

Формирование динамических причинно-следственных зависимостей при управлении поведением интеллектуального агента на основе формализма мультиагентных нейрокогнитивных архитектур*

И. А. Пшенокова^{1,2}, О. В. Нагоева², А. З. Апшев², А. З. Энес²

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

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

Аннотация. Перспективные интеллектуальные системы принятия решений должны обеспечивать построение причинно-следственных связей между событиями в условиях неопределенности, связанной с недостоверными и неполными знаниями, приводящей к невозможности предсказать точные последствия принятого решения при множестве вариантов выбора. В таких случаях очень важно учитывать корреляционную связь между причиной и следствием. Известно, что в основе причинно-следственного вывода лежит представление всех возможных альтернативных сценариев, которое позволяет планировать и манипулировать действиями в процессе принятия решений. В работе представлена имитационная модель формирования динамических причинно-следственных зависимостей для управления поведением автономного интеллектуального агента на основе нейрокогнитивных архитектур. Рассмотрены мультиагентная структура агентов событиями типа и процесс формирования причинно-следственных зависимостей путем заключения или расторжения мультиагентных контрактов. Проведен эксперимент по обучению автономного интеллектуального агента, прогнозированию последствий различных действий в текущих обстоятельствах.

Ключевые слова: интеллектуальный агент, причинно-следственная связь, корреляция, мультиагентные системы, нейрокогнитивная архитектура

Поступила 09.08.2022, одобрена после рецензирования 15.09.2022, принята к публикации 20.09.2022

Для цитирования. Пшенокова И. А., Нагоева О. В., Апшев А. З., Энес А. З. Формирование динамических причинно-следственных зависимостей при управлении поведением интеллектуального агента на основе формализма мультиагентных нейрокогнитивных архитектур // Известия Кабардино-Балкарского научного центра РАН. 2022. № 5 (109). С. 73–80. DOI: 10.35330/1991-6639-2022-5-109-73-80

MSC: 68T42

Original article

Formation of dynamic cause-and-effect relationships when controlling the behavior of an intelligent agent based on the multi-agent neurocognitive architectures formalism*

I.A. Pshenokova^{1,2}, O.V. Nagoeva², A.Z. Apshev², A.Z. Enes²

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

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

© Пшенокова И. А., Нагоева О. В., Апшев А. З., Энес А. З., 2022

* Работа выполнена при поддержке проекта РФФ № 22-19-00787

* The work was supported by the Russian Science Foundation project No. 22-19-00787

Annotation. Promising intelligent decision-making systems should ensure the construction of cause-and-effect relationships between events under conditions of uncertainty associated with unreliable and incomplete knowledge, leading to the inability to predict the exact consequences of the decision taken with a variety of choices. In such cases, it is very important to take into account the correlation between cause and effect. It is known that the basis of the causal inference is the representation of all possible alternative scenarios, which allows to plan and manipulate actions in the decision-making process. The paper presents a simulation model for the formation of dynamic cause-and-effect relationships to control the behavior of an autonomous intelligent agent based on neurocognitive architectures. The multi-actor structure of event-type agneurons and the process of formation of cause-and-effect dependencies by concluding or terminating multi-agent contracts are considered. An experiment was conducted to train an autonomous intelligent agent, to predict the consequences of various actions in the current circumstances.

Key words: intelligent agent, causation, correlation, multi-agent systems, neurocognitive architecture

Submitted 09.08.2022,

approved after reviewing 15.09.2022,

accepted for publication 20.09.2022

For citation. Pshenokova I.A., Nagoeva O.V., Apshev A.Z., Enes A.Z. Formation of dynamic cause-and-effect relationships when controlling the behavior of an intelligent agent based on the multi-agent neurocognitive architectures formalism. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2022. No. 5 (109). Pp. 73–80. DOI: 10.35330/1991-6639-2022-5-109-73-80

