УДК 004.055, 004.82, 510.62, 510.67
DOI 10.25205/2541-7517-2020-18-2-5-29

О задачном подходе в искусственном интеллекте и когнитивных науках

Е. Е. Витяев, С. С. Гончаров, Д. И. Свириденко

Институт математики им. С. Л. Соболева СО РАН
Новосибирск, Россия

Новосибирский государственный университет
Новосибирск, Россия

Аннотация

Данная работа продолжает серию работ по задачному подходу в искусственном интеллекте. Как было отмечено ранее, агентный подход, описанный в монографии Стюарта Рассела и Питера Норвига «Искусственный интеллект. Современный подход», может быть более аргументированно представлен в рамках задачного подхода. В настоящей работе будет показано, что не только задачи оснований математики и искусственного интеллекта, но и многие познавательные функции, осуществляемые человеком и анализируемые в когнитивных науках, также могут быть описаны и изучены в рамках задачного подхода. В частности, в настоящей работе показано, что аналогом понятия «задача» в когнитивных науках является понятие цели и что теория функциональных систем (ТФС), описывающая целенаправленное поведение, может быть представлена как решение мозгом задач по достижению целей и удовлетворению потребностей. Это дает возможность напрямую сопоставлять задачи искусственного интеллекта с естественными когнитивными процессами и, тем самым, выявлять перечень тех задач «естественного» интеллекта и схем их решения, которые могут успешно использоваться при решении задач искусственным интеллектом.

Ключевые слова

искусственный интеллект, общий искусственный интеллект (AGI), когнитивные науки, агентный подход, задачный подход, целенаправленная деятельность, цель и задача, теория функциональных систем, принятие решений, рациональный агент, моделирование рационального агента

Благодарности

Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 17-11-01176)

Для цитирования

Витяев Е. Е., *Гончаров С. С.*, *Свириденко Д. И.* О задачном подходе в искусственном интеллекте и когнитивных науках // Сибирский философский журнал. 2020. Т. 18, № 2. С. 5–29. DOI 10.25205/2541-7517-2020-18-2-5-29

On the Task Approach in Artificial Intelligence and Cognitive Sciences

E. E. Vityaev, S. S. Goncharov, D. I. Sviridenko

Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russian Federation

Novosibirsk State University
Novosibirsk, Russian Federation

Abstract

This work continues a series of publications on the task approach in artificial intelligence. As noted earlier, the agent-based approach described in the monograph by Stuart Russell and Pieter Norwig "Artificial Intelligence. The Modern Approach", may be more argumentatively presented within the framework of the task approach. This paper will show that not only the problems of the bases of mathematics and artificial intelligence, but also many cognitive functions performed by humans and analyzed in cognitive sciences, can also be described and studied within the framework of the task approach. In particular, this paper shows that the analogue of the concept of task in cognitive sciences is the concept of goal and that the Functional Systems Theory (FST), which describes purposeful behavior, can be presented as the brain's solution of tasks to achieve goals and satisfaction of needs. It gives the chance to compare directly the tasks of artificial intellect with natural cognitive processes and, thereby, to reveal the list of those tasks of "natural" intellect and schemes of their solution which can be successfully used for the solution of artificial intelligence tasks.

Keywords

artificial intelligence, general artificial intelligence (AGI), cognitive science, agent-based approach, task approach, purposefulness activity, purpose and task, Functioning System Theory (FST), decision making, rational agent, rational agent modeling

Acknowledgements

The study was supported by the Russian Science Foundation (research project no. 17-11-01176)

For citation

Vityaev E. E., Goncharov S. S., Sviridenko D. I. On the Task Approach in Artificial Intelligence and Cognitive Sciences. *Siberian Journal of Philosophy*, 2020, vol. 18, no. 2, p. 5–29. (in Russ.) DOI 10.25205/2541-7517-2020-18-2-5-29

1. Введение

Данная работа продолжает серию работ по задачному подходу в искусственном интеллекте [Витяев, Гончаров, Свириденко, 2019]. Как было отмечено ранее, наиболее продвинутым подходом в современном искусственном интеллекте, носящем синтетический характер, является *агентный подход*, описанный в монографии Стюарта Рассела и Питера Норвига «Искусственный интеллект. Современный подход» [Рассел, Норвиг, 2006], позволяющий с единых методологических пози-

ций и, что очень важно, в единых терминах, где ключевыми понятиями являются понятия «*рациональный агент*» и «*среда*», обсуждать различные направления в искусственном интеллекте. Особенно полезным агентный подход, как нам представляется, оказался для такого бурно развивающегося направления искусственного интеллекта, каковым является так называемый общий искусственный интеллект (AGI), о чем говорят попытки дать содержательные определения AGI его современными «апостолами» (см.: [Goertzel, 2010]) [1].

Бен Гёрцел (Ben Goertzel): «*Общий интеллект – это способность достигать сложных целей в сложных средах*»;

Шейн Легге (Shane Legg) и Маркус Хуттер (Marcus Hutter): «*Интеллект измеряет способность агента успешно действовать в широком диапазоне сред*»;

Пей Ванг (Pei Wang): «*Интеллект – это способность системы адаптироваться к своей среде, работая при недостаточных знаниях и ресурсах*».

Однако если внимательно проанализировать содержание указанной выше монографии и результаты обсуждения приведенных определений AGI, то легко можно убедиться в том, что все они опираются прежде всего на понятие «задача» и фактически неявно следуют основным положениям задачного подхода, представленного в [Витяев, Гончаров, Свириденко, 2019].

В настоящей работе мы покажем, что не только проблемы оснований математики и искусственного интеллекта, но и многие познавательные функции, осуществляемые человеком и анализируемые в когнитивных науках, также могут описываться и изучаться в рамках задачного подхода. Это обстоятельство дает возможность напрямую сопоставлять задачи искусственного интеллекта с естественными когнитивными процессами и, тем самым, выявлять тот перечень задач *естественного интеллекта,* схемы решения которых могут быть успешно использованы в искусственном интеллекте. Именно по этой причине мы считаем, что изучение и непосредственное формальное моделирование когнитивных систем живых существ и, прежде всего, человека с позиций задачного подхода должно представлять несомненный интерес для современного искусственного интеллекта, особенно для такой его области, как AGI. Кстати, этот тезис полностью согласуется с мнением многих ведущих специалистов в области AGI, которые отмечают, что, скажем, изучение и формальное моделирование таких свойств когнитивных систем, как целеполагание, рефлексия, эмоции и другие субъективные переживания и введение их в интеллектуальные системы способствует успешной реализации многих ожидаемых от таких систем свойств и качеств. Так, например, ряд авторов

[1] Другим примером такой попытки является «Собрание определений Искусственного интеллекта» Ш. Легга и М. Хуттера. См.: Legg S., Hutter M. A *Collection of Definitions of Intelligence.* URL: http://www.hutter1.net/ai/idefs.pdf (дата обращения 07.05.2020).

