

УДК 004.8

DOI: 10.36871/2618-9976.2023.07.001

МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ СОЗДАНИЯ РЕГУЛЯРИЗИРУЮЩИХ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Светлана Васильевна Прокопчина¹

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

ИНФОРМАЦИЯ

Ключевые слова:

устойчивость решений
регуляризирующий байесовский подход
искусственный интеллект

АННОТАЦИЯ

В статье показана актуальность обеспечения устойчивости решений систем искусственного интеллекта. Предложен методологический подход и рассмотрены теоретические основы для создания методов искусственного интеллекта, обладающих регуляризирующими свойствами и обеспечивающих получение устойчивых решений задач в условиях значительной неопределенности. Методология базируется на регуляризирующем байесовском подходе и технологиях шкалирования на его основе. Приведены теоремы и аналитические зависимости для создания регуляризирующих алгоритмов искусственного интеллекта.

ДЛЯ ЦИТИРОВАНИЯ: Прокопчина С.В. Методологические основы создания регуляризирующих методов искусственного интеллекта // Мягкие измерения и вычисления. 2023. № 7. Т. 68. С. 5–23; <https://doi.org/10.36871/2618-9976.2023.07.001>.

METHODOLOGICAL FOUNDATIONS FOR THE CREATION OF REGULARIZING ARTIFICIAL INTELLIGENCE METHODS TO ENSURE THE SUSTAINABILITY OF SOLUTIONS

Svetlana V. Prokopchina¹

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

ARTICLE INFO

Keywords:

Decision stability
Regularizing Bayesian approach
Artificial intelligence

ABSTRACT

The article shows the relevance of ensuring the sustainability of solutions of artificial intelligence systems. A methodological approach is proposed and the theoretical foundations for the creation of artificial intelligence methods with regularizing properties and providing stable solutions to problems in conditions of significant uncertainty are considered. The methodology is based on the regularizing Bayesian approach and scaling technologies based on it. Theorems and analytical dependencies for the creation of regularizing algorithms of artificial intelligence are given.

FOR CITATION: Prokopchina S.V. (2023) Methodological foundations for the creation of regularizing artificial intelligence methods to ensure the sustainability of solutions. *Soft measurements and computing*, vol. 68, no. 7, pp. 5–23 (In Russ.); <https://doi.org/10.36871/2618-9976.2023.07.001>.

**Введение. Постановка задачи обеспечения устойчивости решений
в системах искусственного интеллекта**

После первой волны повышенного интереса к системам искусственного интеллекта (ИИ), прошедшей в предыдущие десятилетия, в настоящее время к решениям ИИ стали предъявляться требования по обеспечению их качества. Проявился интерес к свойствам исходной информации, в результате чего активно разрабатывается направление DATA SCIENCE. Выдвинуты требования прозрачности и прослеживаемости решений (*Transparency Artificial Intelligence (AI) Traceability*), интерпретируемости и объяснений (*Interpretability, Explainable AI (XAI)*), доверительности (*Trustworthiness AI*). Но первостепенным, основополагающим требованием является требование устойчивости решений (*Sustainability of solutions*) систем ИИ. А это требование практически не обозначается в работах по ИИ. Несоблюдение его ведет к неединственности и ненадежности получаемых решений. Таким образом, одним из основных требований к алгоритмам систем искусственного интеллекта является требование обеспечения устойчивости получаемых решений.

Причинами неустойчивости являются высокая зашумленность исходной информации, отсутствие контроля и обеспечения точности и достоверности входных данных, непредсказуемое влияние внешних воздействий на объект изучения или управления, неструктурированность и нечеткость информации, малые выборки статистических данных и другие причины.

Устойчивость решений может быть достигнута путем применения различных методов регуляризации как пространства исходной информации, так и пространства решений.

В алгоритмах интеллектуальной обработки информации существенна также задача связи двух указанных выше регуляризованных пространств.

Таким образом, задача обеспечения устойчивости делится на три составляющие:

- обеспечение регуляризации пространства исходной информации;
- обеспечение регуляризации пространства решений;
- обеспечение методологии и технологий функциональных преобразований регуляризованного пространства исходной информации в регуляризованное пространство решений.

Самыми известными методами регуляризации пространства решений являются методы решения некорректных задач на основе методов А.Н. Тихонова [14] и его научной школы. Однако эти методы требуют определенной формализации представления информационных технологий функциональных преобразований, в частности в виде дифференциальных уравнений, что не всегда возможно.

В качестве альтернативного подхода можно рассмотреть регуляризацию пространства решений по В.С. Иванову [15], которая состоит в шкалировании информации.

В исходном виде методы шкалирования не ориентированы на регуляризацию решений ИИ, которым присуща разнотипность информации и ее распределенность во времени и в пространстве. Однако, сама идея регуляризации на основе шкалирования достаточно плодотворна, что будет показано в данной статье. Далее будет показан пример решения задачи обеспечения устойчивости решений ИИ на основе регуляризирующего байесовского подхода.

Для понимания похода рассмотрим некоторые важные методологические аспекты байесовской методологии.

1. Методологические аспекты классической байесовской идеологии

Идейная основа обобщения информации для получения решений заложена в принципах классического байесовского подхода в работах ученых-байесианцев Джеффриса, Де Гроота, Гуда [4, 5], Савчука [1] и других как индуктивная (обобщающая)

логика. В отличие от широко распространенной дедуктивной логики принятия решений, при которой поиск решения осуществляется по пути от общего к частному, в выводах индуктивной логики байесовского подхода возможно получение решений h_k на основе обобщения частных разнообразных сведений, обобщаемых в интегральное решение, соответствующее цели, причем, при неполной и неточной информации с определенной степенью сомнения, выражаемой количественной мерой апостериорной байесовской вероятности (достоверности) решения P_K . Значение вероятности вычисляется по известной формуле Байеса в виде:

$$P_K(h_k|x_i) = \frac{P_K^{(a)}(h_k) \cdot l(x_i|h_k)h_k}{\sum_{j=1}^K P_j^{(a)}(h_j) \cdot l(x_i|h_j)}, \quad (1)$$

где P_K – апостериорная байесовская достоверность решения h_k при условии вновь поступившей информации x_i ; $P_K^{(a)}$ – априорная вероятность решения h_k ; $l(\cdot)$ – функция правдоподобия выборки, отражающая вероятность появления информации $x_i \in X_I$ при условии истинности решения h_k .

Байесовскому подходу, его истории, методологии и конкретным приложениям посвящено немало научных изысканий и работ. В настоящее время к этим направлениям привлечено значительное число ученых в России и за рубежом. Сформировались отдельные ветви этих направлений, реализующие разработки модификаций классического байесовского подхода. Но основные принципы его остаются незыблемыми и могут быть определены в виде следующих положений [1]:

- свойства объектов изучения, их характеристики и параметры считаются неопределенными или случайными, им соответствуют априорные и апостериорные распределения;
- результаты наблюдений эксперимента и априорная информация, выраженная в виде априорных распределений, объединяются на основании основной теоремы Байеса для получения апостериорного распределения определяемого, свойствами объекта изучения в соответствии с формулой (1);
- вывод или решение принимается на основе правила, минимизирующего риск принятия полученного решения или максимизирующего полезность и безопасность решения.

Эти принципы, развитые позднее в работах ученых-байесианцев, позволили на основании первого из приведенных принципов прийти к понятию субъективной (фидуциальной) вероятности, что создало научную основу для привлечения неформализованного опыта специалистов к принятию решений по объективным данным экспериментов. Таким образом возникла возможность учета как статистической неопределенности экспериментальных данных, так и неопределенности за счет неадекватности модельного представления объекту изучения. Кроме того, эта возможность в свою очередь обусловила возможность эффективной работы с выборками малого объема, так как по выражению Эванса [3], "знания могут компенсировать отсутствие данных".