ВВЕДЕНИЕ

Основной частью процесса мышления является возможность принимать решения в условиях неопределенности [1–3]. Под неопределенностью понимается невозможность предсказать точные последствия своего решения при множестве вариантов выбора. Поэтому важно учитывать корреляционную связь между действиями и последствиями или причиной и следствием. Согласно теории принятия решений фон Неймана-Моргенштерна, если рациональному агенту, принимающему решение, известна эта корреляционная связь, он выбирает действие, которое максимизирует ожидаемую полезность [4]. С другой стороны, если агент не знает вероятности наступления определенных следствий на некоторые действия, ему необходимо, опираясь на собственный опыт, выбрать такое действие, которое максимизирует ожидаемую полезность с большей вероятностью [5–7]. Далее, взаимодействуя с окружающей средой, агент может обновить свои количественные оценки неопределенностей в соответствии с наблюдаемыми данными для улучшения решения [6, 8]. В свою очередь знание причинно-следственных связей наделяет агента способностью к превентивному планированию, поскольку он может предсказать, к чему приведут определенные действия [9, 10].

В работах [11–14] утверждается, что человеческий мозг на базовом уровне представляет собой аппарат каузального вывода, который использует следствия для выяснения причин. В [3] указывается, что в основе причинно-следственного вывода лежит представление всех возможных альтернативных сценариев, которое позволяет планировать и манипулировать действиями в процессе принятия решений [15]. В работе [16] предполагается, что автономный агент может обучаться и строить причинно-следственные связи, взаимодействуя с неструктурированной средой до тех пор, пока не запомнит все неизвестные связи. Такой способ обучения состоит в том, чтобы записать ему предположения о неструктурированной среде. Эти предположения принимаются за истинную причинно-следственную модель, а затем в процессе наблюдения и обучения обновляются и закрепляются.

Согласно [3] существует три уровня причинно-следственных рассуждений. Первый – наблюдение, предполагает восприятие закономерностей или паттернов во входных данных, выраженных в виде корреляций. Второй уровень – манипуляции, предсказывает последствия преднамеренных действий, выраженных в виде причинно-следственных связей. Третий уровень – контрфактуал, включает в себя построение теории, которая объясняет, почему конкретные действия имеют конкретные эффекты и что происходит в отсутствие таких действий. Важно отметить, что хотя обычно в причинно-следственных связях причина предшествует во времени следствию, это не обязательно должно быть общим правилом.

Причинно-следственное знание поддерживает принятие решений двумя способами: позволяет прогнозировать последствия различных действий в текущих обстоятельствах и выбрать в дереве решений путь, субоптимальный по критерию максимизации суммарной энергии агента [17].

Цель работы состоит в повышении обоснованности принимаемых решений интеллектуальной системой за счет построения причинно-следственных зависимостей между событиями в условиях неопределенности.

Задача исследования – разработать подход к формированию динамических причинно-следственных зависимостей для управления поведением автономного агента.

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ МУЛЬТИАГЕНТНОЙ НЕЙРОКОГНИТИВНОЙ АРХИТЕКТУРЫ

В [17, 18] представлен мультиагентный нейрокогнитивный подход к проектированию интеллектуальных систем принятия решений и введено понятие интеллектуального агента. Интеллектуальный агент, представленный в виде мультиагентной нейрокогнитивной архитектуры, состоит из множества агентов-нейронов (агнейроны), которые в свою очередь состоят из агентов-акторов (акторы), объединенных в функциональные узлы, взаимодействующих друг с другом и синтезирующих собственное поведение.

В соответствии с [19] структуру поведенческого акта различной степени сложности формируют следующие функциональные узлы: распознавания, моделирования, эмоциональной оценки, целеполагания, синтеза плана действий, управления выполнением плана. Следовательно, для управления поведением автономного интеллектуального агента нужно в нейрокогнитивной архитектуре создать и обучить функциональные узлы, которые выполнят синтез поведения этого агента в различных проблемных ситуациях [18].

Для прогнозирования своего поведения и поиска оптимального пути в дереве решений интеллектуальный агент должен уметь устанавливать причинно-следственные связи между событийными нейронами. Когда на вход интеллектуальный агент получает некоторую информацию о своем текущем состоянии, в архитектуре формируются концептуальные агнейроны, которые констатируют мультиагентный факт наступления данного события в виде агнейрона событийного типа. Этот агнейрон должен запросить желаемое событие у других агнейронов этого типа путем массовой рассылки сообщений вида «Кто купит информацию?» для установления своей причины или следствия. После такой рассылки агент формирует множество мультиагентных связей, и не обязательно каждая из них является его причиной или следствием. Поиск причины и следствия агнейроном осуществляется путем заключения или расторжения мультиагентных контрактов.

