УДК 519.6
DOI 10.34822/1999-7604-2020-1-61-68

АЛГОРИТМЫ ПОДГОТОВКИ ДАННЫХ К ПАРАМЕТРИЧЕСКОМУ СИНТЕЗУ ПРИ ПРЕДВАРИТЕЛЬНОМ ПРОЕКТИРОВАНИИ ТЕХНИЧЕСКИХ ОБЪЕКТОВ НА ОСНОВЕ ОБРАБОТКИ ЭКСПЕРТНОЙ ИНФОРМАЦИИ

Г. С. Вересников
Институт проблем управления им. В. А. Трапезникова
Российской академии наук, Москва, Россия
E-mail: veresnikov@mail.ru

В статье рассмотрены вопросы подготовки экспертных данных к параметрическому синтезу при предварительном проектировании технических объектов в условиях неопределенности параметров. В рамках теории неопределенности Б. Лю разработаны алгоритмы формирования критических вариантов функций распределения неопределенности, отражающих достижение «наилучших» и «наихудших» для лица, принимающего решения (далее – ЛПР), числовых характеристик функций, зависящих от неопределенных параметров. Предложено выполнять формирование функций распределения неопределенности, обеспечивающих компромисс между мнениями экспертов, в два этапа, включающих группировку неопределенных величин, являющихся отражением мнений отдельных экспертов, и формирование компромиссных функций распределения неопределенности для каждой выделенной группы экспертов. Разработан алгоритм кластеризации неопределенных величин на фиксированное количество классов, который основан на методе *k*-средних (*k*-means). Предложен интерактивный алгоритм группировки неопределенных величин. В рамках этого алгоритма неопределенные величины выступают в качестве узлов графа, а пороговое значение, задаваемое ЛПР, и расстояние между неопределенными величинами определяют наличие или отсутствие связи между этими узлами. Посредством использования предложенных алгоритмов обработки экспертных данных формируются функции распределения неопределенности, которые являются входной информацией для алгоритмов вычисления числовых характеристик функций, зависящих от недетерминированных параметров.

Ключевые слова: эпистемическая неопределенность, оптимизационная модель, неопределенный параметр, параметрический синтез.

ALGORITHMS OF DATA PREPARATION FOR PARAMETRIC SYNTHESIS IN PRELIMINARY DESIGN OF THE TECHNICAL OBJECTS BASED ON EXPERT INFORMATION PROCESSING

G. S. Veresnikov
V. A. Trapeznikov Institute of Control Sciences
of Russian Academy of Sciences, Moscow, Russia
E-mail: veresnikov@mail.ru

The article describes the issues of preparation of the expert information for parametric synthesis at the preliminary design of the technical objects under uncertain parameters. Within B. Liu's uncertainty theory, the algorithms for generating critical variants of the uncertainty distribution functions reflecting the achievement of the “best” and “worst” for a decision-maker (DM) of the numerical characteristics of functions that depend on the uncertain parameters are developed. It is proposed to carry out the formation of uncertainty distribution functions providing a compromise between the expert views, in two stages including the grouping of uncertain values reflecting the views of some experts, and establishment of the compromise functions of uncertainty distribution for each selected expert group. The algorithm of uncertain values clustering on the fixed number of classes based on the k-means method is developed. An interactive algorithm of uncertain values grouping is suggested. In terms of this algorithm, uncertain values act as graph nodes, whereas the threshold value determined by DM and the distance between uncertain values define the presence or absence of connection between these nodes. Using the proposed expert information processing algorithms, the uncertainty distribution functions are formed, which are the input to algorithms for calculating the numerical characteristics of functions that depend on non-deterministic parameters.

Keywords: epistemic uncertainty, optimization model, uncertain parameter, parametric synthesis.

Введение. Параметрический синтез при предварительном проектировании технических объектов (далее – ТО) представляет собой процесс определения (синтеза) параметров в условиях недостатка исходной информации. Параметр ТО является детерминированным, если может быть представлен точным значением – детерминированной величиной. Параметр является недетерминированным вследствие отсутствия знаний о его точном значении.

Алеаторная (объективная) и эпистемическая (субъективная) неопределенности – два типа неопределенности, отражающие недетерминированность параметров ТО [1]. Если информация о случайных (стохастических) параметрах содержится в статистических данных, то имеет место алеаторная неопределенность. В этом случае недетерминированные параметры представляются случайными величинами с известными функциями распределения вероятности. Если информация о параметрах ТО формируется с участием экспертов, то имеет место эпистемическая неопределенность, при этом для детерминированных параметров могут быть неизвестны точные значения на момент принятия проектных решений или для случайных параметров недостаточно статистических данных. В этом случае параметры ТО моделируются эпистемическими величинами.

