УДК 004.8 DOI: 10.36871/2618-9976.2023.08.001

МЕТОДОЛОГИЯ СОЗДАНИЯ "ДОВЕРИТЕЛЬНОГО" ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СИСТЕМАХ НА ОСНОВЕ РЕГУЛЯРИЗИРУЮЩЕГО БАЙЕСОВСКОГО ПОДХОДА

Светлана Васильевна Прокопчина[1]

[1] *Доктор технических наук, профессор, Финансовый университет при Правительстве Российской Федерации, Москва, Россия, e-mail: svprokopchina@mail.ru*

ИНФОРМАЦИЯ

Ключевые слова:
искусственный интеллект
метрология
регуляризирующий байесовский подход

АННОТАЦИЯ

В статье предлагается методология создания системы оценки качества решений искусственного интеллекта (ИИ), в частности оценки достоверности решений, определяющей степень доверия к получаемым решениям. Методология построена на основе регуляризирующего байесовского подхода и реализована в виде системы метрологического сопровождения решений ИИ. В рамках этой методологии и технологий метрологического сопровождения предложены комплексы метрологических характеристик, определяющих точность, надежность (уровень ошибок 1-го и 2-го рода), достоверность, риск, энтропию, количество информации для каждого этапа алгоритмов, что обеспечивает прослеживаемость и прозрачность получаемых решений. Приведены практические примеры определения метрологических комплексов для решений прикладных систем искусственного интеллекта.

ДЛЯ ЦИТИРОВАНИЯ: *Прокопчина С.В.* Методология создания "доверительного" искусственного интеллекта в системах на основе регуляризирующего байесовского подхода // Мягкие измерения и вычисления. 2023. № 8. Т. 69. С. 5–21; https://doi.org/10.36871/2618-9976.2023.08.001.

METHODOLOGY FOR CREATING "TRUSTWORTHINESS" ARTIFICIAL INTELLIGENCE IN SYSTEMS BASED ON THE REGULARIZING BAYESIAN APPROACH

Svetlana V. Prokopchina[1]

[1] *Doctor of Technical Sciences, Professor, Financial University under the Government Russian Federation, Moscow, Russia, e-mail: svprokopchina@mail.ru*

ARTICLE INFO

Keywords:
Artificial intelligence
Metrology
Regularizing Bayesian approach

ABSTRACT

The article proposes a methodology for creating a system for evaluating the quality of artificial intelligence (AI) solutions, in particular, evaluating the trustworthiness of solutions, which determines the degree of confidence in the solutions obtained. The methodology is based on the regularizing Bayesian approach and implemented as a system of metrological support of AI solutions. Within the framework of this methodology and technologies of metrological support, complexes of metrological characteristics are proposed that determine accuracy, reliability (level of errors of the 1st and 2nd kind), reliability, risk, entropy, amount of information for each stage of the algorithms, which ensures traceability and transparency of the solutions obtained. Practical examples of determining metrological complexes for solving applied artificial intelligence systems are given.

FOR CITATION: *Prokopchina S.V.* (2023) Methodology for creating "trustworthiness" artificial intelligence in systems based on the regularizing bayesian approach. *Soft measurements and computing,* vol. 69, no. 8, pp. 5–21 (In Russ.); https://doi.org/10.36871/2618-9976.2023.08.001.

Введение

Понятие доверительности (англ. *trustworthiness*) решений искусственного интеллекта пришло из английского языка, где оно обозначает в переводе высокую степень надежности или достоверности решений.

Оно неразрывно связано с понятиями прозрачности (*transparency*) и прослеживаемости (*traceability*) решений. Действительно, когда мы, следуя алгоритму решения задачи, можем убедиться и убеждаемся в его правильности, правильности последовательных вычислений и решений, получаемых на каждом этапе алгоритма, мы доверяем получаемым решениям, считаем их надежными.

Однако многие современные системы искусственного интеллекта этими качествами не обладают. Международное научное сообщество ищет пути создания систем искусственного интеллекта с указанными свойствами, а практики ждут их появления для решения прикладных задач, о чем свидетельствуют многочисленные публикации и доклады на различных конференциях.

Но если посмотреть на эти требования с позиций измерительного подхода, то легко понять, что эти требования являются требованиями оценки качества получаемых решений. В теории и практике измерений все они давно и хорошо известны.

Решаются задачи оценки указанных свойств на основании метрологического подхода. Потому задача оценки сводится к созданию методологии определения метрологических характеристик решений искусственного интеллекта.

Далее в статье покажем, как это реализуется в интеллектуальных измерительных системах на основе регуляризирующего байесовского подхода (РБП) [6].

1. Метрологическое обоснование решений РБП

В работе [8] показано, что интеллектуальные прикладные системы, построенные на основе методологии РБП, являются сверточными байесовскими нейронными сетями. Поэтому методология и технологии оценки качества решений РБП и определенные показатели качества, рассмотренные в настоящей статье, можно отнести к решениям ИИ и применить для других методов ИИ.

Методы и алгоритмы РБП основаны на метризации и шкалировании пространства решений, что дает возможность разработки их метрологического обоснования. Результатом метрологического обоснования решений РБП является комплекс показателей метрологических характеристик качества полученных решений. К числу основных вопросов, связанных с формированием такого комплекса, относятся:

– выбор вида метрологических характеристик методов, алгоритмов и результатов, учитывающих их специфику и обеспечивающих наиболее полную оценку всех составляющих погрешности решений и ее интегральную оценку;

– определение метрологических свойств получаемых результатов и алгоритмов на основе созданного комплекса метрологических характеристик;

– функциональные преобразования метрологических комплексов в соответствии с алгоритмами технологии обработки информации.

Эти вопросы решены в 90-х гг. прошлого столетия [4, 6], и их решения описываются далее в статье.

Решением систем искусственного интеллекта могут быть: значение параметра или характеристики реальной системы, оценки динамики, трендов, ситуаций, классифицированный тип объекта, прогнозы, выводы, рекомендации. Такие решения в условиях значительной неопределенности можно рассматривать как реализацию (оценку) случайной величины или процесса, которые представлены выборкой экспериментальных данных.

В качестве экспериментальных могут быть применены данные приборных измерений, статистических временных рядов, анкетные данные опросов, видео-информация, картографические данные и др. Полученные на их основе оценки обладают статистическими свойствами и являются случайными. Точное выборочное распределение этих оценок определяется свойствами типа решения и объемом экспериментальных данных, по которым ведется оценивание.

Как уже отмечалось выше, РБП основан на применении шкалирования. Шкалирование производится на шкалах специального типа, называемых шкалами с динамическими ограничениями (ШДО) [4, 9]. Каждый элемент носителя такой шкалы может выступать в роли условно-истинного значения определяемой величины в рамках существующих в момент обработки ограничений. При определенных условиях элементы носителя шкалы могут быть представлены как центры условных точных выборочных распределений $f(\tilde{h}_s|h_s)$ статистик $\tilde{h}_s$.