На рисунке 1 представлена мультиакторная структура агнейронов событийного типа.

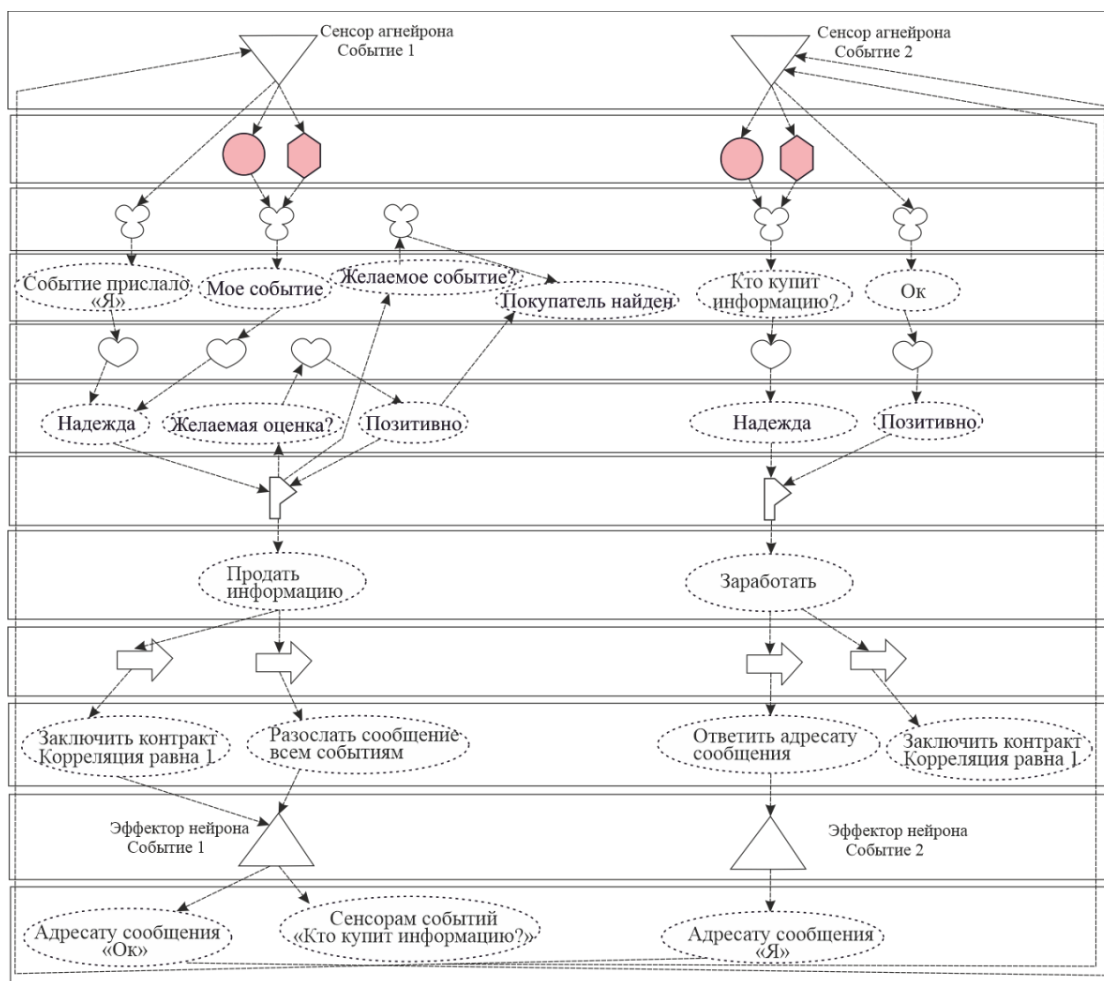
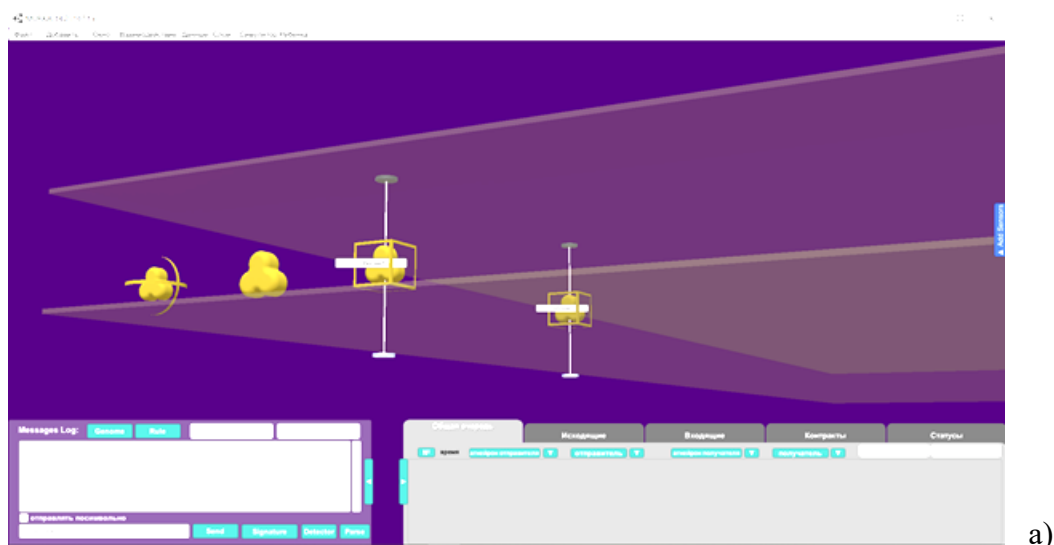