утверждают (см. [Rosenbloom, Gratch, Ustun, 2015] и [Wang, Talanov, Hammer, 2016]), что введение и использование в интеллектуальных системах «искусственных» эмоций позволяет успешно решать проблему оптимизации распределения ресурсов между моделями разных когнитивных процессов и повышает эффективность процесса принятия интеллектуальной системой решений

Приступая к проектированию интеллектуальной системы как *агента*, воспринимающего некую среду и *рационально* действующего в ней, мы должны прежде всего определиться с устройством этой системы. Поскольку поведение агента во многом определяется его целями, то для их достижения интеллектуальная система должна уметь воспринимать окружающую среду, воздействовать на нее, обучаться, хранить опыт взаимодействия с ней и на основе этого опыта предсказывать реакцию среды на свои действия и планировать их последовательность. Очевидно, что такое многообразие функциональных способностей интеллектуальной системы предполагает нетривиальность ее устройства. И здесь знание того, как устроены когнитивные системы живых существ и, в частности, человека, может оказаться чрезвычайно полезным. Особенно в том случае, когда эти знания допускают математическую формализацию и/или компьютерную реализацию. Подобное обстоятельство и объясняет тот вывод, что изучение и моделирование когнитивных систем представляет несомненный интерес для искусственного интеллекта и особенно для AGI. Нам представляется, что именно по этой причине направление «*когнитивные архитектуры*», ставящее своей целью практическое воплощение теоретических знаний об устройстве человеческого мышления, признано в AGI одним из центральных направлений его развития, поскольку использование этих знаний позволяет надеяться на реальную возможность интеграции в единую систему большой части необходимых для систем искусственного интеллекта свойств, функций и качеств, таких как целеполагание, рассуждение, обучение, хранение и представление знаний и т.д. (см.: [Cassimatis, 2006] и [Langley, 2006]). Об одной из таких перспективных когнитивных архитектур, основу которой составляет так называемая теория функциональных систем (ТФС) работы мозга (см.: [Anokhin, 1973], [Anokhin, 1974], [Анохин, 1976] и [Судаков, 1984]), и пойдет речь в настоящей статье. И начнем мы с обсуждения и уточнения процесса целеполагания в когнитивных системах.

2. Понятия цели и целенаправленной деятельности

Авторы исходят из того, что аналогом понятия «задача» в когнитивных науках является понятие *цели* (см.: [Витяев, 2014] и [Vityaev, 2015]). Как и в случае задачи цель нельзя достичь, не имея *критерия её достижения*, иначе всегда можно считать, что цель уже достигнута. Заметим также, что любое *действие* всегда является

целенаправленным, поскольку если у действия нет цели, то непонятно – когда, как и чем его надо завершить. *Цель действия* – изменить текущее состояние и/или внешнее воздействие нужным для организма образом и тем самым удовлетворить некую *потребность* организма. Таким образом, удовлетворение той или иной потребности и есть реальная *цель*, которая ставится перед организмом. Напомним, что еще Ю.Л. Ершовым и К.Ф. Самохваловым при анализе понятия «задача» в основаниях математики была отмечена связь между понятиями цели, потребности и задачи: «"Я хочу пить" (потребность. – *Авт.*) – что это значит? Нет, конечно, никакой ошибки полагать, что слова "я хочу пить" означают просто вот это, где это – определенное состояние сознания, которое я переживаю сейчас и которое я именую жаждой. Но тогда возникает новый вопрос: как ощущение жажды (хотения) связано с фактическим питьем? Откуда я знаю, что удовлетворить жажду можно питьем? Содержится ли в самом переживании жажды сознание того, чем эту жажду можно удовлетворить? … Знать желание не означает знать желаемое, а означает знать способность узнать желаемое (т.е. речь идет о *критерии решения задачи* или *достижения цели*. – *Авт.*), как только этому представится случай. Иными словами, вы понимаете какое-либо свое желание (т.е. потребность. – *Авт.*) … только тогда, когда этому желанию вы сопоставили чувство уверенности в том, что любое будущее состояние сознания вы сумеете убедительным и безошибочным образом распознать (пользуясь критерием решения задачи. – *Авт.*) как состояние удовлетворения желания или состояние неудовлетворения... (т.е. как результат решения задачи. – *Авт.*). Хотя … при этом я не обязательно знаю, чем это утоление будет достигнуто. По прошлому опыту ожидаю, что водой, но, быть может, какая-нибудь таблетка тоже утолит мою жажду» [Ершов, Самохвалов, 1984. С. 142–143].

Установим теперь более точное соответствие между понятиями «задача» и «цель». Для этого несколько иначе сформулируем понятие «задача», приведенное ранее в [Витяев, Гончаров, Свириденко, 2019], и проинтерпретируем составляющие этого понятия применительно к понятию «цель». Итак, далее будем считать, что мы имеем дело с *задачей* (в нашем случае – с *целью*), если в нашем распоряжении имеются:

- *модель предметной области* / представление о текущей ситуации и окружении;
- *запрос* / цель как необходимость удовлетворить возникшую потребность;
- *критерий решения задачи* / критерий достижения цели;
- *каким видится результат решения задачи и последствия ее решения* / каким должен быть результат в случае достижения цели.

Как же решаются задачи живыми организмами, т.е. как ставятся и достигаются цели, понимаемые как необходимость удовлетворения потребностей? Сразу отме-

тим, что среди целого ряда достаточно развитых физиологических теорий целенаправленной деятельности, пытающихся ответить на этот вопрос, мы выделяем и далее будем придерживаться ТФС работы мозга, которая подробно анализирует физиологические механизмы решения мозгом задачи по достижению целей, удовлетворению потребностей и организации целенаправленного поведения.

Согласно ТФС, удовлетворение возникшей у организма потребности фиксируется специальным рецепторным аппаратом, который по своим свойствам точно соответствует свойствам подкрепляющего раздражителя, удовлетворяющего эту потребность. Фактически этот рецепторный аппарат отвечает за наличие и применение организмом *критерия достижения цели*. Выше мы отмечали, что цель не имеет смысла, если не указан критерий ее достижения. При этом утверждалось, что знать потребность/желание организма не означает знать желаемое (способ удовлетворения потребности), а означает знать *способность узнать* желаемое, которое и фиксируется критерием достижения цели в лице соответствующего рецепторного аппарата, как только представится возможность это осуществить в результате целенаправленной деятельности.

Обратим внимание читателя на некую парадоксальность определения цели (этот эффект имеет место и в случае формулировки задачи). Дело в том, что процесс достижения цели (как и процесс решения задачи), являясь целенаправленной деятельностью, в то же время принципиально не содержит знание о том, как, где и когда можно достичь цели, – для задания цели необходимо определить только потребность и критерий её удовлетворения. Это обстоятельство назовем *парадоксом цели*. И как утверждает ТФС, поведение организма как целенаправленная деятельность постоянно направлено на разрешение этого парадокса путем определения того, чем, как и когда можно достичь цели. Такое понимание цели позволяет определить *результат достижения цели* как то, что организм получает при достижении цели, подкрепленное критерием того, что потребность удовлетворяется: мы утоляем жажду (результат), когда пьем воду; утоляем голод (результат), когда едим пищу, и т.д. Таким образом, между понятиями «цель» и «результат» имеется следующая связь: результат получен, когда цель достигнута и положительно «сработал» критерий достижения цели. Однако когда цель только ставится, мы имеем дело лишь с формулировкой цели и критерием ее достижения, но не имеем еще результата.

Перейдем теперь к изложению теории функциональных систем, в которой подробно исследовано, как мозг в процессе целенаправленной деятельности постоянно разрешает парадокс цели и определяет, чем, как и когда можно достичь цели.

3. Анализ целенаправленной деятельности в теории функциональных систем (ТФС)

Отличие ТФС от других существующих в настоящее время когнитивных физиологических теорий П.К. Анохин описывает следующим образом: «Пожалуй, одним из самых драматических моментов в истории изучения мозга как интегративного образования является фиксация внимания на самом действии, а не на его результатах... мы можем считать, что результатом "хватательного рефлекса" будет не само хватание как действие, а та совокупность афферентных раздражений, которая соответствует признакам "схваченного" предмета (результат действия)» [Anokhin, 1974. P. 27]. Далее, излагая ТФС, мы будем в основном следовать положениям монографии одного из ведущих учеников П.К. Анохина – К.В. Судакова [Судаков, 1984] (отметим, что в этой работе подводится итог работ не только самого П.К. Анохина, но и всей его школы).

