j=1}^{m} r_{ij} \right)}, \quad i = 1, n.$$  

(8)

where $r_{ij}$ ($j = 1, m$) – rank set $j$-s an expert, moreover $\sum_{i=1}^{n} w_i = 1$. 

4
Since the sum of all weighting coefficients \( w_i \) is 1, and the indicators included in the PI are normalized or evaluated on a k-point scale, the maximum possible value of the integrated indicator is equal to the upper limit of normalization (usually +1) or the maximum possible score (k), and the minimum - the lower limit of normalization (usually 0) or the lowest possible score (usually +1).

The proposed integral indicator allows you to comprehensively assess the state of the simulated system.

To assess the effectiveness of the functioning of the system, taking into account the nature of the change in several indicators, an efficiency criterion is used, based on the assumption that the importance of changing the controlled indicators changes exponentially, starting from the beginning of the observation. For each indicator \( P_{kn}^k \) the performance criterion is calculated, which is a convolution that takes into account the nature of the indicator \( k ( k = 1, K) \):

\[
E_n^k = \sum_{i=1}^{I-1} (P_n^k - P_n^{k+1}) x_i, \quad E_n^k = \sum_{i=1}^{I-1} (P_n^k - P_n^{k+1}) x_i, \quad k = 1, K
\]

where \( n = 1, N \) — serial number of the object; \( x_i, (i = 1, I - 1) \) — coefficients characterizing the importance of changing indicators in the desired direction between \( i \)-st and \( i+1 \)-st by measurement.

Odds \( x_i, (i = 1, I - 1) \) calculated on the assumption that the importance of improving performance decreases exponentially over time \( (e^{-at}) \), and for the last observation period \( (t_I) \) she makes up \( \Delta \) percent of the original:

\[
x_i = e^{-a t_i}, \quad a = \frac{\ln(\Delta/100)}{t_I} - \frac{t_I}{0.5(t_I + t_{I+1})}.
\]

The value \( \Delta \) is set by the expert. Due to the change \( \Delta \), various situations are simulated: the performance criterion directly depends on the rate of decline in the importance of improving indicators over time. From formulas (9) - (10) it follows that the convolution obtained is more dependent on the value \( P_{n}^{k_i} \forall n, k \), those. to enable comparison of the functioning efficiency of individual groups of objects of the medical system, it is necessary that the mathematical expectation \( M[P_{n}^{k_i}] \) was the same in the compared groups:

\[
M[P_{1}^{k_i}] \approx M[P_{2}^{k_i}] \approx ... \approx M[P_{n}^{k_i}] \approx ... \approx M[P_{N}^{k_i}] (k = 1, K)
\]

the case when, as a result of the impact, it is important to achieve some functional boundaries, for indicators that do not reach the required boundaries at \( t_I \) - day (year) of observation, the following penalty function is introduced into the expression:

\[
-e^{v(P_{extr}^{k_i} - P_{norm}^{k_i})}, \quad v = \frac{s}{|P_{extr}^{k_i} - P_{norm}^{k_i}|},
\]

where \( v \) is the normalization coefficient; \( s \) is a coefficient characterizing the importance of getting the value of the \( k \)-th indicator on the \( t_I \)-th day (year) of observation in the functional boundaries \( (s = 0 \ldots 3, \text{recommended } s = 0.7); P_{norm}^{k_i} \) — nearest functional boundary for indicator value \( P_{n}^{k_i} \); \( P_{extr}^{k_i} \) — the maximum (minimum) value of the indicator \( k \) on \( t_I \)-day (year) of measurement that goes beyond functional boundaries. At the next stage, using the analysis of variance for each indicator, the hypothesis of the presence of a processing effect is tested. As a result, many of the most informative indicators are formed, for which the hypothesis has been confirmed. For ease of comparison, the relative effectiveness of each indicator is also calculated.
\[ E_{o}^{k} = \frac{E_{n}^{k}}{\min(E_{n}^{k})} \]  \hspace{1cm} (13)

To calculate the generalized performance criterion, a convolution of the form is used:

\[ E_{g}^{n} = \frac{1}{K} \sum_{k=1}^{K} E_{n}^{k} \cdot w_{k} \]  \hspace{1cm} (14)

where \( w_{k} \) - the importance of the k-th indicator.

5. Conclusion
The accuracy of the classification, prognostic and optimization models of enterprises using the IoT technology is largely determined by the quality of the source data. In this regard, the organization of purposeful collection, pre-processing and structuring of archival and experimental information is necessary. In order to increase the accuracy of models and reduce the complexity of their use, preliminary processing of statistical information is necessary, namely: information filtering, which enables the selection of reliable data and the elimination of sharply distinguished observations; the exclusion of parametric redundancy, which allows to minimize the number of measured parameters, provided that the selected parametric system is sufficiently informative.

References
[1] Minerva R, Biru A and Rotondi D 2015 Towards a Definition of the Internet of Things (IoT) (IEEE Internet Initiative, Torino, Italy)
[2] Stankovic J A 2014 Research directions for the internet of things IEEE Internet of Things Journal 1 3-9
[3] Broadband Internet Technical Advisory Group Internet of Things (IoT) Security and Privacy Recommendations 2016
[4] Mahesh V, Swathi Lekshmi P S, Dilip Ananda Pawar, Daliamol, Alok Kumar and Prakash P 2017 Matrix Ranking - An important PRA tool to assess farmers preferences and priorities Agric. Sci. Digest. 37(2) 100-5
[5] Maria-Pilar Jarabo-Amores, David de la Mata-Moya, Roberto Gil-Pita and Manuel Rosa-Zurera 2013 Radar detection with the Neyman-Pearson criterion using supervised-learning-machines trained with the cross-entropy error Emerging radar techniques Journal on Advances in Signal Processing 1 44
[6] Lee S, Liao Y, Seo M H, and Shin Y 2018 Oracle estimation of a change point in highdimensional quantile regression Journal of the American Statistical Association 113(523) 1184-94
[7] Riza L S, Bergmeir C, et al. 2018 Fuzzy Rule-Based Systems for Classification and Regression Journal of statistical software 65(6) 1-30
[8] Odu G O and Charles-Owaba O E 2013 Review of Multi-criteria Optimization Methods Theory and Applications 3 1-14
[9] Shah A and Ghahramani Z 2015 Parallel predictive entropy search for batch global optimization of expensive objective functions Advances in Neural Information Processing Systems 2 3330-8
[10] Neittaanmki P, Repin S and Tuovinen T 2016 Mathematical Modeling and Optimization of Complex Structures; Series: Computational Methods in Applied Sciences (Springer International Publishing AG, Switzerland)
[11] Yao Y and Chen J 2010 Global optimization of a central air-conditioning system using decomposition-coordination method Energy and Buildings 42(5) 570-83