

«Что делает мозг?», или Решение априорно не определенных проблем на основе синтеза поведения автономного агента в реальной среде

З. В. Нагоев^{1,2}, О. В. Нагоева^{✉2}

¹Кабардино-Балкарский научный центр Российской академии наук
360010, Россия, г. Нальчик, ул. Балкарова, 2

²Институт информатики и проблем регионального управления –
филиал Кабардино-Балкарского научного центра Российской академии наук
360000, Россия, г. Нальчик, ул. И. Арманд, 37-а

Аннотация. Актуальность работы определяется необходимостью разработки теоретических основ создания систем универсального искусственного интеллекта.

Цель исследования – создание концептуальной модели системной сущности мозга.

Методы исследования. Принципиальное значение имеет то, что как метаалгоритмы онтологизации, так и алгоритмы идентификации и решения проблем выполняют синтез точек и отрезков в траектории движения агента в пространстве поведения. Это означает, что онтологизация, идентификация и решение всех проблем, с которыми сталкивается агент в реальной среде в течение своей экзистенции, выполняются с помощью синтеза и реализации поведения агента в этой среде.

Результаты. Построена гипотеза о системной целевой функции мозга, разработана вычислительная абстракция базовых процессов ее обеспечения на основе метафоры проектирования нейрокогнитивной системы управления поведением автономного агента в реальной среде. Предложена концепция экзистенции агента как траектории его движения в пространстве поведения, являющемся математической абстракцией пространства состояний системы «агент – среда – наблюдатель». Предложена концепция адаптации агента к условиям реального мира на основе реализуемых мозгом метаалгоритмов онтологизации и алгоритмов идентификации и решения априорно не определенных проблем на базе метафоры проектирования автономного синтеза траектории движения агента в пространстве поведения.

Заключение. Разработанные концепции обосновывают представление об онтологизации, идентификации и решении проблем агентом в реальной среде как о процессах автономного синтеза его поведения, что создает теоретический базис для создания систем универсального искусственного интеллекта.

Ключевые слова: универсальный искусственный интеллект, мозг человека, модели мышления, мультиагентные системы, нейрокогнитивные архитектуры, решение априорно не определенных проблем

Поступила 07.11.2025, одобрена после рецензирования 24.11.2025, принята к публикации 10.12.2025

Для цитирования. Нагоев З. В., Нагоева О. В. «Что делает мозг?», или Решение априорно не определенных проблем на основе синтеза поведения автономного агента в реальной среде // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 6. С. 172–185. DOI: 10.35330/1991-6639-2025-27-6-172-185

“What does the brain do?”, or Solving a priori undefined problems based on an autonomous agent behavior in real environment

Z.V. Nagoev^{1,2}, O.V. Nagoeva^{✉2}

¹Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
2, Balkarov street, Nalchik, 360010, Russia

²Institute of Computer Science and Problems of Regional Management – branch of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
37-a, I. Armand street, Nalchik, 360000, Russia

Abstract. The relevance of the work is determined by the need to develop theoretical foundations for the creation of universal artificial intelligence systems.

Aim. This work is to create a conceptual model of the systemic essence of the brain.

Research methods. Of fundamental importance is that both ontology meta-matching technique and problem-solving algorithms synthesize points and segments along an agent's movement in a behavioral space. This means that ontologization, identification, and solution of all problems encountered by the agent in the real environment during its existence are accomplished through the synthesis and implementation of the agent's behavior in that environment.

Results. A hypothesis about the systematic target function of the brain has been formulated, and a computational model of the basic processes underlying its operation has been developed, based on the analogy of designing a neurocognitive system to control the behavior of an autonomous agent in a real-world environment. A concept of an agent's existence as a trajectory in its behavioral space was proposed, which is a mathematical abstraction of the state space of the "agent – environment – observer" system. A concept for adapting agents to real-world conditions has been proposed, based on brain-implemented meta-algorithms for ontologization and algorithms for identifying and solving a priori undefined problems, grounded on a metaphor for designing an autonomous synthesis of the agent's trajectory in behavioral space.

Conclusion. The developed concepts support the idea of ontologization, identification, and problem-solving by an agent in a real-world environment as processes of autonomous behavior synthesis, which provides a theoretical foundation for the development of universal artificial intelligence systems.

Keywords: general artificial intelligence, human brain, thought models, multi-agent systems, neurocognitive architectures, solving a priori undefined problems

Submitted 07.11.2025,

approved after reviewing 24.11.2025,

accepted for publication 10.12.2025

For citation. Nagoev Z.V., Nagoeva O.V. “What does the brain do?”, or Solving a priori undefined problems based on an autonomous agent behavior in real environment. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2025. Vol. 27. No. 6. Pp. 172–185. DOI: 10.35330/1991-6639-2025-27-6-172-185

ВВЕДЕНИЕ

Вопрос «Что делает мозг?» в контексте знаменитого высказывания Марвина Минского: «The Mind is What the Brain Does!» («Разум есть то, что делает мозг!») [1], аргументирующего его позицию по вопросу об объекте моделирования искусственного интеллекта, адресуясь в некоторой степени и к результатам деятельности мозга, и к характеру ее протекания, прежде всего, по нашему мнению, относится к системному назначению мозга, к самой его сущности. Создав удачной метафорой контекст нового конструктивного научно-популярного нарратива в искусственном интеллекте, М. Минский, тем не менее, как и большинство современных исследователей и визионеров, признавал, что единого взгляда ни на системное назначение мозга, ни на результаты, ни на способы его функционирования современная наука пока не сформировала [2].

В попытке определить сущность мозга рассмотрим его в системном контексте и как систему. Оговоримся также, что, как и Минского, вопрос «Что делает мозг?» интересует нас прежде всего в контексте решения задачи синтеза *in silico* имитационной модели функциональных феноменологических комплексов мозга.

Целью исследования является создание концептуальной модели системной сущности мозга.

Объектом исследования является системная сущность мозга.

Предмет исследования – выявление основной системной целевой функции мозга.

Задача исследования – разработка вычислительной абстракции базовых процессов обеспечения основной целевой функции мозга.