В классическом байесовском подходе выделяют четыре составные части, основанные на аксиоматике теории вероятностей [1]:

1. Вероятностное пространство мгновенных значений СВ или СЭСП (X, X_I, P_x) , где X – множество всех возможных значений СВ или СЭСП;

$X_I = \{x_i\}$ – σ -алгебра случайных событий, заключающихся в появлении набора данных из совокупности X_I ; P_x – вероятностная мера, определенная на (X, X_I) .

2. Вероятностное пространство (H, E_H, P_H) для оцениваемого параметра h , который считается случайным с априорным распределением определяемым P_H на (H, E_H) , где E_H σ -алгебра для H .

3. Множество возможных решений D , такое, что любой элемент из $D, d \in D$ измеримая функция на X .

4. Функция штрафа (платежная функция) $L(h, \delta)$, определенная на $H \times H$, отражающая выигрыши и проигрыши при правильном и ошибочном решении относительно оцениваемого параметра.

Важным шагом в развитии байесовской идеологии явились работы по созданию иерархического байесовского подхода [2], в которых для исследуемых свойств объекта изучения устанавливались вероятностные связи, отражаемые в виде условных априорных и апостериорных распределений или их композиций согласно следующему выражению:

$$P[(q_{i_1}, q_{i_2})|x_i] = \frac{P^{(a)}(q_{i_2}|q_{i_1}) \cdot P_{(q_{i_1})}^{(a)} \cdot l[x_i|(q_{i_1}, q_{i_2})]}{\sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} P^{(a)}(q_{j_2}|q_{j_1}) P_{(q_{j_1})}^{(a)} \cdot l[x_i|(q_{j_1}, q_{j_2})]}, \quad (2)$$

где q_{i_1}, q_{i_2} – контролируемые свойства объекта изучения.

Как видно из уравнения (1), в байесовском подходе использованы вероятностные представления свойств изучаемых объектов, что делает его особенно привлекательным для работы с неполными и неточными экспериментальными данными.

Методы принятия решений относительно свойств изучаемого объекта на основе байесовского подхода минимизируют средний риск решения, который с учетом принятых обозначений можно записать для дискретного пространства H_K в виде:

$$r_k = \sum_{k=1}^K \sum_{i=1}^I L(h_k, \delta_{kj}) P_k^{(a)}(h_k) \cdot P(x_i|h_k), \quad (3)$$

$h_k \in H_K;$

где $L(h_k, \delta_{kj})$ – функция штрафа, определенная на $X_I \times H_K$, отражающая потери и выигрыши при принятии решения h_k , если истинно решение h_j . Для непрерывного пространства контролируемого свойства выражение для среднего риска может быть переписано в виде:

$$r = \iint_{HX} L(h, \delta) f^{(a)}(h) \cdot f(x|h) dx dh, \quad (4)$$

где $f^{(a)}(h); f(x|h)$ – плотности вероятности априорного и условного распределений параметра и основной СВ X .

Если контролируемое свойство представляет собой сложное свойство, являющееся композицией других свойств $q_1 \in Q_1$ и $q_{i+1} \in Q_2$, то согласно принципу иерархичности, выражение для среднего риска может быть переписано с учетом двумерного закона распределения:

$$f(q_i, q_{i+1}) = [f(q_{i+1})|q_i] \cdot f(q_i); \quad (5)$$

в следующем виде:

$$r = \iiint_{Q_2 Q_1 X} L(h, \delta) f_{(q_i, q_{i+1})}^{(a)} \cdot f[x|(q_i, q_{i+1})] dx dq_i dq_{i+1}. \quad (6)$$

Таким образом, возможно построение сколь угодно сложных иерархических структур байесовских правил. Это обуславливает возможность реализации байесовского подхода для прямых, косвенных, совокупных и совместных измерений.

Схема классического байесовского вывода отражена на [рисунке 1](#).

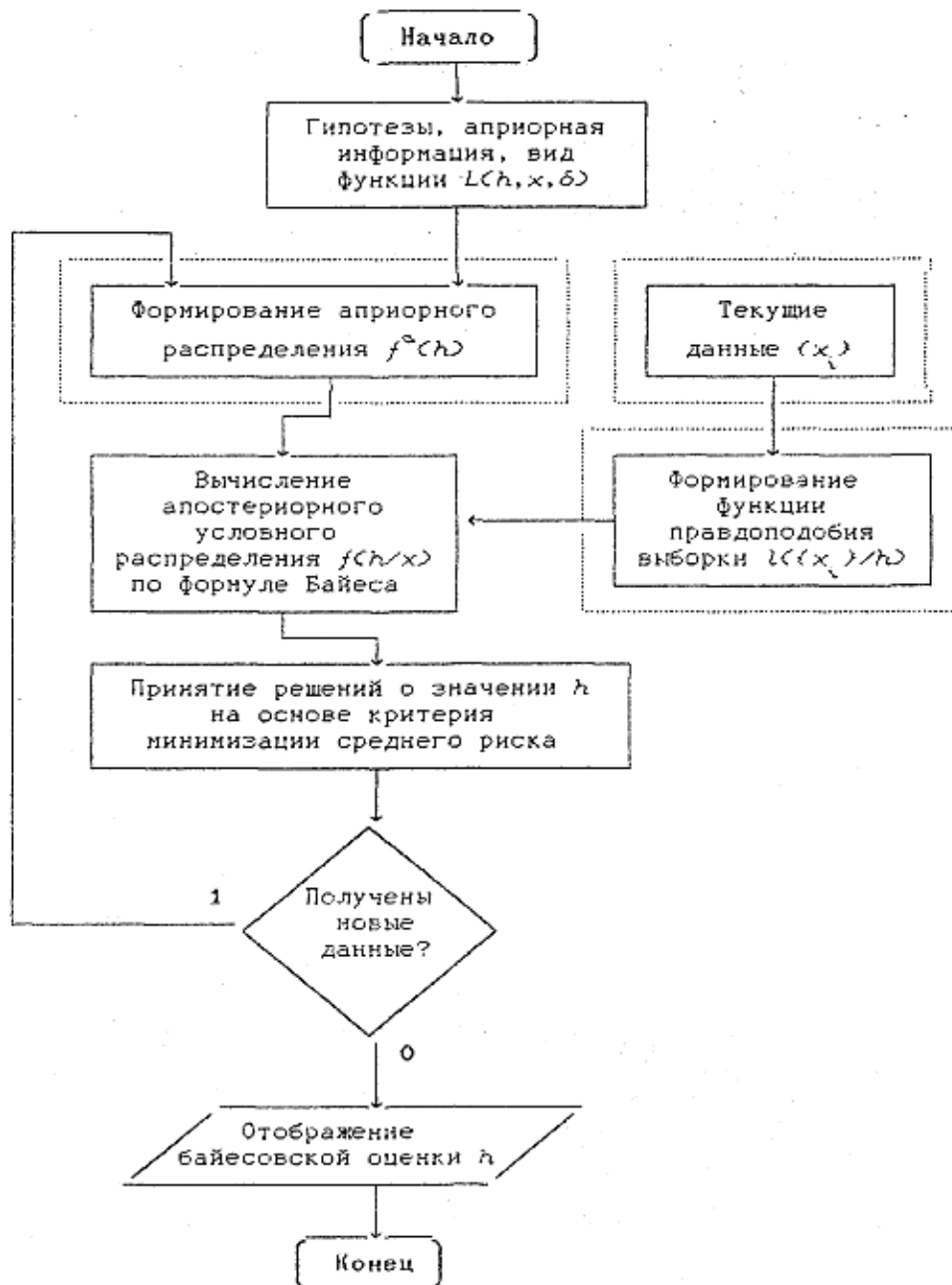


Рис. 1. Структурная схема классического байесовского вывода

С методической точки зрения задачи байесовского оценивания делятся на классы по трем признакам [1].

Первый из них определяет сложность структур признаков оцениваемого свойства. По этому признаку все задачи байесовского оценивания делятся на одноуровневые и задачи со сложной иерархией признаков определяемого показателя.