С позиций теории измерений и метрологии общая погрешность решения может определяться мерой отличия оценки определяемой величины от ее истинного значения. Тогда общая погрешность результата БИИ может быть определена в виде:

$$\Delta h_s = \rho\left(h_s - \tilde{h}_s\right)$$

Являясь, как и сама оценка, случайной величиной или процессом, погрешность Δh_s полностью определяется законом распределения вероятностей значений погрешности $f(\Delta h_s)$, который, в свою очередь, зависит от точного условного выборочного распределения оценки. Плотность вероятности погрешности решения на основе РБП может быть записана в виде:

$$f\left(\Delta h_{s,l}\right) = f\left(\Delta h_{s,l}|\varphi_{j,l}|x_l\right) * f(\Delta x_l|x_l)|Y_l, \tag{1}$$

где $f\left(\Delta h_{s,l}|\varphi_{j,l}|x_l\right), f(\Delta x_l|x_l)$ – трансформированная (привнесенная в процессе обработки при трансформации согласно преобразованию $\varphi_{j,l}$ погрешности исходной информации) и собственная погрешности исходной информации.

Полная погрешность решений Δh_s может быть представлена суммой трех составляющих: погрешности неадекватности модели объекту $\Delta h_s^{(\text{на})}$; погрешности, обусловленной статистическими свойствами оценки $\Delta h_s^{(\text{ст})}$; погрешности инструментальной реализации алгоритма средствами $\Delta h_s^{(\text{си})}$, в виде:

$$\Delta h_s = \Delta h^{(\text{на})} * \Delta h^{(\text{ст})} * \Delta h^{(\text{си})} \tag{2}$$

Условная погрешность неадекватности на компакте шкалы конечна и может быть выражена численно при определенных условиях (например, при известных характеристиках модели и ограничениях). Так, для шкалы ШДО условная погрешность неадекватности определяется максимальным расстоянием между элементами носителя шкалы:

$$\Delta h_s^{(\text{на})}|Y_i = \max_{\substack{h_s \in H_K \\ S=\overline{1,K}}} \rho(h_{s+1}; h_s) \le \rho_{\text{доп.}}(\xi_{\text{доп.}}) \tag{3}$$

Для ШДО эта погрешность является функцией времени и определяется в виде:

$$\Delta h_{s,l}^{(\text{на})}|Y_{i,l} = \max_{\substack{h_s \in H_{K(l)} \\ S=1,K(l)}} \rho\left(h_{(s+1),l}; h_{s,l}\right) \le \rho_{\text{доп.}}\left(\xi_{\text{доп.},l}\right) \tag{4}$$

Статистическая погрешность за счет конечности объема данных определяется условным законом распределения статистики РБО, который для ШДО может быть представлен в виде нестационарного закона распределения $f\left(\tilde{h}_{s,l}^{(\text{и})}\right)$:

$$f\left(\tilde{h}_{s,l}^{(\text{ст})}\middle|Y_l\right) = f\left(h_{s,l}^{(\text{ст})}\middle|h_{s,l-1}\right) \cdot f\left(h_{s,l-1}^{(\text{ст})}\right)|Y_l^{(\text{ст})} \tag{5}$$

Закон распределения этой погрешности с учетом (5) может быть записан в виде:

$$f\left(\Delta h_{s,l}^{(\text{ст})}\middle|Y_l\right) = f\left(\Delta h_{s,l}^{(\text{ст})}\right)|f\left(\Delta h_{s,l-1}^{(\text{ст})}\right) \cdot f\left(\Delta h_{s,l-1}^{(\text{ст})}\right)|f^{(a)}\left(h_{s,l}\right) \cdot f^{(a)}\left(h_{s,l}\right)|Y_l^{(\text{ст})} \tag{6}$$

Погрешность неидеальности для статистики РБО представляется в виде композиции законов распределения погрешностей реализации (неидеальности) алгоритма БИИ и представления набора исходных данных:

$$f\left(\Delta h_{s,l}^{(\text{ни})}\right)\left|\varphi_{j,l}; x_l; Y_l = f\left(\Delta h_{\varphi_{j,l}}^{(\text{ни})}\right) * f\left(\Delta h_{x_l}^{(\text{ни})}\right)\right| Y_l^{(\text{ни})} \tag{7}$$

При зависимых составляющих погрешностей (1) + (7) выражение для полной погрешности результата БИИ можно переписать в виде их условного закона распределения:

$$\begin{aligned} &f\left(\Delta h_{s,l}\middle|h_{s,l}; Y_l\right) = \\ &= \left[\left(f\left(\Delta h_{s,l}^{(\text{ни})}\right)|f\left(\Delta h_{s,l}^{(\text{ст})}\right)|f\left(\Delta h_{s,l}^{(\text{на})}\right) \times f\left(\Delta h_{s,l}^{(\text{ст})}\right)|f\left(\Delta h_{s,l}^{(\text{ни})}\right) \times f\left(\Delta h_{s,l}^{(\text{на})}\right)\right)\right]|Y_l \end{aligned} \tag{8}$$

Качество статистической оценки определяется свойствами несмещенности, состоятельности и эффективности. Рассмотрим соответствие этим свойствам свойств статистик РБО.

Соответствие этим свойствам классических байесовских оценок подробно рассмотрено в [2, 11]. Как указывается в [11], применительно к байесовским оценкам эти свойства оказываются неадекватными им. Отличительным свойством классических байесовских оценок является их оптимальность. Однако в ряде работ свойства байесовских оценок рассматриваются с позиций классической теории статистики.

Так, в работах [1, 11] показана достаточность байесовских оценок в смысле традиционного определения достаточности [3], поскольку по байесовской идеологии апостериорная плотность распределения параметра, порождаемая достаточной статистикой, эквивалентна апостериорной плотности распределения параметра, построенной по первоначальному наблюдению:

$$|z(x_i) \sim h_s|x_i \tag{9}$$

Этот вывод важен для построения алгоритмов интеллектуализации косвенных измерений.

Методологические вопросы, связанные с определением свойства состоятельности байесовских оценок в обобщенном и детализированном вариантах решены в работе [2] и др. В этих работах при трактовке параметра как фиксированного, но неизвестного, была доказана состоятельность байесовских оценок в смысле стремления оценки параметра при значительных объемах выборки концентрироваться около истинного его значения. Причем апостериорная дисперсия стремится к нулю, а само апостериорное распределение стягивается в функцию, сосредоточенную в точке, равной истинному значению параметра. Этот результат важен еще и потому, что указывает на преобладание в байесовской оценке при больших объемах выборочных данных объективности, уменьшающей влияние субъективно сформированного априорного распределения. Вывод обобщен для векторных данных в [11] и строго доказан Бернштейном и Мизесом в виде теоремы, которая гласит, что по мере возрастания объема выборочных данных апостериорная плотность распределения стремится к пределу, не зависящему от априорного распределения. Как указано в [11, 13] в качестве точного выборочного распределения можно использовать его асимптотическое приближение с погрешностью аппроксимации не более $\frac{1}{\sqrt{N}}$.