Прежде всего рассмотрим физиологические механизмы постановки целей организмом. Для этого нам понадобится понятие *функциональной системы*, под которой мы будем понимать «комплекс нервных образований с соответствующими им периферическими рабочими органами, объединенный на основе выполнения какой-либо вполне очерченной и специфической функции организма. К таким очерченным функциям можно отнести, например, локомоцию, дыхание, глотание, плавание и т.д.». [Судаков, 1984. С. 19]. Как было сказано выше, цель (задача) осмысленна, если есть критерий достижения цели (критерий решения задачи). Понятно, что отправляемые функции организма должны приводить к достижению некоторых целей, что фиксируется как некоторый результат. «Основным постулатом теории функциональных систем является положение о том, что ведущим системообразующим фактором, организующим функциональную систему любого уровня организма, служит полезный для организма и системы в целом приспособительный результат. Именно результат благодаря постоянной обратной афферентации о его состоянии производит своеобразную "мобилизацию" центральных и исполнительных образований в функциональную систему» [Судаков, 1984. С. 34–35].

Мы уже говорили, что достижение результата должно обязательно фиксироваться некоторым критерием. Но какова физиологическая природа критерия, фиксирующего достижение организмом некоего результата? Оказывается, что физиологически он реализуется «специальным рецепторным аппаратом»: «Каждая *потребность*, даже при незначительном отклонении жизненно важной функции от оптимального для метаболизма уровня, немедленно воспринимается специальными рецепторными аппаратами» [Судаков, 1984. С. 43]. И далее: «Сигнализация о потребности несет двоякую функцию. С одной стороны, она играет пусковую роль,

возбуждая специальные аппараты саморегуляции, а с другой, она постоянно информирует эти же центры о результатах действий, совершенных функциональной системой. Поскольку эта сигнализация заключает в себе информацию о конечном результате, о его отклонениях от оптимального для метаболизма уровня или восстановлении... она была названа обратной афферентацией» [Там же. С. 45]. Теперь мы способны объяснить, как организмом физиологически осуществляется постановка целей.

Как мы уже говорили, ТФС целью организма объявляет необходимость удовлетворения его *потребности.* Такая цель носит двойной характер – во-первых, перед организмом ставится подцель по восстановлению нарушенного метаболизма, а, во-вторых, одновременно ставится подцель энергетически обеспечить процесс достижения первой подцели некоей целенаправленной деятельностью. Заметим, что если у организма нет опыта по достижению цели, то у него нет и ни малейшей информации о том, как ее достичь. В этом случае у организма есть единственная возможность – воспользоваться *методом проб и ошибок*: «Возникшее на основе той или иной биологической потребности поведение новорожденного животного строится в полном смысле слова методом "проб и ошибок"... Поражает направленный поиск новорожденным специальных раздражителей внешней среды, с которыми он никогда не встречался. Следовательно, они должны иметь врожденные модели, в которых запрограммированы свойства удовлетворяющих эти потребности раздражителей, с которыми осуществляется постоянное сравнение достигнутых результатов... непосредственно после рождения первой целенаправленной деятельностью лосенка является освоение вертикальной позы, затем движение в сторону матери, поиск соска, сосание и, наконец, реакция следования» [Судаков, 1984. С. 74].

Согласно ТФС, взаимодействие различных результатов и целей организмом осуществляется несколькими способами: по «принципу доминанты», «иерархией результатов» и «моделями результатов». Рассмотрим, например, «принцип доминанты». Этот принцип говорит о том, что две цели организмом одновременно достигаться не могут и всегда доминирует только некоторая одна потребность. Тем самым, наиболее важные для организма цели (так называемые *доминирующие потребности*) всегда линейно упорядочены во времени.

Следующий принцип, принцип «*иерархии результатов*», позволяет упорядочить взаимодействие разных функциональных систем организма в некоторый момент времени – по отношению к доминирующей функциональной системе остальные функциональные системы выстраиваются в определенную иерархию: «По отношению к каждой доминирующей функциональной системе все другие функциональные системы выстраиваются в определенном иерархическом поряд-

ке… Иерархия функциональных систем..., прежде всего, включает иерархическое взаимодействие результатов их действий, когда результат деятельности одной функциональной системы входит в качестве компонента в результат деятельности другой… Так, у голодного кролика доминирует функциональная система, деятельность которой направлена на поиск пищи. В это время другие функциональные системы, определяющие, например, кровяное давление, дыхание, выделение, направлены на лучшее обеспечение доминирующей пищедобывательной функциональной системы» [Судаков, 1984. С. 54].

Согласно ТФС П.К. Анохина, центральные механизмы функциональных систем, обеспечивающих целенаправленное поведение, имеют однотипную архитектуру. Рассмотрим более детально архитектуру целенаправленной деятельности и физиологические механизмы разрешения организмом парадокса цели.

Афферентный синтез. Начальную стадию поведенческого акта любой степени сложности составляет афферентный синтез, включающий в себя синтез *мотивационного возбуждения, памяти, обстановочной и пусковой афферентации.*

Мотивационное возбуждение. Как мы знаем, постановка цели определяется возникшей потребностью. Но в случае целенаправленного поведения она трансформируется в мотивационное возбуждение. «Ведущим возбуждением…, определяющим целенаправленную деятельность даже животных, является мотивационное возбуждение, формирующееся на основе ведущей внутренней потребности» [Судаков, 1984. С. 73]. При целенаправленной деятельности достижение результата и действие подкрепляющего стимула субъективно ощущается появлением положительной эмоции (ликвидацией отрицательной эмоции). Целенаправленному поведению надо обучаться, поэтому надо запомнить ту последовательность возбуждений, которая привела к достижению результата. Положительные эмоции (ликвидация отрицательных) имеют поэтому ещё и подкрепляющую (санкционирующую) роль, которая фиксирует в памяти всю последовательность действий, приведшей к достижению цели.

Память. Память – второй компонент афферентного синтеза. «Извлечение прошлого опыта из памяти происходит по той же нейрохимической трассе, по которой он был зафиксирован в момент приобретения опыта (подкрепления)» [Судаков, 1984. С. 91].

Обстановочная афферентация. При фиксации следа в памяти фиксируется и та обстановка, в которой удалось получить результат. Эта обстановка фиксируется как набор необходимых условий, наряду с мотивацией, которые требуются для достижения результата. Поэтому мотивационное возбуждение в данной обстановке «извлекает из памяти» только те способы достижения цели, которые возможны в данной обстановке. Таким образом, обстановочная афферентация

при взаимодействии с извлеченным из памяти опытом определяет, что и как можно делать в данной обстановке для достижения цели.

Пусковая афферентация. Четвертым компонентом афферентного синтеза является пусковая афферентация. По смыслу она также является обстановочной афферентацией, только связанной не со стимулами обстановки, а со временем и местом достижения результата. Поэтому пусковая афферентация отвечает на вопрос: когда и где можно достичь результат.

Таким образом, на стадии афферентного синтеза в значительной степени разрешается парадокс цели и определяется, что, как, где и когда следует сделать для достижения цели. Именно мотивационное возбуждение как цель с учетом имеющегося опыта и обстановки автоматически разрешает парадокс цели. «Вытягивая» из памяти весь накопленный опыт, мотивационное возбуждение как цель преобразуется в *конкретную цель*, определяющую способ своего достижения. Конкретная цель называется в ТФС «*высшей мотивацией*».

Принятие решений. На стадии афферентного синтеза мотивационным возбуждением может быть извлечено из памяти несколько способов достижения цели. Но на стадии принятия решения должен быть выбран только один способ – конкретный *план действий*. Принятие решений – очень тонкий процесс и должен учитывать (см., например: [Анохин, 1976], [Симонов, 1975], [Симонов, 1981]):

- вероятность достижения цели в данной ситуации;
- суммарные энергетические затраты того или иного способа достижения цели с учетом информационной определенности возможности достижения цели (переключающая функция эмоций);
- объем извлеченного из памяти опыта, включая доминантные (генетически определенные) формы поведения в случае, когда имеющегося опыта недостаточно для принятия решения (компенсаторная функция эмоций).