По второму признаку определяется степень определенности знаний об априорном распределении возможных решений из пространств H или H_K . Выделяются классы с полной априорной определенностью, когда априорное распределение известно точ-

но; с частичной априорной неопределенностью, когда распределение не задано, но заданы ограничения, допущения или значения функции априорного распределения; с полной априорной неопределенностью, когда имеется только выборка значений x_i или конечное число оценок определяемого параметра или характеристики. Третий тип задач решается, как правило, на основе эмпирического байесовского подхода Роббинса, адаптивного байесовского подхода, или методами имитационного моделирования.

Целесообразно отметить, что указанные классы задач полностью соответствуют трем типам информационных ситуаций, приведенных в [6]: с полной определенностью компонентов информационного процесса обработки информации; с частичной неопределенностью и с полной неопределенностью. Поэтому третий класс задач байесовского оценивания соответствует классу задач ИИ.

Третий признак делит задачи байесовского оценивания на задачи параметрические, когда задано параметрическое пространство решений, а модель распределения случайной величины X известна с точностью до определяемого параметра и непараметрические, когда для модели распределения известны лишь ограничения и допущения, характеризующие класс распределения (например, класс всех унимодальных распределений). Методы решения таких задач основываются на методах параметрической и структурной идентификации.

Уже приведенные свойства байесовского подхода могли бы послужить рекомендацией для выбора его в качестве критериальной основы принятия решений ИИ. Однако перечисленные возможности далеко не исчерпывают всех достоинств байесовского подхода, актуальных для решения задачи ИИ. К ним, несомненно, можно отнести следующие важные свойства подхода:

- подход обладает свойством внутреннего единства методологии, что делает его универсальным для применений в параметрической, функциональной и системной задачах ИИ и позволяет строить соответствующие шкалы иерархической структуры на основе единой методологии, обеспечивающей формализованную связь всех видов ИИ, системность и комплексность их результатов;

- подход максимизирует объективность и полноту решений, что важно для реализации задач ИИ, так как дает возможность учета всей совокупности данных и знаний, представленных в разнообразных формах, при получении результата ИИ;

- принципы подхода обеспечивают последовательное накопление и коррекцию знаний об объекте изучения, что позволяет реализовать основную функцию ИИ познания объекта и создает основу для самообучения средств ИИ;

- при использовании подхода в процессе длительного изучения объекта, характерного для задач мониторинга, имеется принципиальная возможность получения все более точных результатов, так как по мере накопления знаний байесовские решения концентрируются вокруг истинного решения (заметим, что данное положение вполне согласуется с принципом теории измерений о возможности постоянного увеличения точности измерений);

- на основе байесовского подхода имеется возможность перспективного и ретроспективного изучения свойств объекта на основе индуктивной логики, что особенно важно для получения объективных и достоверных прогнозов и восстановления истории развития объекта;

- в связи с предыдущими свойствами подход может быть эффективным для реализации его средствами интеллектуальных технологий на основе средств получения и обобщения знаний, в частности, в виде экспертных систем и интеллектуальных измерительных комплексов.

Однако при всех достоинствах байесовского подхода применить его в классическом виде или в виде его модификаций реализуемых на базе указанных методов, в средствах ИИ не удастся в связи с рядом причин.

Основная из них касается невозможности использования основной формулы байесовского подхода при отсутствии знаний о какой-либо составляющей этой формулы, то есть в условиях неопределенности. Вторая причина состоит в том, что хотя алгоритмы байесовского оценивания и учитывают статистический характер данных, используемых при принятии решений в виде вероятностной характеристики решения в соответствии с (1), но не обеспечивают устойчивости получаемых решений. В связи с тем, что задачам ИИ присуща значительная априорная неопределенность используются эмпирические или имитационные оценки априорной плотности распределения параметра h и условной плотности распределения случайной величины X . Как известно, такие оценки неустойчивы, что вносит в результаты байесовского оценивания неконтролируемую в известных методах дополнительную погрешность. Кроме того, точность оценивания контролируемого свойства также не определяется ввиду отсутствия меры ее определения и соответственно метрического представления функционального пространства байесовских решений. Показатели этих погрешностей должны входить в комплексы метрологических характеристик (КМХ) результатов байесовского оценивания. Таким образом, была необходима модификация байесовского подхода в направлении обеспечения возможности полного метрологического обоснования решений, что позволит реализовать основные принципы измерительного подхода на основе байесовской идеологии.

2. Теоретические основы регуляризирующего байесовского подхода.

Необходимые и достаточные условия реализации регуляризации методом шкалирования

Как уже указывалось, традиционные методы решения задач байесовского оценивания третьего типа основываются на применении методов эмпирического (экспериментального или экспертного) определения основных составляющих байесовского подхода. При этом в основном уравнении байесовского подхода, которое можно переписать в приближенном виде:

$$\tilde{P}(h_k|x_i) = \frac{\tilde{P}_{(h_k)}^{(a)} \cdot \tilde{l}(x_i|h_k)}{\sum_{j=1}^K \tilde{P}_{(h_j)}^{(a)} \cdot \tilde{l}(x_i|h_j)}; \quad (7)$$

по экспериментальным данным оценивается функция правдоподобия $\tilde{l}(x_i|h_k)$, а априорное распределение $\tilde{P}_{(h)}^{(a)}$ определяется по экспертным оценкам. По полученному апостериорному распределению $\tilde{F}(h|x_i)$ затем получают оценку изучаемой характеристики $\tilde{h}_k$, которая является решением приближенного уравнения вида:

$$h_k = \underset{h_k \in H_K}{\operatorname{argmin}} \tilde{r}(h, x_i), \quad (8)$$

где знак $\sim$ указывает на эмпирическое определение составляющей. Тогда задачу определения оптимального байесовского оценивания можно рассматривать как задачу восстановления измеряемой характеристики h по экспериментальной неполной информации, что, как показано в [6] делает задачу байесовского оценивания в математической постановке некорректной. Эта некорректность постановки задачи как обратной задачи восстановления значения параметра, вида функции, решения, полученных по неполной, неточной информации, присуща и всем вышеуказанным задачам ИИ. Таким образом, методы решения задач байесовского оценивания третьего типа, как и все методы решения задач ИИ, должны иметь регуляризирующие свойства.

Согласно фундаментальным работам по теории некорректных обратных задач Тихонова [14], некорректность постановки задачи состоит в нарушении хотя бы одного из трех условий ее корректности задачи по Адамару: 1) существования, 2) единственности 3) устойчивости решения. Для получения однозначного и устойчивого решения (8) задача байесовского оценивания должна быть введена в класс корректных, что определяет необходимость модификации подхода в этом направлении.

Привлечение измерительного подхода в процесс байесовского оценивания производится путем применения шкалирования, метризирующего пространство байесовских решений. Тогда байесовское решение может быть получено в виде одного из элементов такой шкалы. Таким образом, при синтезе шкал в системах ИИ необходимо учитывать указанные свойства решений некорректных задач ИИ.

Одним из способов обеспечения устойчивости решений является формирование системы дополнительных ограничений пространства решений. Очевидно, процесс шкалирования в ИИ, обуславливая дополнительные ограничения в пространстве результатов, будет способствовать регуляризации решений и введению задачи ИИ в класс условно-устойчивых по Тихонову [14].

