UDC 658.5:004.94
DOI: 10.15587/1729-4061.2022.265782

MODELS OF ADAPTIVE INTEGRATION OF WEIGHTED INTERVAL DATA IN TASKS OF PREDICTIVE EXPERT ASSESSMENT

Igor Ruban
Doctor of Technical Sciences, Professor, First Vice-Rector*

Hennadii Horenskyi
Department of Electronic Computers*

Yuri Romanenkov
Corresponding author
Doctor of Technical Sciences, Professor
Department of Management**
E-mail: romanenkov.yu.a@gmail.com

Daniil Revenko
Doctor of Economic Sciences, Associate Professor
Department of Economics, Marketing and International Economic Relations**

*Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166
**National Aerospace University «Kharkiv Aviation Institute»
Chkalova, str., 17, Kharkiv, Ukraine, 61070

The problem of improving the methodological and algorithmic support of the complexing process by developing models of adaptive redundant complexing of weighted interval data has been solved. The object of this study is the process of complexing interval data obtained from several independent sources; the subject is the algebraic methods of excessive complexing of weighted interval data. The relevance of the task is due to the severity of the problem of consolidating homogeneous data in order to obtain more accurate and relevant information about the object or process under study. Models have been developed that, unlike the existing ones, make it possible a posteriori to take into account the accuracy of experts at the preliminary stage of expert evaluation. A single analytical form of the model for processing weighted interval and point estimates with the possibility of structural and parametric tuning is proposed. It allows one to increase the degree of automation of processing expert assessments under conditions of interval uncertainty. Recommendations for the practical application of the proposed models have been formulated. Options for parametric configuration of preference functions were indicated depending on the characteristics of weighted interval estimates. The commonality of the limiting cases of the proposed models with previously known ones is proved. The example shows the shift of the integrated assessment to the side of more accurate assessments at the previous stage of source assessment. The adaptability of the proposed models is illustrated. At the same time, a slight, on average, about 10 %, expansion of the complexed interval relative to the primary ones was registered. The built models and algorithms can be used in automated expert systems, as well as in cascade models of information processing and compression

Keywords: predictive assessment, complexing, alternative data, interval analysis, confidence probability

Received date 19.07.2022
Accepted date 28.09.2022
Published date 30.10.2022

***How to Cite:** Ruban, I., Horenskyi, H., Romanenkov, Y., Revenko, D. (2022). Models of adaptive integration of weighted interval data in tasks of predictive expert assessment. Eastern-European Journal of Enterprise Technologies, 5 (4 (119)), 6–15. doi: https://doi.org/10.15587/1729-4061.2022.265782*

1. Introduction

The analytical work of hedge funds and various financial companies has traditionally focused, as a rule, on the analysis of fundamental market data [1], which are used to determine the parameters of trading and investment strategies. The rapid development of information and communication technologies has provided analysts with access to a wide range of new, alternative data sources. An example is StockTwits [2], a social network and microblogging platform where several hundred thousand investment professionals exchange information and trade.

In addition to data on individuals, data generated by various business processes are also accumulated and become available. These include transaction data, bank records, cashier scanner data, supply chain orders, company financial statements, press releases, etc. For example, customers of Eagle Alpha, one of the well-known business data providers, received JP Morgan's transaction data sets. At the same time, they confirmed the effectiveness of using such data in the investment process, as well as data on social media sentiment, press releases, and any other messages of companies [3]. A number of researchers note the impact of sentiment, informativeness, innovation, and variability of the financial statements from SEC (US Securities and Exchange Commission) on the stability, profit, and value of the company's shares [4].

The heterogeneity and novelty of sources of this kind of data contributed to the emergence of the term «alternative data» (alt-data) and gave rise to a fast-growing service industry [5, 6]. At the same time, in the arsenal of providers of such services there are modern methods of quantitative and text analysis [7].

The process of data processing and analysis in modern information systems, in particular, financial and analytical ones, is characterized by the uncertainty of data of various nature, both objective and subjective. Taking into account the continuous increase in the volume of data, the relevance of the task of consolidation (integration, data fusion) of homogeneous data increases, in order to obtain more accurate and relevant information about the object or process under study.

The task of data integration can also arise in the process of predictive expert assessment when information from several sources (or obtained by different methods) is available to the expert. In this case, experts are faced with the need to form consolidated estimates, taking into account the uncertainty of primary data [8]. This task belongs to the technologies of the so-called «gray» analysis [9, 10], which is characterized by partial uncertainty of information.

The mathematical apparatus for combining such data should make it possible to take into account various kinds of uncertainty objectively inherent in the data. In this regard, promising areas are the creation of new and adaptation of existing methods of information processing to take into account the features of primary data [11]. They can be used to improve the methodological base of specialized expert systems and decision support systems, including investment ones.

Thus, a relevant task is to develop new and adapt existing methods of integration, taking into account the various forms and features of data uncertainty for the generation of consolidated assessments of the indicators of the state of the object of expertise. The solution to this urgent scientific and applied problem will increase the degree of automation of predictive assessment processes, as well as ensure the relevance of consolidated assessments of data on the object of expertise.

2. Literature review and problem statement

Consolidation of data from different sources in order to obtain a more accurate and reliable description of the object of expertise requires the use of specially developed methods of integration. They can be divided into three groups, depending on the type of complexation [12].

In complementary integration, information from several sources is fragments of some information about an object and is used to obtain more complete information about it. An example is a network of cameras with different fields of view to observe a certain space [13, 14].

In redundant aggregation, sources provide data on the same characteristic of an object in order to obtain a more reliable and relevant assessment of that characteristic. This type of integration is used to improve the reliability, accuracy, and reliability of data, and is often used to ensure fault tolerance and robustness of systems [15]. It is used, for example, to combine images coming from different cameras with a partially matching field of view to reduce noise levels [16].

In cooperative integration, information received from different sources is combined into new information, usually more complex than the original. A typical example is the determination of the location of an object based on azimuth and distance data from the reference point [17].

The process of data integration is characterized by objective problems associated with the peculiarities of data.

Inaccurate data. Data provided by different sources are often low accurate due to the influence of the procedure for obtaining them, the intrinsic properties of the sources, external factors, or the nature of the data itself. In such cases, the uncertainty that arises during the evaluation of data leads to a deliberately incorrect result.

Contradictory data. Multiple sources provide data that are inconsistent. The degree of inconsistency can range from low («release», partial contradiction) to extremely high (inconsistency of most of the data reported). Such a conflict may be a consequence of the internal properties of the data source (for example, the subjectivity of the expert or the error of the sensor) or the influence of external factors (for example, environmental conditions).

Heterogeneous data. Data from sources can be represented in the form of qualitatively different information characterizing the same object of examination. The assessment is complicated by the fact that in addition to quantitative values, the parameters can also be expressed by qualitative indicators.

