Бабич М.Ю. Понятие рационального агента и многоагентные системы. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей XVII Междунар. научно-техн. конф. – Пенза: ПДЗ, 2017. – С. 11-16.

УДК 004.89

ПОНЯТИЕ РАЦИОНАЛЬНОГО АГЕНТА И МНОГОАГЕНТНЫЕ СИСТЕМЫ

М.Ю. Бабич

THE CONCEPT OF RATIONAL AGENT AND MULTIAGENT SYSTEMS

M. Yu. Babich

Аннотация. В работе рассматриваются понятие рационального или интеллектуального агента и понятие многоагентной системы. Выделяются противоречивые свойства рационального агента, являющиеся следствием принадлежности агента нескольким системам, необходимостью планирования и достижения цели в отведенный отрезок времени, функционированием системы в области незначительного горизонта прогноза.

Ключевые слова: рациональный агент, многоагентные системы, организационно-технические системы, странный аттрактор.

Abstract. The work consider the concept of rational or intelligent agent and the concept of multiagent system. The contradictory properties of the rational agent are distinguished, which are the consequence of the agent's belonging to several systems, the need to plan and achieve the goal within the allotted time interval, the functioning of the system in the area of a minor forecast horizon.

Keywords: rational agent, multiagent systems, organizational-technical systems, strange attractor.

1. Одним из важных понятий в теории искусственного интеллекта является понятие рационального или интеллектуального агента. С этим понятием связаны многоагентные системы, то есть системы, в контуре которых функционирует множество взаимодействующих рациональных агентов.

Самое общее определение агента и рационального агента приведено в [1, с. 103, с. 39]. Агентом является нечто воспринимающее и действующее в определенной среде. Рациональным агентом называется агент, который действует таким образом, чтобы можно было достичь наилучшего результата или в условиях неопределенности наилучшего ожидаемого результата. Очень близко к понятию рационального агента (практически синонимы) понятие интеллектуального агента. Интеллектуальность агента связана с рациональной деятельностью. В идеальном случае интеллектуальный агент в любой ситуации предпринимает наилучшее возможное действие [1, с. 71]. Со временем понятие рационального агента и то, что может обладать его свойствами, уточнялось [2 – 6]. Например, в [7] под рациональным агентом понимается *человек* как целеустремленный агент, то есть рациональный агент, для которого выделено свойство хранения информации о желательных ситуациях.

Проанализируем противоречия в определении и свойствах рационального агента и тем самым определим возможности многоагентных систем; условия, при

которых имитация функционирования реальных систем является адекватной; влияние поведенческих свойств агентов на функционирование системы; взаимодействие систем через агентов, работающих в их контуре.

Ограничимся лишь организационно-техническими системами. Особенностью таких систем является обязательное наличие в них человека, а также их неоднородность, то есть наличие как технических компонентов, так и людей. В качестве рациональных агентов будем понимать человека или человека, управляющего техническим устройством.

Обозначим систему, в которой функционирует агент, через S, среду – через C. Из системы S выделим множество агентов A, состоящее из рациональных агентов a. Эти же обозначения будем использовать, говоря о состоянии системы, среды, агентов в момент времени t – $S(t)$, $C(t)$, $A(t)$. Через P_a и P_S обозначим цели агента и системы. Если агент принадлежит системе, то есть $a \in A \subset S$, то будем считать, что агент функционирует в интересах системы, которой он принадлежит. Всегда ли цели агента и системы, которой он принадлежит, совпадают?

2. Введем понятие суперсистемы W, которое ограничивает множество систем сверху. Суперсистем может быть несколько, системы S могут принадлежать только одной из суперсистем, но некоторая глобальная система, состоящая из нескольких суперсистем, невозможна, точнее, нами не рассматривается.

Любой живой организм обладает следующим свойством. Он никогда не принадлежит только одной системе. Выполняется аксиома [8]:

$$\forall a \in A,\ \exists\, S_1, S_2,\ \text{что}\ (a \in S_1) \wedge (a \in S_2) \wedge (S_1 \neq S_2) \wedge\ (S_1 \not\subset S_2) \wedge (\, S_2 \not\subset S_1) \wedge$$
$$\wedge (S_1 \subseteq W_1) \wedge ((S_2 \subseteq W_1) \vee (S_2 \subseteq W_2)) \wedge (P_{S1} \neq P_{S2}). \qquad (1)$$

Аксиома (1) утверждает, что любой агент принадлежит как минимум двум разным системам из суперсистем, одна из которых не является подмножеством другой. Цели систем, которым принадлежит агент, не совпадают.