При решении задач параметрического синтеза на основе обработки экспертных данных с использованием теории неопределенности [2] параметры с эпистемической неопределенностью интерпретируются как неопределенные параметры с функциями распределения неопределенности.

Функция распределения неопределенности неопределенного параметра ξ есть функция $\Phi\colon R \to [0,1]$, определяемая как:

$$\Phi(x) = M\{\xi \le x\},$$

где x – детерминированное значение;

$M\{\bullet\}$ – мера неопределенности, определяемая в теории неопределенности как степень уверенности эксперта, что событие $\bullet$ произойдет.

Функция $f(\bar{\xi})$, зависящая от вектора неопределенных величин $\bar{\xi}$, является неопределенной величиной, которая не может использоваться для решения оптимизационных задач и принятия решений ЛПР.

Для решения этой проблемы в теории неопределенности выведены аналитические выражения [3]:

- для ожидаемого значения $E[f(\bar{\xi})]$;

- квантилей неопределенных величин $inf_{\alpha}[f(\bar{\xi})]$ и $sup_{\alpha}[f(\bar{\xi})]$, где $inf_{\alpha}[f(\bar{\xi})]$ = $= inf\{r \mid M\{f(\bar{\xi}) \le r\} \ge \alpha\}$, $sup_{\alpha}[f(\bar{\xi})] = sup\{r \mid M\{f(\bar{\xi}) \ge r\} \ge \alpha\}$, где α – уровень меры неопределенности (степени уверенности эксперта);

- дисперсии $V[f(\bar{\xi})]$.

В этих аналитических выражениях, позволяющих обеспечить высокую вычислительную эффективность решения задач параметрического синтеза проектных решений, используются обратные функции распределения неопределенности, которые должны быть получены в результате сбора и обработки экспертных данных.

Функции распределения неопределенности для неопределенных параметров эксперты задают непосредственно в аналитическом виде или набором эмпирических данных, которые затем аппроксимируются функцией распределения неопределенности.

При сборе эмпирической информации о ТО эксперты для каждого неопределенного входного параметра ξ_q указывают:

- область определения функции распределения неопределенности $\varPhi_{\xi_q}$;

- отдельные значения функции распределения неопределенности, сопоставленные значениям, которые может принимать неопределенная величина, соответствующая данному неопределенному параметру.

В процессе опроса экспертов для детерминированных значений $x_1, x_2, \ldots, x_n$ неопределенного параметра ξ_q эксперт определяет степени уверенности, равные $\alpha_1, \alpha_2, \ldots, \alpha_n$, $i = 1, \ldots, n$, в том, что $\xi_q \le x_1$, $\xi_q \le x_2$, …, $\xi_q \le x_n$. В результате формируются эмпирические данные $(x_1, \alpha_1), (x_2, \alpha_2), \ldots, (x_n, \alpha_n)$, где $x_1 < x_2 < \ldots < x_n$, $0 \le a_1 \le a_2 \le \ldots \le a_n \le 1$.

Для формирования функций распределения неопределенности в аналитической форме на основе данных, полученных от одного эксперта, используются известные алгоритмы аппроксимации. В самом простом случае может применяться линейная аппроксимация функции распределения неопределенного параметра ξ_q, которая дает эмпирическую функцию распределения [2]:

$$\varPhi_{\xi_q}(x) = \begin{cases} 0, \text{ если } x \le x_1, \\ \alpha_i + \dfrac{(\alpha_{i+1} - \alpha_i)(x - x_i)}{x_{i+1} - x_i}, \text{ если } x_i \le x \le x_{i+1}, 1 \le i \le n, \\ 1, \text{ если } x > x_n\ . \end{cases} \tag{1}$$

Если ЛПР не устраивает низкая точность линейной аппроксимации, то могут использоваться различные модели нелинейной регрессии, полиномы высоких порядков.

Задача параметрического синтеза при предварительном проектировании ТО нередко осложняется тем, что при получении информации об отдельных недетерминированных параметрах используются мнения группы экспертов. В связи с этим требуется решение проблемы подготовки данных к параметрическому синтезу, связанной с анализом и интеграцией экспертной информации об отдельных неопределенных параметрах ТО.