Incomplete data. Often there is a situation when the data provided are not enough to obtain satisfactory knowledge about the object of examination. It is connected, first of all, with the impossibility of obtaining information from all sources. For example, when measuring with a wireless sensor network, the failure of the sensors, the radio module of the node, or the overload of the radio channel may be the reasons that some of the required data will not be delivered to the central node.

Large amount of data. The data supplied by the sources may differ in a significant volume, which makes it difficult to transfer and process them further. This problem is especially relevant when it comes to information obtained with the help of technical means [18]. The transfer of large amounts of data is always associated with an increase in the time, energy and computational resources of the system. Integration procedures should reduce the size of the data while retaining meaningful information.

Since the solution to the problem of combining predictive estimates by definition allows for a multiplicity of solutions [19, 20], in modern publications you can find different approaches to the solution. At the same time, they can be characterized as static [21, 22] or dynamic [23].

In the case when the assessments of experts are formed in interval form [24], the application of classical methods of complexation becomes very difficult. This is due, first of all, to the fact that attempts to reduce interval estimates to point estimates do not always adequately take into account the specificity of the task [25]. Direct comparison of interval quantities by means, for example, of the utility function [26], does not make it possible to achieve the necessary commonality of the solution.

The main approaches to the integration of interval data include methods based on the Dempster-Schafer obvious theory [27, 28] but they cannot simultaneously ensure the uniqueness of the result of integration and its resistance to the inconsistency of the input data.

The application of methods based on probability theory and mathematical statistics [29, 30] is difficult since in many cases there is a contradiction between the postulates of probability theory and interval analysis.

Methods of fuzzy logic [31] are in many cases more in demand. However, additional input on fuzzy data membership functions is needed to use them effectively.

The group of methods of approving voting [32], despite its simplicity, has a significant drawback. In the case of poorly coordinated intervals, the probability of the appearance of so-called paradoxes is high. This leads to false results.

A statistical approach is also used to determine the smallest consistent subset of intervals and then calculate the weighted average value [33, 34] but it is effective only with a high level of data consistency. The same feature is also characteristic of methods based on interval ordinal numbers [35].

The application of the methods mentioned above becomes even more difficult if we take into account another feature inherent in the data. This is additional expert information expressed in the form of weights attributed to intervals.

Despite the fact that the problems of processing interval data have been intensively studied for a long time, their final solution is still far from being completed. Thus, the development of new methods of integration and the search for ways to adapt existing data to the interval form is a promising means of solving the problem associated with data inaccuracy.

3. The aim and objectives of the study

The purpose of this study is to improve the methodological and algorithmic support of the process of integrating interval data, taking into account additional expert information. This will provide an opportunity to expand the range of input data and increase the degree of automation of expert information and analytical systems.

To accomplish the aim, the following tasks have been set:

– to consider the prerequisites for the formalization of expert data in an interval form;

– to construct models of adaptive integration of weighted interval data.

4. The study materials and methods

The concept of «interval» or «interval number» $[x]$ within the framework of this problem will be interpreted in the classical sense [36–40], as a set of possible values of an unknown true value x, that is, as a limited interval of its uncertainty:

$$[x]=[\underline{x},\ \overline{x}]=\{x\,|\,\underline{x}\le x\le\overline{x}\}, \tag{1}$$

where $\underline{x},\ \overline{x}$ is the lower (left) and upper (right) boundary of the interval, respectively.

The width or window of the interval $[x]$ is the value

$$\mathrm{wid}[x]=\overline{x}-\underline{x}. \tag{2}$$

In interval arithmetic, real numbers are identified with zero-width intervals (degenerate intervals) $[x]=[x,\ x]$.

For two intervals $[a]=[\underline{a},\ \overline{a}]$ and $[b]=[\underline{b},\ \overline{b}]$ in classical interval arithmetic, the following operations are specified:

$$[a]+[b]=\left[\underline{a}+\underline{b},\ \overline{a}+\overline{b}\right], \tag{3}$$

$$[a]-[b]=\left[\underline{a}-\overline{b},\ \overline{a}-\underline{b}\right], \tag{4}$$

$$[a]\cdot[b]=\left[\min\left\{\underline{a}\underline{b},\underline{a}\overline{b},\overline{a}\underline{b},\overline{a}\overline{b}\right\},\max\left\{\underline{a}\underline{b},\underline{a}\overline{b},\overline{a}\underline{b},\overline{a}\overline{b}\right\}\right], \tag{5}$$

$$[a]/[b]=[a]\cdot\left[1/\overline{b},\ 1/\underline{b}\right],\ 0\notin[b]. \tag{6}$$

Operations similar to arithmetic and applicable to intervals are considered by a special field of mathematics – interval analysis [41].

The use of interval analysis makes it possible to enclose in the intervals those solutions to problems whose input data are known only that they lie in some intervals. As is known, there are several ways to process interval quantities: classical interval arithmetic [42], generalized interval arithmetic [43], Kaucher arithmetic [44], interval analysis [36, 39–41, 43, 44].

In classical interval arithmetic, symbolic transformations become impossible since the distributive law is not fulfilled and there are no inverse elements [36]. Interval analysis, by contrast, is devoid of this drawback. It pursues the goal of finding the area of possible values of the result, taking into account the data structure and functions specified in symbolic form. This makes it possible to save information about the set of possible values of the model parameters during the transformations [45].

5. Improvement of methodological support for the process of interval data integration, taking into account additional expert information

5. 1. Prerequisites for the formalization of expert data in interval form

The prerequisites for the formalization of expert assessments in interval form may be the following factors [46]:

– in the process of predictive expert evaluation, the interval form of assessments may arise directly as a result of the execution of the examination task [47];

– the results of measurements of the parameters of the object of examination, direct or indirect, performed with errors, can naturally be represented in interval form [36];

– interval models are preferable to probabilistic-statistical ones in the case of one-time decisions [48];

– the apparatus of interval analysis has proved its effectiveness in solving various scientific and practical problems [37];

– algorithms for processing interval data, as a rule, do not require specialized tools for software implementation.

5. 2. Construction of models of adaptive integration of weighted interval data

Problem statement. It is necessary to synthesize a consolidated interval weighted assessment by combining particular, in the general case poorly coordinated, weighted interval estimates of experts of the form:

$$[x_i]=\left[\underline{x_i},\ \overline{x_i}\right],\ \alpha_i=P\left(X\in[x_i]\right),\ i=1,\dots,N, \tag{7}$$

where $[x_i]$ are the interval estimates of experts with a total number of N, α_i are the confidence probabilities attributed to them by experts [29, 30].

