УДК 007:519.816
DOI: 10.15827/0236-235X.122.239-245
Дата подачи статьи: 06.04.18
2018. Т. 31. № 2. С. 239–245

О РЕАЛИЗАЦИИ СРЕДСТВ МАШИННОГО ОБУЧЕНИЯ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ РЕАЛЬНОГО ВРЕМЕНИ

А.П. Еремеев [1], *д.т.н., профессор, eremeev@appmat.ru*
А.А. Кожухов [1], *аспирант, saaanchezzz@yandex.ru*
В.В. Голенков [2], *д.т.н., профессор, golen@bsuir.by*
Н.А. Гулякина [2], *к.ф.-м.н., доцент, guliakina@bsuir.by*

[1] *Национальный исследовательский университет «МЭИ», ул. Красноказарменная, 14, г. Москва, 111250, Россия*
[2] *Белорусский государственный университет информатики и радиоэлектроники (БГУИР), ул. П. Бровки, 6, г. Минск, 220013, Республика Беларусь*

В работе дан анализ методов обучения с подкреплением (RL-обучения) в плане их использования в интеллектуальных системах реального времени на примере интеллектуальных систем поддержки принятия решений реального времени. Описана реализация алгоритмов обучения с подкреплением на основе временных (темпоральных) различий и рассмотрены основные преимущества использования гибких алгоритмов, которые могут оказывать значительное влияние на эффективность и производительность интеллектуальных систем реального времени.

Гибкие алгоритмы могут иметь решающее значение для интеллектуальных систем поддержки принятия решений реального времени, так как они способны находить приемлемые решения в условиях жестких временных ограничений и улучшать их (вплоть до получения оптимальных) при увеличении предоставляемых ресурсов (особенно временных). Предложен гибкий алгоритм, включающий в себя статистический модуль прогнозирования и мультиагентный модуль RL-обучения. Рассмотрены возможности внедрения разработанного гибкого алгоритма в подсистему прогнозирования интеллектуальных систем реального времени типа интеллектуальных систем поддержки принятия решений реального времени для управления и мониторинга сложного технологического объекта.

Описываются подход к реализации перехода от обучения интеллектуальных систем, основанных на знаниях, к обучению средств их разработки (при этом архитектура такой интеллектуальной системы рассматривается как основа обеспечения ее гибкости и обучаемости), а также направления обучения и самообучения интеллектуальных систем, их способность приобретать знания и навыки из различных источников.

Дается обоснование применения развиваемой в работе технологии OSTIS для разработки интеллектуальных систем, основанных на знаниях, включая интеллектуальные системы реального времени.

Ключевые слова*: искусственный интеллект, интеллектуальная система, реальное время, поддержка принятия решений, машинное обучение, самообучение, обучение с подкреплением, гибкие алгоритмы, технология разработки интеллектуальных систем, программное средство.*

Для реализации Стратегии научно-технологического развития Российской Федерации в плане перехода к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, создания систем обработки больших объемов данных, машинного обучения и искусственного интеллекта, а также аналогичной стратегии Республики Беларусь, актуальной является задача разработки эффективных методов машинного обучения в составе современных *интеллектуальных систем* (ИС), особенно ИС *реального времени* (РВ), функционирующих при достаточно жестких временных ограничениях и неопределенности (зашумленности) поступающей информации, а также включения средств машинного обучения в инструментальные средства разработки ИС и ИС РВ.

Интеграция методов обучения с подкреплением и гибких алгоритмов в ИС РВ

Применительно к ИС и особенно к ИС РВ активно используются методы обучения с подкреплением (reinforcement learning, RL) [1, 2], основанные на использовании довольно большого количества информации для обучения в произвольной окружающей среде. К основным достоинствам RL-обучения в плане использования в ИС РВ относятся:

– использование простой обратной связи на основе скалярных платежей;

– поддержка режима оперативного реагирования при необходимости быстрой адаптации агента к изменениям внешней среды;

– интерактивность и возможность изменения (пополнения) анализируемых данных;

– действенность в недетерминированных средах;

– эффективность в сочетании с темпоральными моделями для задач нахождения последовательных решений;

– открытость к модификации и сравнительная простота включения в ИС различного назначения (планирования, управления, обучения и т.д.).