Акцептор результатов действия. Понятно, что цель может быть достигнута, только если будет достигнут каждый из промежуточных результатов соответствующего этой цели плана действий. При этом выбранный план действий не всегда способен гарантировать не только достижимость конечного результата, но даже и любого промежуточного результата. Как мы знаем, за определение всей последовательности и иерархии результатов, которые должны быть получены при выполнении плана действий по достижению цели, отвечает мотивационное возбуждение – именно оно «извлекает из памяти нужную последовательность. Эта последовательность и иерархия результатов и называется в ТФС *акцептором результатов действия*. Именно доминирующая мотивация "вытягивает" в аппарате акцептора результатов действия весь накопленный опыт до конечного, удовлетворяющего лежащую в ее основе потребность результата, создавая определенную *модель* или

программу поведения. С этих позиций модель акцептора результатов действия представляет собой доминирующую потребность организма, трансформированную в форме опережающего возбуждения мозга, как бы в своеобразный *комплексный "рецептор"* соответствующего подкрепления» [Судаков, 1984. С. 84].

Заметим, что мотивационное возбуждение, преобразуясь в конкретную цель, одновременно извлекает из памяти также и *конкретный результат* этой конкретной цели, которым является вся последовательность и иерархия результатов, которые должны быть получены в процессе достижения этой конкретной цели и выполнения плана действий, т.е. акцептор результатов действия. «Формирование "цели" в центральной архитектуре поведенческого акта связано с построением следующей стадии системной организации поведенческого акта аппарата предвидения будущего результата (всей последовательности и иерархии результатов), удовлетворяющего доминирующую потребность, – аппарата акцептора результатов действия» [Там же. С. 81].

Важно отметить, что преобразование мотивационного возбуждения в конкретную цель, а плана действий – в конкретный результат (акцептор результатов действия) значительно снижает уровень парадоксальности первоначальной цели, для которой не было определено, чем, как и когда достигать цель. В результате мы получаем уже не парадоксальную конкретную цель, в которой конечная цель (и результат) разбиты на подцели (и подрезультаты) так, что для каждой подцели уже известно, чем, как и когда её можно достичь. Но, тем не менее, парадоксальность определения цели полностью не снимается, так как даже если мы знаем по прошлому опыту, что цель (результат) достигается такой-то и такой-то последовательностью действий, то у нас нет (и в принципе не может быть) никакой гарантии того, что и в этот раз данная последовательность действий приведет к этому же результату. А потому приведет ли или не приведет некоторая последовательность действий к результату, всё равно должно быть проверено некоторым критерием, который в нашем случае есть акцептор результатов действия.

Эффекторные механизмы функциональных систем – выполнение плана действий. «Стадия формирования акцептора результатов действия динамически последовательно сменяется формированием самого целенаправленного действия. Однако ему предшествует стадия, когда действие уже сформировано как центральный процесс, но внешне еще не реализуется... По-видимому, наиболее удачно отражает семантический смысл этой стадии название "стадия эфферентного синтеза"» [Судаков, 1984. С. 88].

Так как реальная ситуация всегда чем-то отличается от тех ситуаций, которые были извлечены из памяти и учтены в процессе принятия решений, то неизбежно могут возникать «рассогласования» между ожидаемыми результатами и реаль-

но поступающей обратной афферентацией о результатах совершенных действий. «Оценка результата действия происходит с помощью активной *ориентировочно-исследовательской деятельности.* Ориентировочно-исследовательская реакция возникает и усиливается во всех случаях, когда результат совершенного действия неожиданно не соответствует свойствам сформированного на основе афферентного синтеза акцептора результатов действия, т.е. при возникновении "рассогласования" в поведенческой деятельности. Благодаря включению такой реакции немедленно перестраивается афферентный синтез, принимается новое решение, строится новая программа действия [Судаков, 1984. С. 90–91]… Целенаправленный поведенческий акт, таким образом, заканчивается последней санкционирующей стадией. На этой стадии при действии раздражителя, удовлетворяющего ведущую потребность, – подкрепления в общепринятом смысле – параметры достигнутого результата через раздражения соответствующих рецепторов... вызывают потоки обратной афферентации, которая по всем своим свойствам соответствует ранее запрограммированным свойствам подкрепляющего раздражителя в акцепторе результатов действия. При этом удовлетворяется ведущая потребность (срабатывает критерий достижения цели – *зам. авторов*) и поведенческий акт заканчивается» [Судаков, 1984. С. 89–91]. Важно отметить, что при подкреплении каждый раз фиксируется «след» всех возбуждений, приведших к достижению результата, и тем самым успешно реализованный план действий «заносится» в память.

4. Моделирование рациональных агентов в соответствии с ТФС

Приведем формальную модель рационального агента, который, с одной стороны, достаточно точно отражает описание целенаправленной деятельности в соответствии с ТФС, а с другой стороны, представляет собой формальную модель одного из наиболее сложных видов рациональных агентов, указанных в книге С. Рассела и П. Норвига [Рассел, Норвиг, 2006] – *обучающегося рационального агента, основанного на цели*.

Следуя предварительно разработанным моделям агентов и экспериментам, проведенным одним из авторов этой статьи, определим специальный тип рационального агента следующим образом [2]. Пусть рациональный агент имеет некоторый набор сенсоров $S_1, ..., S_n$, показания которых характеризуют состояние как самого агента, так и внешней среды. Пусть каждый сенсор S_i имеет некоторое множество возможных показаний VS_i. Агент также располагает множеством воз-

[2] См.: [Витяев, 2006; Демин, Витяев, 2008; Мухортов, Хлебников, Витяев, 2012; Vityaev, 2013; Vityaev, Perlovsky, Kovalerchuk, Speransky, 2013; Demin, Vityaev, 2014].

можных действий в среде $A=\{a_1, ..., a_m\}$. Любое действие агента, совершаемое в момент времени t_i, может приводить в следующий момент времени t_i+1 к какому-то изменению среды, и, как следствие, к изменению показаний его сенсоров.

Поскольку агент «воспринимает» окружающий мир только через свои сенсоры, то с точки зрения агента состояние его самого и окружающей среды как системы в каждый конкретный момент времени может быть записано вектором показаний всех сенсоров $V(t)=(v_1, ..., v_n)$, где $v_i \in VS_i$ – показание *i*-го сенсора в момент времени *t*, причем состояния с одинаковыми показаниями сенсоров для агента неразличимы. Множество всех возможных состояний системы обозначим через $S=(VS_1 \times VS_2 \times ... \times VS_n)$.

Поскольку в общем случае сенсоры агента не могут учитывать всех физических законов среды и имеют собственные физические ограничения (например, по чувствительности, радиусу действия и т.п.), то при совершении агентом некоторого действия в состоянии $s \in S$ система с точки зрения агента может переходить в одно или несколько других возможных состояний. Тогда действие a_i агента можно определить как функционал, переводящий систему «агент – внешняя среда» из одного состояние в другое с некоторой вероятностью:

$$a_i : (S_i) \to (S_i \times S \times P),$$

где S_i – подмножество S состояний системы, в которых действие a_i имеет смысл (осуществимо), $S_i \times S \times P$ – множество троек (s_0, s, p), где $s \in S$ – полученное в результате действия состояние, $p \in [0,1]$ – вероятность его достижения из состояния $s_0 \in S_i$ при совершении действия a_i, вычисляемая в соответствии с объективными факторами осуществления действия во внешнем мире.

Определим понятие события и истории событий. Под *событием* $e=(s_0, s_e, a)$ будем понимать единичный факт перевода системы из состояния $s_0 \in S_0$ в состояние $s_e \in S$ в результате совершения действия a. Тогда *историей событий* H назовем множество пар (e, t), где *e* – событие, *t* – момент времени, когда произошло данное событие.