It is logical to interpret the par$_i$ameter α as the degree of confidence of the expert in finding the parameter X inside the corresponding interval $[x_i]$. Given that the interval $[x_i]$ with the «weight» α_i is a three-parameter data model, within the framework of this paper we shall use the term «weighted interval». It does not matter how the weight of the interval is called in the procedure of expert estimation: «confidence interval», «degree of confidence of the expert», or something else. The only requirement for α_i parameter is that it is limited:

$$0<\alpha_i\le 1. \tag{8}$$

The limit value of parameters $\alpha_i=1$ reduces model (7) to the classic two-parameter interval data model.

The main approaches to finding the integrated interval are the Marzullo method [49] and its modification. For example, in [30] it is proposed to select subsets of the maximum or in advance of a given power on the combined set of intervals. In [50, 51], a method for combining interval data based on preference aggregation is proposed. These methods are widely used in the field of information and communication networks, in particular in synchronization tasks. However, these methods give the best results when the set of primary data for the complex is predominantly consistent, that is, most of the intervals overlap in pairs. Within the framework of the task, the interval assessments of experts are poorly coordinated, in contrast to the data from automatic monitoring systems.

Formulations of assumptions and limitations of the set problem in accordance with [46]:

1. The paradigm of interval analysis is taken as the basic one. It takes into account, in addition to the rules of classical interval arithmetic, the physical meaning and logic of analytical transformations of the mathematical model of complexing.

2. The main hypothesis is that the complexed interval estimation belongs to the set of superpositions of the original partial interval data:

$$\left[x^*\right]=\sum_{i=1}^{N} w_i\left[x_i\right], \tag{9}$$

where w_i are the weightings of the aggregation model.

3. It should be ensured that each of the interval methods of integration being developed in the extreme case (when the interval estimates are narrowed to point estimates with unit weights) would be reduced to an appropriate method of integrating point estimates.

4. Data from experts shall be considered unbiased until the contrary is justified.

5. The history of multiple evaluation is available for accumulation and statistical processing.

The study was conducted by analytical modeling. The simplest algebraic model of integration can be considered the arithmetic mean of weighted interval data (7):

$$\left[x^*\right]=\frac{\sum_{i=1}^{N}\alpha_i\left[x_i\right]}{\sum_{i=1}^{N}\alpha_i} \tag{10}$$

or, in designations of (9)

$$w_i=\frac{\alpha_i}{\sum_{i=1}^{N}\alpha_i}. \tag{11}$$

In this case, we shall search for the weight of the complexed interval for the following reasons. Let the measure of the quality of the result of the expert's work be, in addition to accuracy, the degree of proximity of the interval weighted estimate to the point with a single confidence probability. In this case, the greater the confidence probability α_i and the smaller the width of the wid[x] assessment, the more preferable the expert's assessment can be considered. Then it is acceptable and logical to assume that the complexed interval will average the selected quality assessment for all experts:

$$\frac{wid\left[x^*\right]}{\alpha^*}=\frac{1}{N}\sum_{i=1}^{N}\frac{wid\left[x_i\right]}{\alpha_i},\ \alpha_i\neq 0. \tag{12}$$

Hence,

$$\alpha^*=N\frac{wid\left[x^*\right]}{\sum_{i=1}^{N}\frac{wid\left[x_i\right]}{\alpha_i}}. \tag{13}$$

Thus, on the basis of the initial data in the form of (7), it is possible to obtain a complexed interval (10) with weight (13), that is, a model of integration of the following type:

$$\begin{cases}\left[x^*\right]=\sum_{i=1}^{N} w_i\left[x_i\right],\\ w_i=\frac{\alpha_i}{\sum_{i=1}^{N}\alpha_i},\\ \alpha^*=N\frac{wid\left[x^*\right]}{\sum_{i=1}^{N}\frac{wid\left[x_i\right]}{\alpha_i}},\\ i=1,\ldots,N,\ \alpha_i\neq 0.\end{cases} \tag{14}$$

Adaptive aggregation. In the case when the results of the previous assessment are known, that is, the magnitude of absolute deviations in the interval form for the time $t=T$:

$$\left[\Delta_i\right]=x\big|_{t=T}-\left[x_i\right]_{t=T}=\left[\underline{\Delta}_i,\ \overline{\Delta}_i\right],\ i=1,\ldots,N, \tag{15}$$

there is a problem of quantitative comparison of interval numbers.

The direct use of classical interval arithmetic in this case is unacceptable [36]. Nevertheless, there are several options for formalizing such a task. For example, in [36] a graph-analytical formalization of the problem of comparing interval numbers is proposed. It allows us to express the measure of the reliability of hypotheses about the mutual arrangement of two numbers within the intervals through the relations between the characteristic areas inside the rectangle formed by the interval numbers themselves.

Another variant of formalization is proposed in [52, 53] and is associated with the construction of adjusted interval logic. The assumptions in this model are not absolutely strict, which leads to failures in some special cases.

To formalize such a problem, one can also use the magnitude of the distance between the interval numbers $dist\left([a],[b]\right)=\max\left\{\left|\underline{a}-\underline{b}\right|,\left|\overline{a}-\overline{b}\right|\right\}$, however, the distributional logic is not strictly observed, which casts doubt on the possibility of full-fledged practical application of this option.

The formalization option proposed in [46, 54] was used to determine the quantitative measure of the proximity of interval errors to zero. This choice is subjective and due to the wide possibilities of parametric tuning of the model in this case.

According to [46], the quantitative measure of the proximity of the interval error $\left[\Delta\right]=\left[\underline{\Delta},\overline{\Delta}\right]$ to zero can be determined as follows:

$$u^*=\frac{1}{wid\left[\Delta\right]}\int_{\underline{\Delta}}^{\overline{\Delta}}u\left(\Delta\right)\mathrm{d}\Delta, \tag{16}$$

where $u(\Delta)$ is a preference function, even, monotonically decreasing, non-negative on the entire real axis.

The choice of the type of function $u(\Delta)$ depends on the characteristics of the particular study. For example, the use of inversely proportional:

$$u=\frac{1}{|\Delta|}, \tag{17}$$

inversely quadratic:

$$u=\frac{1}{\Delta^2} \tag{18}$$

or piecewise-linear

$$u(\Delta)=\begin{cases} k\left(\Delta^*-|\Delta|\right), \ |\Delta|<\Delta^* \\ 0, \ |\Delta|>\Delta^* \end{cases} \tag{19}$$

functions. For example, the choice of form (19) ensures the cutting off of sources that do not provide a predetermined allowable error level Δ^*. This translates the integration model into the category of selective [30].

Having standardized the indicatorα α_i and u_i^*, we obtain expressions for the weighting coefficients w_i of the integration model (9):

$$w_i=\frac{\alpha_i u_i^*}{\sum_{i=1}^{N}\alpha_i u_i^*}, \ i=1,\ldots,N. \tag{20}$$

The weight of the complexed interval will be found by analogy with (12).