Как показано в [2], одним из наиболее эффективных в плане использования в ИС РВ типа *ИС поддержки принятия решений РВ* (ИСППР РВ) относительно интегрированного критерия «качество

обучения – временные затраты» является RL-обучение на основе темпоральных различий (temporal-difference, TD), когда процесс обучения основывается непосредственно на получаемом опыте без предварительных знаний о модели поведения окружающей среды. Отмечено, что ключевой особенностью TD-алгоритмов является обучение на основе различий во временных последовательных предсказаниях. TD-методы, предназначенные для многомерных временных рядов, способны обновлять расчетные оценки, не дожидаясь окончательного результата, то есть являются самонастраиваемыми (и в определенном смысле самообучаемыми). Последнее свойство весьма важно для динамических ИС семиотического типа, способных адаптироваться и подстраиваться к изменениям в управляемом объекте и окружающей среде [3, 4]. С использованием TD-методов в ИС РВ можно решать задачи как предсказания значений некоторых переменных в течение нескольких временных шагов, так и управления, основанные на RL-обучении агента влиянию на окружающую среду.

Для возможности обучения и адаптации к изменениям внешней среды агент должен обладать памятью для хранения предыстории. Использование мультиагентного подхода, базирующегося на применении групп автономных взаимодействующих между собой субъектов (агентов), имеющих общую интеграционную среду и способных получать, хранить, обрабатывать и передавать информацию в интересах решения как собственных, так и корпоративных (общих для группы агентов) задач анализа и синтеза информации, является перспективным подходом для динамических ИС РВ. Мультиагентные системы характеризуются возможностью параллельных вычислений, обмена опытом между агентами, отказоустойчивостью, масштабируемостью и т.д. [5].

При разработке современных ИС РВ, а также инструментальных средств создания таких систем особое внимание необходимо уделить средствам прогнозирования развития ситуации на объекте и последствиям принимаемых решений, экспертным методам и средствам обучения, оптимальному использованию доступных ресурсов системы и возможности работы в среде с достаточно жесткими ограничениями по времени. Для решения поставленных задач была разработана инструментальная программная среда с использованием параллельных алгоритмов для статистических методов и методов RL-обучения и спроектирована модель гибкого (anytime) алгоритма, который способен находить приемлемые решения в условиях жестких временных ограничений и доступных ресурсов [2, 6], что необходимо для ИС РВ и, в частности, ИСППР РВ, предназначенных для помощи оперативному персоналу при управлении сложными технологическими объектами типа энергообъектов (энергоблоков) в пределах допустимых отклонений, с учетом текущего состояния окружающей среды и использования параллельных и фоновых вычислений. Разработанный алгоритм способен улучшать свои решения (вплоть до нахождения оптимального) при увеличении временных интервалов. Использование гибких алгоритмов открывает возможности оперативной модификации и адаптации ИС РВ к различным изменениям на объекте и во внешней среде, расширяет область применения и способность увеличивать производительность таких систем в целом.

Гибкие алгоритмы все чаще используются в ряде практических областей, включая планирование и поиск решения, глубокие сети доверия, оценку диаграмм влияния, обработку запросов к БД, контроль и сбор информации и т.д. Данный подход может иметь решающее значение для ИС РВ при наличии большого количества датчиков, способных к анализу, и большой вычислительной сложности алгоритмов планирования для нахождения приемлемых в случае ограниченного времени или оптимальных при наличии достаточных временных ресурсов решений и может значительно повысить производительность и расширить область применения ИС РВ.

На примере ИСППР РВ для ИС РВ была разработана подсистема прогнозирования на основе интеграции методов прогнозирования с использованием статистических и экспертных методов, алгоритмов TD-методов обучения и гибкого алгоритма поиска решения, способная получать результаты в условиях достаточно жестких временных ограничений [2]. В идеальном варианте работы алгоритма при наличии достаточного времени и ресурсов будут выполнены параллельное прогнозирование комбинированным статистическим методом, обучение на основе RL-обучения различными методами, анализ полученных результатов и нахождение оптимального решения. В условиях жестких временных ограничений применяется технология разбиения на этапы [7], согласно которой алгоритм выбирает наиболее перспективный относительно точности прогноза и времени выполнения путь и рассчитывает результат только наиболее адекватными в текущий момент методами. При этом все остальные этапы могут исполняться в фоновом режиме с целью включения их в анализ на последующих шагах.

На рисунке 1 изображена функциональная схема подсистемы прогнозирования для ИС РВ, интегрирующая методы машинного RL-обучения на основе темпоральных различий (от наиболее простого метода TD(0) до более сложных TD-методов: SARSA – с интегрированной оценкой ценности стратегий, Q-обучение – с разделенной оценкой ценности стратегий, TD(λ) – с временным различием протяженностью n шагов) [2] с гибким алгоритмом поиска решения. На схеме пунктиром обозначены необязательные этапы, исполнение