СИСТЕМНАЯ СУЩНОСТЬ МОЗГА

Для понимания системной сущности мозга необходимо прежде всего определить основной функционал его предназначения. В силу того, что мозг является эволюционно сформировавшимся органом в составе тела человека, очевидно, его назначение необходимо рассматривать в контексте основной функции, выполняемой им в организме. В этом проявляется системный контекст, в котором функционирует мозг, – он часть организма и что бы он ни делал, он делает это для организма, выполняя некоторую базовую целевую функцию.

Без сомнения, одной из центральных базовых функций мозга является управление поведением человека в реальной среде, в которую этот человек погружен. В этом смысле проблема системной сущности мозга неотделима от проблемы системной сущности поведения человека.

Здесь необходимо вспомнить буквальный смысл высказывания Минского: по нему получается, что системная сущность того, что делает мозг, есть системная сущность того, что есть разум. Следует также оговориться, что в контексте дела исследования и создания искусственного интеллекта, которому Минский посвятил свою научную жизнь, мы предполагаем, что под разумом он понимал именно интеллект и именно общего типа («общий» интеллект). Таким образом, в данной работе мы отождествляем понятия «разум» и «общий» интеллект, обращая внимание на необходимость учитывать некоторые противоречия, в которые такой подход вступает с прецедентами использования частных значений полисемичных терминов «разум» и «интеллект», отождествления либо связывания феноменологических комплексов, обозначаемых этими терминами, отношениями вложенности (интеллект как разум, разум как часть интеллекта, интеллект как часть разума), а также сложившуюся в психологии традицию под «общим» интеллектом понимать общие когнитивные способности.

В [3] выдвигается гипотеза о том, что с целью построения вычислительной абстракции общего интеллекта необходимо рассматривать человека как материального (телесного) агента, погруженного в материальную внешнюю среду и стремящегося максимизировать некоторую сложную целевую функцию, вид которой существенным образом зависит от состояния базы знаний, используемой системой управления поведением такого агента. Мозг же (как «главная» часть нервной системы), в свою очередь, как раз является такой системой управления. В этом смысле прямым ответом на вопрос «Что делает мозг?» является утверждение о том, что мозг управляет поведением биологического агента в реальном мире.

Вероятнее всего, такое утверждение большинство читателей справедливо посчитает тривиальным, однако для нас здесь важно, что в контексте отождествления сущности интеллекта и функционала мозга (*The Mind is What the Brain Does*) его можно перефразировать: «Сущность общего интеллекта есть управление поведением биологического агента в реальной среде». В [3] дается классификация таких агентов, построенная на их отношении к некоторым наборам структурно-функциональных свойств, определяемых качественными изменениями

в производительности и эффективности системы управления поведением агента в реальной среде, детерминированных метасистемными переходами в эволюции видов.

Таким образом, мы рассматриваем сущность общего интеллекта (системную сущность мозга) в контексте целевого назначения управления поведением интеллектуального биологического агента в реальной среде. Эта функция реализуется путем синтеза закона управления с помощью системы управления – мозга этого агента.

АГЕНС, БИОС, НАБЛЮДАТЕЛЬ, ЭКЗИСТЕНЦИЯ

Рассматривая системные цели биологического агента в эволюционном аспекте, необходимо решить вопрос о системных основаниях *биоса* и *агенса* как таковых. В общем случае отличительным свойством *агентных систем* (*агентов*) является способность самостоятельно совершать действия во внешней для них среде, в которую они погружены с помощью своих *сенсоров* и *эффекторов*, или, иными словами, способность синтезировать свое поведение в этой среде. Возникновение таких систем в процессе эволюции материи, в соответствии с нашей гипотезой, связано с обособлением в их структуре подсистемы энергетического обеспечения синтеза и выполнения поведения, а также самой подсистемы синтеза поведения, или, более точно, – подсистемы управления поведением агента, способной выбирать варианты поведения и тем самым опосредованно управлять направлениями и способами диссипации собственной (аккумулированной в резервуаре энергии, являющейся частью тела агента) энергии [4].

Биологические системы, в свою очередь, в нашем понимании являются агентными системами, поведение которых подчинено целевому критерию максимизации продолжительности времени своего собственного существования (экзистенции). В контексте вышеприведенных рассуждений *биос* и *агенс* неразрывно взаимосвязаны и объединены представлением об *экзистенции агента*, разворачивающейся в условиях реального мира.

Сложность концептуального анализа здесь проявляется в том, что экзистенция как предмет онтологической проблемы философии в контексте разработки теоретических оснований создания систем искусственного интеллекта требует определения в явном виде субъекта принятия решения о своей «объективности»: в систему необходимо ввести некоего *наблюдателя*, способного сформировать критерии для вынесения вердикта о наделении части внешней для него среды свойствами *биоса* и *агенса*. Следует признать, что все критерии, предложенные в науке для онтологического обоснования феноменологии экзистенции (наличие системы управления и энергетического базиса автономного поведения, метаболизма, раздражимости, особенностей диссипации энергии, способов производства энтропии, констатации интроспективных ощущений и т. п.), пригодны для дифференциации части материи, атрибутированной свойствами *биоса* и *агенса*, из внешней среды только в субъективном аспекте, т.е. только в той степени, в которой наблюдатель считает эти критерии достаточными для того, чтобы признать, что эта часть материи демонстрирует принципиально отличное поведение.

Таким образом, оппозиция «*агент – среда*», обосновывающая назначение мозга как системы управления автономным поведением агента в среде, возникает только в системе «*агент – среда – наблюдатель*».

Следует также отметить, что вплоть до возникновения видов, способных к интроспекции, филогенетический ряд включает в себя биологические системы, не способные к субъективной констатации собственной экзистенции (не могут выступить в роли наблюдателя по отношению к самим себе). Поэтому дифференцирующий критерий максимизации продолжительности собственной экзистенции также «присваивается» таким системам в атрибутивном аспекте внешним наблюдателем.

Принципиальное значение вышеприведенных рассуждений, с нашей точки зрения, проявляется в необходимости определения непротиворечивых оснований баланса физикалистских и функционалистских воззрений в процессе концептуализации интеллектуального синтеза разумного поведения автономных агентов в реальной среде как сущности целевого назначения мозга. В таком контексте поведение, направленное на максимизацию экзистенции агента в реальной среде, признается наблюдателем разумным, а усилия мозга по его синтезу – системно обоснованными, характеризующимися этологической целостностью. В таком аспекте объектом имитационного моделирования интеллектуального синтеза разумного поведения автономного агента в реальной среде должны стать не физические, а функциональные свойства процессов, протекающих в мозге.