Предположим, что рациональный агент a принадлежит одновременно двум системам с различными целями. $(a \in S_1) \wedge (a \in S_2) \wedge (P_{S1} \neq P_{S2})$. Например, агент принадлежит системе «Предприятие» с целью системы «Завершить создание изделия» и подсистеме системы ВУЗ – «Заочная аспирантура» с целью подсистемы «Выпуск кандидата наук». Пусть цели агента совпадают с целями систем. Для достижения цели P_{S1} агент осуществляет последовательность действий $g_1, g_2, \ldots, g_m$, а для достижения цели P_{S2} – последовательность действий $h_1, h_2, \ldots, h_m$. Последовательности действий должны быть рациональны. Если последовательности имеют непересекающиеся временные периоды реализации действий, то с точки зрения наблюдателей из систем S_1 и S_2 агент рационален. Предположим, что временные периоды пересекаются, и агент вынужден сдвинуть начало периода реализации действия g_3 после периода реализации действия h_3, так как не может выполнить в один период сразу два действия. Получаем следующую подпоследовательность: g_1, g_2, h_3, g_3, которая с точки зрения наблюдателя в системе S_1 является нерациональной, а с точки зрения наблюдателя системы S_2 – по-прежнему рациональной. Например, действия в системе «Предприятие»: разработка проекта, изготовление изделия, испытание изделия; действия в системе «Заочная аспирантура»: сдача кандидатских экзаменов, предварительная защита, защита кандидатской диссертации. Агент

имеет все возможности защититься на очередном заседании диссертационного совета, но откладывает защиту, что не рационально, так как ему необходимо провести испытание изделия на предприятии.

Таким образом, возможна ситуация, когда с точки зрения агента и наблюдателя в системе S_1 агент рационален, для наблюдателя в системе S_2 агент нерационален, для наблюдателя в суперсистеме W агент не определен. Условия, при которых агент не может выполнить действия по достижению нескольких целей одновременно, и его поведение рассматривались в [9, 10].

3. Важным параметром достижения цели является интервал времени, в течение которого агент должен достичь выбранной цели. Цель может быть составной, и время является одной из частей составной цели: $P_a = P^* \wedge [t_{min}, t_{max}]$, где P^* – сама цель, $[t_{min}, t_{max}]$ – период, за который она должна быть достигнута. Предположим, что цель P^* достигается при состоянии системы и среды $(S^*(t_{max}), C^*(t_{max}))$ и не может быть достигнута при других состояниях, то есть можно определить $P_a = [t_{min}, t_{max}] \wedge (S^*(t_{max}), C^*(t_{max}))$. К заданному состоянию система и среда в результате действия агентов приходят по реальной траектории: $(S(t_{min}), C(t_{min}))$, $(S(t_1), C(t_1))$,..., $(S^*(t_{max}), C^*(t_{max}))$. Агентом прогнозируются возможные промежуточные состояния $(S'(t_j), C'(t_j))$, где $t_{min} \le t_j \le t_{max}$, S', C' – прогнозируемое состояние системы и среды, то есть агент предполагает, что движение происходит по траектории

$$(S(t_{min}), C(t_{min})), (S'(t_1), C'(t_1)), (S'(t_2), C'(t_2)),..., (S^*(t_{max}), C^*(t_{max})) \quad (2)$$

Считается, что ошибка прогноза будет незначительной, то есть

$$\rho((S(t_j), C(t_j)), (S'(t_j), C'(t_j))) < \varepsilon, \quad (3)$$

где ρ – мера близости. Понятно, что чем больше разность $t_j - t_{min}$, тем больше ошибка (3). Прогнозируемое состояние определяется на основе имеющейся информации и модели системы, среды. На основе прогноза происходит выбор очередных действий агента. Можно предположить, что не всегда выполняется условие

$$\forall\, t' > \tau_2, \exists\, t \in [\tau_1, \tau_2], \text{ что } \rho(S(t), S(t')) < \varepsilon, \rho(C(t), C(t')), \rho(A(t), A(t')) < \varepsilon, \quad (4)$$

то есть в процессе функционирования системы не существует времени t', при котором состояние системы, среды, агентов повторялось бы для ранее известного интервала $[\tau_1, \tau_2]$, то есть каждый раз агенту необходимо адаптироваться к ранее неизвестным состояниям системы и среды. Не зная точного состояния системы и среды, после выбранных и реализованных действий агент должен полагаться только на свой прогноз. После каждого изменения состояния системы и среды он корректирует свою модель системы и среды, осуществляет следующее прогнозирование и выбирает действие из возможных, соответствующее наилучшему из спрогнозированных. Каждый раз агент выбирает интервал предсказания $t_{j+k} - t_j$, исходя, скорее, не из рациональности, а из осторожности, так как ошибка может привести к точке траектории состояний, из которой нет пути в требуемую точку $(S^*(t_{max}), C^*(t_{max}))$. Возможно, что, если бы выбранный интервал прогноза $t_{j+k} - t_j$ был бы чуть шире, то действия агента были бы более эффективными, то есть рациональность отсутствует.