Рис. 1. Мультиакторная структура агнейронов, условно названных «Событие 1» и «Событие 2»

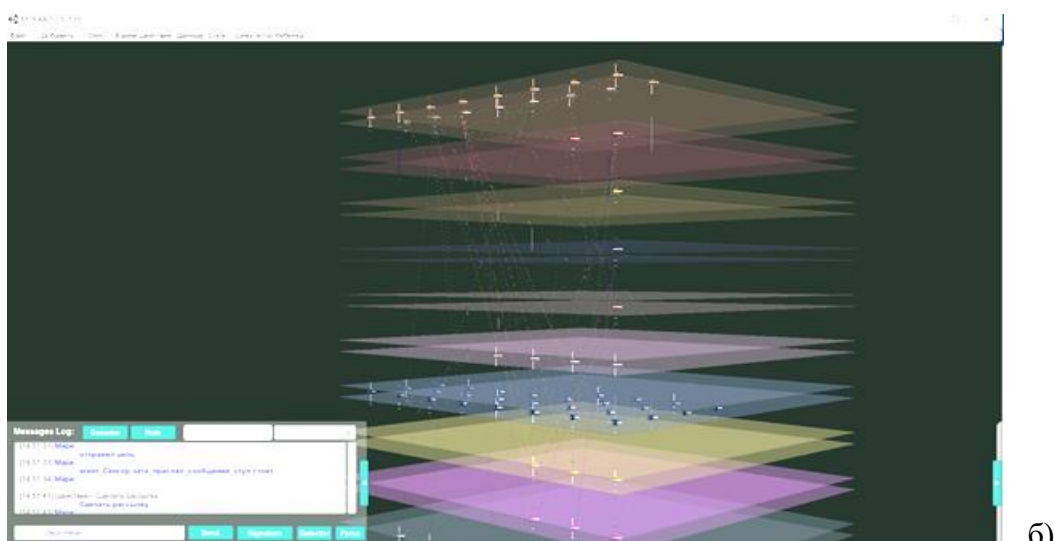
Fig. 1. Multi-actor structure of agneurons, conventionally named "Event 1" and "Event 2"