Теперь от общей модели «агент-внешняя среда» перейдем к более конкретной дискретной модели. На множестве состояний системы $S=(VS_1 \cup VS_2 \cup ... \cup VS_n)$ определим множество предикатов $PS=\{P_1, ..., P_k\}$, каждый из которых вычисляется на основе показаний сенсоров. Это дает нам возможность каждое состояние системы записать в виде вектора значений истинности предикатов из PS, $s=(p_1, ..., p_k)$, $p_i \in \{0,1\}$, где 1 означает истинность соответствующего предиката, а 0 – его ложность. Будем считать, что задачей агента является достижение некоторой цели. Определим цель Goal как состояние системы $s_{Goal}=(p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$,

$p_{i_1}^{goal} = 1, ..., p_{i_{goal}}^{goal} = 1$, которое требуется достичь. Запись $(p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$ означает, что предикаты $P_{i_1}^{goal}, ..., P_{i_{goal}}^{goal}$ при достижении цели становятся истинными.

Уточним теперь понятие события и истории. Под *событием* $e = (s_0, s_e, a)$, как и раньше, будем понимать единичный факт перевода системы из состояния $s_0 = (p_1^0, ..., p_k^0)$ в состояние $s_e = (p_1^e, ..., p_k^e)$ в результате совершения действия a, а под *историей событий* H – множество пар (e_t, t), где $e_t = (s_t, s_{t+1}, a)$ – событие, t – момент времени, когда произошло данное событие.

Правила R, предсказывающие изменение состояния после осуществления действия *a*, определим как преобразование $R = (s_0 \xrightarrow[p]{a} s_e)$, где:

s_0 – начальное состояние системы $(p_{i_1}^0, ..., p_{i_0}^0)$;

s_e – конечное состояние системы $(p_{i_1}^e, ..., p_{i_e}^e)$;

a – действие, которое переводит начальное состояние в конечное;

p – вероятность правила *R*, с которой действие *a* переводит начальное состояние в конечное.

Вероятность правила R рассчитывается следующим образом: если n – число случаев, когда начальным состоянием было s_0 и выполнялось действие *a*, а m – число тех случаев из *n*, когда действие a переводило состояние s_0 в состояние s_e, тогда $p = m/n$. Заметим, что вероятности правил R (предсказывающие переход из состояния s_0 в состояние s_e после осуществлении действия *a*) и вероятности из множества P (предсказывающие переход из состояния s_0 в состояние s_e при осуществлении действия *a* и вычисляемые в соответствии с объективными факторами осуществления действия во внешнем мире) – различные величины. Можно сказать, что задачей обучения является максимальное приближение «субъективных» вероятностей правил R, оцениваемых агентом, к объективным вероятностям P, характеризующим взаимодействие агента с внешней средой. Обнаружение правил осуществляется нейронами в соответствии с семантическим вероятностным выводом (см., например: [Vityaev, 2013]).

Определим функциональную систему FSC, достигающую цель одним действием, как набор $FSC = (s_{Goal}, R_1, ..., R_n, p_{FSC})$. Функциональная система FSC осуществляет преобразование $s_0 \xrightarrow[p_{FSC}]{R_1, ..., R_n} s_{Goal}$, где $s_{Goal} = (p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$ – целевое состояние функциональной системы, $R_1, ..., R_n$ – правила вида $s_0 \xrightarrow[p]{a} s_{Goal}$, с помощью которых из различных начальных состояний s_0 с помощью некоторого действия *a* можно попасть в целевое состояние s_{Goal} (рис. 1). Цель s_{Goal} функ-

циональной системы ставится соответствующим мотивационным возбуждением. Способ вычисления вероятности p_{FSC} приведен ниже.

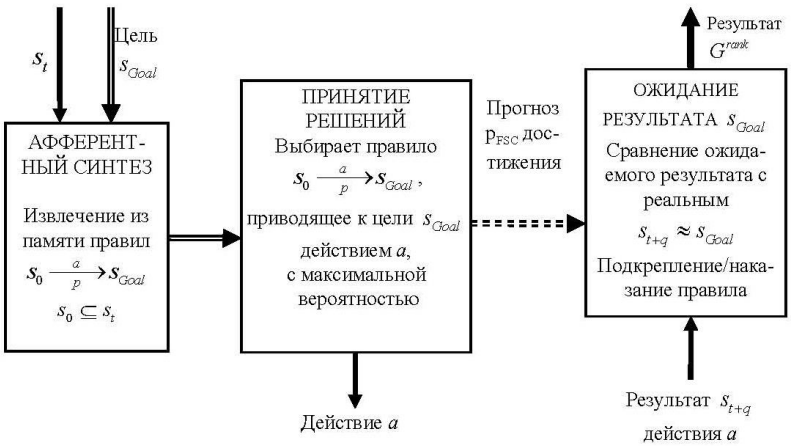

Рис. 1. Схема функциональной системы, реализующей сенсорные коррекции

В соответствии с принципом сенсорных коррекций, сформулированным Н. А. Бернштейном [1997], принципиально нельзя знать заранее точный результат предыдущего движения. Поэтому выбрать максимально вероятное правило $s_0 \xrightarrow[p]{a} s_{Goal}$, приводящее к достижению цели, в текущем состоянии $s_t = (p_1^t, ..., p_k^t)$ можно только после поступления афферентации о завершении предыдущего действия, приводящего к этому состоянию. Затем можно выбрать правило с начальным состоянием $s_0 = (p_{i_1}^0, ..., p_{i_0}^0)$, соответствующим текущему состоянию $\{p_{i_1}^0, ..., p_{i_0}^0\} \subset \{p_1^t, ..., p_k^t\}$ (обозначим это как $s_0 \subseteq s_t$ на рис. 1).

Функциональные системы в общем случае являются последовательностями и иерархией функциональных систем FSC. Когда функциональной системой верхнего уровня, удовлетворяющей некоторую потребность, принимается решение и перебираются различные последовательности/иерархии действий по достижению цели подчиненными функциональными системами FSC, то мы принципиально не можем знать тех состояний s_t, которые возникнут в результате реального осуществления этой последовательности/иерархии действий. Мы также не можем знать, какие будут выбраны правила и действия для достижения каждой конкретной цели каждой функциональной системой FSC в этой последовательности / иерархии. Тем не менее, для принятия решения необходим прогноз вероятности достижения цели каждой функциональной системой FSC. Оценку ве-

роятности тогда надо подсчитывать уже не на основании выбранных правил и действий, а на основании статистики достижения целей. Если n – число случаев, когда в некоторую функциональную систему FSC поступил запрос на достижение цели s_{Goal}, а m – число случаев, когда выбранные правила и действия привели к достижению цели s_{Goal}, то $p_{FSC} = m/n$. Поэтому на рис. 1 прогноз достижения цели осуществляется с вероятностью p_{FSC} достижения цели функциональной системой.

Когда в момент времени t пришел запрос на достижение цели s_{Goal} функциональной системой FSC в текущем состоянии $s_t = (p_1^t, ..., p_k^t)$, то она

1) выбирает правило $s_0 \xrightarrow[p]{a} s_{Goal}$ из набора $R_1, ..., R_n$, которое:
   a) применимо в текущей ситуации $s_0 \subseteq s_t$;
   b) может достичь цели s_{Goal} с максимальной вероятностью p;
2) если для текущего состояния s_t нет подходящего правила, то функциональной системой в ответ на запрос возвращается, что цель не достижима и принятый план достижения цели пересматривается;
3) ожидает в акцепторе результатов действия достижение цели s_{Goal} после осуществления действия a;
4) сравнивает с акцептором результатов действия достигнутое состояние $s_{t+q} = (p_1^{t+q}, ..., p_k^{t+q})$ в момент $t+q$ в результате осуществления действия a с целью $s_{Goal} \approx s_{t+q}$. Если $s_{Goal} \subset s_{t+q}$, то цель достигнута, и правило $s_0 \xrightarrow[p]{a} s_{Goal}$ подкрепляется (его статистика увеличивается). Если целевое состояние s_{Goal} не достигнуто, то выбранное правило наказывается (его статистика уменьшается).