В этом смысле нам представляется, что удобной метафорой проектирования целевого назначения системы интеллектуального синтеза поведения, выполняемого мозгом агента, погруженного в реальную среду, является *вычислительная абстракция* алгоритма, определяющего траекторию движения этого агента в системе «агент – среда». Так как агент погружен в среду с помощью своих сенсоров и эффекторов, система «агент – среда» определяется наблюдателем как некоторое пространство состояний агента, которые задаются в каждый момент дискретного времени (шаг которого выбран наблюдателем), как множества значений состояний всех сенсоров и эффекторов агента.

Математической абстракцией такого пространства состояний, введенной в [5], является т.н. *пространство поведения* агента, оси которого представляют собой множество всех значений сенсоров и эффекторов агента, а также множество всех дискретных шагов времени на протяжении всей экзистенции агента. Мощность пространства поведения равна прямому произведению множеств всех значений всех его осей. *Экзистенция агента, таким образом, представляется траекторией его движения в пространстве поведения.*

Множество точек такой траектории, непосредственно воспринимаемых (измеряемых) наблюдателем, считаются моментами *настоящего* (времени).

Точки траектории, рассматриваемые наблюдателем в настоящем и относящиеся к моментам времени, предшествующим его наступлению, считаются *прошлым* агента. Точки траектории, рассматриваемые наблюдателем в настоящем и относящиеся к моментам времени, следующим за его наступлением, считаются *будущим* агента.

Математической абстракцией алгоритма синтеза траектории агента в системе «агент – среда», в свою очередь, является *функция поведения*, вычисляющая точки траектории движения агента в пространстве поведения в будущем на основании информации о точках такой траектории в прошлом и настоящем. Как следует из данного определения, пространство поведения является как областью определения, так и областью значения такой функции.

ЧТО ЕСТЬ ПРОБЛЕМА?

Задача синтеза траектории движения агента в пространстве поведения связана с необходимостью преодоления многочисленных трудностей, связанных со структурно-функциональными свойствами агента и реальной среды. С одной стороны, агент, являясь весьма малой частью реального мира, всегда ограничен в ресурсах и обстоятельствах (энергия, вычислительная мощность, пространство, время). С другой стороны, среда, в которую он погружен, огромна, многообразна и в общем случае непостижима агентом, так как описывается множеством знаний, мощность которого на многие десятки порядков превосходит мощность устройства (системы) управления, реализующего принятие решений о формировании и выборе точек траектории движения агента в пространстве поведения.

С точки зрения наблюдателя, описывающего процесс синтеза агентом траектории своего движения в пространстве поведения, реальная среда характеризуется свойствами *неструктурированности, неопределенности, нелинейности, многомодальности, стохастичности, динамичности, частичной наблюдаемости, эпизодичности, физической корректности, активности*, каждое из которых существенно затрудняет процесс принятия решения [6–8].

В общем случае система управления синтезом траектории агента в пространстве поведения должна стремиться таким образом построить эту траекторию, чтобы агент не попадал в состояния, в которых он теряет энергию, подвергается повреждениям или разрушению, частичному либо полному (*отрицательные точки траектории*). Напротив, траектория должна доставлять максимум целевой функции агента, что предполагает в ее составе только такие точки, в которых агент приобретает дополнительную энергию (*положительные точки траектории*), либо точки, нейтральные по изменению энергии.

Под *энергией* в общем случае мы понимаем описательную величину, характеризующую значения целевой функции агента.

Точки траектории, в которых агент прекращает свое существование, назовем *терминальными*. Положительные, отрицательные и терминальные точки траектории агента в пространстве поведения, как предложено в [5], будем называть *экзистенциальными точками*.

Предположим, что агент, выполняя движение по некоторой траектории в пространстве поведения, попадает в экзистенциальную точку (рис. 1).

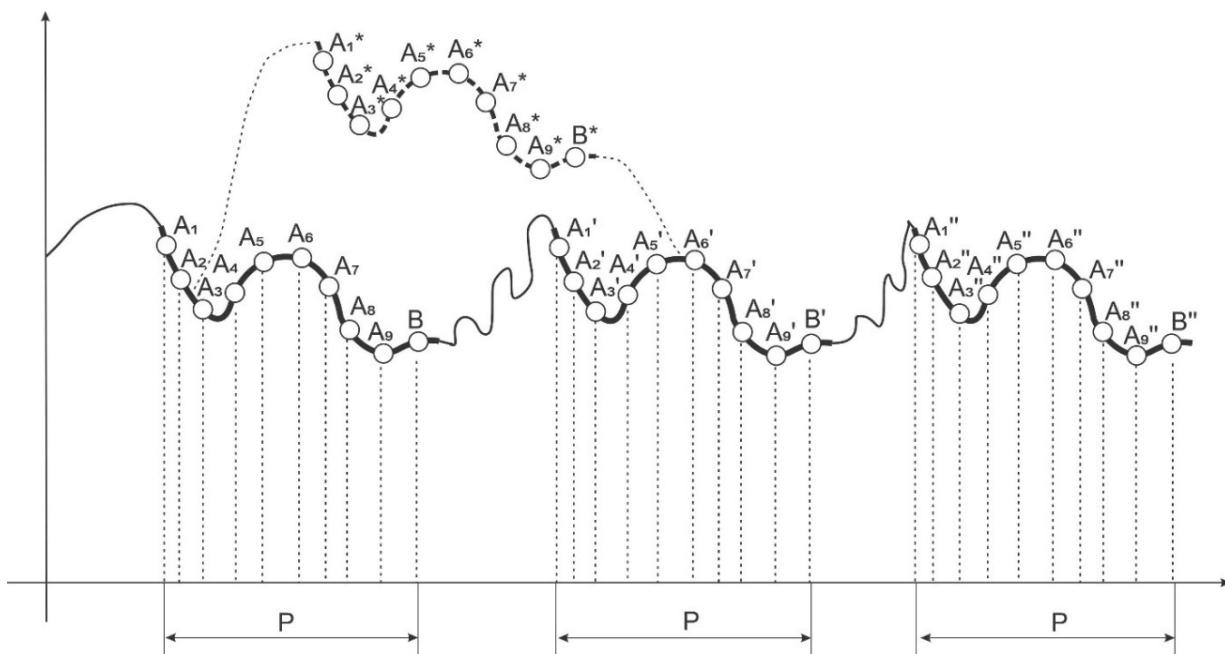


Рис. 1. Схематическое изображение паттернов проблемы агента в упрощенном пространстве поведения