Свойства несмещенности для байесовских оценок определяются в виде условной несмещенности [13], которая применительно к результатам БИИ имеет вид:

$$E\big(\tilde{h}_s(x_i)\big|h_s\big) = E(h);\ \forall h_s \in H_K \tag{10}$$

При таком определении несмещенности байесовские оценки считаются несмещенными и точными, если соблюдается условие [11]:

$$P\big[\tilde{h}(x_i) \neq h\big] = 0 \tag{11}$$

Как показано в работе [13], эффективность байесовской оценки определить нельзя, поскольку при одинаковых условиях эксперимента (объеме выборки и виде априорного распределения) двух различных байесовских оценок не существует.

Для РБО эти свойства детально рассматриваются в [4, 6]. Если принять концепцию несмещенности в виде (10), можем показать, как это сделано в работах [4, 6], что условие (11) для РБО не всегда выполняется. Регуляризация и связанный с ней процесс шкалирования значений параметра ведут к появлению запланированного смещения РБО, регулируемого параметром регуляризации $\xi_{\text{доп.}}$. Однако, так как это смещение не превосходит заданных пределов, очевидно, РБО можно считать квазинесмещенными оценками.

Если же рассматривать эти свойства для РБО, получаемых на ШДО, то можно показать, что при условии накопления информации об ОИ и при повышении метрологических требований такие оценки будут асимптотически несмещенными. Действительно, если неограниченно уменьшать значение параметра регуляризации $\xi_{\text{доп.}}$, то допустимое расстояние между элементами носителя шкалы неограниченно уменьшается, что ведет к уменьшению смещения РБО, которое в пределе будет стремиться к нулю. Этому выводу соответствует запись:

$$\forall\big(h_{s,l}, h_{s+1,l}\big) \in H_{K,l};\ \lim_{\substack{N\to\infty \\ t\to\infty}}\big(h_{s,l}, h_{s+1,l}\big) \to 0 \sim \lim_{\substack{N\to\infty \\ t\to\infty}}\Big[\Delta h_{s,l}^{(\text{на})}|Y_l\Big] \to 0 \tag{12}$$

Поскольку смещение в рамках данного компакта H_K определяет условную потенциальную точность алгоритма, очевидно, при условиях (12) можно достичь на ШДО неограниченной условной потенциальной точности измерений.

При выполнении условия (12) и неограниченном возрастании объема выборочных данных N распределение $f\left(\Delta_{s,l}^{(\text{ст})}\right)$ погрешности $\Delta_{s,l}^{(\text{ст})}$ измерений (6) стягивается в функцию, сосредоточенную в точке, соответствующей элементу носителя шкалы. Такая ситуация соответствует состоятельным РБО и может быть отражена записью:

$$\lim_{N\to 0} \rho\left(\Delta h_{s,l}^{(\text{ст})} \middle| Y_l\right) \to 0 \tag{13}$$

Таким образом, оценки, полученные на ШДО, асимптотически состоятельны.

Однако в реальной статистической практике достичь выполнения условия (13) невозможно, и оценки остаются оценками с конечной дисперсией. Эта ситуация приводит к появлению альтернативных решений. Очевидно, чем ближе элементы носителя ШДО друг к другу, тем больше вероятность альтернативного решения и тем большее число альтернативных решений окажется в списке полного решения. Оценки на основе РБП являются квазиустойчивыми оценками, и требуется количественная оценка степени остаточной неустойчивости решений БИИ, определяемой статистическими свойствами погрешности оценки при конечных объемах выборочных данных.

Относительно свойства эффективности оценки остается справедливым вывод, приведенный выше для классических байесовских оценок.

Уникальная характеристика байесовских оценок – их достоверность, определяемая апостериорной вероятностью их появления на компакте H_K. Носитель шкалы при этом рассматривается как полная группа несовместных событий, каждое из которых состоит в появлении одного из элементов носителя шкалы и имеет собственную вероятность появления в данной измерительной ситуации на данном компакте. При выходе за рамки компакта (например, при расширении пространства носителя ШДО) вероятность переходит в условную вида:

$$P(h_s|x_i) = P(h_s|x_i|H_K) \cdot P(H_K), \tag{14}$$

где $P(H_K)$ – вероятность принадлежности результата компакту H_K из множества компактов $H_K^{(O)}$.

В свою очередь, вероятность реализации шкалы G_S равна:

$$P(H_K) = P\left(H_K|H_K^{(O)}\right) \cdot P\left(H_K^{(O)}\right) \tag{15}$$

Таким образом, эта характеристика связывает данный компакт с внешним по отношению к нему миром – расширенным пространством решений $H_K^{(O)}$, причем $H_K \subset H_K^{(O)}$. Выражения (14), (15) дают формальную основу для расширения пространства решений на основе ШДО.

2. Комплекс метрологических характеристик и показатели качества решений РБП

На основании рассмотрения свойств РБО можно определить основные характеристики метрологического комплекса, которые должны отражать:

– смещенность оценки, характеризующую точность (погрешность) решения;

– устойчивость оценки, определяющую надежность полученного результата БИИ;

– апостериорную вероятность оценки среди альтернативных частных решений полного решения, идентифицирующую достоверность результата БИИ.

Выбор показателей качества решений РБП должен обеспечивать оценку степени проявления всех метрологических свойств оценки, которая может представлять собой решение. Как было показано выше, к таким свойствам относятся ограниченная смещенность оценки, надежность при квазиустойчивости оценки и ее условная достоверность среди возможных альтернативных решений ШДО.

В настоящее время существуют два основных направления исследования погрешности результатов измерений. Первое из них состоит в представлении полной погрешности результата измерения как случайной величины, которая полностью обусловлена своим законом распределения. Второе направление предполагает разделение погрешности результата измерения на составляющие и изучение их в отдельности.

Первое направление наиболее полно определяет все свойства погрешности измерений, поскольку эти свойства устанавливаются законом ее распределения вида (8). Однако в связи со сложностью процессов обработки информации закон распределения погрешности не всегда можно идентифицировать с достаточной точностью. Кроме того, частные показатели характеризуют вклад тех или иных метрологических характеристик оценки, что позволяет на основе их изучения компенсировать их частично или полностью в процессе оптимизации эксперимента. Поэтому целесообразным является представление погрешности оценки в виде комплекса их основных метрологических характеристик (КМХ), который должен отвечать поставленным выше требованиям.

В качестве показателей смещения могут быть, вообще говоря, выбраны любые применяемые оценки точности результатов измерений, например абсолютная, относительная или приведенная систематическая погрешность измерений. Причем расстояние $\rho(h_s, h_{s+1})$ на рис. 1 характеризует абсолютную погрешность, равную для каждой пары элементов носителя шкалы расстоянию между ними. При равномерной шкале эта величина будет постоянной и иметь ту же размерность, что и основное свойство. При неравномерной шкале БИИ названная составляющая равна максимальному расстоянию между соседними элементами носителя шкалы:

$$\Delta_{max} = \max_{h_s \in H_K} \rho(h_s, h_{s+1}) \tag{16}$$

Однако для шкалы ШДО, а особенно универсальной (безразмерной), наиболее удобной формой представления точности решения является приведенная систематическая погрешность:

$$\xi_s = \xi_{max} = \frac{\Delta_{max}}{\rho(h_K - h_1)} \tag{17}$$

Отметим, что показатель ξ_{max} одновременно и является параметром регуляризации алгоритма РБП, и определяет потенциальную точность оценки на данной шкале. Точность решения можно оценить по формуле:

$$\forall h_s \in H_K, \xi_s = \frac{\max[\rho(h_s, h_{s+1}); \rho(h_s; h_{s-1})]}{\rho(h_K; h_1)} \tag{18}$$

Надежность решения характеризует устойчивость решения, то есть вероятность того, что принятое квазиустойчивое решение будет принадлежать интервалу, соответствующему выбранному элементу носителя шкалы.