Если эмпирические данные о неопределенном параметре получены от группы экспертов, то формируется множество пар $(x_{i1}, a_{i1}), (x_{i2}, a_{i2}), \ldots, (x_{in_i}, a_{in_i})$, где $0 \le x_{i1} \le x_{i2} \le \ldots \le x_{in_i} \le 1$, $0 \le a_{i1} \le a_{i2} \le \ldots \le \alpha_{in_i} \le 1$; $i = 1, \ldots, m$, m – количество экспертов, n_i – количество значений неопределенного параметра ξ_q, рассматриваемых i-м экспертом.

В статье предлагаются алгоритмы, позволяющие обеспечить обработку данных о неопределенных параметрах, полученных от группы экспертов, что обусловлено необходимостью:

- формирования критических – оптимистических и пессимистических – вариантов функции распределения неопределенности;

- формирования функций распределения неопределенности, обеспечивающих компромисс между мнениями экспертов.

Алгоритм формирования критических вариантов функций распределения неопределенности. Критические варианты функций распределения неопределенности определяются для каждого неопределенного параметра ξ_q и отражают достижение «наилучших» и «наихудших» для ЛПР числовых характеристик функций от неопределенных параметров.

Разработан следующий алгоритм формирования критических вариантов функций распределения неопределенности:

Алгоритм 1:

Шаг 1. Создается упорядоченное множество, которое формируется из детерминированных значений x_k, $k = 1, .., n$, неопределенного параметра ξ_q.

$$x_k = \min_i x_{i1} + (k-1)\frac{(\max\limits_i x_{in_i} - \min\limits_i x_{i1})}{(n-1)},$$

где n задается проектировщиком ТО.

Шаг 2. Для каждого эксперта i, $i = 1, .., m$ на основе заданных им пар (x_{i1}, a_{i1}), (x_{i2}, a_{i2}), …, (x_{in_i}, a_{in_i}) находится $\Phi_{\xi_q i}(x)$ − аппроксимация функции распределения неопределенности.

Шаг 3. Формируются два упорядоченных множества пар $A_{q\min}$ и $A_{q\max}$, состоящие из пар $(x_k, \alpha_{k\min})$ и $(x_k, \alpha_{k\max})$, $k = 1, .., n$. Затем на их основе выполняется аппроксимация функций распределения неопределенности, т. е. для каждого неопределенного параметра ξ_q будет получено две функции − $\Phi_{\xi_q \min}(x)$ и $\Phi_{\xi_q \max}(x)$.

Значения $\alpha_{k\min}$ и $\alpha_{k\max}$ для каждого k ищутся в интервале $[\min\limits_i \Phi_{\xi_q i}(x_k),\ \max\limits_i \Phi_{\xi_q i}(x_k)]$ таким образом, чтобы при построении функций распределения неопределенности на основе $A_{q\min}$ и $A_{q\max}$ используемая для решения задачи синтеза параметров ТО числовая характеристика функции от неопределенных параметров достигала своих минимальных и максимальных значений. При этом должно выполняться условие, что если $x_1 \le x_2 \le \ldots \le x_n$, то $\alpha_{1\min} \le \alpha_{2\min} \le \ldots \le \alpha_{n\min}$ и $\alpha_{1\max} \le \alpha_{2\max} \le \ldots \le \alpha_{n\max}$.

Множества пар $A_{q\min}$ и $A_{q\max}$, используемые при формировании критических вариантов функций распределения неопределенности для числовых характеристик $E[f(\bar{\xi})]$, $inf_\alpha[f(\bar{\xi})]$, $sup_\alpha[f(\bar{\xi})]$ при условии монотонности функции $f(\bar{\xi})$ по неопределенным параметрам, могут быть построены более простым способом.