Рассмотрим процесс формирования причинно-следственной зависимости путем заключения или расторжения мультиагентных контрактов. В ответ на входное значение когнитивный узел распознавания (2-й слой рис. 1) формирует представление в виде концептуальных акторов, отвечающих за обработку поступившей информации. Далее когнитивный узел моделирования формирует текущее состояние в виде мультиагентного факта (3-й слой). Данный факт получает эмоциональную оценку на следующем функциональном уровне (5-й слой), и агнейрон, отвечающий за текущее состояние, рассматривается как событие-причина, для которой нужно найти агнейрон событие-следствие. Если событие-причина получило эмоциональную оценку в сторону увеличения собственной энергии, активизируется когнитивный узел целеполагания (7-й слой), в котором формируются желаемая оценка, соответствующее ей конечное событие и горизонт планирования. Эти данные передаются на вход когнитивного узла плана действий (9-й слой), который формирует действия, необходимые для того, чтобы связать текущее событие с конечным событием. Но таких событий может быть множество. И агнейрон путем рассылки массовых сообщений формирует множество динамических связей с возможными агнейронами событий. В следующий раз для прохождения по каждому из путей в образовавшемся множестве вводится оценка степени корреляции знаний, полученных в результате формирования динамических связей. Коэффициент корреляции вычисляется отношением числа позитивных срабатываний событий-контрагентов на общее число событий.

Если некоторому входному событию соответствует несколько событий-следствия, агнейроны, отвечающие за эти следствия, сигнализируют для получения вознаграждения. Когда агнейрон, отвечающий за событие-причина, объявляет о вознаграждении, он находит своих контрагентов, затем формируются моделирующие агнейроны, содержащие в себе событие-причину и соответствующее ему событие-следствие. В результате экспериментов, тестирующих эти решения, происходят обучение агента и оценка степени корреляции полученного знания. При увеличении или уменьшении коэффициента корреляции ослабевает или укрепляется мультиагентный контракт и соответственно добавляются или исключаются логические условия в разных частях знания, происходит динамическое укрепление или разрыв связей между агентами. Важно отметить, что такое свойство мультиагентной нейрокогнитивной архитектуры сходно со свойством нейропластичности структур головного мозга, при котором происходит формирование или разрыв аксо-дендрональных связей [22]. На рисунке 2 представлены примеры визуализации мультиагентной нейрокогнитивной архитектуры при проведении эксперимента по обучению автономного агента и построению причинно-следственных зависимостей между событиями в разработанном программном модуле имитационного моделирования. В частности, на рис. 2 а представлены два агнейрона типа события, на рис. 2 б – мультиакторная структура агнейрона «Событие 1», на рис. 2 в – процесс формирования желаемой оценки, соответствующего ей конечного события и горизонт планирования.



а)



б)