На рис. 1 указанный интервал обозначен как $\left[h_{\gamma_s}, h_{\gamma_{s+1}}\right]$. Это событие можно рассматривать как совместное появление двух событий: нахождение решения в обозначенном интервале и отсутствие в нем оценки при истинности любого другого элемента шкалы.

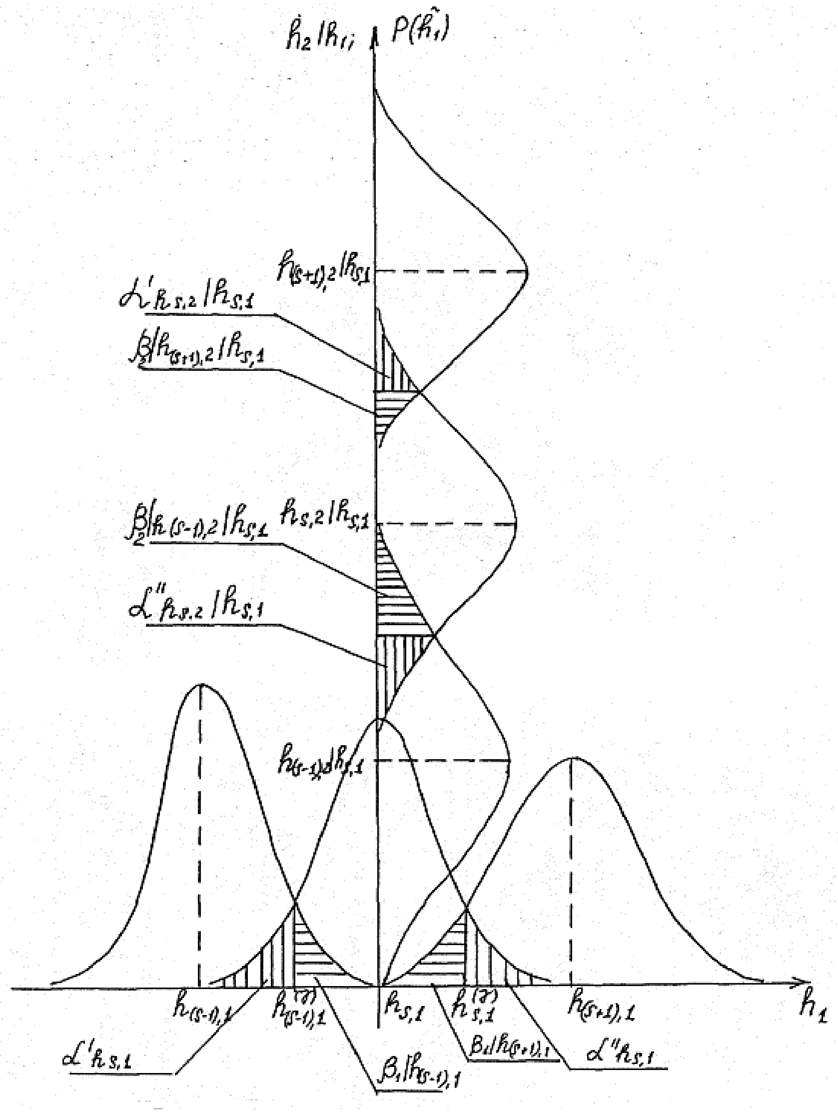

Рис. 1. Иллюстрация двумерной ШДО

Очевидно, что первое из этих событий будет характеризоваться доверительной вероятностью $P_{\alpha s}$:

$$P_{\alpha_s} = \int_{h_{\gamma_s}}^{h_{\gamma_{s+1}}} f\left(\tilde{h}|h_s|Y_i\right) d\tilde{h}, \tag{19}$$

а второе – "мощностью" отвержения альтернативного решения h_j:

$$\mathcal{D}_{s_j} = \int_{h_{\gamma_s}}^{h_{\gamma_{s+1}}} f\left(\tilde{h}|h_j|Y_i\right) d\tilde{h} \tag{20}$$

С учетом выражений (19) и (20) выражение для надежности решения может быть записано в виде:

$$V_s = \left(P_{\alpha_s} \cdot \mathcal{D}_s\right)|Y_i;\ \mathcal{D}_s = \min_{\substack{h_j \in H_K; \\ h_j \neq h_s}} \mathcal{D}_{s_j}\,|Y_i \tag{21}$$

Вероятности P_{α_s} и $\mathcal{D}_s$ удобно характеризовать уровнем ошибок первого рода α_s и уровнем ошибок второго рода (β_s) для каждого интервала шкалы. Эти величины связаны следующими соотношениями:

$$\alpha_s = \left(1 - P_{\alpha_s}\right)|Y_i, \tag{22}$$

$$\beta_s = \max_{\substack{h_j \in H_K;\\ h_j \neq h_s}} \left[\int_{h_{\gamma_s}}^{h_{\gamma_{s+1}}} f\left(\tilde{h}|h_j|Y_i\right)dh\right] = 1 - \min_{\substack{h_j \in H_K;\\ h_j \neq h_s}} \mathcal{D}_{s_j}\,|Y_i \tag{23}$$

Тогда надежность результата БИИ может быть определена в виде:

$$V_s = (1 - \alpha_s)(1 - \beta_s)|Y_i \tag{24}$$

В нормативных документах, используемых в измерительной практике, для характеристики неустойчивости решений применяется случайная составляющая погрешности измерений, которая, однако, не учитывает влияния альтернативных решений на устойчивость результата измерений, что особенно важно для некорректных задач.