Если $f(\bar{\xi})$ возрастает по неопределенному параметру ξ_q для $E[f(\bar{\xi})]$, $inf_\alpha[f(\bar{\xi})]$ или убывает для $sup_\alpha[f(\bar{\xi})]$:

$$A_{q\min} = \{(x_k, \alpha_{k\min}) \mid \alpha_{k\min} = \max_i[\Phi_{\xi_q i}(x_k)],\ k = 1 \ldots n,\ i = 1, \ldots, m\},$$
$$A_{q\max} = \{(x_k, \alpha_{k\max}) \mid \alpha_{k\max} = \min_i[\Phi_{\xi_q i}(x_k)],\ k = 1 \ldots n,\ i = 1, \ldots, m\}.$$

Если $f(\bar{\xi})$ убывает по неопределенному параметру ξ_q для $E[f(\bar{\xi})]$, $inf_\alpha[f(\bar{\xi})]$ или возрастает для $sup_\alpha[f(\bar{\xi})]$:

$$A_{q\min} = \{(x_k, \alpha_{k\min}) \mid \alpha_{k\min} = \min_i[\Phi_{\xi_q i}(x_k)],\ k = 1 \ldots n,\ i = 1, \ldots, m\},$$
$$A_{q\max} = \{(x_k, \alpha_{k\max}) \mid \alpha_{k\max} = \max_i[\Phi_{\xi_q i}(x_k)],\ k = 1 \ldots n,\ i = 1, \ldots, m\}.$$

Такой подход является обоснованным вследствие основного свойства монотонности определенного интеграла и условия монотонности вычисляемой функции по неопределенным параметрам.

Конец алгоритма 1.

Полученные в результате аппроксимации функции $\Phi_{\xi_q \min}(x)$ и $\Phi_{\xi_q \max}(x)$ используются при расчете числовых характеристик функции от неопределенных параметров. Критические варианты функций распределения неопределенности $\Phi_{\xi_q \min}(x)$ и $\Phi_{\xi_q \max}(x)$ являются пессимистическими или оптимистическими в зависимости от вида вычисляемых на их основе числовых характеристик и соответствующих им субъективных оценок – предпочтений ЛПР.

Алгоритм формирования функций распределения неопределенности, обеспечивающих компромисс между мнениями экспертов. Для поиска функций распределения неопределенности, обеспечивающих компромисс между мнениями экспертов, в работах [4, 5] предлагается применять подход, основанный на методе Дельфи. Его суть заключается в том, что для каждого неопределенного параметра ξ_q методом наименьших квадратов производится аппроксимация функции распределения неопределенности на основе данных, полученных от всех экспертов группы. Затем полученные в результате аппроксимации функции используются в работе с экспертами для итерационного сближения их позиций. Конечный вариант функции распределения неопределенности считается компромиссным и используется для решения задачи синтеза параметров ТО.

Такой упрощенный подход к определению компромиссных функций распределения неопределенности не всегда подходит при решении задач синтеза параметров ТО, например, в случае, когда мнения экспертов могут значительно отличаться, но при этом объединяться в несколько однородных групп.

Предлагается формирование функций распределения неопределенности, обеспечивающих компромисс между мнениями экспертов, проводить в два этапа, включающих:

- группировку неопределенных величин, являющихся отражением мнений отдельных экспертов;
- формирование компромиссных функций распределения неопределенности для каждой выделенной группы экспертов.

Пусть каждый эксперт *i* задал неопределенную величину $\xi_i, i = 1, ..., m,$ соответствующую некоторому неопределенному параметру. Тогда группировка данных о неопределенных параметрах может выполняться с использованием алгоритмов, основанных на известном способе вычисления расстояния между неопределенными величинами из теории неопределенности [2]:

$$d(\xi_i, \xi_j) = E[|\xi_i - \xi_j|],$$
$$d(\xi_i, \xi_j) = \int_0^1 \left|\Phi_{\xi_i}^{-1}(\alpha) - \Phi_{\xi_j}^{-1}(1-\alpha)\right| d\alpha, \quad (2)$$

где ξ_i и ξ_j – неопределенные величины.

Разработан алгоритм кластеризации неопределенных величин $\xi_i, i = 1 ... m$ на фиксированное количество классов $C_1, C_2, ..., C_N$, который основан на идеях метода *k*-means. В качестве центров кластеров используются неопределенные величины.

Алгоритм 2:

Шаг 1. Выбор ЛПР количества кластеров *N*.

Шаг 2. Выбор начальных центров кластеров. Центром каждого кластера является неопределенная величина μ_j, $j = 1 ... N,$ заданная функцией распределения неопределенности $\Phi_{\mu_j}(x)$.

Для каждой неопределенной величины μ_j, являющейся центром кластера C_j, $j = 1 ... N$, выбирается вид функции и вектор коэффициентов $p_j = (p_{j1}, p_{j2}, ..., p_{jn})$, который задает поведение данной функции в области ее определения.