Fig. 1. Schematic representation of agent problem patterns in a simplified behavior space

На рисунке 1, на котором приведено схематическое, существенно упрощенное изображение траектории (для двумерного случая), такая точка обозначена буквой *B*. Как видно из рисунка, попаданию в точку *B* предшествует последовательный проход агента через точки A_1, A_2, \dots, A_9 . Если предположить также, что всякий раз, когда траектория агента включает в себя эту последовательность точек, как показано на рисунке, для последовательностей, обозначенных $A'_1, A'_2, \dots, A'_9, B'$ и $A''_1, A''_2, \dots, A''_9, B''$, агент попадает в ту же самую экзистенциальную точку (B' и B'' на рисунке 2), то можно составить представление о *проблеме агента* как об отрезке траектории этого агента в пространстве поведения, оканчивающемся в экзистенциальной точке.

Если такая точка позитивная, то проблему назовем *позитивной проблемой*, а если отрицательная, то – *отрицательной проблемой* (в терминальной точке – *терминальной проблемой*).

Допустим, что проблема P , включающая в себя, как показано на рисунке 1, точки A_1, A_2, \dots, A_k, B , – отрицательная. Примем также, что агент обладает возможностью произвольно влиять на изменение траектории своего движения – например, так, как это демонстрирует последовательность точек траектории $A_1, A_2, A_3, A_1^*, A_2^*, A_3^*, A_4^*, A_5^*, G$. Из рисунка, в частности, следует, что агент, попав последовательно в начальные точки A_1, A_2, A_3 проблемы P , затем меняет паттерн траектории, свойственный этой проблеме, и вместо точек A_4, A_5, \dots, A_9, B попадает в точки $A_1^*, A_2^*, \dots, A_5^*, G$. Если при этом точка G не является ни отрицательной, ни терминальной экзистенциальной точкой траектории агента в пространстве поведения, то будем считать, что агент, синтезировав некоторые управляющие сигналы, изменил состояния своих эффекторов (что повлекло за собой изменение системы «агент – среда»), и, соответственно, изменив свою траекторию таким образом, что она вместо негативной экзистенциальной точки B привела в нейтральную (или даже в позитивную экзистенциальную) точку G , *решил* тем самым проблему P .

Если агент попадает в начальную точку паттерна траектории проблемы, то с учетом стохастичности среды в общем случае совпадение каждой последующей точки фактической траектории агента с точками этого паттерна увеличивает вероятность того, что агент «вошел» именно в данную проблему и, соответственно, движется к экзистенциальной (финальной) точке этой проблемы. С позиции наблюдателя, обладающего представлением о времени и способностью прогнозировать, сразу после идентификации нескольких начальных точек паттерна проблемы можно предсказать, что фактическая траектория агента пройдет по этому паттерну и агент окажется в экзистенциальной точке этой проблемы. Такой прогноз может строить и сам агент, если в его системе управления уже представлены структурно-функциональные компоненты, обеспечивающие репрезентацию времени, способность моделировать и прогнозировать будущее.

Будем считать, что в ситуации, когда агент вошел в некоторую отрицательную или терминальную проблему, возникает угроза его перехода в отрицательную или терминальную экзистенциальную точку. В случае же, когда агент входит в позитивную проблему, возникает надежда на его переход в позитивную экзистенциальную точку.

Следует обратить внимание на то, что для решения отрицательной или терминальной проблемы необходимо *снять угрозу* путем изменения фактической траектории движения агента в пространстве поведения таким образом, чтобы она перестала совпадать с паттерном соответствующей отрицательной или терминальной проблемы, в которую вошел агент.

Для решения же позитивной проблемы, напротив, необходимо *реализовать надежду*, сохранив паттерн траектории неизменным.

Точки траектории, в которых снята угроза для отрицательной или терминальной проблемы либо реализована угроза для позитивной проблемы, будем считать *точками решения проблем*.

Тогда под *решением проблемы* в системе «агент – среда» (в пространстве поведения агента) будем понимать такой отрезок траектории этого агента, проход по которому перемещает агента в точку решения этой проблемы.

В общем случае и для снятия угрозы, и для реализации надежды агенту необходимо предпринимать активные действия, что требует как решения задачи синтеза плана поведения, так и затрат энергии на синтез и выполнение этого плана.

Таким образом, и сама проблема, и ее решение представляют собой части траектории агента в пространстве поведения. Для того чтобы построить траекторию решения проблемы, в общем случае агенту необходимы знания о том, какие именно действия необходимо предпринять для перехода в точку ее решения. Такие знания должны представлять собой корреляции между паттернами траекторий проблем и паттернами траекторий их решений. Будем считать, что их совокупность образует базу знаний (базу решений) агента.

Проблемы, представленные в базе знаний агента в начальный момент его экзистенции, будем считать *aприорно определенными*.

Если бы все проблемы, с которыми сталкивается агент на протяжении своей экзистенции, были априорно определены, то агент с высокой вероятностью всегда успешноправлялся бы с их решением. Сложность, однако, состоит в том, что количество проблем в реальной среде весьма велико, а база решений агента, как правило, ограничена в силу технических причин.

Кроме того, совершенно точно агент на протяжении своей экзистенции столкнется с необходимостью решения лишь незначительного количества всех проблем, которые потенциально ожидают его в реальной среде. Поэтому нецелесообразно хранить в базе знаний решения всех проблем, гораздо эффективнее было бы составить стартовый набор из относительно небольшого множества априорно определенных проблем, которые встречаются агенту с наибольшей вероятностью, и снабдить его алгоритмическим аппаратом, обеспечивающим избирательное пополнение базы знаний за счет решений тех проблем, с которыми он фактически сталкивается в тех частях реальной среды, где протекает его экзистенция.

Назовем проблемы, которые в начальный момент экзистенции агента не были представлены в его базе решений, *априорно не определенными проблемами*. Рассмотрим способы обработки таких проблем, по нашему мнению, составляющие концептуальную основу цлевого функционала мозга, как системы управления поведением агента в реальной среде.