С помощью выражений (19) + (24) можно определить надежность шкалы V_s, а также ее верхнюю $\overline{V}_s$ и нижнюю $\underline{V}_s$ границы в виде:

$$V_s^{(s)} = \underline{V}_s^{(s)} = \left(\underline{P}_s^{(s)} \cdot \underline{\mathcal{D}}_s^{(s)}\right)\Big|Y_i = \left[\min_{h_s \in H_K} P_{\alpha_s} \cdot \left(\mathcal{D}_s\Big|\min_{h_s \in H_K} P_{\alpha_s}\right)\right]\Big|Y_i, \tag{25}$$

$$\overline{V}_s^{(s)} = \left(\overline{P}_s^{(s)} \cdot \overline{\mathcal{D}}_s^{(s)}\right)\Big|Y_i = \left(\max_{h_s \in H_K;} P_{\alpha_s} \cdot \left(\max_{h_s \in H_K;} \mathcal{D}_s\,|\max_{h_s \in H_K} P_{\alpha_s}\right)\right)\Big|\,Y_i, \tag{26}$$

где $\overline{P}_s^{(s)}$, $\underline{P}_s^{(s)}$ и $\overline{\mathcal{D}}_s^{(s)}$, $\underline{\mathcal{D}}_s^{(s)}$ – верхние и нижние границы доверительной вероятности и мощности шкалы, которые могут быть определены выражениями:

$$\overline{P}_s^{(s)} = \max_{h_s \in H_K} P_{\alpha_s}\Big|Y_i = \left(1 - \underline{\alpha}_s^{(s)}\right)\Big|\,Y_i, \tag{27}$$

$$\underline{P}_s^{(s)} = \min_{h_s \in H_K} P_{\alpha_s}|Y_i = \left(1 - \overline{\alpha}_s^{(s)}\right)Y_i, \tag{28}$$

$$\overline{\mathcal{D}}_s^{(s)} = \max_{\substack{h_s \in H_K;\\ h_j \in H_K;\\ h_s \neq h_j}} \mathcal{D}_{s_j}\Big|Y_i = \left(1 - \underline{\beta}_s^{(s)}\right)\Big|\,Y_i, \tag{29}$$

$$\underline{\mathcal{D}}_s^{(s)} = \min_{\substack{h_s \in H_K;\\ h_j \in H_K;\\ h_j \neq h_s}} \mathcal{D}_{s_j}\Big|Y_i = \left(1 - \overline{\beta}_s^{(s)}\right)\Big|\,Y_i \tag{30}$$

Аналогично определяются верхние и нижние границы для уровней ошибок первого $\overline{\alpha}_s^{(s)}$, $\underline{\alpha}_s^{(s)}$ и второго $\overline{\beta}_s^{(s)}$, $\underline{\beta}_s^{(s)}$ рода шкалы:

$$\overline{\alpha}_s^{(s)} = \max_{h_s \in H_K} \alpha_s\,|Y_i, \tag{31}$$

$$\underline{\alpha}_s^{(s)} = \min_{h_s \in H_K} \alpha_s\,|Y_i, \tag{32}$$

$$\bar{\beta}_s^{(s)} = \max_{\substack{h_s \in H_K; \\ h_j \in H_K; \\ h_j \neq h_s;}} \beta_s \,|Y_i, \tag{33}$$

$$\underline{\beta}_s^{(s)} = \min_{\substack{h_s \in H_K; \\ h_j \in H_K; \\ h_j \neq h_s.}} \beta_s \,|Y_i \tag{34}$$

Показателем достоверности решений на данной шкале может быть апостериорная вероятность его появления, а достоверность компакта решений, то есть шкалы, определяется по формуле (25).

Таким образом, комплекс метрологических характеристик (КМХ), определяющий качество решений РБП, имеет вид:

$$\{MX\}_s = \{\xi_s; V_s; P_s\} \tag{35}$$

КМХ для шкалы имеет вид:

$$\{MX\}_s^{(s)} = \left\{\xi_s^{(s)}; V_s^{(s)}; P_s^{(s)}\right\}, \tag{36}$$

в котором $P_s^{(s)}$ – априорная вероятность реализации данной шкалы РБП (априорная вероятность появления результата БИИ на этой шкале), связывающая компактное пространство конкретной шкалы с пространством объектов или моделей.

При этом если имеется обобщающая иерархическая шкала РБП, в которую данная шкала входит как составляющая, то показатель $P_s^{(s)}$ определяет вероятность появления решения на соответствующем уровне иерархии, что может быть отражено следующим образом:

$$P_s^{(s)} = P_{s_{i_\pi}} = P_{s_{i_\pi}} \left| \left(P_{s_{(i_\pi - 1)}}, \cdots \right) \cdot P_{s_{(i_\pi - 1)}} \right| \cdots \cdot P_{s_1} \tag{37}$$

Если пространство шкалы динамически "погружается" в пространство объекта (при реализации ШДО), КМХ будет иметь вид динамического комплекса:

$$\{MX\}_{s,l}^{(s)} = \left\{\xi_{s,l}^{(s)}; V_{s,l}^{(s)}; P_{s,l}^{(s)}\right\} \tag{38}$$

КМХ шкалы ШДО характеризует качество и метрологические характеристики алгоритма ИИ при его идеальной реализации. Таким образом, КМХ содержит показатели методических погрешностей решений ИИ.

3. Функциональные преобразования комплексов метрологических характеристик БИИ

Показатели метрологических характеристик (MX), входящие в комплексы, могут быть преобразованы в процессе обработки информации согласно структуре алгоритмов. При этом показатели КМХ преобразуются отдельно по правилам, соответствующим каждому типу показателей и виду функционального отношения.

При одноместных функциональных отношениях вид КМХ определяется видом функционального преобразования F_f. Причем преобразование показателей производится на основании правил вычисления функционалов вида:

$$\xi_F = F_f^{\xi}\big(\xi_s(h_s)\big); V_F = F_f^{V}\big(V_s(h_s)\big); P_F = F_f^{(P)}\big(P_s(h_s)\big) \tag{39}$$

КМХ для одноместного функционального преобразования может быть записан в виде функционального преобразования КМХ решения:

$$\{MX\}_F = F_f^{(MX)}(\{MX\}_s) \tag{40}$$

Любому функциональному преобразованию, имеющему вид двухместного отношения $W_f\left(h_s, h_j\right)$ результатов БИИ h_s и h_j, соответствует обобщенный КМХ $\{MX\}_o$, представляющий композицию комплексов $\{MX\}_s$ и $\{MX\}_j$:

$$\{MX\}_o = \{MX\}_s * \{MX\}_j | Y^{(o)}, \tag{41}$$

при обобщенных условиях измерений $Y_i^{(o)} = Y_i^{(s)} * Y_i^{(j)}$, $Y_i^{(s)}, Y_i^{(j)}$ – условия измерений для i-го и j-го решения на ШДО.

Составляющие комплекса преобразуются при независимых результатах $\left(h_s, h_j\right) \in H_K$ следующим образом:

$$\xi_o = \left(\xi_s * \xi_j\right)\Big|Y_i^{(s)}\Big|Y_i^{(j)}, \tag{42}$$

$$V_o = V_s \cdot V_j | Y_i^{(o)}, \tag{43}$$

$$P_o = \frac{P_s \cdot P_j | Y_i^{(o)}}{P_{\sum(s,j)}}, \tag{44}$$

$$P_{\sum(s,j)} = \sum_{s=1}^{K} \sum_{\substack{j=1 \\ j \neq s}}^{K} P_s \cdot P_j | Y_i^{(s)} \, | Y_i^{(j)} \tag{45}$$

Отношение L – местного обобщения $\begin{pmatrix} L \\ * \\ l = 1 \end{pmatrix}$ реализуется на практике в виде алгоритмов синтеза l-мерных условных шкал.