Первоочередным вопросом синтеза измерительных шкал является вопрос формирования носителя шкалы, представляемого дискретным множеством результатов измерений $\{h_k\} \subset H_K$, принятых в качестве представителей градаций $\{q_i^*\}$ измеряемого свойства q_i , что позволяет рассматривать множество H_K как подмножество множества нешкалированных возможных результатов. Множество H_K реперов шкалы неплотно покрывает множество H , а система носителя шкалы $G_S^{(H)}$ является фрагментом или подсистемой системы G_H , соответствующей множеству результатов. Для реализации измерительной шкалы пространство ее носителя H_K должно быть метрическим, для которого определено понятие меры близости между отдельными элементами носителя шкалы. Тогда каждой паре h_s, h_{s+1} , элементов носителя шкалы должно соответствовать вещественное число $\rho_s = \rho(h_s, h_{s+1})$ удовлетворяющее аксиомам расстояния:

- неотрицательности и тождественности:

$$\forall h_s \in H_K; \rho(h_s, h_{s+1}) \geq 0;$$

$$\rho(h_s, h_{s+1}) = 0 \Leftrightarrow h_s = h_{s+1};$$

- симметрии:

$$\rho(h_s, h_{s+1}) = \rho(h_{s+1}, h_s);$$

- треугольника:

$$\rho(h_s, h_i) \leq \rho(h_s, h_{s+1}) + \rho(h_{s+1}, h_i).$$

Функция $\rho(h_s, h_{s+1})$ – метрика в пространстве H_K . Различные виды метрик достаточно полно изучены для различных типов пространства и представлены в литературе. Выбор конкретного вида метрики определяется свойствами пространства модельного представления измеряемого свойства и характеристиками задачи в целом.

Таким образом, формирование носителя шкалы сводится к представлению функционального пространства измеряемого свойства метрическим пространством, соответствующим множеству H_K , с метрикой, определяющей расстояние между соседними элементами носителя шкалы. При поставленных метрологических требованиях в эксперименте это расстояние не должно превышать некоторой величины $\rho_{\text{доп.}}$, определя-

емой требуемой точностью измерений $\xi_{\text{доп.}}$. Элементы множества H_K формируются в соответствии с условием:

$$\forall(h_s, h_{s+1}) \in H_K; \rho(h_s, h_{s+1}) \leq \rho_{\text{доп.}}(\xi_{\text{доп.}}). \quad (9)$$

Для регуляризации в условиях неопределенности был разработан новый тип шкал с динамическими ограничениями по диапазону, виду и числу реперов шкалы. Он подробно рассмотрен в ряде монографий статей и учебных пособий [6, 7].

Специфическим свойством шкалы с динамическими ограничениями (ШДО), как показано в указанных выше источниках, является свойство ее развития при меняющихся информационной ситуации, метрологических требованиях, допущениях и ограничениях. Для обеспечения этого носитель такой шкалы должен представлять собой систему с динамическими ограничениями вида. Реализация принципа "динамического погружения" в систему изучаемого или управляемого сложного объекта, как основного для ШДО, возможна только при обеспечении возможности непрерывного изменения расстояния $\rho_{\text{доп.}}(t)$ при изменении метрологических требований, например, относительной погрешности $\xi_{\text{доп.}}(t)$. Это приведет к изменению числа элементов множества $H_K(t)$; $K = K(t)$. Добавление новых контролируемых свойств при накоплении знаний возможно при построении системы носителя по иерархическому принципу, что, в свою очередь, определяет необходимость увеличения мерности B пространства H^B и множества носителя H_K^B , что для процесса интеллектуальной обработки информации, развивающегося во времени можно записать в виде:

$$H_K^B(t) \subset H^B(t); K = K(t, B); B = B(t). \quad (10)$$

Тогда выражение (9) для системы носителя шкалы с динамическими ограничениями может быть переписано в виде:

$$\forall(h_s^B(t) \in H_K^B(t); h_{s+1}^B(t) \in H_K^B(t)), \rho(h_s^B(t), h_{s+1}^B(t)) \leq \rho_{\text{доп.}}(\xi_{\text{доп.}}(t) \leq \varepsilon), \quad (11)$$

где ε – сколь угодно малое, наперед заданное число.

Определены необходимые условия для реализации (11) при синтезе носителя ШДО.

Для реализации (11), очевидно, необходимо, чтобы функциональное пространство измеряемого свойства $H^B(t)$ обладало свойствами:

- равномерной непрерывности – для обеспечения возможности монотонного изменения расстояния $\rho_{\text{доп.}}$ при изменении метрологических требований и ограничений, накладываемых на диапазон шкалы;
- равностепенной ограниченности – для обеспечения возможности представления функционального пространства шкалы конечным множеством элементов носителя шкалы.

Последнее требование, вообще говоря, для ШДО должно выполняться в каждый отдельный момент времени, а в течении времени для шкал типа ШДО при неограниченном "погружении" их системы в систему ОИ должно выполняться требование замкнутости при динамических ограничениях функционального пространства, то есть замкнутости на расширяющемся функциональном пространстве.

Как известно, указанные требования определяют компактное пространство (компакт). Очевидно, добавление третьего требования замкнутости на динамическом пространстве позволяет определить его как требование "динамического" компакта, под которым в данной работе понимается компактное пространство, границы которого изменяются во времени.

По известной теореме А.Н.Колмогорова компакт может быть покрыт конечной -сетью. Конечной -сетью называется множество H_K^ε , содержащее элементы $h_s^\varepsilon \in H_K^\varepsilon$, для которых для любого сколь угодно малого $\varepsilon > 0$ и элементе множества h_s^ε справедливо условие:

$$\rho(h_s^\varepsilon, h_{s+1}^\varepsilon) \leq \varepsilon, \forall h_s^\varepsilon \in H_K^\varepsilon, h_{s+1}^\varepsilon \in H_K^\varepsilon. \quad (12)$$

Очевидно, условие (12) для "динамического" компакта, при котором меняющееся его пространство сохраняет свойства компактности, может быть записано в виде:

$$\rho[h_s^\varepsilon(t); h_{s+1}^\varepsilon(t)] \leq \varepsilon, \forall (h_s^\varepsilon(t) \in H_K^\varepsilon(t), h_{s+1}^\varepsilon(t) \in H_K^\varepsilon(t)). \quad (13)$$

Для определения необходимых условий построения ШДО БИИ в [*] доказана следующая теорема.

Теорема 1.

Синтез шкалы с динамическими ограничениями возможен тогда и только тогда, когда функциональное пространство результатов измерений является компактом в смысле (13) при любых ограничениях области его определения во времени и в пространстве.

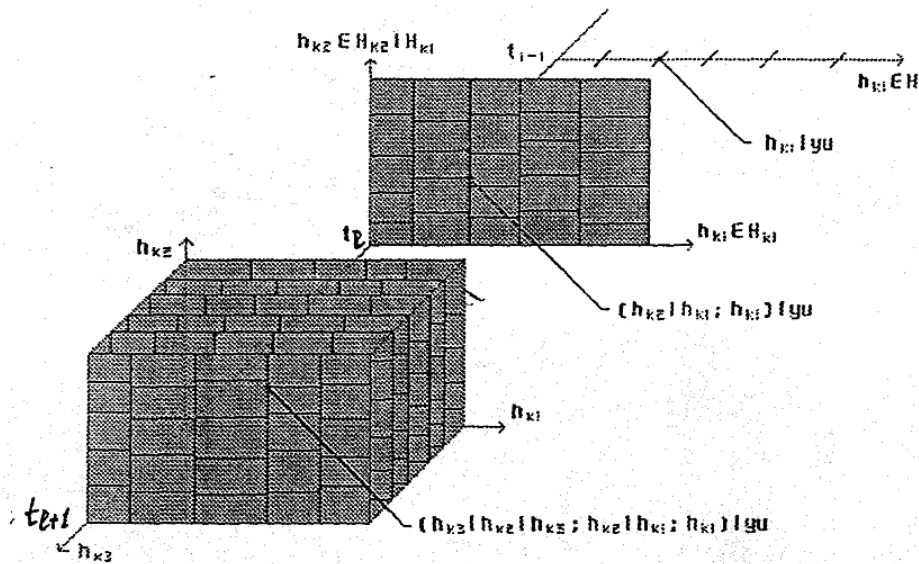


Рис. 2. Иллюстрация ШДО ИИ

Доказательство.