Если выбирается линейная аппроксимация функции распределения неопределенности (1), то $p_j = (\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jn})$, где n – количество точек, для которых определяются значения функции распределения неопределенности $\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jn}$ в зафиксированных значениях $x_{j1},\ x_{j2}, ..., x_{jn}$ неопределенного параметра.

В качестве центра кластера могут быть использованы нормальная, логнормальная функции распределения неопределенности, полиномы высоких порядков.

Шаг 3. Распределение неопределенных величин $\xi_i,\ i = 1...m$ по кластерам, которое выполняется следующим образом:

$$\forall \xi_i, i = 1...m : \xi_i \in C_j \Leftrightarrow j = \arg\min_k d(\mu_k, \xi_i),$$

где $d(\mu_k, \xi_i) = \int_0^1 \left| \Phi_{\mu_k}^{-1}(\alpha) - \Phi_{\xi_i}^{-1}(1-\alpha) \right| d\alpha$ (следует из (2)).

Шаг 4. Изменение центров кластеров. Определяются функции распределения неопределенности величин $\mu_j,\ j = 1...N,$ которые становятся центрами кластеров. Для этого определяются коэффициенты $(p_{j1}, p_{j2}, ..., p_{jn})$ на основе всех экспертных данных $\xi_i \in C_j$ с использованием метода наименьших квадратов.

Шаг 5. Проверка условия завершения алгоритма. Алгоритм прекращает свою работу на итерации номер t, если $\forall j = 1... N : d(\mu_j^t, \mu_j^{t-1}) = 0$, иначе переход к шагу 3.

Данное условие обеспечивает прекращение работы алгоритма, если не происходит изменения центров кластеров.

Конец алгоритма 2.

Для формирования кластеров также разработан интерактивный алгоритм 3, позволяющий структурировать экспертные данные на основе информации о расстоянии между неопределенными величинами. Этот алгоритм основан на методе «корреляционных плеяд» [6], в котором в качестве меры близости между анализируемыми объектами вместо коэффициента корреляции выступает расстояние между неопределенными величинами. Аналогичным образом можно использовать другие методы, позволяющие объединить параметры по наличию схожих зависимостей, например «вроцлавскую таксономию». Данный метод также позволяет построить связный граф, но обладает меньшими возможностями для визуализации близости неопределенных величин.

Алгоритм 3:

Шаг 1. Строится матрица расстояний между неопределенными величинами, каждая ячейка которой на пересечении i-го столбца и j-ой строки содержит вычисленное по формуле (2) расстояние между неопределенными величинами ξ_i и ξ_j (табл.).

Таблица

Матрица расстояний между неопределенными величинами

	ξ_1	ξ_2	…	ξ_i	…	ξ_m
ξ_1	$d(\xi_1, \xi_1)$	$d(\xi_1, \xi_2)$	…	$d(\xi_1, \xi_i)$	…	$d(\xi_1, \xi_m)$
ξ_2	$d(\xi_2, \xi_1)$	$d(\xi_2, \xi_2)$	…	$d(\xi_2, \xi_i)$	…	$d(\xi_2, \xi_m)$
…	…	…	…	…	…	…
ξ_j	$d(\xi_j, \xi_1)$	$d(\xi_j, \xi_2)$	…	$d(\xi_j, \xi_i)$	…	$d(\xi_j, \xi_m)$
…	…	…	…	…	…	…
ξ_m	$d(\xi_m, \xi_1)$	$d(\xi_m, \xi_2)$	…	$d(\xi_m, \xi_i)$	…	$d(\xi_m, \xi_m)$

Примечание: авторская разработка.

Шаг 2. ЛПР задает пороговое значение от $\min_{i,j}(d(\xi_i, \xi_j))$ до $\max_{i,j}(d(\xi_i, \xi_j))$, и те неопределенные величины, расстояние между которыми не превышает заданный порог, объединяются в связный граф. Фактически неопределенные величины выступают в качестве узлов, а расстояние между ними определяет наличие или отсутствие связи. Посредством постепенного увеличения порогового значения строится последовательность графов (плеяд) (рис.).

По мере увеличения порогового значения количество связей между узлами будет увеличиваться, что позволяет ЛПР принять решение о группах неопределенных величин.

Конец алгоритма 3.