Если решения коррелированы между собой, то формулы преобразования составляющих комплексов принимают следующий вид:

$$\xi_o = \left(\xi_s | \xi_j * \xi_j\right) | Y_i^{(o)}, \tag{46}$$

$$V_o = \left(V_s | V_j \cdot V_j\right) | Y_i^{(o)}, \tag{47}$$

$$P_o = \frac{P_s | P_j \cdot P_j | Y_i^{(o)}}{P_{\sum(s,j)}}, \tag{48}$$

где $\xi_s\big|\xi_j, V_s\big|V_j, P_s|P_j$ – условные составляющие комплекса.

Полученные формульные зависимости достаточно просто распространяются на многоместные функциональные преобразования, при которых обобщенный КМХ может быть выражен следующей записью:

$$\{MX\}_o = \left(\underset{i=1}{\overset{I}{*}} \{MX\}_i \right) |Y_i^{(o)}, \tag{49}$$

а составляющие КМХ имеют вид:

$$\xi_o = \underset{i=1}{\overset{I}{*}} \xi_i | \xi_{i-1} \cdots | \xi_1 | Y_i^{(o)}, \tag{50}$$

$$V_o = \prod_{i=1}^{I} V_i | V_{i-1} \cdots | V_1 | Y_i^{(o)}, \tag{51}$$

$$P_o = \frac{\prod_{i=1}^{I} P_i | P_{i-1} \cdots | P_1 | Y_i^{(o)}}{P_{\sum(i,i-1,\cdots,1)}} \tag{52}$$

Если рассматривать решение РБП как случайное событие, состоящее в совместном появлении конкретной реализации алгоритма φ_i из набора Φ_i и конкретного набора данных x_i из совокупности данных X_I, то можно определить КМХ этого результата как композицию КМХ вида:

$$\{MX\}_s = \{MX\}_{\varphi_j} * \{MX\}_{x_i} | Y_{\varphi_j} | Y_{x_i}, \tag{53}$$

где $\{MX\}_{x_i}$ – собственные и $\{MX\}_{\varphi_j}$ трансформированные комплексы данных по алгоритму φ_j при условиях измерений Y_{x_i}; Y_{φ_j}.

Для всего множества X_I КМХ $\{MX\}_s$:

$$\{MX\}_s = \underset{i=1}{\overset{I}{*}} \left(\{MX\}_{\varphi_j} | x_i * \{MX\}_{x_i} \right) | Y_i^{(o)};\ Y_i^{(o)} = Y_{x_i} * Y_{\varphi_j}, \tag{54}$$

может быть записан в виде обобщения условных комплексов, причем $\{MX\}_{\varphi_j}$ – условный КМХ алгоритма при условии конкретного набора данных x_i, $\{MX\}_{x_i}$ – собственный комплекс данных.

Если алгоритм ИИ представляет собой цепь преобразований в виде совокупности взаимозависимых этапов, где $l = \overline{1, L}$ – номер этапа преобразований, число которых L. КМХ для такого измерительного процесса можно записать в виде:

$$\{MX\}_{S,L} = \underset{l=1}{\overset{L}{*}} \left[\left(\{MX\}_{\varphi_{j,l}} | \{MX\}_{\varphi_{j,l-1}} | \cdots | \{MX\}_{\varphi_{j,1}} | \{MX\}_{x_i} \right) * \{MX\}_{x_i} \right] | Y_{i,L}^{(o)}, \tag{55}$$

$$Y_{i,L}^{(o)} = \underset{l=1}{\overset{L}{*}} Y_{i,l}^{(o)} \tag{56}$$

В выражениях (55) ÷ (56) $\{MX\}_{x_i}$ – собственный комплекс КМХ подэтапов, а $\{MX\}_{\varphi_{j,l}}$ – трансформированный комплекс данных для l-го этапа алгоритма БИИ.

Таким образом, на каждом этапе алгоритма ИИ можно проследить и процесс получения решения, и его качество. Этим обеспечивается реализация принципов ***прослеживаемости и прозрачности***.

При определении решения по признакам согласно (6) или по достаточным статистикам $\tilde{h} = z(x_i)$ комплексы КМХ имеют вид:

$$\{MX\}_o = \mathop{*}_{i_\pi = 1}^{I_\pi} \left(\{MX\}_{i_\pi} | Y_{i_\pi}\right) \tag{57}$$

$$\{MX\}_o = \{MX\}_{z(x_i)} | \{MX\}_{x_i} * \{MX\}_{x_i} | Y_z^{(o)}, \tag{58}$$

$$Y_z^{(o)} = Y_{z(x_i)} * Y_{(x_i)} \tag{59}$$

Приведенные выше зависимости позволяют рассматривать качество решений РБП с позиции качества алгоритмического и информационного обеспечения ИИ, а также оценивать отдельно качество алгоритма ИИ и наборов исходных данных, что позволяет производить необходимые мероприятия по повышению качества соответствующих составляющих процесса интеллектуальной обработки информации.

Исходя из вышеприведенных формул можно определить КМХ для отдельных составляющих погрешности измерений. Например, можно определить КМХ для систематической составляющей $\Delta h_s^{(\text{с})}$ в виде КМХ оценки математического ожидания погрешности. Для характеристики случайной составляющей можно определить КМХ для дисперсии погрешности. Для оценки $\Delta h_s^{(\text{д})}$ динамической составляющей систематической погрешности КМХ представляется в виде кортежа, компоненты которого зависят от времени:

$$\{MX\}_{s,l}^{(\text{д})} = \left\{\xi_{s,l}^{(\text{д})}; V_{s,l}^{(\text{д})}; P_{s,l}^{(\text{д})}\right\} | Y_{l,s} \tag{60}$$

Таким образом, для полной группы погрешностей можно определить КМХ, характеризующие не только средние значения составляющих погрешности, но и надежность этих оценок, и их достоверность, что традиционные (регламентированные) показатели обеспечить не могут. Очевидно, КМХ есть обобщенный вариант комплекса метрологических характеристик, включающий и регламентированные показатели, для определения метрологических свойств составляющих погрешности решений ИИ.

Для обеспечения эффективного функционирования алгоритма ИИ необходима оптимизация процесса измерений в соответствии с метрологическими требованиями $M_{i,l}$ эксперимента, направленная на обеспечение требуемого качества решений. Оптимальная стратегия измерительного эксперимента является результатом решения обратной задачи метрологического обоснования, в которой при заданных значениях составляющих комплекса определяются основные параметры и характеристики алгоритма ИИ, при которых достигается требуемое качество решений (задачи метрологического синтеза алгоритма ИИ).

К числу основных параметров алгоритмов ИИ можно отнести объем выборочных данных, определяющий продолжительность эксперимента для обеспечения требуемых значений метрологических показателей, входящих в КМХ. Минимальный объем выборочных данных, удовлетворяющий условиям теоремы 2 [7], определяется по формуле:

$$N_{min} = \max_{\substack{h_s^{(\pi)} \in H_K^{(\pi)};\\ i_\pi = \overline{1, I_\pi};\\ s = \overline{1, K_\pi}}} N_{s,i_\pi} | Y_{s,i_\pi}, \tag{61}$$

где K_π – число элементов носителя шкалы i_π-го признака или i_π-й составляющей.