Thus, the obtained model of interval data integration:

$$\begin{cases} \left[x^*\right]=\sum_{i=1}^{N} w_i\left[x_i\right], \\ w_i=\dfrac{\alpha_i u_i^*}{\sum_{i=1}^{N}\alpha_i u_i^*}, \\ u_i^*=\dfrac{1}{wid\left[\Delta_i\right]}\int_{\underline{\Delta_i}}^{\overline{\Delta_i}} u(\Delta)\mathrm{d}\Delta, \\ \alpha^*=N\dfrac{wid\left[x^*\right]}{\sum_{i=1}^{N}\dfrac{wid\left[x_i\right]}{\alpha_i}}, \\ i=1,\ldots,N, \ \alpha_i\neq 0 \end{cases} \tag{21}$$

is adaptive to the results of the previous assessment, expressed in the form of absolute deviations of past assessments of experts (15).

If the confidence probabilities α_i', $i=1, \ldots, N$ used by the experts in the previous assessment are also known, the quality of the expert's assessment μ_i can be specified as a function of two variables:

$$\mu_i\left(\alpha_i', \ u_i^*\right)=\alpha_i' u_i^*. \tag{22}$$

It is not difficult to verify that function (22) turns to zero at $\alpha_i'=0$ or $u_i^*=0$.

Then, taking into account (22), model (21) can be transformed as follows:

$$\begin{cases} \left[x^*\right]=\sum_{i=1}^{N} w_i\left[x_i\right], \\ w_i=\dfrac{\alpha_i\alpha_i' u_i^*}{\sum_{i=1}^{N}\alpha_i\alpha_i' u_i^*}, \\ u_i^*=\dfrac{1}{wid\left[\Delta_i\right]}\int_{\underline{\Delta_i}}^{\overline{\Delta_i}} u(\Delta)\mathrm{d}\Delta, \\ \alpha^*=N\dfrac{wid\left[x^*\right]}{\sum_{i=1}^{N}\dfrac{wid\left[x_i\right]}{\alpha_i}}, \\ i=1,\ldots,N, \ \alpha_i\neq 0. \end{cases} \tag{23}$$

Model (23) is adaptive to the results of the previous weighted interval estimation with weights α_i', $i=1, \ldots, N$, expressed as absolute deviations of past estimates of experts (15).

Note that in the obtained models (21) and (23) it is possible to process both interval estimates and point estimates simultaneously in one set of expert data. In this case, expression (16) for point data will be simplified and will take the following form:

$$u_i^*=u\left(\Delta_i\right), \tag{24}$$

that is, the degree of proximity of the error to zero for point data is defined as the value of the preference function at the point Δ_i.

It can be verified that at $\alpha_i'\to 1$, $i=1, \ldots, N$, model (23) degenerates into model (21), and, at the limit values of the parameters,

$$\begin{cases} wid\left[x_i\right]\to 0, \\ \alpha_i, \ \alpha_i'\to 1, \\ u(\Delta)=\dfrac{1}{\Delta^2} \end{cases} \tag{25}$$

both models coincide with the aggregation model for point estimates given, for example, in [20].

Adaptive aggregation models in the form of (14), (21), (23) make it possible a posteriori to account for the accuracy of experts. Another distinctive feature of the proposed models is the absence of a requirement for a high degree of consistency of expert assessments.

The analytical form of the built models makes it possible to directly process both interval and point estimates at the same time (Fig. 1).

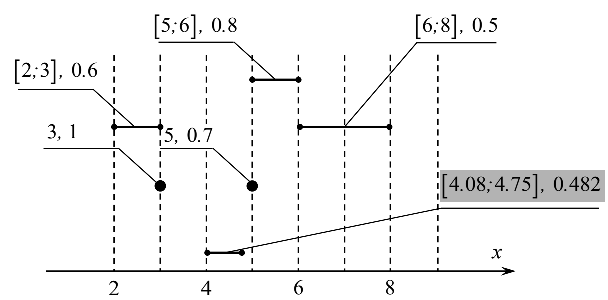

Fig. 1. Example of algebraic aggregation of three interval and two point weighted estimates

The adaptability of the built models is ensured by retrospective comparison of errors in expert assessments. Under conditions of interval uncertainty, the comparison of intervals is carried out using a preference function (Fig. 2).

Table 1 gives an example of adaptive data aggregation from Fig. 1.

Fig. 3 graphically presents the results of integration – gray rectangles with heights proportional to the weights of the obtained intervals are built over the combined intervals obtained using models (14), (21), and (23).

As can be seen from Fig. 3, the expansion of the integrated intervals relative to the initial intervals is on average about 10 %, which can be considered a satisfactory result.

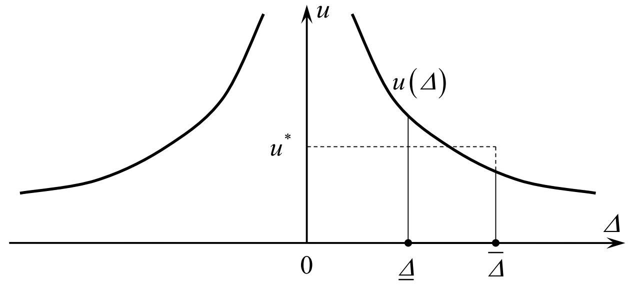

Fig. 2. Graphical interpretation of the measure of the proximity of interval estimation to zero using the preference function $u(\Delta)$