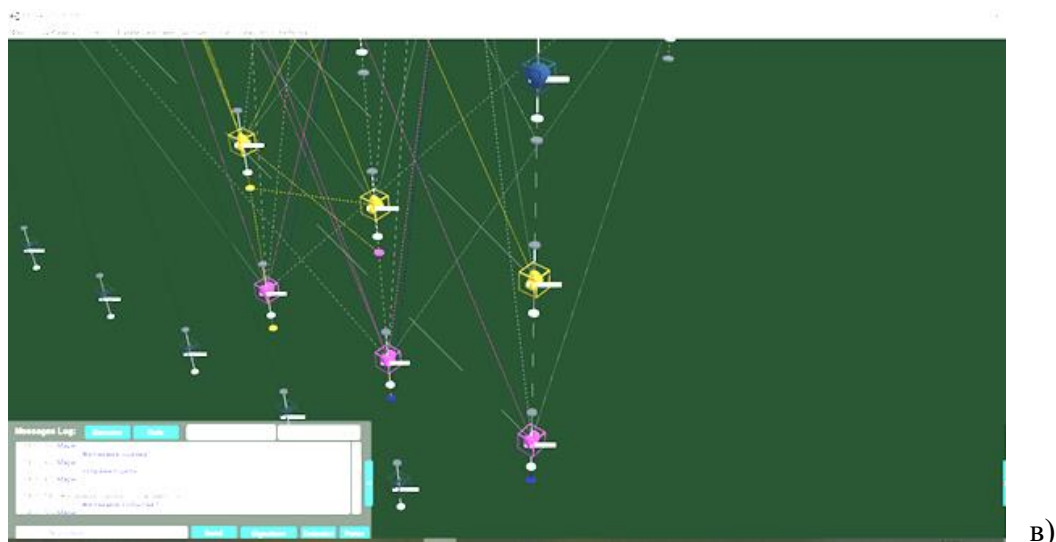


Рис. 2 (а-в). Имитационный эксперимент по обучению автономного интеллектуального агента
Fig. 2 (a-b). Simulation experiment for training an autonomous intelligent agent

Таким образом, событие-причина и событие-следствие представлены агнейронами, каждый из которых обладает знаниями, описывающими соответствующую причинно-следственную зависимость. Эти агнейроны при повторяемости причины и/или следствия объединяют свои части знаний путем заключения взаимовыгодного мультиагентного контракта. Если наступило текущее событие и на это событие есть контрагент агнейрон-следствие, интеллектуальный агент обучается, что позволяет прогнозировать последствия различных действий в текущих обстоятельствах и выбрать в дереве решений субоптимальный путь.

ЗАКЛЮЧЕНИЕ

В работе представлен подход к формированию динамических причинно-следственных зависимостей для управления поведением автономного агента. Разработана мультиакторная структура агнейронов событийного типа, представлен процесс формирования причинно-следственной зависимости путем заключения или расторжения мультиагентных контрактов. В результате проведенных экспериментов видно, что интеллектуальный автономный агент способен к прогнозированию последствий различных действий в текущих обстоятельствах.

Информация об авторах

Пшенокова Инна Ауесовна, канд. физ.-мат. наук, зав. лаб. «Интеллектуальные среды обитания», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

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

вед. науч. сотр. лаборатории «Нейрокогнитивные автономные интеллектуальные системы», Кабардино-Балкарский научный центр РАН;

360002, Россия, Нальчик, ул. Балкарова, 2;

pshenokova_inna@mail.ru, ORCID: <https://orcid.org/0000-0003-3394-7682>

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

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

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

Апшев Артур Заурбиевич, стажер-исследователь лаб. «Интеллектуальные среды обитания», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

360000, Россия, Нальчик, ул. И. Арманд, 37-а;
artur.apshev@gmail.com

Энес Ахмед Зюлфикар, стажер-исследователь лаб. «Компьютерная лингвистика», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

360000, Россия, Нальчик, ул. И. Арманд, 37-а;
ahmedenes@mail.ru, ORCID: <https://orcid.org/0000-0003-3633-4910>

REFERENCE

1. Danks D. Unifying the mind: Cognitive representations as graphical models. MIT Press. 2014. 304 p.
2. Lake B.M., Ullman T.D., Tenenbaum J.B., Gershman S.J. Building machines that learn and think like people. *Behavioral and Brain Sciences*. 2017. Vol. 40. DOI: <https://doi.org/10.1017/S0140525X16001837>
3. Pearl J., Mackenzie D. The Book of Why: The New Science of Cause and Effect. Basic Books. 2018. 432 p.
4. Von Neumann J., Morgenstern O. Theory of games and economic behavior. Princeton University Press. 1944. 776 p.
5. Savage L. The Foundations of Statistics. New York: John Wiley & Sons. 1954. 310 p.
6. Bernardo J. M., Smith A. F. M. Bayesian theory. Wiley Series in Probability and Statistics. 2000. 608 p.
7. Gilboa I. Theory of Decision under Uncertainty. Cambridge University Press. 2009. 230 p.
8. Peterson M. An Introduction to Decision Theory. Cambridge University Press. 2017. 348 p.
9. Spirtes P., Glymour C. N., and Scheines R. Causation, prediction, and search. MIT press. 2000. 546 p.
10. Pearl J. Theoretical impediments to machine learning with seven sparks from the causal revolution. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 2018. Vol. 3. <https://doi.org/10.48550/arXiv.1801.04016>
11. Woodward J. Making things happen: A theory of causal explanation. Oxford Studies in Philosophy of Science. Oxford University Press. 2003. 432 p.
12. Friston K. The free-energy principle: a unified brain theory? *Nature Reviews Neuro science*. 2010. Vol. 11(2). Pp. 127–138.
13. Hohwy J. The predictive mind. Oxford University Press. 2013. 288 p.
14. Clark A. Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press. 2015. 424 p.
15. Danks D. Unifying the mind: Cognitive representations as graphical models. MIT Press. 2014. 304 p.
16. Gonzalez-Soto L. E., Sucar H. J. Escalante Playing against Nature: causal discovery for decision making under uncertainty. *arXiv:1807.01268v1 [Artificial Intelligence (cs.AI)]*. 2018. <https://doi.org/10.48550/arXiv.1807.01268>
17. Nagoev Z.V. *Intellektika, ili Myshleniye v zhivykh i iskusstvennykh sistemakh* [Intelligence, or thinking in living and artificial systems]. Nalchik: Publishing House of KBSC of RAS. 2013. 213 p. (in Russian)
18. Nagoev Z., Pshenokova I., Nagoeva O., Sundukov Z. Learning algorithm for an intelligent decision making system based on multi-agent neurocognitive architectures. *Cognitive Systems Research*. 2021. Vol. 66. Pp. 82–88. <https://doi.org/10.1016/j.cogsys.2020.10.015>.