При нормальном точном выборочном распределении статистики измеряемого свойства или признака и простой матрице штрафов минимальный объем выборочных данных можно определить по формуле:

$$N_{s,i_\pi} = \left[2 \ln \left[\frac{P_{(s+1),i_\pi}^{(a)} \cdot \varkappa_{s,i_\pi}}{P_{s,i_\pi}^{(a)} \cdot \varkappa_{(s+1),i_\pi}}\right] \varkappa_{(s+1),i_\pi}^2\right] : h_{(s+1),i_\pi}, \tag{62}$$

где: $P_{s,i_\pi}^{(a)}$ – априорная вероятность появления i-го элемента шкалы i_π-го признака;
h_{s,i_π} – значение i-го элемента шкалы i_π-го признака;
$\varkappa_{s,i_\pi}$ – коэффициент дисперсии, $\varkappa_{s,i_\pi} = \sigma_{s,i_\pi} \cdot \sqrt{N}$;
σ_{s,i_π} – с.к.о. оценки признака;
N – объем выборки.

При решении обратной задачи метрологического обоснования (задачи метрологического синтеза) необходимо определить требуемый объем выборочных данных для обеспечения заданных уровней ошибок 1-го и 2-го рода. Объем выборочных данных для определения требуемой мощности алгоритма для указанных выше условий может быть установлен по формуле:

$$N_{s,i_\pi}^{(\beta)} = \mu_{1,s,i_\pi}^2 \left[\varkappa_{(s+1),i_\pi} + \varkappa_{s,i_\pi} \cdot \sqrt{1 - \frac{2}{\mu_{1,s,i_\pi}^2} \ln \frac{P_{(s+1),i_\pi}^{(a)} \cdot \varkappa_{s,i_\pi}}{P_{s,i_\pi}^{(a)} \cdot \varkappa_{(s+1),i_\pi}}}\right], \tag{63}$$

где: $\mu_{1,s,i_\pi} = \arg \Phi | \beta_д$;
Φ – функция Лапласа;
$\beta_д$ – требуемый уровень ошибок 2-го рода;
$\mathcal{D}_д = 1 - \beta_д$ – мощность алгоритма ИИ.

Для обеспечения требуемой достоверности алгоритма ИИ или уровня ошибок 1-го рода, объем выборочных данных определяется по формуле:

$$N_{s,i_\pi}^{(\alpha)} = \mu_{2,s,i_\pi}^2 \left[\varkappa_{s,i_\pi} + \varkappa_{(s+1),i_\pi} \sqrt{\mu_{2,s,i_\pi}^2 + 2 \ln \frac{P_{(s+1),i_\pi}^{(a)} \cdot \varkappa_{s,i_\pi}}{P_{s,i_\pi}^{(a)} \cdot \varkappa_{(s+1),i_\pi}}}\right], \tag{64}$$

где $\alpha_д$ – заданный уровень ошибок 1-го рода;

$$\mu_{1,s,i_\pi} = \frac{h_{\gamma_{s,i_\pi}} - h_{s,i_\pi}}{\sigma_{s,i_\pi}}; \ \mu_{2,s,i_\pi} = \frac{\left|h_{\gamma_{s,i_\pi}} - h_{s+1,i_\pi}\right|}{\sigma_{s+1,i_\pi}} \tag{65}$$

Оптимальный объем выборочных данных для обеспечения метрологических требований N вычисляется как максимальный из N_{s,i_π}:

$$N^{(o)} = \max_{\substack{i_\pi=\overline{1,I_\pi};\\ s=\overline{1,K_\pi}}} \left(\max N_{s,i_\pi}^{(o)}\right);\ N_{s,i_\pi}^{(o)} = \max\left(N_{s,i_\pi}^{(\alpha)}; N_{s,i_\pi}^{(\beta)}\right) \tag{66}$$

Риск получаемых решений ИИ определяется как разность между единицей и достоверностью (апостериорной вероятностью решения) в виде:

$$\mathrm{R} = 1 - \int_{h_{\gamma_s}}^{h_{\gamma_{s+1}}} f\left(\tilde{h}|h_s|Y_i\right) d\tilde{h} \tag{67}$$

Таким образом, на основании приведенной методологии и при использовании РБП и данных аналитических зависимостей можно обеспечить требуемое качество решений ИИ.

Методология метрологического сопровождения реализована в программной среде "Инфоаналитик". В каждой прикладной системе, созданной на базе компьютерной платформы "Инфоаналитик", функционирует подсистема метрологического сопровождения. Динамическая модель показателя давления водопотока сети горячего водоснабжения представлена на рис. 2. На рис. 3 приведен комплекс метрологических характеристик для этого решения.