4. В своих прогнозах агент не учитывает всю существующую информацию о системе и среде. Он ее просто не знает. Однако он также не использует всю имеющуюся у него информацию. Для моделирования развития текущей ситуации агент

определяет значимые, по его мнению, параметры системы и среды: (s_1, s_2, …, s_n), (c_1, c_2, …, c_m). Обозначим через S_n и C_m наборы параметров системы и среды – S_n = (s_1, s_2, …, s_n), C_m = (c_1, c_2, …, c_m). Агент отслеживает изменение только этих параметров. Последовательность (2) в представлении агента выглядит следующим образом:

$$(S_n(t_{min}), C_m(t_{min})), (S_n'(t_1), C_m'(t_1)), (S_n'(t_2), C_m'(t_2)),..., (S_n^*(t_{max}), C_m^*(t_{max})), \quad (5)$$

в предположении, что

$$\rho((S(t_j), C(t_j)), (S_n'(t_j), C_m'(t_j))) < \varepsilon. \quad (6)$$

Считается, что прогноз, учитывающий незначащие параметры s_{n+1}, c_{m+1}, незначительно отличается от прогноза без учета этих параметров, то есть

$$\rho((S_n'(t_j), C_m'(t_j)), ((S_{n+1}'(t_j), C_{m+1}'(t_j))) < \varepsilon. \quad (7)$$

Таким образом, агент учитывает только часть доступной информации. Если неучтенные параметры существенным образом изменяют прогнозируемые значения, то есть имеет место:

$$\rho((S_n'(t_j), C_m'(t_j)), (S_{n+1}'(t_j), C_{m+1}'(t_j))) >> \varepsilon, \quad (8)$$

тогда возможно, что

$$\rho((S(t_j), C(t_j)), (S_n'(t_j), C_m'(t_j))) >> \varepsilon. \quad (9)$$

Это означает, что агент ошибается в прогнозе и это не рационально.

В настоящее время идет активное исследование открытых, нелинейных систем. Результаты исследований можно перенести на системы рассматриваемого класса. Некоторые системы при определенных условиях достаточно часто могут входить в область странного аттрактора. Область странного аттрактора – это область, для которой горизонт прогноза очень мал. Наличие возможных разнообразных факторов, от которых зависят будущие состояния системы, среды, затрудняет прогнозирование следующей точки траектории [11]. Наличие неучтенных параметров s_{n+1}, c_{m+1} в области странного аттрактора осуществляет выполнение неравенства (8), (9). Следовательно, в области странного аттрактора не существует рациональных агентов.

Выводы.

Таким образом, понятие рационального агента является относительным. Необходимо определить, где находится наблюдатель агента, включая систему наблюдения и временной интервал, определить условия, при которых агент не может выполнить действия по достижению нескольких целей одновременно, определить горизонт прогноза, множество доступной и значащей информации для агента, определить возможность его ошибочных действий. Все перечисленное, будучи не определенным, может внести неточность в функционирование модели многоагентной системы, в поведенческие свойства рациональных агентов.

Библиографический список

1. Рассел С., Норвиг П. Искусственный интеллект. Современный подход. М.: Вильямс, 2006. 1048 с.

2. Городецкий В.И., Грушинский М.С., Хабалов А.В. Многоагентные системы (обзор) // Новости искусственного интеллекта. 1998. №2. С. 64 – 116.

3. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. М.: УРСС, 2002. 352 с.

4. Михеенкова М.А., Финн В.К. Об одном подходе к распознаванию рациональности в коллективах агентов // Искусственный интеллект и принятие решений. 2010. № 2. С. 22 – 32.

5. Википедия. Интеллектуальный агент. URL: https://ru.wikipedia.org/wiki/интеллектуальный_агент

6. Бабич М.Ю., Ковалева В.С., Нисенбаум Е.Э., Ползунов Н.В., Рыбин С.А. Имитационная модель обнаружения нарушителей границы охраняемой области // Вопросы радиоэлектроники. Серия ЭВТ. 2008. Вып. 5. С. 112 – 120.

7. Виноградов Г.П., Кузнецов В.Н. Моделирование поведения агента с учетом субъективных представлений о ситуации выбора // Искусственный интеллект и принятие решений. 2011. № 3. С. 58 – 72.

8. Бабич М.Ю., Кузнецов В.Е. Развитие теории моделирования организационно-технических систем на основе аксиомы принадлежности агентов нескольким системам // Вопросы радиоэлектроники. Серия СОИУ. 2015. Вып. 2. С. 44 – 53

9. Бабич М.Ю., Бабич А.М. Алгоритмы достижения целей адаптивных агентов в контуре нескольких организационно-технических систем // XXI век: итоги прошлого и проблемы настоящего плюс. Серия: «Технические науки. Информационные технологии». 2015. Вып. № 3 (25). С. 77 – 81.

10. Бабич М.Ю.Общий алгоритм функционирования систем и агентов в случае выполнения условий аксиом принадлежности // XXI век: итоги прошлого и проблемы настоящего плюс. Серия: «Технические науки. Информатика, вычислительная техника и управление». 2016. Вып. № 3 (31). С. 85 – 89.

11. Князева Е.Н., Курдюмов С.П. Основания синергетики. Синергетическое мировоззрение. М.: Книжный дом ЛИБРОКОМ, 2014. 256 с.

Бабич Михаил Юрьевич
АО «НПП «Рубин»,
г. Пенза, Россия
E-mail: babichmj@mail.ru

Babich M.Yu.
Scientific-Industrial Enterprice
Joint-Stock Company "Rubin",
Penza, Russia