19. Anokhin P.K. *Uzlovyye voprosy teorii funktsional'nykh sistem* [Key questions of the theory of functional systems]. Moscow: Science. 1980. 203 p. (in Russian)

Анохин П. К. Узловые вопросы теории функциональных систем. Москва: Наука, 1980. 203 с.

20. Nagoev Z., Pshenokova I., Nagoeva O., Kankulov S. Situational analysis model in an intelligent system based on multi-agent neurocognitive architectures. *Journal of Physics: Conference Series*. 2021. Vol. 2131. <https://doi.org/10.1088/1742-6596/2131/2/022103>

21. Nagoev Z.V. Ontoneuromorphogenetic modeling. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2013. No. 4(54). Pp. 56–63. (in Russian)

Нагоев З. В. Онтонейроморфогенетическое моделирование // Известия Кабардино-Балкарского научного центра РАН. 2013. № 4(54). С. 56–63.

22. Nagoev Z.V., Pshenokova I.A., Kankulov S.A., Atalikov B.A., Airan A.A. Formal model of multi-agent search for the optimal plan of behavior of an intelligent agent based on self-organization of distributed neurocognitive architectures. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2021. No. 3(101). Pp. 21–31. DOI: 10.35330/1991-6639-2021-3-101-21-31 (in Russian)

Нагоев З. В., Пшенокова И. А., Канкулов С. А., Аталиков Б. А., Айран А. А. Формальная модель мультиагентного поиска оптимального плана поведения интеллектуального агента на основе самоорганизации распределенных нейрокогнитивных архитектур // Известия Кабардино-Балкарского научного центра РАН. 2021. № 3(101). С. 21–31. DOI: 10.35330/1991-6639-2021-3-101-21-31

Information about the authors

Pshenokova Inna Auesovna, Candidate of Physical and Mathematical Sciences, Head of lab. “Intellectual Habitats”, the Institute of Computer Science and Regional Management Problems – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

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

Leading Researcher laboratory “Neurocognitive autonomous intelligent systems”, Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

360002, Russia, Nalchik, 2 Balkarov street;

pshenokova_inna@mail.ru, ORCID: <https://orcid.org/0000-0003-3394-7682>

Nagoeva Olga Vladimirovna, Research Associate of the Department of the multiagent systems, the Institute of Computer Science and Regional Management Problems – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

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

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

Apshev Artur Zaurbievich, Research Assistant, lab. “Intellectual Habitats”, the Institute of Computer Science and Regional Management Problems – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

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

artur.apshev@gmail.com

Enes Akhmed Zulfikar, Research Assistant, lab. “Computer Linguistics”, the Institute of Computer Science and Regional Management Problems – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

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

ahmedenes@mail.ru, ORCID: <https://orcid.org/0000-0003-3633-4910>