В общем случае функциональные системы, обозначаемые как *FS*, объединяют последовательности / иерархии функциональных систем вида *FSC*. Далее функциональную систему *FS* мы будем мыслить как набор

$$FS = (s_{Goal}, FSC_1, ..., FSC_n, p_{FS}),$$

реализующий преобразование

$$FS = s_0 \xrightarrow[\rightarrow s_1 \rightarrow s_2 \rightarrow ... \rightarrow s_{goal} \quad p_{FS} = p_{FSC_1} \cdot ... \cdot p_{FSC_n}]{FSC_1, ..., FSC_n} s_{goal},$$

где

$$FSC_1 = (s_0 \xrightarrow[p_{FSC_1}]{R_1^1, ..., R_{v_1}^1} s_1), \quad FSC_2 = (s_0 \xrightarrow[p_{FSC_2}]{R_1^2, ..., R_{v_2}^2} s_2), \ldots, \quad FSC_n = (s_0 \xrightarrow[p_{FSC_n}]{R_1^n, ..., R_{v_n}^n} s_{goal}).$$

Цель функциональной системы FS состоит в последовательном достижении целей $s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow ... \rightarrow s_{goal}$ функциональными системами $FSC_1, ..., FSC_n$ с суммарной вероятностью $p_{FS} = p_{FSC_1} \cdot ... \cdot p_{FSC_n}$. Такие функциональные системы могут образовываться автоматически, что будет описано ниже.

Функциональные системы FS могут объединять не только последовательности / иерархии функциональных систем FSC, но и функциональных систем FS. Тогда функциональная система $FS = (s_{Goal}, FS_1^1, ..., FS_n^1, p_{FS})$ есть последовательность функциональных систем, реализующих преобразование

$$FS = s_0 \xrightarrow[\rightarrow s_1 \rightarrow s_2 \rightarrow ... \rightarrow s_{goal} \quad p_{FS} = p_{FS_1^1} \bullet ... \bullet p_{FS_n^1}]{FS_1^1, ..., FS_n^1} s_{goal},$$

где FS_i^1 – либо FS, либо FSC. Например, если $FS_i^1, FS_j^1 \in \{FS_1^1, ..., FS_n^1\}, i < j$ реализуют преобразования

$$FS_i^1 = \xrightarrow[\rightarrow s_1^i \rightarrow s_2^i \rightarrow ... \rightarrow s_i \quad p_{FS_i^1}]{FS(i)_1^2, ..., FS(i)_{n_i}^2} s_i, \quad FS_j^1 = \xrightarrow[\rightarrow s_1^j \rightarrow s_2^j \rightarrow ... \rightarrow s_j \quad p_{FS_j^1}]{FS(j)_1^2, ..., FS(j)_{n_j}^2} s_j,$$

то функциональные системы $FS(i)_1^2, ..., FS(i)_{n_i}^2, FS(j)_1^2, ..., FS(j)_{n_j}^2$ находятся уже на уровне 2 и реализуют преобразование:

$$FS = s_0 \xrightarrow[\rightarrow s_1 \rightarrow s_2 \rightarrow ... \rightarrow [\rightarrow s_1^i \rightarrow s_2^i \rightarrow ... \rightarrow s_i] ... \rightarrow [\rightarrow s_1^j \rightarrow s_2^j \rightarrow ... \rightarrow s_j] ... \rightarrow s_{goal} \quad p_{FS}]{FS_1^1, ..., FS_i^1[FS(i)_1^2, ..., FS(i)_{n_i}^2], ..., FS_j^1[FS(j)_1^2, ..., FS(j)_{n_j}^2], ..., FS_n^1} s_{goal}.$$

Каждая функциональная система представляет собой тот или иной способ достижения цели s_{Goal}, В соответствии с теорией организации движений Н. А. Бернштейна (см.: [Бернштейн, 1997]), ведущим уровнем организации движений является верхний уровень $FS = (s_{Goal}, FS_1^1, ..., FS_n^1, p_{FS})$ ранга 1, соответствующий смыслу решаемой задачи. Функциональная система верхнего уровня может вызывать функциональные системы более низких уровней. Когда приходит запрос на достижение цели s_{Goal} функциональной системой FS, то она:

1) выбирает правила, применимые в текущей ситуации, для первой из функциональных систем FSC, входящих в данную функциональную систему. Если для текущего начального состояния s_0 первой FSC нет подходящего правила, то функциональная система FS не применима к данной ситуации и принятый план достижения цели пересматривается;

2) формирует «конкретную цель» (высшую мотивацию) в виде последовательности и иерархии целей всех, входящих в нее, функциональных подсистем. Например, для приведенной выше функциональной системы это будет последовательность

$$s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow ... \rightarrow [\rightarrow s_1^i \rightarrow s_2^i \rightarrow ... \rightarrow s_i] ... \rightarrow [\rightarrow s_1^j \rightarrow s_2^j \rightarrow ... \rightarrow s_j] ... \rightarrow s_{goal};$$

3) прогнозирует достижение цели s_{Goal} с вероятностью p_{FS};

4) ожидает (акцептором результатов действия) достижение всей последовательности и иерархии целей всех входящих в нее *FSC* после выполнения соответствующих действий;

5) запускает последовательное выполнение действий в функциональных подсистемах FSC;

6) если в какой-либо функциональной подсистеме цель не достигнута, то акцептором результатов действий этой функциональной системы фиксируется, что цель не достигнута и возникает ориентировочно-исследовательская реакция, которая выбирает другую функциональную систему FS для достижения цели s_{Goal}. Правила этой функциональной подсистемы наказываются;

7) достижение результата каждой функциональной подсистемой фиксируется акцептором результатов действия, и вся последовательность действий подкрепляется.

Опишем все элементы архитектуры функциональных систем, используя введенные определения.

Афферентный синтез включает в себя синтез мотивационного возбуждения, памяти, обстановочной и пусковой афферентации, а также обратную афферентацию об осуществленных действиях. Вся эта афферентация может быть задана набором сенсоров $S_1, ..., S_n$, включая сенсоры *мотивационного возбуждения*, *обстановочной и пусковой афферентаций*. Мотивационным возбуждением также задается цель $Goal = (p_{i_1}^{goal}, ..., p_{i_{goal}}^{goal})$.

Память. Каждая цель может достигаться различными последовательностями действий, реализуемыми различными функциональными системами. Поэтому мотивация извлекает из памяти все функциональные системы

$$FS = (s_{Goal}, FS_1^1, ..., FS_n^1, p_{FS}),$$

приводящие к достижению этой цели.

Обстановочная и пусковая афферентации задают текущее состояние системы $s_t = (p_1, ..., p_k)$ в каждый момент времени *t*. Начальные состояния $s_0 = (p_{i_1}^0, ..., p_{i_0}^0)$ применяемых в этот момент правил $s_0 \xrightarrow[p]{a} s_e$ должны соответствовать текущему состоянию системы $s_0 \subseteq s_t$.