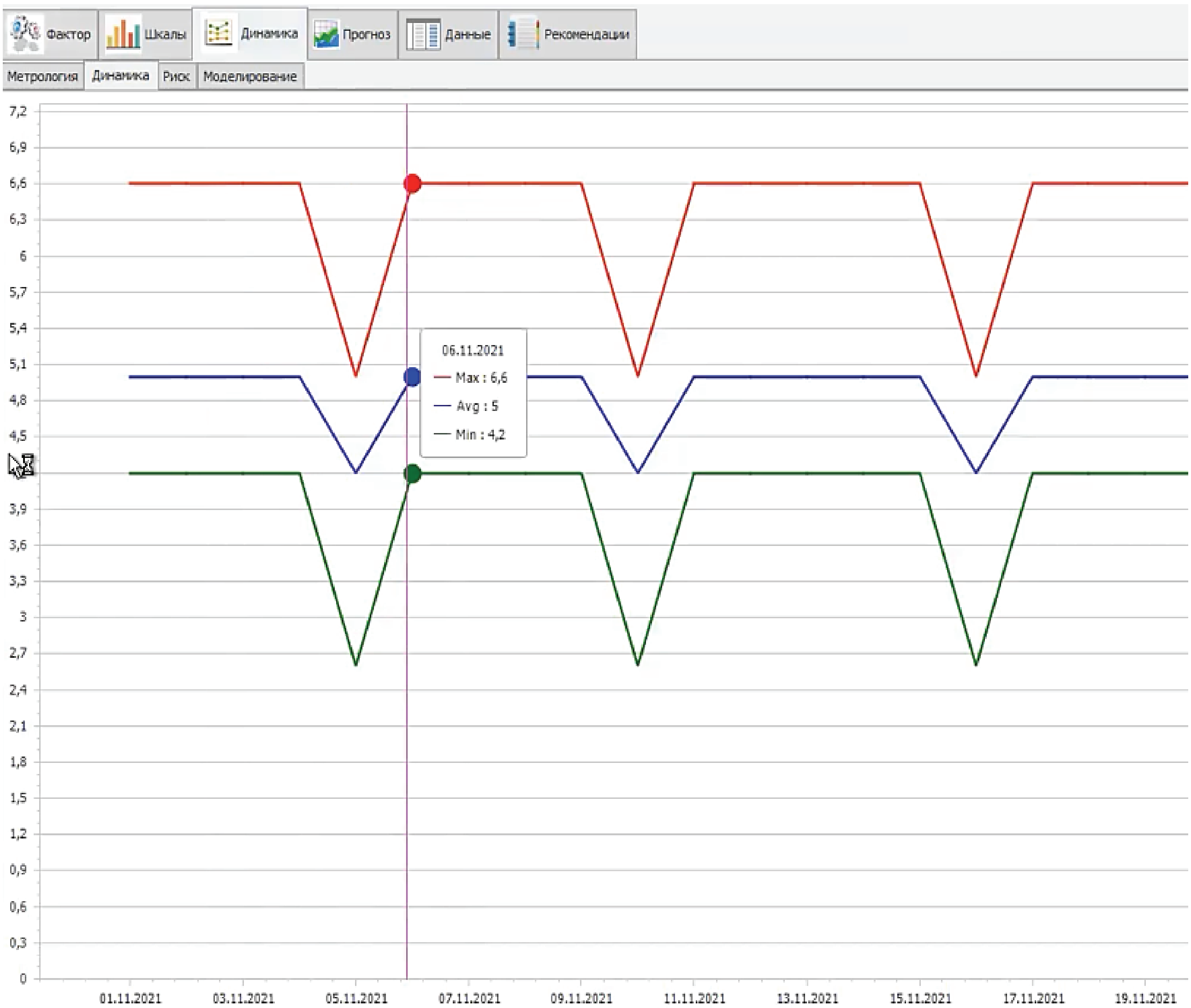

Рис. 2. Модель динамики показателя давления водопотока сети ГВС

Рис. 3. Комплекс метрологических характеристик модели давления водопотока сети ГВС

По значениям показателей видно, что в условиях неопределенности выбрать какое-либо одно решение неэффективно из-за его низкой достоверности. В условиях значительной неопределенности необходимо учитывать весь набор альтернативных решений, достоверность которых в сумме приблизительно равна 1.

Заключение

Методология создания системы оценки качества решений искусственного интеллекта (ИИ), в частности оценки достоверности решений, определяющей степень доверия к получаемым решениям, разработана на основе регуляризирующего байесовского подхода и реализована в виде системы метрологического сопровождения решений ИИ. В рамках данной методологии и технологий метрологического сопровождения предложены комплексы метрологических характеристик, определяющих точность, надежность (уровень ошибок 1-го и 2-го рода), достоверность, риск, энтропию, количество информации для каждого этапа алгоритмов, что обеспечивает прослеживаемость и прозрачность получаемых решений.

Приведен практический пример определения метрологических комплексов для решений прикладных систем искусственного интеллекта.

Методология рекомендована для использования в системах искусственного интеллекта для оценки показателей качества решений ИИ.

Список источников

[1] *Де Гроот М.* Оптимальные статистические решения. М.: Мир, 1974. 492 с.
[2] *Закс Ш.* Теория статистических выводов. М.: Мир, 1975. 776 с.
[3] *Крамер Г.* Математические методы статистики. М.: Мир, 1975. 647 с.
[4] *Недосекин Д.Д., Прокопчина С.В.* Метрологическое обеспечение моделей и алгоритмов байесовской идентификации вероятностных характеристик // Депонированные научные работы. 1991. № 11.
[5] *Прокопчина С.В.* Байесовские интеллектуальные технологии как методологическая основа обработки больших данных в условиях неопределенности // Экономика и управление: проблемы, решения. 2019. Т. 11, № 3. С. 105–109.
[6] *Прокопчина С.В.* Интеллектуальные измерения на основе регуляризирующего байесовского подхода. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2019. 508 с.

[7] *Прокопчина С.В.* Методы и средства моделирования закона распределения в условиях неопределенности. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2018. 252 с.
[8] *Прокопчина С.В.* Новый тип нейронных сетей: байесовские измерительные нейронные сети (БИН) на основе методологии регуляризирующего байесовского подхода // Мягкие измерения и вычисления. 2020. Т. 35. № 10. С. 17–24.
[9] *Прокопчина С.В.* Основы теории шкалирования в экономике. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2021. 299 с.
[10] *Прокопчина С.В.* Разработка методов и средств интеллектуализации байесовских измерений в задачах комплексного мониторинга объектов: диссертация на соискание ученой степени доктора технических наук. СПб., 1996.
[11] *Савчук В.П.* Байесовские методы статистического оценивания. Надежность технических объектов. М.: Наука, 1989. 328 с.
[12] *Good I.J.* (1965) The estimation of probabilities, an essay on modern Bayesian methods. Wiley, 110 p.
[13] *Jeffreys H.* (1966) Theory of Probalility. Oxford, Clarenson Press.

References

[1] *De Groot M.* (1974) Optimal statistical solutions. Moscow, Mir, 492 p. (In Russ.).
[2] *Zaks Sh.* (1975) Theory of statistical inference. Moscow, Mir, 776 p. (In Russ.).
[3] *Kramer G.* (1975) Mathematical methods of statistics. Moscow, Mir, 647 p. (In Russ.).
[4] *Nedosekin D.D., Prokopchina S.V.* (1991) Metrological support of models and algorithms for Bayesian identification of probabilistic characteristics. *Deponirovannyye nauchnyye raboty*, no. 11. (In Russ.).
[5] *Prokopchina S.V.* (2019) Bayesian intelligent technologies as a methodological basis for processing big data under uncertainty. *Ekonomika i upravleniye: problemy, resheniya*, vol. 11, no. 3, pp. 105–109. (In Russ.).
[6] *Prokopchina S.V.* (2019) Intelligent measurements based on the regularizing Bayesian approach. Moscow, Publishing house "NAUCHNAYA BIBLIOTEKA", 508 p. (In Russ.).
[7] *Prokopchina S.V.* (2018) Methods and tools for modeling the distribution law under uncertainty. Moscow, Publishing house "NAUCHNAYA BIBLIOTEKA", 252 p. (In Russ.).
[8] *Prokopchina S.V.* (2020) A new type of neural networks: Bayesian measuring neural networks (BIN) based on the methodology of the regularizing Bayesian approach. *Myagkiye izmereniya i vychisleniya*, vol. 35, no. 10, pp. 17–24. (In Russ.).
[9] *Prokopchina S.V.* (2021) Fundamentals of the theory of scaling in economics. Moscow, Publishing house "NAUCHNAYA BIBLIOTEKA", 299 p. (In Russ.).
[10] *Prokopchina S.V.* (1996) Development of methods and means of intellectualization of Bayesian measurements in the problems of complex monitoring of objects: dissertation for the degree of Doctor of Technical Sciences. St. Petersburg. (In Russ.).
[11] *Savchuk V.P.* (1989) Bayesian methods of statistical estimation. Reliability of technical objects. Moscow, Nauka, 328 p. (In Russ.).
[12] *Good I.J.* (1965) The estimation of probabilities, an essay on modern Bayesian methods. Wiley, 110 p.
[13] *Jeffreys H.* (1966) Theory of Probalility. Oxford, Clarenson Press.

Статья поступила в редакцию 11.08.2023; одобрена после рецензирования 18.08.2023; принята к публикации 23.08.2023.

The article was submitted 11.08.2023; approved after reviewing 18.08.2023; accepted for publication 23.08.2023.