Существование компакта при любых условиях тождественно возможности покрытия в любой момент времени всего функционального пространства исходной информации или решений систем ИИ конечной -сетью вида (13). Выражение (13) полностью совпадает с выражением (11) для записи условия синтеза носителя шкалы с динамическими ограничениями при размерности его пространства B , равного единице, что и служит доказательством теоремы 1.

Из теоремы 1 следует, что существование ШДО доказывает компактность функционального пространства результатов измерений, которое может быть покрыто конечной -сетью.

Очевидно, что теорема 1 может быть распространена на многомерный вариант ШДО.

Многомерные шкалы ШДО ИИ имеют иерархическую структуру, и их существование доказывает возможность существования многомерного "динамического" компакта

как многомерного замкнутого монотонного функционального пространства с динамическими ограничениями, что, в свою очередь, доказывает возможность существования иерархической многомерной динамической -сети в виде:

$$\forall \left(h_s^{b_\varepsilon}(t); h_{s+1}^{b_\varepsilon}(t) \right) \in H_K^{B_\varepsilon}(t), \forall b \in B, \rho \left(h_s^{b_\varepsilon}(t), h_{s+1}^{b_\varepsilon}(t) \right) \leq \varepsilon. \quad (14)$$

Когда определяемое свойство имеет сложную структуру или определяется состояние объекта, характеризуемое рядом определяемых в процессе обработки признаков $\{\pi_i\} \in \pi, i = 1, I_\pi$, то необходимым условием синтеза носителя такой шкалы при коррелированных признаках является условие (14), в котором мерность пространства шкалы B определяется числом признаков I_π . При некоррелированных признаках ШДО представляет собой комплект одномерных шкал ШДО, число которых равно I_π .

Условие (14) справедливо при любой форме представления решения ИИ: параметрической, функциональной, системной или лингвистической.

3. Синтез обобщенного метода регуляризации(обеспечения устойчивости) решений ИИ на основе РБП

Для создания обобщенного метода обеспечения устойчивости решений ИИ применяется общая теория оптимальных решений и вышеприведенные рассуждения и зависимости. На основе развитой Вальдом общей теории решений учеными-байесовцами Де Гроотом, Де Финетти, Эвансом, и другими была создана теория оптимальных статистических байесовских решений, базирующаяся на основной идее теории Вальда допустимого решающего правила, максимизирующего полезность принимаемого решения. Для байесовских решающих правил максимальная полезность решений достигается при минимальном среднем риске r решения, который может быть выражен в виде:

$$R(h_k, x) = E \left(r_k \left(h_k, x, L \left(h_k, \delta_{k_j} \right) \right) \right) = \sum_{K=1}^K \int_X L \left(h_{k_1}, \delta_{k_j} \right) P_{(h_k)}^{(a)} f(x|h_k) dx,$$

где $f(x|h_k) \cdot P_{(h_k)}^{(a)} = f(h_k, x)$; $E(r)$ – символ математического ожидания случайной величины r .

Двумерный закон распределения выборочных данных $f(h, x)$ может быть представлен в виде композиции условных законов распределения выборочных данных или достаточных статистик $Z(x)$ при условии $h = h_k, f_{(h_k)}^{(a)}$:

$$f(h, x) = \prod_{K=1}^K f(x|h_k) P_{(h_k)}^{(a)}; \quad (15)$$

$$f(h, x) = \prod_{K=1}^K f(z(x)|h_k) P_{(h_k)}^{(a)}; \quad (16)$$

для независимых элементов носителя шкалы:

$$f(h, x) = \prod_{K=1}^K f(x|h_k) \cdot P_{(h_k)}^{(a)}; \quad (17)$$

$$f(h, z(x)) = \prod_{k=1}^K f(z(x)|h_k) \cdot P_{(h_k)}^{(a)}; \quad (18)$$

С учетом выражений (15) + (18) выражение для среднего риска решения принимает вид:

$$R(h_K, x) = \sum_{k=1}^K L(h_{k_1} \delta_{k_j}) f(z(x)|h_k) P_{(h_k)}^{(a)} dx. \quad (19)$$

Байесовские оценки определяемых свойств изучаемого объекта представляют собой результаты оптимальных решений или выводов, при которых минимизируются ожидаемые проигрыши, если полученная оценка не будет соответствовать истинному ее значению.

Для получения таких оценок синтезируются оптимальные решающие правила, обеспечивающие выполнение условия:

$$r_k \left(P_{h_k}^{(a)}, \gamma^*(x, h_k, \delta_{k_j}) \right) \leq r_j \left(P_{h_j}^{(a)}, \gamma(x, h_j, \delta_{j_i}) \right), \quad (20)$$

$$\forall j \neq K; j = \overline{1, (K-1)}; i = \overline{1, K},$$

где $\{\gamma^*(x, h_j, \delta_{j_i})\}$ – решающая функция, являющаяся равномерно лучшей для всего множества H_K и всей совокупности решающих функций $\{\gamma(x, h_j, \delta_{j_i})\}$ определенных для класса решающих функций Γ , осуществляет отображение экспериментальной информации X в решение h_k , что может быть отражено записью:

$$H_K = \{h_k\} = \Gamma \left(X | P_{(h)}^{(a)}, L(h, \delta) \right) = \left\{ \gamma(x_i | P_{h_k}^{(a)}), L(h_k, \delta_{k_j}) \right\}. \quad (21)$$

Это позволяет определить класс решающих функций Γ как класс алгоритмов ИИ, а результатом ИИ считать байесовское решение или вывод, полученные при использовании оптимальной решающей байесовской функции γ^* и удовлетворяющие условию (20).

При такой интерпретации уравнения (21) оптимальному решающему правилу $\gamma^*(h_k, x, \delta_{k_j})$ соответствует оптимизирующий алгоритм ИИ $\varphi_j^{(\text{опт})}$, принадлежащий множеству алгоритмов Φ_j , которое определится классом решающих правил Γ . Тогда процесс получения решений ИИ можно рассматривать как процесс нахождения и принятия оптимального решения на множестве решений H_K по критерию минимума среднего риска (19).

При отсутствии знаний о законах распределения их аналитические зависимости в формулах (15) ÷ (19) заменяются эмпирическими оценками $\tilde{f}(x|h_k), \tilde{f}(z(x)|h_k)$, что приводит к неустойчивости получаемых решений и вносит дополнительные погрешности в получаемые решения или результаты байесовского оценивания. Величины этих погрешностей зависят от закона распределения основной СВ или СЭСП $f(x)$, условных распределений выборочных значений $f(\tilde{x}|h_k)$, точных выборочных распределений оцениваемых достаточных статистик $f[\tilde{z}(x_i)]$.

Условная устойчивость решений БИИ на компакте шкалы, обеспечиваемая принципами синтеза шкал ШДО БИИ позволяет назвать методы БИИ и алгоритмы, построенные на их основе регуляризирующими с параметром регуляризации ξ , определяю-

щим семейство регуляризирующих алгоритмов $\{Y_j^\xi\}$. Параметр ξ определяет потенциальную точность решений БИИ, что подробно рассмотрено в [6]. Это означает, что, если применить данный подход к регуляризации исходной информации систем ИИ и к пространству получаемых решений на основе заданного класса функциональных преобразований, то использование вышеприведенной теории позволяет создавать методы и алгоритмы ИИ, обладающие регуляризирующими свойствами, а решения получаемые при этом будут квази-устойчивыми или устойчивыми в пределах принятых ограничений и допущений, а также созданных динамических компактов.