ЦИКЛ ОБРАБОТКИ АПРИОРНО НЕ ОПРЕДЕЛЕННЫХ ПРОБЛЕМ

Рассмотрим существенно упрощенный случай, когда размерность пространства поведения агента равна двум (рис. 2). Пусть отрицательная проблема, которая начинается для агента в точке *В* его траектории в этом пространстве и завершается в отрицательной экзистенциальной точке этой траектории *Д*, является для агента априорно не определенной. Следовательно, на момент первой встречи (*дебют проблемы*) агента с этой проблемой в его базе знаний отсутствует ее решение. Более того, до момента перехода в отрицательную экзистенциальную точку этой проблемы *Д* агент даже не знает о том, что он уже вошел в паттерн проблемы, которая завершится негативной экзистенциальной точкой и потерей энергии.

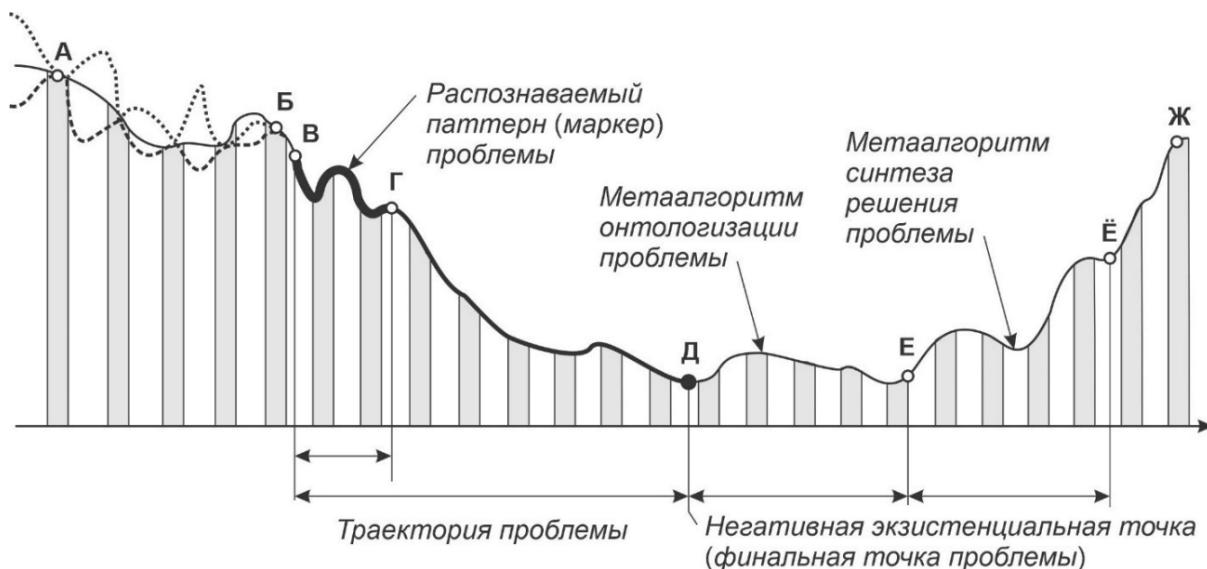


Рис. 2. Дебют, онтологизация и решение проблемы

Fig. 2. Debut, ontologization and problem solving

Более того, если такая проблема встретится агенту в следующий раз, при условии отсутствия информации о ней или ее решении в его базе знаний он опять перейдет в негативную экзистенциальную точку *Д*.

Концептуализируя системную сущность мозга как органа, реализующего функционал системы управления синтезом и выполнением поведения агентов в столь сложных условиях реальной среды, мы исходим из убеждения, что в процессе эволюции, направленной на оптимизацию конструкции агента в интересах максимизации своей экзистенции, мозг претерпел адаптационные изменения, направленные на компенсацию действия факторов, связанных со свойствами среды, затрудняющими принятие решений, такими, например, как значительная вариативность и стохастичность. В соответствии с нашей гипотезой одним из центральных патентов природы, направленных на компенсацию этих свойств реальной среды, является система онтологизации проблем и их решений.

Сущностный смысл этой системы состоит в обеспечении агента возможностью запоминать паттерны априорно не определенных проблем (*онтологизировать такие проблемы*), синтезировать *алгоритмы идентификации* таких проблем по начальным частям их паттернов, а также синтезировать и запоминать *алгоритмы их решения*.

В случае успешной работы такой системы некоторая априорно не определенная проблема (например, проблема на отрезке траектории *ВД* на рис. 2) становится определенной в том смысле, что в базе знаний агента появляются все необходимые записи, содержащие информацию о том, каким образом идентифицировать данную проблему по начальным частям ее паттерна, и о том, каким образом ее решать.

Если же проблема становится определенной в указанном смысле, то в случае, когда она встречается агенту снова (как на участках *БГ* и *ЕЁ* на рис. 3), агент в отличие от поведения, которое он демонстрировал во время дебюта проблемы (участок *ВД* на рис. 3) и которое привело его в негативную экзистенциальную точку траектории *Д*, приступает к последовательному выполнению алгоритма идентификации проблемы (отрезки *БГ* и *ЕЁ* на рис. 3) и ее решения (отрезки *ГД* и *ЁЖ* на рис. 3). Таким образом, как видно на рисунке 3, агент в случае повторного появления той же самой (уже онтологизированной) проблемы теперь не попадает в отрицательные экзистенциальные точки этой проблемы, а, напротив, своевременно применив алгоритмы ранней идентификации и решения проблемы, «выравнивает» свою траекторию и направляет ее в нейтральную область, где не проявляются ни угрозы, ни надежды (отрезки *ДЕ* и *ЖЗ* на рис. 3).