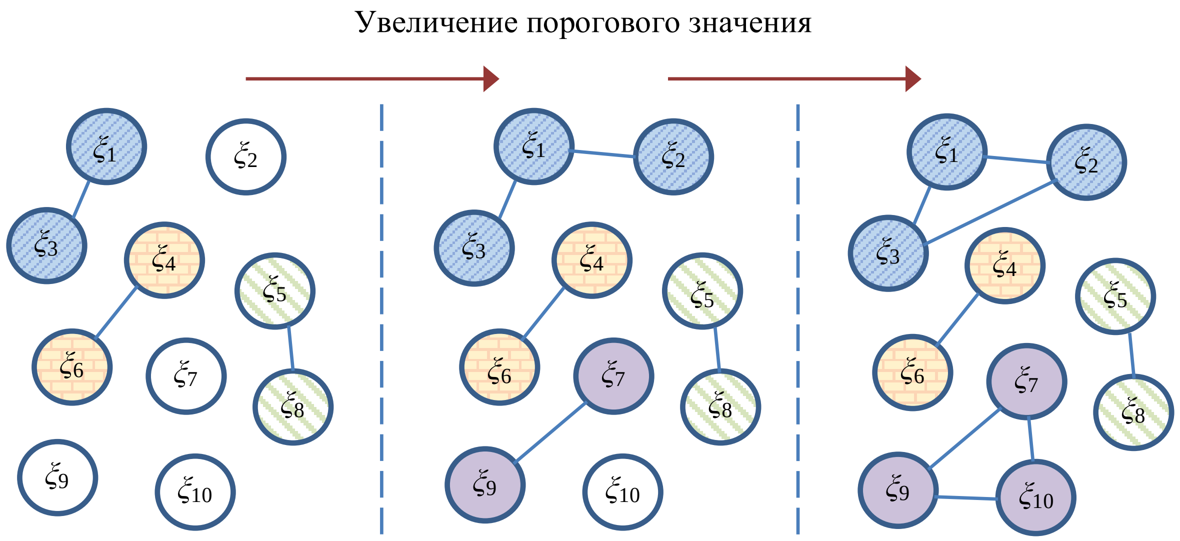

Рисунок. Пример последовательности графов, формируемых при увеличении порогового значения для расстояния между неопределенными величинами

Примечание: авторская разработка.

Применяя предлагаемые алгоритмы, ЛПР может сформировать группы неопределенных величин. Для каждой сформированной группы строятся компромиссные функции распределения неопределенности, которые являются входной информацией для алгоритмов расчета числовых характеристик функций от неопределенных параметров.

Заключение. Разработаны алгоритмы обработки экспертной информации о неопределенных параметрах. С использованием этих алгоритмов эмпирические данные о неопределенных параметрах, полученные от группы экспертов, представляются в форме, необходимой для применения алгоритмов расчета числовых характеристик функций от неопределенных параметров – функции распределения неопределенности. Это позволяет подготовить данные к параметрическому синтезу проектных решений с использованием теории неопределенности [2] и, в частности, разработанных на ее основе оптимизационных моделей [7, 8].

Литература

1. Der Kiureghian A. Aleatory or epistemic? Does it matter? // Special Workshop on Risk Acceptance and Risk Communication, Stanford University, 2007. 13 p.
2. Liu B. Uncertainty Theory. 4th edition. Berlin : Springer-Verlag, 2015. 487 p.
3. Liu B. Theory and Practice of Uncertain Programming. 3rd ed. Berlin : Springer-Verlag, 2009. 201 p.
4. Jinwu G. Delphi Method for Estimating Uncertainty Distributions // Proceedings of the First International Conference on Uncertainty Theory, Urumchi, China. 2010. P. 291–297.

5. Haiying G., Xiaosheng W. Delphi Method for Estimating Membership Function of Uncertain Set // Journal of Uncertainty Analysis and Applications. 2016. Vol. 4 (3). DOI https://doi.org/10.1186/s40467-016-0044-1

6. Архипова М. Ю., Сиротин В. П., Мхитарян В. С., Дуброва Т. А., Миронкина Ю. Н. Анализ данных. 1-ое изд. М. : Юрайт, 2016. 490 с.

7. Veresnikov G. S., Pankova L. A., Pronina V. A. Models of Uncertain-random Programming // Advances in Systems Science and Applications. 2019. Vol. 19, No. 2. P. 8–22.

8. Veresnikov G. S., Pankova L. A., Pronina V. A. Preliminary Design with the Epistemic Uncertainty of Parameters // Advances in Systems Science and Applications. 2018. Vol. 18, No. 3. P. 154–164.