Table 1

Example of adaptive aggregation of interval weighted estimates from five sources

No. of source i	1	2	3	4	5
Estimates $[x_i]$ at time $t=T+1$	[2, 3]	[5, 6]	[6, 8]	3	5
Weight (confidence) α_i	0.6	0.8	0.5	1	0.7
Aggregated estimation $[x^*]$ at moment $t=T+1$, model (14)	[4.08, 4.75]				
Weight (confidence probability) α_i^*, model (14)	0.482				
Forecast estimates $[x_i]$ at time $t=T$	[9, 10]	[7, 8]	[6, 8]	7	6
Weight (confidence probability) α_i point in time $t=T$	0.5	0.9	0.7	0.8	1
Actual value at time $t=T$	8.1				
Absolute interval error Δ_i at time $t=T$	[–1.9, –0.9]	[0.1, 1.1]	[0.1, 2.1]	1.1	2.1
$u(\Delta)=\frac{1}{\lvert\Delta\rvert}$					
Quantitative measure of the proximity of the interval error u_i^*, models (21), (23)	0.75	2.40	1.52	0.91	0.48
Weighting coefficients w_i of the model of complexation (21)	0.103	0.439	0.174	0.208	0.076
Aggregated estimation $[x^*]$ at time $t=T+1$, model (21)	[4.45, 5.34]				
Weight (confidence probability) α_i^*, model (21)	0.643				
Weighting coefficients w_i of the integration model, model (23)	0.038	0.526	0.126	0.197	0.113
Integrated estimation $[x^*]$ at time $t=T+1$, model (23)	[4.619, 5.435]				
Weight (confidence probability) α_i^*, model (23)	0.59				
$u(\Delta)=\frac{1}{\Delta^2}$					
Quantitative measure of the proximity of the interval error u_i^*, models (21), (23)	0.585	9.091	4.762	0.826	0.227
Weighting factors w_i of the aggregation model, model (21)	0.132	0.662	0.217	0.075	0.014
Aggregated estimation $[x^*]$ at time $t=T+1$, model (21)	[4.97, 6.097]				
Weight (confidence probability) α_i^*, model (21)	0.815				
Weighting coefficients w_i of the integration model, model (23)	0.112	0.752	0.149	0.068	0.02
Integrated estimation $[x^*]$ at time $t=T+1$, model (23)	[4.98, 6.041]				
Weight (confidence probability) α_i^*, model (23)	0.767				
$u(\Delta)=\begin{cases}2-\lvert\Delta\rvert, & \lvert\Delta\rvert<2\\ 0, & \lvert\Delta\rvert>2\end{cases}$					
Quantitative measure of the proximity of the interval error u_i^*, models (21), (23)	0.6	1.4	0.95	0.9	0
Weighting coefficients w_i of the model of complexation (21)	0.126	0.392	0.166	0.315	0
Aggregated estimation $[x^*]$ at time $t=T+1$, model (21)	[4.158, 5.009]				
Weight (confidence probability) α_i^*, model (21)	0.615				
Weighting coefficients w_i of the integration model, model (23)	0.05	0.502	0.129	0.319	0
Integrated estimation $[x^*]$ at time $t=T+1$, model (23)	[4.341, 5.151]				
Weight (confidence probability) α_i^*, model (23)	0.585				

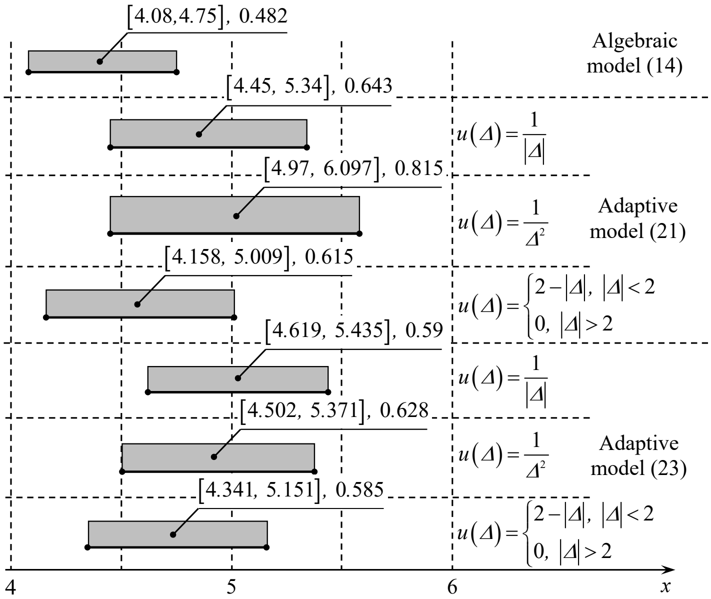

Fig. 3. Graphical interpretation of the results of integration of data from five sources in accordance with the algebraic (14) and adaptive models (21), (23) of integration

6. Discussion of results of the construction of models of adaptive redundant integration of weighted interval data

The main features of the proposed mathematical apparatus, which may be important in its practical use, are as follows:

1. The proposed models of integration are algebraic, built on the paradigm of interval analysis, and in parametrically limiting cases degenerate into known models of integration.

2. The requirements for interval estimates of experts regarding their width do not differ from those in most practical problems of interval analysis. The actual terms of reference for the examination should limit the allowable sizes of the intervals, for example, within 20 % of the average values, which, as a rule, is sufficient for most practical situations. This should prevent the degeneration of the interval model of integration into an analytical one, with complete parametric uncertainty.

3. It should be borne in mind that the integration operation shows an effective result only if the assessments of all experts are not biased, which is extremely difficult to ensure in most practical situations. Nevertheless, the repeated use of the integration procedure makes it possible to ensure a gain in the accuracy of the final estimates.

4. The independence of experts, which is a prerequisite for the correctness of any model of integration, should be provided (or at least assess its degree) by the technology of expert evaluation itself. For this purpose, there are special methods, for example, correlation analysis, which are not the subject of consideration in this seminal paper.

5. When choosing a preference function $u(\Delta)$ in the form of (17) and (18), it should be borne in mind that its use for experts who gave absolutely accurate estimates in the previous assessment will not be informative. The same can be said about interval errors containing zero. In this regard, the form of the preference function (19) seems to be most popular in practical problems. It is devoid of this drawback and has wide possibilities of parametric adjustment.

6. Analytical transformations used in the construction of models do not contradict either the rules of classical interval arithmetic or the postulates of interval analysis. This is a guarantee that there will be no unpleasant effect of artificial expansion of intervals in the model. Therefore, it is not necessary to use non-classical interval arithmetic, for example, Kaucher arithmetic.

7. The proposed models can be included in the methodological support of automated systems for processing expert information.

The advantages of the built models include the following features:

1. The proposed models of integration of interval weighted data are invariant to the type of input data. They allow you to process both interval and point estimates.

2. Adaptive models allow for structural and parametric adjustment by selecting the form and coefficients of the preference function. In a single analytical form, they combine classes of hybrid and selective models.

3. The output parameters of the models have the same set of characteristics as the input parameters, namely, they are weighted intervals. This opens up prospects for their use in cascading models of information processing and compression.

4. Algorithmic and program implementation of the proposed models is not problematic and can be carried out even in a tabular editor. The results of the simulation can be easily presented in graphical form, which provides visibility in the process of processing the results of the examination.

A set of proposed integration models can be included in the circuit of an automated system for processing expert information (Fig. 4).

The practical significance of the built models lies in obtaining an analytically rigorous apparatus for consolidating weighted interval expert assessments. First of all, the built models can be in demand in the tasks of group expert assessment. They can also find applications in automated information compression processes. This is made possible by storing a set of attributes of the output parameters of the model relative to the input parameters. Thus, cascading data collapse becomes possible, which is characterized by a significant reduction in the amount of primary data without loss of relevance.

The variety and volume of data available to modern business analysts objectively indicates that in many practical tasks there may be interval data uncertainty. In particular, in the process of group predictive expert review, data from experts can be obtained in interval form. A feature of this study is the consideration of additional expert information, expressed in the form of weights assigned to interval estimates. To solve the problem of combining such data, an interval analysis apparatus was chosen. It makes it possible to ensure, on the one hand, the mathematical rigor of the model, on the other hand, the uniqueness of solving the problem of integration.