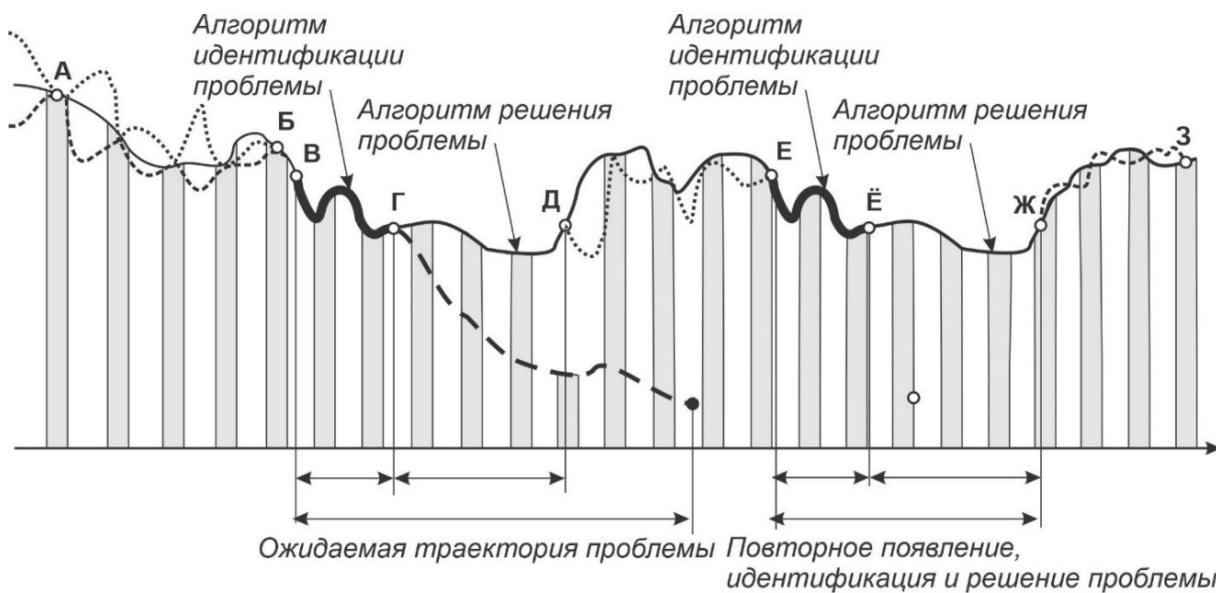


Рис. 3. Онтологизация, идентификация и решение проблем

Fig. 3. Ontologization, identification and problem solving

Для того чтобы таким способом определить новую проблему в своей базе знаний, агент должен выполнить два *метаалгоритма* (алгоритмы, синтезирующие другие алгоритмы) *онтологизации проблемы*, результатом работы которых должны стать *алгоритм идентификации* и *алгоритм решения проблемы*, которые расширят эту базу таким образом, что данная априорно не определенная проблема станет онтологизированной (определенной).

Очевидно, что для успешной работы системы онтологизации априорно не определенных проблем необходимо, чтобы метаалгоритмы онтологизации проблемы в начальный момент работы агента уже были реализованы в его системе управления, чтобы с самого начала его функционирования в среде организовать пополнение базы знаний этого агента, тем самым снижая вероятность прекращения экзистенции агента в результате повторного попадания в негативные экзистенциальные точки негативных проблем и, наоборот, увеличивая вероятность ее продления путем повторного перехода в финальные точки позитивных проблем.

Функциональной основой метаалгоритмов онтологизации, по нашему мнению, являются т.н. алгоритмы *онтонейроморфогенеза*, обеспечивающие ситуативно детерминированное формирование множеств причинно-следственных связей в базе знаний системы управления агента, представляющие собой вычислительную абстракцию процессов роста и деградации аксо-дендрональных связей между нейронами головного мозга [7, 8].

В рамках этой вычислительной абстракции направленный рост связей между нейронами интерпретируется как формирование мягкой функциональной связи между агентами (*агенты-нейроны*) в составе эмерджентно формирующегося *мультиагентного алгоритма*. Именно мультиагентными алгоритмами, роль агентов в которых выполняют агенты-нейроны различных типов, находящиеся в разных функциональных узлах *когнитивной архитектуры* (системы управления) агента, по нашему представлению, и являются метаалгоритмы онтологизации и алгоритмы идентификации и решения проблемы [9–12].

Принципиальное значение для синтеза таких алгоритмов и метаалгоритмов имеет то, что, будучи по своей природе эмерджентными, формирующими и модифицирующими на основе мягких мультиагентных связей, они, тем не менее, строятся в строгом соответствии со структурой и составом т.н. *нейрокогнитивных функциональных систем* [8], представляющих собой динамические структурно-функциональные мультиагентные модули мозга, выполняющие вычисление функции поведения агента, описанной выше.

Для алгоритмов идентификации и решения проблем нейрокогнитивные функциональные системы реализуют функции, для которых пространство поведения агента является и областью определения, и областью значения. Такие мультиагентные функциональные системы организованы на основе т.н. *инварианта нейрокогнитивной архитектуры* [7, 13], являющегося структурной основой, задающей набор шаблонов для эмерджентного формирования алгоритмов и метаалгоритмов на базе сообществ нейронов мозга.

Следует обратить внимание на то, что метаалгоритм синтеза алгоритма идентификации и сам этот алгоритм идентификации, хотя и строятся на основе единообразного подхода на базе мультиагентной функциональной системы в структуре инварианта нейрокогнитивной архитектуры, тем не менее совсем не тождественны друг другу.

Метаалгоритм должен синтезировать такое поведение агента, в результате которого будет сформирован алгоритм идентификации проблемы, т.е. в результате выполнения этого метаалгоритма агент должен научиться тому, каким образом распознавать проблему по самым первым точкам ее паттерна. Способ построения такого алгоритма идентификации в общем случае будет существенно отличаться от самого этого алгоритма. Например, для того чтобы научиться распознавать проблему, агенту может понадобиться совершить какие-то дополнительные действия в среде, провести дополнительные эксперименты, исследовать какие-то области среды, задать вопросы контрагентам по коммуникативной среде и т.п. Выполнение метаалгоритма может занять много времени, возможно, несколько часов (дней, лет). Возможно, он будет построен не сразу, а за несколько итераций.

Важно учесть, что отправной точкой для запуска метаалгоритма синтеза алгоритма идентификации проблемы является момент перехода агента в негативную экзистенциальную точку проблемы, которая является также финальной точкой этой проблемы. Такой порядок связан с тем, что при дебюте проблемы до перехода в такую точку агент не испытывает существенных резких изменений значений целевой функции и воспринимает текущие точки своей траектории в пространстве поведения как нейтральные (не относящиеся к какой-либо проблеме). Следовательно, в основе такого метаалгоритма лежит процесс установления причин наступления такого резкого изменения, которые, очевидно, лежат в прошлом, т.е. метаалгоритм направлен на то, чтобы отыскать в составе траектории агента в пространстве поведения в прошлом такие точки, которые являются маркером паттерна проблемы. Чтобы решить эту задачу, метаалгоритм использует локальную память о недавних событиях, которая в соответствии с нашей гипотезой обеспечивается с помощью механизма синтеза, передачи и регулирования времени жизни нейромедиаторных сообщений, которыми нейроны, описывающие точки траектории агента, обмениваются друг с другом.