Практическая реализация регуляризирующих методов ИИ на основе РБП заключается в формировании такого класса решающих правил Γ_s , который бы в пространстве регуляризирующей шкалы ШДО позволял бы осуществлять преобразование вида (21) для любого набора x_i экспериментальной информации из X_I и условий получения исходной информации и ее свойств $y_i \in Y_I$. Эта задача сводится к построению на основе РБП множества границ $H_K^{(\gamma)}$:

$$H_K^{(\gamma)} = \{h_{\gamma_k}\}, \gamma_k = \overline{1, I_\gamma}; h_{\gamma_k}: h_k \leq h_{\gamma_k} < h_{k+1}; \forall h_k \in H_K. \quad (22)$$

Границы $\{h_{\gamma_k}\}$ разделяют элементы носителя шкалы в соответствии с условиями (20) и (13). При построении разделяющих границ возможны условия, когда разделяющая граница между элементами носителя и может быть неединственной или вообще не может быть определена. Эти варианты иллюстрирует [рисунок 3](#) а), б), в). Первый вариант возможен при слишком большом значении параметра регуляризации $\xi_{\text{доп.}}$ алгоритма БИИ (слишком "грубой" шкале БИИ), а второй – при слишком малом значении $\xi_{\text{доп.}}$.

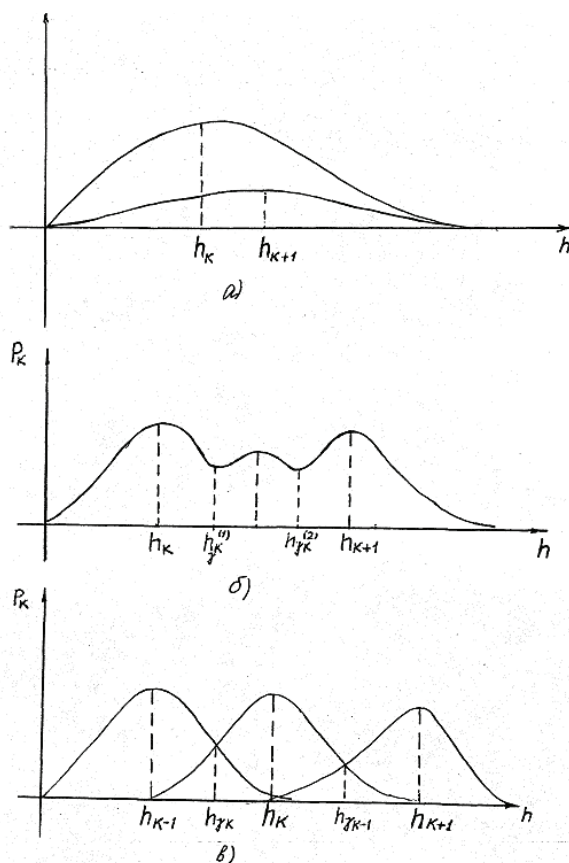


Рис. 3. Варианты ситуаций при построении разделяющих границ решений по алгоритму БИИ

Необходимые и достаточные условия возможности синтеза алгоритма ИИ на основе РБП (построения решающего байесовского правила на измерительных принципах), вводящего задачу ИИ в класс корректных по Тихонову, и, тем самым, обеспечивающего получение условно-устойчивых решений ИИ, определяются Теоремой 2, доказательство которой приведено в [6].

Теорема 2.

Синтез регуляризирующего алгоритма на основе РБП, обеспечивающего введение задачи в класс условно-корректных по Тихонову с заданными метрологическими требованиями M_i с и функцией полезности L в условиях имеющейся априорной информации A_i и ограничений O_i возможен тогда и только тогда, когда для каждой пары соседних элементов носителя шкалы h_k и h_{k+1} найдется только одна точка h_{γ_k} , доставляющая минимум риска решения r_k , что может быть записано в формализованном виде:

$$\forall(h_k, h_{k+1}) \in H_K; \forall y_i \in Y_i; \exists h_{\gamma_k}: r_k[(h_{\gamma_k})|h_k|Y_i] = r_{k+1}[(h_{\gamma_k})|h_{k+1}|Y_i]. \quad (23)$$

Имея в виду асимптотическую нормальность распределения многих статистик при простой платежной матрице $L(h_k, \delta_{kj})$ и равномерном априорном распределении можно записать выражение для разделяющих границ шкалы по РБП в виде:

$$h_{\gamma_k} = \frac{h_{(k+1)}\sigma_k^2 - h_k\sigma_{k+1}^2 - \sigma_k\sigma_{k+1} \sqrt{(h_{k+1} - h_k)^2 + 2(\sigma_k - \sigma_{k+1})^2 \ln \frac{P_{k+1}^{(a)}\sigma_k}{P_k^{(a)}\sigma_{k+1}}}}{h_{k+1} - h_k}, \quad (24)$$

где σ_k – среднеквадратическое отклонение выборочного распределения статистики $\tilde{z}(x)$; $P_{k+1}^{(a)} = P^{(a)}(h_{k+1})$.

Для многомерных шкал БИИ формула (24) будет иметь следующий вид:

$$h_{\gamma_k}^{(b)} = h_{\gamma_k}^{(b)} | h_{\gamma_k}^{(b-1)} \dots | h_{\gamma_k}^{(1)}, \quad (25)$$

где $h_{\gamma_k}^{(b)}$ – условная разделяющая граница на b -ой координате многомерной шкалы, вычисленная по (24) при условии, что разделяющие границы по остальным координатам шкалы вычисляются также в соответствии с (24) и (25).

В литературе дано строгое доказательство асимптотической нормальности апостериорного распределения байесовской оценки, причем погрешность аппроксимации истинного апостериорного распределения нормальным не превосходит величины порядка $N^{-1/2}$, где N – объем выборочных данных, что делает формулы (24), (25) пригодными для широкого класса статистик и дает возможность учесть погрешность такой аппроксимации в общей оценке погрешности БИИ.

Применение РБП для построения разделяющих границ на одномерной и двумерной шкалах БИИ при простой штрафной функции иллюстрирует [рисунок 4](#). Таким образом, по методу БИИ в качестве результата принимается решение о соответствии градации измеряемого свойства q_i^* элементу носителя h_k , если выполняется условие (23) в виде:

$$h_k | P_k = \underset{h_k \in H_K}{\operatorname{argmin}} C[\varphi_j(x_i | \{h_k\}; \{h_{\gamma_k}\}; Y_i)]. \quad (26)$$

Таким образом, решение ИИ в (24) представлено элементом множества носителя шкалы, апостериорная вероятность P_k , для которого максимальна среди всех элементов H_K , что обеспечивает выполнение условия (23).

Однако, уместно заметить, что это не единственно возможный результат в условиях данной информационной ситуации. Учитывая, что решение ИИ является, согласно выводам предыдущего параграфа условно устойчивым, очевидна вероятность существования решений, не являющихся оптимальными в смысле условия (23), но имеющих апостериорную вероятность появления, отличную от нуля. В связи с неопределенностью априорной информации о $P_{(h)}^a$, вида функции правдоподобия, определяемой распределением вероятностей экспериментальных данных $F(x)$ и обусловленной этим некорректности математической постановки задач ИИ, даже в условиях построения методов ИИ на основе РБП может сохраняться многозначность результатов измерений, которая определяется видом апостериорного распределения $F(h|x_i, y_i)$ градаций измеряемого свойства. При этом наиболее вероятными оказываются решения, определяемые соседними с выбранным элементами носителя шкалы БИИ. На рисунке 4 эти решения обозначены как h_{s-1} и h_{s+1} .