As a result, it was possible to obtain an analytical family of integration models in the form of (21) and (23), which has the property of adaptability. The adaptability of the obtained models is manifested in the fundamental ability to take into account the accuracy of experts at the previous stage of assessment. This is an advantage in situations where the assessment is done repeatedly, which is typical for many practical tasks. Another advantage of the obtained models is their performance in the case of poorly agreed intervals, which is typical for predictive expert evaluation. An important property of the adaptability of the proposed models of integration was achieved by using the preference function of a special kind (17) to (19), shown in Fig. 2. This made it possible to strictly formalize the task of comparing interval assessment errors inherent in experts. Unlike, for example, [30], where the problem of aggregation was solved for agreed weighted intervals, the built models have a number of differences. First, they provide the uniqueness of the parameters of the complex interval, namely the values of its width and weight. Second, they are adaptive.

The use of the preference function also made it possible to provide the possibility of structural and parametric adjustment of the integration models. Recommendations for setting up models are formulated on the basis of the features of analytical functions (17) to (19). They prevent the degeneration of models in some special cases. At the same time, it was possible to achieve commonality in the limit cases of the proposed models with known ones, given, for example, in [20].

Thus, the built models can be included in the arsenal of information technologies of expert analysis (Fig. 4), thereby expanding their methodological and algorithmic base. This makes it possible to expand the range of data with which such technologies can work effectively. The presence of weighted weakly agreed interval estimates among the input data in this case will not require pre-processing. The use of highly specialized models and methods makes it possible to take into account the type of input data. This preserves the objective uncertainty inherent in the data until the very moment of decision-making. Thus, the researcher avoids the risk of unnecessarily simplifying the data model in the early stages of the analysis.

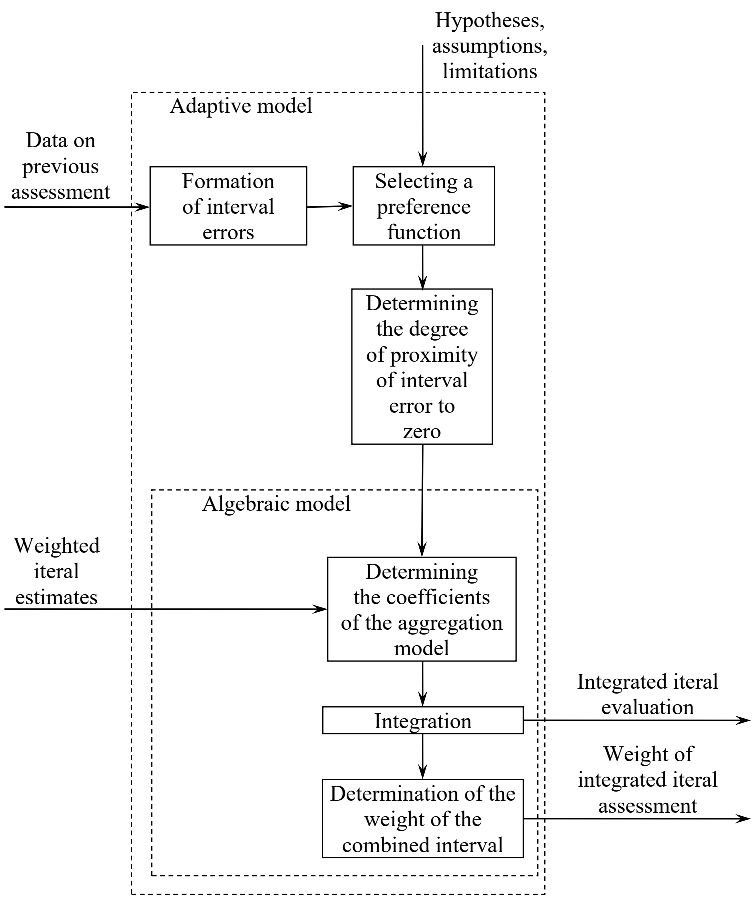

Fig. 4. Process diagram of models for combining weighted interval data in the process of predictive expert evaluation

The main limitation of the proposed models is subjectivity in the process of choosing the type and parameters of the preference function. In addition, the issues of the independence of sources and the non-bias of their assessments remain critical. Nevertheless, they are inherent in all models of integration and are solved by data pre-processing methods.

The disadvantages of the model include a relatively small (about 10 %) expansion of the final integrated intervals relative to the original ones. This can lead to unacceptable spacing in large-dimensional tasks. If, within the framework of the task of predictive expert assessment, it can be assumed that the number of experts does not exceed ten, in other cases it can be measured in hundreds. As this has a direct impact on the accuracy of the final estimates, subsequent seminal papers may investigate this issue. This can serve as a development of the proposed models in terms of estimating the upper limit of adequacy, as well as a whole group of models reported, for example, in [8, 20, 21, 30, 46].

7. Conclusions

1. Within the framework of the task of group predictive expert assessment, the prerequisites for the formalization of expert data in interval form are formulated. The objective uncertainty of primary expert data, formalized in the form of weighted intervals, is proposed to be taken into account in an integrated assessment. Interval analysis was chosen as a tool for solving this problem, taking into account the weak consistency of expert assessments.

2. Models of adaptive integration of weighted interval data have been developed. Their peculiarity, in addition to the analytical form, is the ability to take into account the accuracy of experts at the previous stage of assessment. In addition, a single analytical form of the model makes it possible to process weighted interval and point estimates. The proposed models are analytically rigorous, which ensures the uniqueness of the final integration interval. When solving practical problems, this makes it possible to increase the degree of automation of the processing of expert assessments under conditions of interval uncertainty. An example illustrates the adaptability of the proposed models. It is expressed in the drift of complex assessments towards more accurate ones at the previous stage of expert assessment. The centers of the integration intervals shift by 4–14 % in the direction of more accurate sources. At the same time, a slight, on average about 10 % expansion of the integration interval relative to the original ones was registered.