Задача усложняется тем, что в силу дискретности, эпизодичности и частичной наблюдаемости среды расчет точек траектории в пространстве состояний доступен агенту только с определенным дискретным шагом, что отражено на рисунках 2, 3 в виде точек на заштрихованных столбиках, ширина которых соответствует времени измерения состояния системы «агент – среда», применяемого для определения точки в траектории агента в пространстве состояний. На этих рисунках также показано, что потенциально в силу эпизодичности и частичной наблюдаемости среды точки в пространстве поведения, регистрируемые агентом в дискретные моменты времени, на самом деле могут принадлежать нескольким различным траекториям агента (показаны на рис. 2 и 3 пунктирными линиями), что необходимо также учитывать при принятии решений. Все три вышеуказанные особенности, по нашему мнению, парируются алгоритмом онтонейроморфогенеза, допускающим установление связей между нейронами, выполняющими в когнитивной архитектуре агента функциональную презентацию событий, отстоящих друг от друга на произвольные периоды времени.

Сам же алгоритм идентификации, будучи сформирован метаалгоритмом онтологизации и сохранен в базе знаний агента, должен быть компактен и прост в применении, так как времени на выполнение распознавания проблемы по начальной сигнатуре ее паттерна в интересах реализации эффективного алгоритма решения у агента в реальной среде, как правило, всегда очень мало.

Все вышесказанное для метаалгоритма синтеза алгоритма идентификации проблемы верно и для метаалгоритма синтеза решения этой проблемы.

Принципиальное значение имеет то, что как метаалгоритмы онтологизации, так и алгоритмы идентификации и решения проблем выполняют синтез точек и отрезков в траектории движения агента в пространстве поведения. Это означает, что онтологизация, идентификация и решение всех проблем, с которыми сталкивается агент в реальной среде в течение своей экзистенции, выполняются с помощью синтеза и реализации поведения агента в этой среде.

Таким образом, по нашему мнению, главный ответ на вопрос «Что делает мозг?» состоит в том, что мозг реализует функции системы управления поведением агента в среде [14, 15], направленным на максимизацию экзистенции этого агента, используя для этого метаалгоритмы онтологизации, алгоритмы идентификации и решения априорно не определенных проблем.

ЗАКЛЮЧЕНИЕ

Построена гипотеза о системной целевой функции мозга, заключающейся в максимизации экзистенции агента на основе синтеза управления его поведением в реальной среде. Предложена концепция экзистенции агента как траектории его движения в пространстве поведения, являющемуся математической абстракцией пространства состояний системы

«агент – среда – наблюдатель». Разработана концепция адаптации агента к условиям реального мира на основе реализуемых мозгом метаалгоритмов онтологизации и алгоритмов идентификации и решения априорно не определенных проблем. Разработанные концепции обосновывают представление об онтологизации, идентификации и решении проблем агентом в реальной среде как о процессах автономного синтеза его поведения, что создает теоретический базис для создания *систем универсального искусственного интеллекта*.

СПИСОК ЛИТЕРАТУРЫ

1. *Minsky M.* Steps toward Artificial Intelligence // Proceedings of the IRE. 1961. Vol. 49. No. 1. Pp. 8–30.
2. Минский М. Вычисления и автоматы. М.: Мир, 1971.
3. Нагоев З. В. Основные принципы нейрокогнитивного моделирования сознания агента универсального искусственного интеллекта // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 1. С. 152–170. DOI: 10.35330/1991-6639-2025-27-1-152-170
4. Нагоев З. В., Нагоева О. В. Обоснование символов и мультиагентные нейрокогнитивные модели семантики естественного языка. Нальчик: Издательство КБНЦ РАН, 2022. 150 с.
5. Нагоев З. В. Мультиагентные экзистенциальные отображения и функции // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 63–71. EDN: QZTFLX
6. Russell S., Norvig P. Artificial Intelligence: A Modern Approach (AIMA). 2nd ed. Moscow: Williams, 2007. 1424 р.
7. Нагоев З. В. Интеллектика, или Мышление в живых и искусственных системах. Нальчик: Издательство КБНЦ РАН, 2013. 232 с.
8. Нагоев З. В. Онтонейроморфогенетическое моделирование // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 56–63. EDN: QZTFLN
9. Нагоев З. В., Кудаев В. Ч., Ошхунов М. М., Пшенокова И. А. Онтонейроморфогенетическое моделирование виртуальных прототипов в интегрированных САПР на основе мультиагентных знаний и биоинспирированных алгоритмов // Известия Кабардино-Балкарского научного центра РАН. 2013. № 6-1(56). С. 46–53. EDN: RPXLRL
10. Нагоев З. В., Пшенокова И. А., Нагоева О. В. и др. Имитационная модель нейрокогнитивной системы управления автономным программным агентом, выполняющим кооперативное поведение с целью автоматического пополнения онтологий // Известия Кабардино-Балкарского научного центра РАН. 2023. № 6(116). С. 226–234. DOI: 10.35330/1991-6639-2023-6-116-226-234
11. Нагоев З. В., Анчёков М. И., Бжихатлов К. Ч. и др. Онтоэписоциофилогенетическое развитие систем общего искусственного интеллекта на основе мультиагентных нейрокогнитивных архитектур // Известия Кабардино-Балкарского научного центра РАН. 2022. № 6(110). С. 61–75. DOI: 10.35330/1991-6639-2022-6-110-61-75
12. Апшев А. З., Аталиков Б. А., Канкулов С. А. и др. Онтофилогенетические алгоритмы синтеза фенотипов интеллектуальных программных агентов для применения в задачах многопоколенной оптимизации управляющих нейрокогнитивных архитектур // Известия Кабардино-Балкарского научного центра РАН. 2022. № 6(110). С. 76–91. DOI: 10.35330/1991-6639-2022-6-110-76-91
13. Nagoev Z.V. Multiagent recursive cognitive architecture // Biologically Inspired Cognitive Architectures 2012. Proceedings of the third annual meeting of the BICA Society, in Advances in Intelligent Systems and Computing series, Springer, 2012. Pp. 247–248.
14. Bzhikhatalov K.Ch., Pshenokova I.A., Zammoev A.U. Architecture and algorithm of the information security system for autonomous intelligent systems and complexes // Information Society. 2025. No. 1. Pp. 118–127. DOI: 10.52605/16059921_2025_01_118