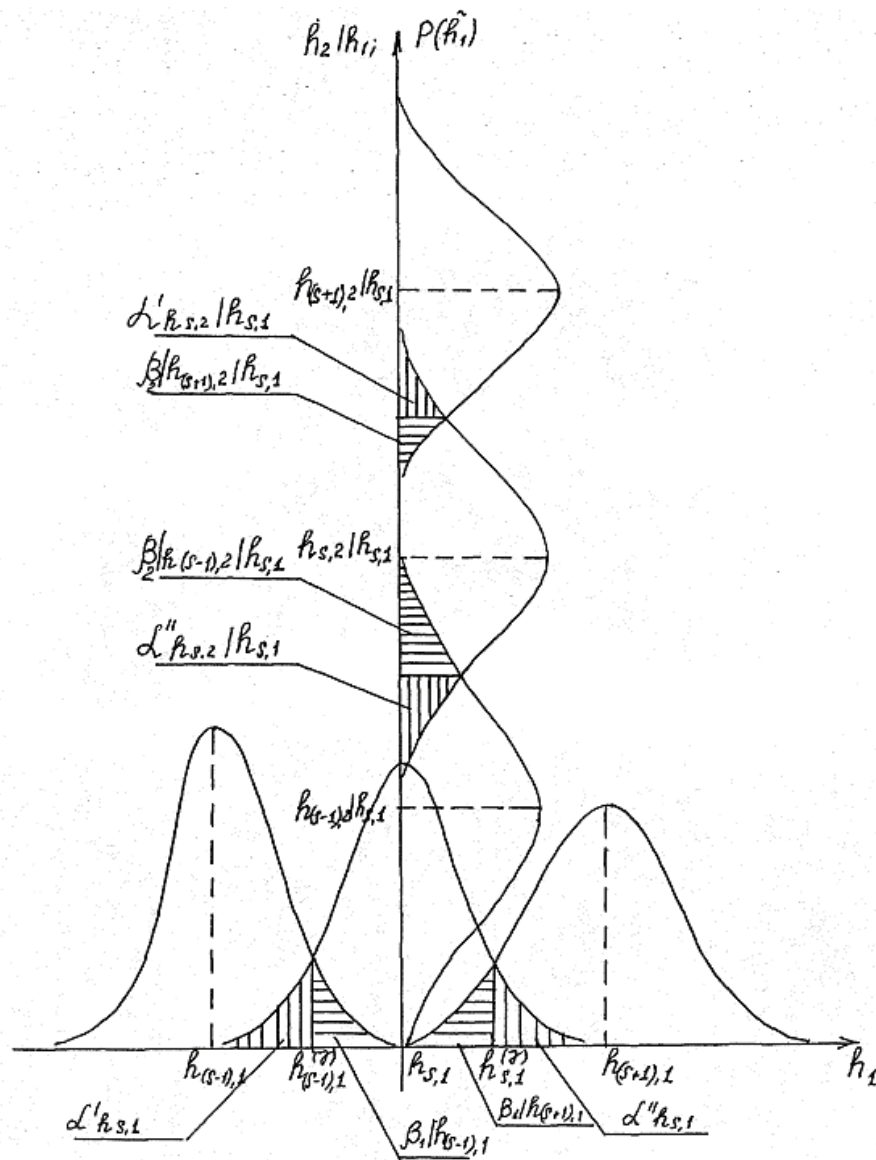


Рис. 4. Фрагмент двумерной ШДО БИИ

Причем, чем больше расстояние между элементами носителя, тем меньше число претендентов на возможные альтернативные решения при тех же условиях.

Формирование возможных решений в данной информационной ситуации ведется по модифицированному БРП на основе критерия, который может быть записан в виде:

$$E(\{r_{i_r}(h_{i_r})\}|x_i, y_i) \leq E(\{r_{j_r}(h_{j_r})\}|x_i, y_i), \quad (27)$$

$$\forall h_{i_r} \in H_{I_r} \subseteq H_K; \forall h_{j_r} \in H_{J_r} \subseteq H_K,$$

где $i_r = \overline{1, I_r}$, I_r – число возможных решений, апостериорная вероятность которых в данных условиях эксперимента отлична от нуля: $j_r = \overline{1, J_r}$; $J_r = C_{H_K}^{I_r}$ – число сочетаний возможных подмножеств решений, исключая подмножество H_{I_r} .

Решающему правилу с критерием (27) соответствует алгоритм ИИ φ_{j, i_r}^* , входящий в подмножество алгоритмов такого типа Φ_{J, I_r} , которое определяется подклассом решающих функций $\Gamma^* \subseteq \Gamma$ класса Γ . Этот алгоритм принадлежит классу алгоритмов ИИ с уравнением измерения в виде:

$$\{h_{i_r}|P_{i_r}\}_{i_r = \overline{1, I_r}} = \left\{ \left(\underset{K_{i_1}=1, K}{\operatorname{argmin}} r_{K_{i_1}} \right); \left(\underset{\substack{K_{i_2}=1, K-1; \\ K_{i_2} \neq i_r}}{\operatorname{argmin}} r_{K_{i_2}} \right); \dots \left(\underset{\substack{K_{i_{I_r}}=1, K-I_r+1 \\ K_{i_{I_r}} \neq i_r \\ \dots \\ K_{i_{I_r}} \neq I_r-1}}{\operatorname{argmin}} r_{K_{i_{I_r}}} \right) \right\} | Y_i. \quad (28)$$

Такому результату БИИ соответствует список возможных решений $\{h_{i_r}|P_{i_r}\}$, длиной I_r , упорядоченный по величине апостериорной вероятности P_K , который является полным решением БИИ. Результат, получаемый по (26) является частным решением БИИ.

Как видно из выражения (26), алгоритм БИИ синтезируется в процессе обработки информации в соответствии с конкретными информационными условиями Y_i , определяемыми A_i, M_i, O_i . Изменение этих составляющих информационной ситуации, следуя уравнению измерений (26) влечет за собой изменение структуры алгоритма, которое адаптирует его к новым условиям, что является основой для его самоорганизации и саморазвития базы алгоритмических знаний систем ИИ.

Практическими примерами реализации алгоритмов искусственного интеллекта, созданными на приведенных в этой статье методологических основах и принципах могут быть прикладные системы, построенные на основе платформы "Инфоаналитик".

Центральная часть платформы "Инфоаналитик" организована в виде байесовской сверточной нейросети, которая настраивается в процессе обучения на решение задач мониторинга, контроля, управления сложными системами и поддержки принятия управленческих решений. Конкретные прикладные системы, построенные на основе данной методологии и указанной платформы разработаны и применяются в управлении промышленными, энергетическими и социально-экономическими системами [13].

На рисунке 5 приведен пример регуляризирующей сопряженной шкалы типа ШДО, где на верхней части шкалы производится регуляризация измерительной информации, поступающей от датчиков, а нижняя часть шкалы используется для регуляризации решений, состоящих в нормировании показаний датчиков.