«Вытягивая» из памяти весь накопленный опыт, мотивационное возбуждение как цель преобразуется в *конкретную цель* «высшую мотивацию», определяющую способ своего достижения. Для каждой функциональной системы

$$FS = (s_{Goal}, FS_1^1, ..., FS_n^1, p_{FS})$$

конкретной целью является вся последовательность и иерархия целей всех входящих в нее функциональных подсистем, например

$$s_0 \to s_1 \to s_2 \to ... \to [\to s_1^i \to s_2^i \to ... \to s_i] ... \to [\to s_1^j \to s_2^j \to ... \to s_j] ... \to s_{goal}.$$

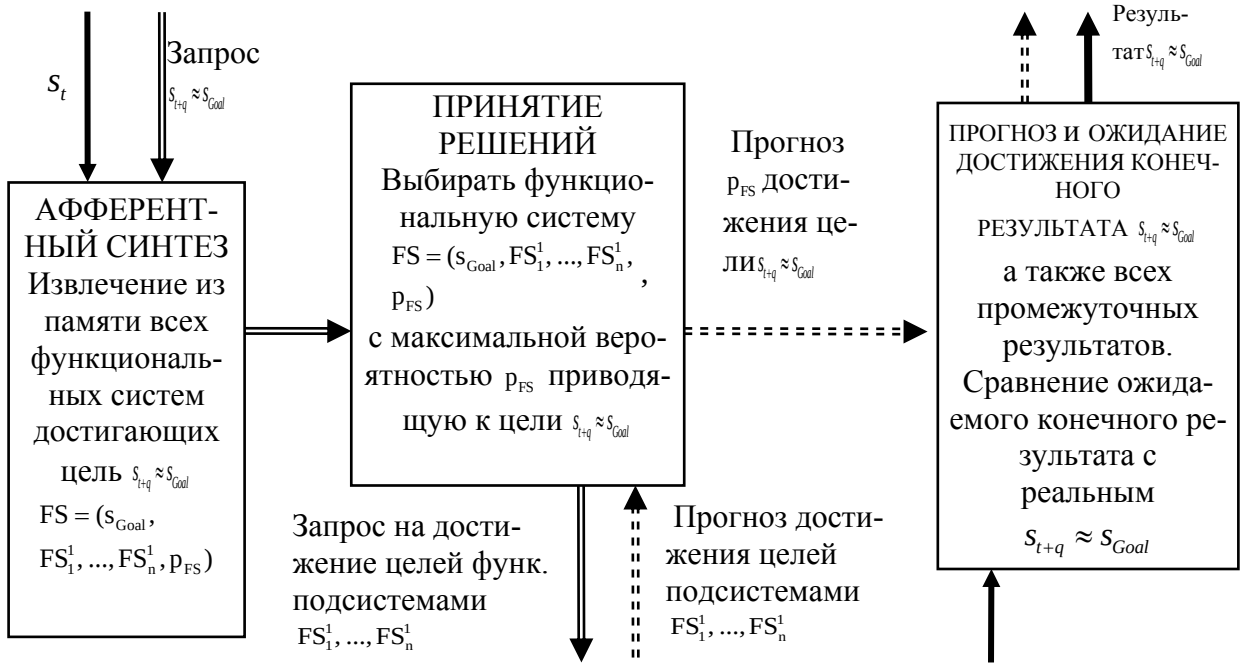

Рис. 2. Схема функциональной системы

Принятие решения. На стадии афферентного синтеза мотивационным возбуждением может быть извлечено из памяти множество функциональных систем $FS = (s_{Goal}, FS_1^1, ..., FS_n^1, p_{FS})$, достигающих цель s_{Goal}. На стадии принятия решения выбирается одна из них и фиксируется *конкретный план действий.* Процесс принятия решений осуществляется переключающей функции эмоций (см. рис. 2).

Акцептор результатов действия. Мотивационное возбуждение, преобразуясь в конкретную цель, извлекает из памяти также и конкретный критерий достижения цели – *акцептор результатов действия*, который состоит из всей совокупности критериев по достижению всей последовательности и иерархии целей

$$s_0 \to s_1 \to s_2 \to ... \to [\to s_1^i \to s_2^i \to ... \to s_i] ... \to [\to s_1^j \to s_2^j \to ... \to s_j] ... \to s_{goal}.$$

Автоматическое формирование новых функциональных систем. Новые функциональные системы FS могут автоматически формироваться путем объединения последовательностей функциональных систем реализующих некоторую устоявшуюся последовательность действий. Последовательность функциональных систем

$$FS_1 = \xrightarrow[\to s_1^1 \to s_2^1 \to ... \to s_1 \quad p_{FS_1}]{FS_1^1, ..., FS_{n_1}^1} s_1, \; FS_2 = \xrightarrow[\to s_1^2 \to s_2^2 \to ... \to s_2 \quad p_{FS_2}]{FS_1^2, ..., FS_{n_2}^2} s_2, \ldots,$$

$$FS_n = \xrightarrow[\to s_1^n \to s_2^n \to ... \to s_n \quad p_{FS_n}]{FS_1^n, ..., FS_{n_n}^n} s_n$$

автоматически объединяются в функциональную систему

$$FS = \xrightarrow[\to s_1 \to s_2 \to ... \to s_n \quad p_{FS} = p_{FS_1} \cdot ... \cdot p_{FS_n}]{FS_1, ..., FS_n} s_n,$$

если последовательность действий не прерывается и не переключается в середине выполнения на другую последовательность действий, так как в этом случае вероятность достижения цели p_{FS} функциональной системой не будет равна произведению $p_{FSC_1} \cdot ... \cdot p_{FSC_n}$ вероятностей входящих в нее функциональных подсистем.

Автоматическое объединение функциональных систем происходит по той же причине, что и формирование правил – замыкаются условные связи между началом выполнения первой функциональной системы FS_1 и результатом всей последовательности действий. Это становится возможным при условии, что FS_1 последовательно получает все результаты $s_1 \to s_2 \to ... \to s_n$, не переключаясь на другую последовательность действий.

5. Заключение

Выше мы показали, как используя задачный подход применительно к когнитивной Теории Функциональных Систем можно построить формальную модель обучающегося рационального агента, основанного на *цели.* Такая модель агента

была построена и проверена в ряде компьютерных экспериментов (см. сноску 2 выше), в которых она продемонстрировала достаточно естественное поведение. Однако, чтобы охватить весь перечень задач, представленный в книге С. Рассела и П. Норвига [Рассел, Норвиг, 2006], желательно дополнительно построить еще и формальную модель рационального агента, основанного на *полезности*. И здесь вновь можно воспользоваться результатами исследований в области когнитивных наук. Так, например, используя теорию эмоций П. В. Симонова, мы уже показали (см: [Витяев, 2006]), что понятие полезности может быть достаточно адекватно и точно определено на основе понятия эмоции. Это дает нам возможность рассматривать формальный процесс принятия решений с использованием критерия полезности как некоторый аналог *естественного* эмоционального процесса. Отсюда, в свою очередь, появляется возможность при моделировании рациональных агентов на основе полезности использовать системы принятия решений, основанных на *эмоциях*. Такой проект также был успешно реализован и результаты экспериментов представлены в [Demin, Vityaev, 2014].

Все вышесказанное убедительно демонстрирует исключительно высокий научный и прикладной потенциал задачного подхода к когнитивным наукам и искусственному интеллекту в целом.

Список литературы / References

Анохин П. К. Проблема принятия решения в психологии и физиологии // Проблемы принятия решения. М.: Наука, 1976. С. 7–16.

Anokhin P. K. Problemy prinyatiya resheniya v psihologii b fiziologii. In: Problemy prinyatiya resheniya. Moscow, Nauka, 1976, p. 7–16 (in Russ.)

Бернштейн Н. А. Биомеханика и физиология движений. М.: МПСИ, 1997.

Bronshtein N. A. Biomehanika i fiziologiya dvizhenii. Moscow, MPSI, 1997. (in Russ.)

Витяев Е. Е. Принятие решений. Переключающая и подкрепляющая функции эмоций // Нейроинформатика-2006: Сб. науч. тр. VIII Всерос. науч.-техн. конф. (Москва, МИФИ, 24–27 января 2006). М: Изд-во МИФИ, 2006. С. 24–30.