Conflict of interests

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

References

1. Preis, T., Moat, H. S., Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. Scientific Reports, 3 (1). doi: https://doi.org/10.1038/srep01684
2. Nasseri, A. A., Tucker, A., de Cesare, S. (2015). Quantifying StockTwits semantic terms' trading behavior in financial markets: An effective application of decision tree algorithms. Expert Systems with Applications, 42 (23), 9192–9210. doi: https://doi.org/10.1016/j.eswa.2015.08.008
3. Kolanovic, M., Krishnamachari, R. T. (2017). Big data and AI strategies: Machine learning and alternative data approach to investing. Global Quantitative & Derivatives Strategy. Available at: https://cpb-us-e2.wpmucdn.com/faculty.sites.uci.edu/dist/2/51/files/2018/05/JPM-2017-MachineLearningInvestments.pdf
4. Cohen, L., Malloy, C., Nguyen, Q. (2020). Lazy Prices. The Journal of Finance, 75 (3), 1371–1415. doi: https://doi.org/10.1111/jofi.12885
5. Fleder, M., Shah, D. (2020). Forecasting with Alternative Data. ACM SIGMETRICS Performance Evaluation Review, 48 (1), 23–24. doi: https://doi.org/10.1145/3410048.3410062
6. Charoenwong, B., Kwan, A. (2021). Alternative Data, Big Data, and Applications to Finance. Blockchain Technologies, 35–105. doi: https://doi.org/10.1007/978-981-33-6137-9_2
7. Bellstam, G., Bhagat, S., Cookson, J. A. (2021). A Text-Based Analysis of Corporate Innovation. Management Science, 67 (7), 4004–4031. doi: https://doi.org/10.1287/mnsc.2020.3682
8. Romanenkov, Y., Vartanian, V. (2016). Formation of prognostic software support for strategic decision-making in an organization. Eastern-European Journal of Enterprise Technologies, 2 (9 (80)), 25–34. doi: https://doi.org/10.15587/1729-4061.2016.66306
9. Buravtsev, A. V. (2017). Grey management analysis. Perspectives of Science & Education, 5 (29), 74–79. Available at: https://cyberleninka.ru/article/n/seryy-upravlencheskiy-analiz
10. Chan, J. W. K., Tong, T. K. L. (2007). Multi-criteria material selections and end-of-life product strategy: Grey relational analysis approach. Materials & Design, 28 (5), 1539–1546. doi: https://doi.org/10.1016/j.matdes.2006.02.016
11. Shostak, I. V., Danova, M. A., Romanenkov, Yu. A. (2015). Informatsionnaya tekhnologiya podderzhki prinyatiya ekspertnykh resheniy v natsional'nykh Forsayt-issledovaniyakh. Komunalne hospodarstvo mist, 123, 58–67. Available at: https://www.researchgate.net/publication/315664668_Informacionnaa_tehnologia_podderzki_prinatia_ekspertnyh_resenij_v_nacionalnyh_forsajt-issledovaniah
12. Durrant-Whyte, H., Henderson, T. C. (2008). Multisensor Data Fusion. Springer Handbook of Robotics, 585–610. doi: https://doi.org/10.1007/978-3-540-30301-5_26
13. Hoover, A., Olsen, B. D. (1999). A real-time occupancy map from multiple video streams. Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C). doi: https://doi.org/10.1109/robot.1999.770442
14. Smelyakov, K., Sandrkin, D., Ruban, I., Vitalii, M., Romanenkov, Y. (2018). Search by Image. New Search Engine Service Model. 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). doi: https://doi.org/10.1109/infocommst.2018.8632117
15. Vartanyan, V. M., Cheranovskiy, O. V., Al Daheri Ali Mohamed (2011). Evaluation of basic model robust stability of unmanned aerial vehicles. Aviatsionno-kosmicheskaya tekhnika i tekhnologiya, 6, 44–51. Available at: http://nbuv.gov.ua/UJRN/aktit_2011_6_9
16. Ruban, I., Khudov, H., Khudov, V., Khizhnyak, I., Makoveichuk, O. (2017). Segmentation of the images obtained from onboard optoelectronic surveillance systems by the evolutionary method. Eastern-European Journal of Enterprise Technologies, 5 (9 (89)), 49–57. doi: https://doi.org/10.15587/1729-4061.2017.109904
17. Abdelgawad, A., Bayoumi, M. (2012). Resource-Aware Data Fusion Algorithms for Wireless Sensor Networks. Lecture Notes in Electrical Engineering. doi: https://doi.org/10.1007/978-1-4614-1350-9
18. Khaleghi, B., Khamis, A., Karray, F. O., Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. Information Fusion, 14 (1), 28–44. doi: https://doi.org/10.1016/j.inffus.2011.08.001
19. Sineglazov, V. M., Chumachenko, E. I., Gorbatyuk, V. S. (2012). Metod resheniya zadachi prognozirovaniya na osnove kompleksirovaniya otsenok. Induktyvne modeliuvannia skladnykh system, 4, 214–223.
20. Bidyuk, P. I., Gasanov, A. S., Vavilov, S. E. (2013). Analiz kachestva otsenok prognozov s ispol'zovaniem metoda kompleksirovaniya. Systemni doslidzhennia ta informatsiyni tekhnolohiyi, 4, 7–16. Available at: http://nbuv.gov.ua/UJRN/sdtit_2013_4_3
21. Vasil'ev, A. A. (2015). Obedinenie prognozov ekonomicheskikh pokazateley na osnove bives-otsenki s vesovoy funktsiey Kh'yubera. Aktual'nye problemy gumanitarnykh i estestvennykh nauk, 10-4, 44–47. Available at: https://cyberleninka.ru/article/n/obedinenie-prognozov-ekonomicheskih-pokazateley-na-osnove-bives-otsenki-s-vesovoy-funktsiey-hyubera
22. Molev, M. D., Zanina, I. A., Stuzhenko, N. I. (2013). Sintez prognoznoy informatsii v praktike otsenki ekologo-ekonomicheskogo razvitiya regiona. Inzhenerniy vestnik Dona, 4, 59. Available at: https://cyberleninka.ru/article/n/cintez-prognoznoy-informatsii-v-praktike-otsenki-ekologo-ekonomicheskogo-razvitiya-regiona
23. Romanenkov, Yu. A., Vartanyan, V. M., Revenko, D. S. (2014). Kompleksirovanie prognoznykh otsenok v sisteme monitoringa pokazateley sostoyaniya biznes-protsessa. Systemy upravlinnia, navihatsiyi ta zviazku, 2 (30). Available at: https://www.researchgate.net/publication/315693915_Kompleksirovanie_prognoznyh_ocenok_v_sisteme_monitoringa_pokazatelej_sostoania_biznes-processa
24. Kovalenko, I., Shved, A. (2018). Development of a technology of structuring group expert judgments under various types of uncertainty. Eastern-European Journal of Enterprise Technologies, 3 (4 (93)), 60–68. doi: https://doi.org/10.15587/1729-4061.2018.133299
25. Sternin, M. Yu., Shepelev, G. I. (2009). Sravnenie poliinterval'nykh otsenok v metode OIO. Intelligent Support of Decision Making. International book series «Information science & computing, 83–88. Available at: http://foibg.com/ibs_isc/ibs-10/ibs-10-p11.pdf