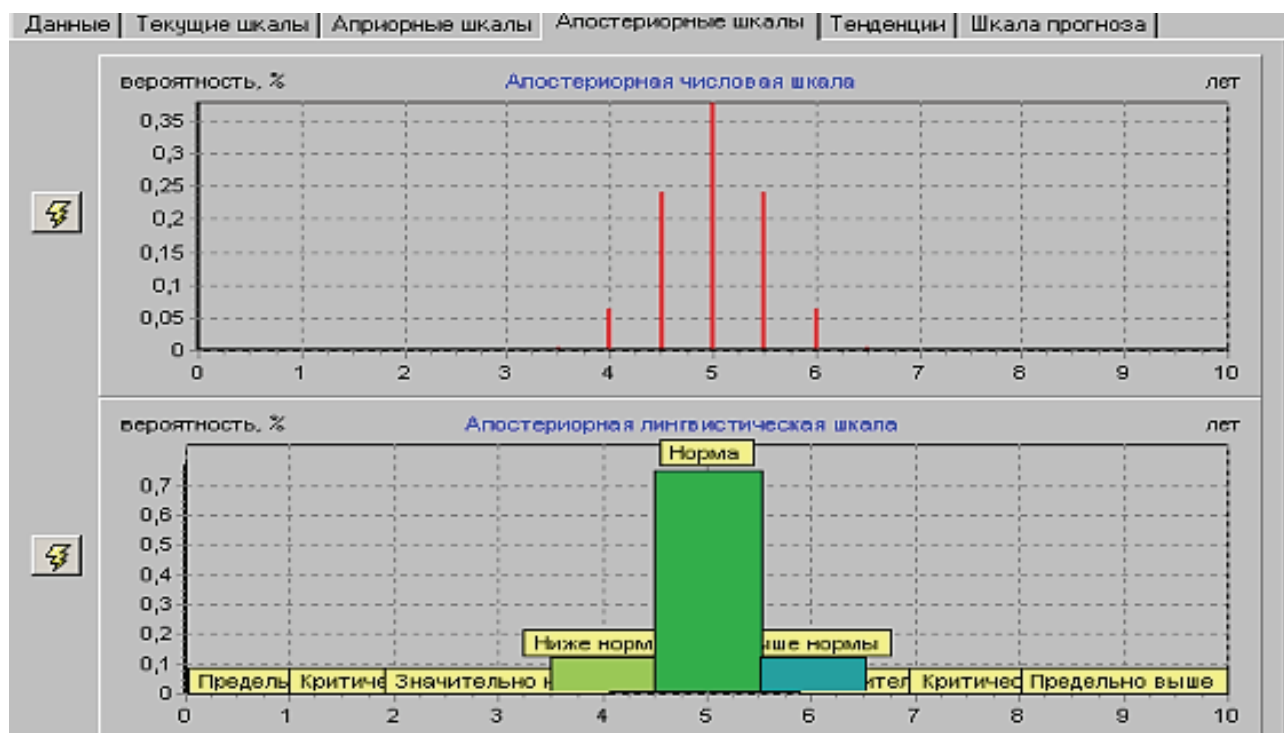


Рис. 5. Пример шкалы ШДО для регуляризации исходной информации и решений

Этот пример иллюстрирует самый простой процесс интеллектуальной обработки информации. Если используются сложные функциональные преобразования исходной информации, то применяется иерархически выстроенное дерево шкал согласно структуре системы объекта.

Заключение

В статье показана актуальность обеспечения устойчивости решений систем искусственного интеллекта. Предложен методологический подход и рассмотрены теоретические основы для создания методов искусственного интеллекта, обладающих регуляризирующими свойствами и обеспечивающих получение устойчивых решений задач в условиях значительной неопределенности. Методология базируется на регуляризирующем байесовском подходе и технологиях шкалирования на его основе. Приведены теоремы и аналитические зависимости для создания регуляризирующих алгоритмов искусственного интеллекта.

Список источников

- [1] Савчук В.П. Байесовские методы статистического оценивания. Надежность технических объектов. М.: Наука, 1989. 328 с.
- [2] Handbook of research on Ambient Intelligence and Smart Environments: Trends and Perspectives. Ed. By N.-Y. Chong and F. Mastrogiovanny. IGI, Global, 2011.
- [3] Evans R.A. Bayes: in theory and practice. *The theory and application of reliability*, vol. 2. New York: Acad. Press. Inc., 1977. Pp. 50–54.
- [4] Good I.J. Some history of the hierarchical Bayesian methodology. *Trab. Est. Invest*, 1980, vol. 31, no. 1, pp. 489–504.
- [5] Good I.J. The estimation of probabilities, an essay on modern Bayesian methods. Wiley, 1965. 110 p.
- [6] Прокопчина С.В. Интеллектуальные измерения на основе регуляризирующего байесовского подхода. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2019. 508 с.

- [7] Прокопчина С.В. Основы теории шкалирования в экономике. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2021. 299 с.
- [8] Prokopchina S. New trends in measurement theory: Bayesian intelligent measurement and its application in the digital economy. *CEUR Workshop Proceedings*, 2020, 2782, pp. 80–88.
- [9] 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.
- [10] Прокопчина С.В. О подходе к измерению социогуманитарных потенциалов с использованием байесовских интеллектуальных технологий // Государственный аудит. Право. Экономика. 2013. № 3. С. 73–82.
- [11] Прокопчина С.В. Реализация принципов концепций DATA SCIENCE, BIG DATA, DATA MINING средствами платформы "ИНФОАНАЛИТИК" в применении к задачам цифровизации // Мягкие измерения и вычисления. 2020. Т. 30. № 5. С. 21–31.
- [12] Прокопчина С.В. Реализация принципов экономики знаний на основе байесовских интеллектуальных технологий // Сб. докладов Междунар. конф. по мягким вычислениям и измерениям "SCM-2008". 23–25 июня 2008. Т. 1. Санкт-Петербург, 2008.
- [13] Прокопчина С.В., Щербаков Г.А., Ефимов Ю.В. Моделирование экономических систем в условиях неопределенности: учебное пособие – практикум, 2-е издание / под. ред. Г.А. Щербакова. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2019. 508 с.
- [14] Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1979. 288 с.
- [15] Иванов В.С., Зуев А.Н. Контроль качества продукции в машиностроении. М.: Машиностроение, 1990. 96 с.

References

- [1] Savchuk V.P. (1989) Bayesian methods of statistical estimation. Reliability of technical objects. М.: Nauka, 328 p.
- [2] Handbook of research on Ambient Intelligence and Smart Environments: Trends and Perspectives. Ed. By N.-Y. Chong and F. Mastrogiovanny. IGI, Global, 2011.
- [3] Evans R.A. (1977) Bayes: in theory and practice. *The theory and application of realibility*, vol. 2. New York: Acad. Press. Inc., pp. 50–54.
- [4] Good I.J. (1980) Some history of the hierarchical Bayesian methodology. *Trab. Est. Invest.*, vol. 31, no. 1, pp. 489–504.
- [5] Good I.J. (1965) The estimation of probabilities, an essay on modern Bayesian methods. Wiley, 110 p.
- [6] Prokopchina S.V. (2019) Intellectual measurements based on the regularizing Bayesian approach. М.: Publishing House "SCIENTIFIC LIBRARY", 508 p.
- [7] Prokopchina S.V. (2021) Fundamentals of scaling theory in economics. М.: Publishing House "SCIENTIFIC LIBRARY", 299 p.
- [8] Prokopchina S. (2020) New trends in measurement theory: Bayesian intelligent measurement and its application in the digital economy. *CEUR Workshop Proceedings*, 2782, pp. 80–88.
- [9] Prokopchina S.V. (2020) A new type of neural networks: Bayesian measuring neural networks (BIN) based on the methodology of the regularizing Bayesian approach. *Soft Measurements and Calculations*, vol. 35, no. 10, pp. 17–24.
- [10] Prokopchina S.V. (2013) On the approach to measuring socio-humanitarian potentials using Bayesian intelligent technologies. *State Audit. Right. Economy*, no. 3, pp. 73–82.
- [11] Prokopchina S.V. (2020) Implementation of the principles of the concepts of DATA SCIENCE, BIG DATA, DATA MINING by means of the INFOANALYTIC platform in application to digitalization tasks. *Soft measurements and Calculations*, vol. 30, no. 5, pp. 21–31.

-
- [12] *Prokopchina S.V.* (2008) Implementation of the principles of the knowledge economy based on Bayesian intelligent technologies. *Collection of reports of the International Conference on Soft Computing and Measurements "SCM-2008"*, June 23–25, 2008, vol. 1, St. Petersburg.
- [13] *Prokopchina S.V., Shcherbakov G.A., Efimov Yu.V.* (2019) Modeling of economic systems in conditions of uncertainty: textbook, workshop, 2nd edition ; edited by G.A. Shcherbakov. M.: Publishing House "SCIENTIFIC LIBRARY", 508 p.
- [14] *Tikhonov A.N., Arsenin V.Ya.* (1979) Methods of solving ill-posed problems. M.: Nauka, 288 p.
- [15] *Ivanov B.C., Zuev A.N.* (1990) Product quality control in mechanical engineering. M.: Mechanical Engineering, 96 p.
-

Статья поступила в редакцию 10.07.2023; одобрена после рецензирования 17.07.2023; принята к публикации 24.07.2023.

The article was submitted 10.07.2023; approved after reviewing 17.07.2023; accepted for publication 24.07.2023.