Vityaev E. E. Prinyatie reshenii. Perekluchaushchaya i podkreplyaushchaya funktsii emotsii. In: Neyroinformatika-2006 (Proc. of the VIII Conference "Neyroinformatika", Moscow, January 24–27, 2006). Moscow, MIFI Publishing, 2006, p. 24–30. (in Russ.)

Витяев Е. Е. Логика работы мозга // Подходы к моделированию мышления / Под ред. В. Г. Редько. М.: УРСС, 2014. С. 120–153.

Vityaev E. E. Logica raboty mozga. In: Redko V. G. (ed.). Podhody k Modelirovanoyu Myshleniya. Moscow, URSS, 2014, p. 120–153. (in Russ.)

Витяев Е. Е., Гончаров С. С., Свириденко Д. И. О задачном подходе к искусственному интеллекту // Сибирский философский журнал. 2019. Т. 17, № 4. С. 5–25.

Vityaev E. E., Goncharov S. S., Sviridenko D. I. On the task approach to artificial intelligence. *Siberian Journal of Philosophy*, 2019, vol. 17, no. 4, p. 5–25. (in Russ.)

Демин А. В., Витяев Е. Е. Логическая модель адаптивной системы управления // Нейроинформатика. 2008. Т. 3, № 1. С. 79–107.

Demin A. V., Vityaev E. E. Logicheskaya model adaptivnoi sistemy upravleniya. *Neyroinformatika*, 2008, vol. 3, no. 1, p. 79–107. (in Russ.)

Ершов Ю. Л., Самохвалов К. Ф. О новом подходе к философии математики // Вычислительные системы. 1984. Вып. 101. С. 141–148.

Ershov Yu. L., Samokhvalov K. F. O novom podkhode k filosofii matematiki. *Vycheslitelnye Sistemy*, 1984, vol. 101, p. 141–148. (in Russ.)

Мухортов В. В., Хлебников С. В., Витяев Е. Е. Улучшенный алгоритм семантического вероятностного вывода в задаче 2-мерного анимата // Нейроинформатика. 2012. Т. 6, № 1. С. 50–62.

Muhortov V. V., Khlebnikov S. V., Vityaev E. E. Uluchshennyi algoritm semanticheskogo veroyatnostnogo vyvoda v zadache 2-mernogo animata. *Neyroinformatika*, 2012, vol. 6, no. 1, p. 50–62. (in Russ.)

Рассел С., Норвиг П. Искусственный интеллект: современный подход. М.: Вильямс, 2006.

Russell S., Norvig P. Artificial Intelligence: A Modern Approach. Transl. into Russian. Moscow, Williams Publishing House, 2006. (in Russ.)

Симонов П. В. Высшая нервная деятельность человека. М.: Наука, 1975.

Simonov P. V. Vysshaya nervnaya deyatelnost cheloveka. Moscow, Nauka, 1975. (in Russ.)

Симонов П. В. Эмоциональный мозг. М.: Наука, 1981.

Simonov P. V. Emotsionalnii mozg. Moscow, Nauka, 1981. (in Russ.)

Судаков К. В. Общая Теория Функциональных Систем М.: Медицина, 1984.

Sudakov K. V. Obshchaya funktsionalnaya sistema. Moscow, Meditsina, 1984. (in Russ.)

Anokhin P. K. Functional system. In: Wolman B. (ed.). Dictionary of Behavioral Science. New York, Van Nostrand Reinhold, 1973, p. 151–154.

Anokhin P. K. Biology and Neurophysiology of the Conditioned Reflex and Its Role in Adaptive Behaviour. Oxford, Pergamon Press, 1974.

Cassimatis N. A Cognitive Substrate for Achieving Human-Level Intelligence. *AI Magazine*, 2006, vol. 27, no 2, p. 45–56.

Demin A. V., Vityaev E. E. Learning in a Virtual Model of the C. Elegans Nematode for Locomotion and Chemotaxis. *Biologically Inspired Cognitive Architectures*, 2014, vol. 7, p. 9–14.

Goertzel B. Toward a Formal Characterization of Real-World General Intelligence. In: Baum E., Hutter M., Kitzelmann E. (eds.). Advances in Intelligent Systems Research. (Proc. of the 3rd Conf. on Artificial General Intelligence, AGI 2010, Lugano, Switzerland, March 5–8, 2010). Springer, 2010, p. 19–24.

Langley P. Cognitive Architectures and General Intelligent Systems. *AI Magazine*, 2006, vol. 27, no. 2, p. 33–44.

Rosenbloom P., Gratch J., Ustun V. Towards Emotion in Sigma: From Appraisal to Attention. In: Bieger J., Goertzel B., Potapov A. (eds.). Artificial General Intelligence (Proc. of the 8th International Conference, AGI 2015, Berlin, Germany, July 22–25, 2015). Springer, 2015, p. 142–151.

Vityaev E. E. A Formal Model of Neuron that Provides Consistent Predictions. In: Chella A., Pirrone R., Sorbello R., Johannsdottir K. R. (eds.). Biologically Inspired Cognitive Architectures 2012. Proceedings of the Third Annual Meeting of the BICA Society. Springer, 2013, p. 339–344.

Vityaev E. E. Purposefulness as a Principle of Brain Activity. In: Nadin M. (ed.). Anticipation: Learning from the Past. Springer, 2015, p. 231–254.

Vityaev E. E., Perlovsky L. I., Kovalerchuk B. Y., Speransky S. O. Probabilistic Dynamic Logic of Cognition. *Biologically Inspired Cognitive Architectures*, 2013, vol. 6, p. 159–168.

Wang P., Talanov M., Hammer P. The Emotional Mechanisms in NARS. In: Steunebrink B., Wang P., Goertzel B. (eds.). Artificial General Intelligence (Proc. of the 9th International Conference, AGI 2016, New York, NY, USA, July 16-19, 2016). Springer, 2016, p. 150–159.

Материал поступил в редколлегию
Received
27.01.2020

Сведения об авторах / Information about the Authors

Витяев Евгений Евгеньевич

доктор физико-математических наук

[1] ведущий научный сотрудник Института математики им. С. Л. Соболева СО РАН (Новосибирск, Россия)

[2] профессор кафедры дискретной математики и информатики Новосибирского государственного университета (Новосибирск, Россия)

Evgeny E. Vityaev

Doctor of Sciences (Physics and Mathematics)

[1] Leading researcher, Sobolev Institute of Mathematics SB RAS (Novosibirsk, Russian Federation)

[2] Professor, Department of Discrete Mathematics and Computer Science, Novosibirsk State University (Novosibirsk, Russian Federation)

vityaev@math.nsc.ru

Гончаров Сергей Савостьянович

доктор физико-математических наук, академик РАН

[1] директор Института математики им. С. Л. Соболева СО РАН (Новосибирск, Россия)

[2] заведующий кафедрой дискретной математики и информатики Новосибирского государственного университета (Новосибирск, Россия)

Sergey S. Goncharov

Doctor of Sciences (Physics and Mathematics), Academician of the Russian Academy of Sciences

[1] Director of the Sobolev Institute of Mathematics SB RAS (Novosibirsk, Russian Federation)

[2] Head of the Department of Discrete Mathematics and Computer Science, Novosibirsk State University (Novosibirsk, Russian Federation)

gonchar@math.nsc.ru

Свириденко Дмитрий Иванович

доктор физико-математических наук

[1] советник директора Института математики им. С. Л. Соболева СО РАН (Новосибирск, Россия)

[2] профессор кафедры общей информатики Новосибирского государственного университета (Новосибирск, Россия)

Dmitry I. Sviridenko

Doctor of Sciences (Physics and Mathematics)

[1] Advisor to the Director of the Sobolev Institute of mathematics SB RAS (Novosibirsk, Russian Federation)

[2] Professor, Department of General informatics, Novosibirsk State University (Novosibirsk, Russian Federation)

dsviridenko47@gmail.com