15. Pshenokova I., Anchakov M., Kankulov S. et al. Simulation model of the decentralized control of autonomous software agents' teams and mobile robots // In: Samsonovich A.V., Liu T. (eds) Biologically Inspired Cognitive Architectures 2024. BICA. Studies in Computational Intelligence. Vol. 477. Pp. 1–8. DOI: 10.1007/978-3-031-76516-2_31

REFERENCES

1. Minsky M. Steps toward Artificial Intelligence. *Proceedings of the IRE*. 1961. Vol. 49. No. 1. Pp. 8–30.
2. Minsky M. *Vychisleniya i avtomaty* [Computations and Automata]. Moscow: Mir, 1971. (In Russian)
3. Nagoev Z.V. Basic principles of neurocognitive modeling of consciousness of a universal artificial intelligence agent. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2025. Vol. 27. No. 1. Pp. 152–170. DOI: 10.35330/1991-6639-2025-27-1-152-170. (In Russian)
4. Nagoev Z.V., Nagoeva O.V. *Obosnovaniye simvolov i mul'tiagentnyye neyrokognitivnyye modeli semantiki yestestvennogo yazyka* [Justification of Symbols and Multi-Agent Neurocognitive Models of Natural Language Semantics]. Nalchik: Izdatel'stvo KBNTS RAN, 2022. 150 p. (In Russian)
5. Nagoev Z.V. Multi-agent existential mappings and functions. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2013. No. 4(54). Pp. 63–71. EDN: QZTFLX. (In Russian)
6. Russell S., Norvig P. Artificial Intelligence: A Modern Approach (AIMA). 2nd ed. Moscow: Williams, 2007. 1424 p.
7. Nagoev Z.V. *Intellektika, ili myshleniye v zhivykh i iskusstvennykh sistemakh* [Intellectics, or thinking in living and artificial systems]. Nalchik: Izdatel'stvo KBNTS RAN, 2013. 232 p. (In Russian)
8. Nagoev Z.V. Ontoneuromorphogenetic modeling. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2013. No. 4(54). Pp. 56–63. EDN: QZTFLN. (In Russian)
9. Nagoev Z.V., Kudayev V.Ch., Oshkhunov M.M., Pshenokova I.A. Ontoneuromorphogenetic modeling of virtual prototypes in integrated cad systems based on multi-agent knowledge and bioinspired algorithms. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2013. No. 6-1(56). Pp. 46–53. EDN: RPXLRL. (In Russian)
10. Nagoev Z.V., Pshenokova I.A., Nagoeva O.V. et al. Simulation model of a neurocognitive control system for an autonomous software agent performing cooperative behavior for the purpose of automatic ontology replenishment. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2023. No. 6(116). Pp. 226–234. DOI: 10.35330/1991-6639-2023-6-116-226-234. (In Russian)
11. Nagoev Z.V., Anchakov M.I., Bzhikhatlov K.Ch. et al. Ontoepisociophylogenetic development of general artificial intelligence systems based on multi-agent neurocognitive architectures. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2022. No. 6(110). Pp. 61–75. DOI: 10.35330/1991-6639-2022-6-110-61-75. (In Russian)
12. Apshev A.Z., Atalikov B.A., Kankulov S.A. et al. Ontophylogenetic algorithms for synthesizing phenotypes of intelligent software agents for application in multigenerational optimization problems of control neurocognitive architectures. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2022. No. 6(110). Pp. 76–91. DOI: 10.35330/1991-6639-2022-6-110-76-91. (In Russian)
13. Nagoev Z.V. Multiagent recursive cognitive architecture. Biologically Inspired Cognitive Architectures 2012. Proceedings of the third annual meeting of the BICA Society, in Advances in Intelligent Systems and Computing series, Springer, 2012. Pp. 247–248.
14. Bzhikhatlov K.Ch., Pshenokova I.A., Zammoev A.U. Architecture and algorithm of the information security system for autonomous intelligent systems and complexes. *Information Society*. 2025. No. 1. Pp. 118–127. DOI: 10.52605/16059921_2025_01_118

15. Pshenokova I., Anchokov M., Kankulov S. et al. Simulation model of the decentralized control of autonomous software agents' teams and mobile robots. In: Samsonovich A.V., Liu T. (eds) *Biologically Inspired Cognitive Architectures 2024. BICA. Studies in Computational Intelligence*. Vol. 477. Pp. 1–8. DOI: 10.1007/978-3-031-76516-2_31

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflict of interest.

Финансирование. Исследование проведено без спонсорской поддержки.

Funding. The study was performed without external funding.

Информация об авторах

Нагоев Залимхан Вячеславович, канд. техн. наук, генеральный директор Кабардино-Балкарского научного центра Российской академии наук;

вед. науч. сотр. отдела «Мультиагентные системы», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра Российской академии наук;

360000, Россия, Нальчик, ул. И. Арманд, 37-а;

zaliman@mail.ru, ORCID: <https://orcid.org/0000-0001-9549-1823>, SPIN-код: 6279-5857

Нагоева Ольга Владимировна, науч. сотр. отдела «Мультиагентные системы», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра Российской академии наук;

360000, Россия, г. Нальчик, ул. И. Арманд, 37-а;

nagoeva_o@mail.ru, ORCID: <https://orcid.org/0000-0003-2341-7960>

Information about the authors

Zalimkhan V. Nagoev, Candidate of Technical Sciences, General Director of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

Leading Researcher of the Department “Multi-Agent Systems”, Institute of Computer Science and Problems of Regional Management – branch of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

37-a, I. Armand street, Nalchik, 360000, Russia;

zaliman@mail.ru, ORCID: <https://orcid.org/0000-0001-9549-1823>, SPIN-code: 6279-5857

Olga V. Nagoeva, Researcher, Multiagent Systems Department, Institute of Computer Science and Problems of Regional Management – branch of the Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

37-a, I. Armand street, Nalchik, 360000, Russia;

nagoeva_o@mail.ru, ORCID: <https://orcid.org/0000-0003-2341-7960>