26. Shepelev, G., Sternin, M. (2014). Evaluating Expected Effectiveness of Interval Alternatives. International Journal «Information Theories and Applications», 21 (3), 263–274. Available at: http://www.foibg.com/ijita/vol21/ijita21-03-p06.pdf
27. Baohua Li, Yunmin Zhu, Rong Li, X. (2002). Fault-tolerant interval estimation fusion by Dempster-Shafer theory. Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997). doi: https://doi.org/10.1109/icif.2002.1021010
28. Liu, H., Wang, B., Liu, Z., Zhou, Y. (2009). Data Fusion Based on Interval Dempster-Shafer Theory for Emitter Platform Identification. 2009 International Conference on Information Engineering and Computer Science. doi: https://doi.org/10.1109/iciecs.2009.5364227
29. Liang, H., Li, C.-C., Dong, Y., Jiang, Y. (2016). The fusion process of interval opinions based on the dynamic bounded confidence. Information Fusion, 29, 112–119. doi: https://doi.org/10.1016/j.inffus.2015.08.010
30. Zhu, Y., Li, B. (2006). Optimal interval estimation fusion based on sensor interval estimates with confidence degrees. Automatica, 42 (1), 101–108. doi: https://doi.org/10.1016/j.automatica.2005.07.008
31. Ramalingam, S., Maheswari, U. (2016). A fuzzy interval valued fusion technique for multi-modal 3D face recognition. 2016 IEEE International Carnahan Conference on Security Technology (ICCST). doi: https://doi.org/10.1109/ccst.2016.7815709
32. Carlson, R. J. (2013). Voter compatibility in interval societies. HMC Senior Theses. Available at: https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1044&context=hmc_theses
33. Collett, M. A., Cox, M. G., Esward, T. J., Harris, P. M., Sousa, J. A. (2007). Aggregating measurement data influenced by common effects. Metrologia, 44 (5), 308–318. doi: https://doi.org/10.1088/0026-1394/44/5/007
34. Duta, M., Henry, M. (2005). The fusion of redundant SEVA measurements. IEEE Transactions on Control Systems Technology, 13 (2), 173–184. doi: https://doi.org/10.1109/tcst.2004.840448
35. Dopazo, E., Martínez-Céspedes, M. L. (2017). Rank aggregation methods dealing with ordinal uncertain preferences. Expert Systems with Applications, 78, 103–109. doi: https://doi.org/10.1016/j.eswa.2017.01.051
36. Voschinin, A. P. (2002). Interval'niy analiz dannykh: razvitie i perspektivy. Zavodskaya Laboratoriya, 68 (1), 118–126.
37. Orlov, A. I. (2013). Osnovnye idei statistiki interval'nykh dannykh. Nauchniy zhurnal KubGAU, 94, 55–70. Available at: https://cyberleninka.ru/article/n/osnovnye-idei-statistiki-intervalnyh-dannyh
38. Alefel'd, G., Khertsberger, Yu. (1987). Vvedenie v interval'nye vychisleniya. Moscow: Mir, 360.
39. Kalmykov, S. A., Shokin, Yu. I., Yuldashev, Z. Kh. (1986). Metody interval'nogo analiza. Novosibirsk: Nauka, 223. Available at: https://hi.b-ok.xyz/book/445084/5ae6b7
40. Shokin, Yu. I. (1981). Interval'niy analiz. Novosibirsk: Nauka, 112. Available at: https://read.in.ua/book156021/
41. Shariy, S. P. (2016). Konechnomerniy interval'niy analiz. Novosibirsk: XYZ, 606.
42. Hansen, E. R. (1993). Computing zeros of functions using generalized interval arithmetic. Interval Computations, 3, 3–27. Available at: https://interval.louisiana.edu/reliable-computing-journal/1993/interval-computations-1993-3-pp-003-028.pdf
43. Moore, R. E. (1979). Methods and Applications of Interval Analysis. Philadelphia: SIAM, 125. doi: https://doi.org/10.1137/1.9781611970906
44. Kaucher, E. (1980). Interval analysis in the extended interval spase IR. Computing Suppl, 2, 33–49. Available at: http://www.math.bas.bg/~epopova/Kaucher-80-CS_33-49.pdf
45. Bardachev, Yu. N., Kryuchkovskiy, V. V., Malomuzh, T. V. (2010). Metodologicheskaya predpochtitel'nost' interval'nykh ekspertnykh otsenok pri prinyatii resheniy v usloviyakh neopredelennosti. Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina. Seriya: Matematychne modeliuvannia. Informatsiyni tekhnolohiyi. Avtomatyzovani systemy upravlinnia, 890, 18–28. Available at: http://mia.univer.kharkov.ua/13/30045.pdf
46. Romanenkov, Y., Danova, M., Kashcheyeva, V., Bugaienko, O., Volk, M., Karminska-Bielobrova, M., Lobach, O. (2018). Complexification methods of interval forecast estimates in the problems on shortterm prediction. Eastern-European Journal of Enterprise Technologies, 3 (3 (93)), 50–58. doi: https://doi.org/10.15587/1729-4061.2018.131939
47. Podruzhko, A. A., Podruzhko, A. S., Kiritsev, P. N. (2009). Interval'nye metody resheniya zadach kalibrovki i klassifikatsii. Trudy Instituta sistemnogo analiza Rossiyskoy Akademii nauk. Dinamika neodnorodnykh sistem, 44, 173–186.
48. Kochkarov, R. A. (2015). Interval'nye zadachi na predfraktal'nykh grafakh. Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh, 18, 255–264. Available at: https://cyberleninka.ru/article/n/intervalnye-zadachi-na-predfraktalnyh-grafah
49. Marzullo, K. (1990). Tolerating failures of continuous-valued sensors. ACM Transactions on Computer Systems, 8 (4), 284–304. doi: https://doi.org/10.1145/128733.128735
50. Muravyov, S. V., Khudonogova, L. I., Emelyanova, E. Y. (2018). Interval data fusion with preference aggregation. Measurement, 116, 621–630. doi: https://doi.org/10.1016/j.measurement.2017.08.045
51. Khudonogova, L. I., Muravyov, S. V. (2021). Interval Data Fusion with Preference Aggregation for Balancing Measurement Accuracy and Energy Consumption in WSN. Wireless Personal Communications, 118 (4), 2399–2421. doi: https://doi.org/10.1007/s11277-021-08132-9
52. Levin, V. I. (2004). Uporyadochenie intervalov i zadachi optimizatsii s interval'nymi parametrami. Kibernetika i sistemnyy analiz, 3, 14–24. Available at: http://ec.lib.vntu.edu.ua/DocDescription?doc_id=112446
53. Levin, V. I. (2002). Intervals comparison and optimisation under the conditions of uncertainty. Vestnik Tambovskogo universiteta. Seriya: Estestvennye i tekhnicheskie nauki, 7 (3), 383–389. Available at: https://cyberleninka.ru/article/n/sravnenie-intervalov-i-optimizatsiya-v-usloviyah-neopredelennosti
54. Romanenkov, Yu., Kosenko, V., Lobach, O., Grinchenko, E., Grinchenko, M. (2019). The Method for Ranking Quasi-Optimal Alternatives in Interval Game Models Against Nature. Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2019). Available at: https://www.researchgate.net/publication/333488747_The_Method_for_Ranking_Quasi-Optimal_Alternatives_in_Interval_Game_Models_Against_Nature