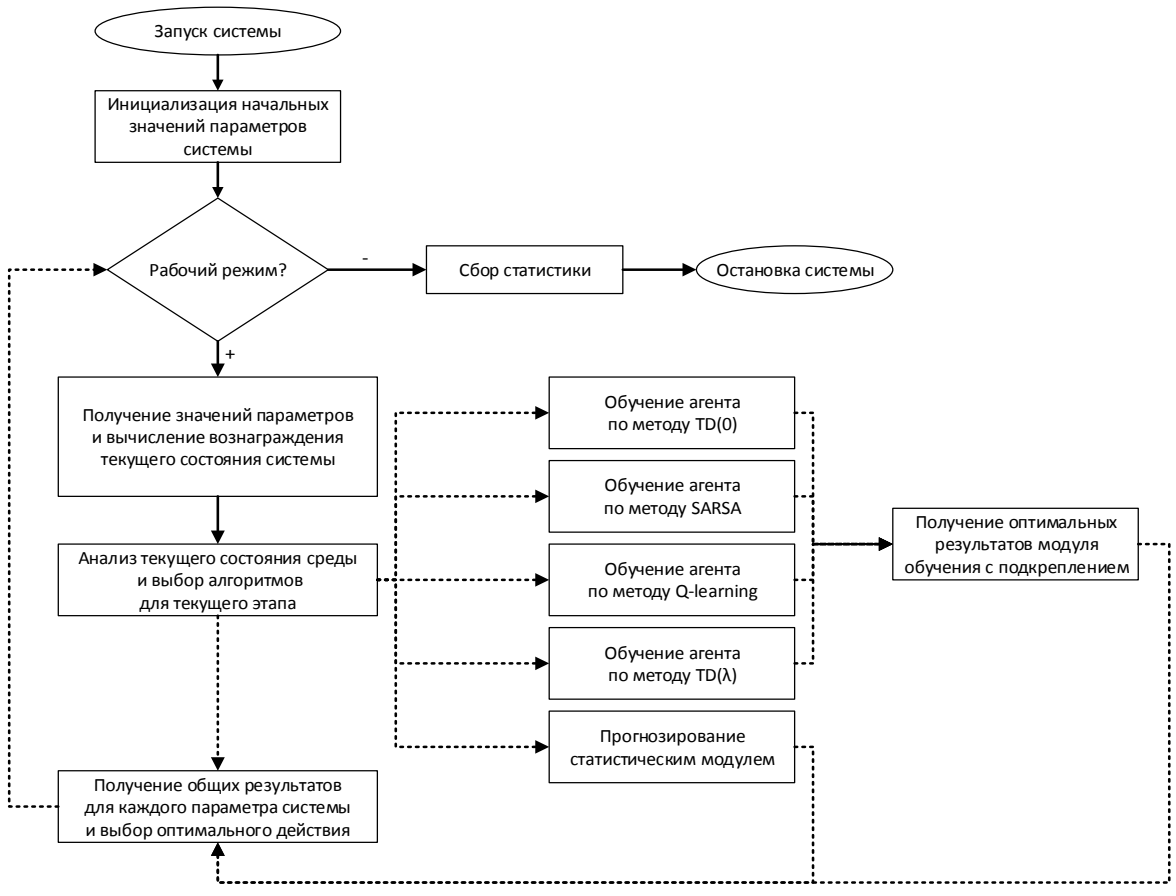

Рис. 1. Функциональная схема подсистемы прогнозирования ИС РВ

Fig. 1. Functional diagram of a real-time intelligent system forecasting subsystem

которых может варьироваться относительно текущего состояния системы.

Интеграция методов позволила реализовать следующие преимущества:

– гибкость и адаптируемость: в зависимости от состояния среды могут использоваться различные параллельные алгоритмы обучения и прогнозирования и их комбинации;

– возможность расчета предполагаемого действия независимо от доступных времени и ресурсов памяти;

– возможность практически немедленной выдачи решения по запросу и продолжение расчетов в фоновом режиме;

– выявление наиболее эффективных алгоритмов для текущей среды и нахождение наилучших (локально или глобально оптимальных) решений.

От обучения ИС к обучению средств их разработки на основе технологии OSTIS

Обучаемость ИС создает необходимые условия для обеспечения быстрых темпов их эволюции, для расширения множества решаемых ими задач и для повышения качества решения, для быстрой адаптации ИС к изменениям внешней среды и условий эксплуатации [8–12]. Уровень обучаемости ИС определяется уровнем развития ее средств самообучения и прежде всего уровнем ее рефлексивности, способности к самоанализу. Важнейшим критерием качества предлагаемых технологий разработки ИС и ИС РВ является то, какой уровень обучаемости разрабатываемых ИС они обеспечивают. ИС, разрабатываемые на основе технологии OSTIS (которые будем называть OSTIS-системами), обладают высоким уровнем гибкости и обучаемости [8].

Гибкость OSTIS-систем определяется следующими факторами:

– смысловой характер внутреннего представления знаний;

– развитые средства систематизации хранимых знаний, структуризации *базы знаний* (БЗ);

– развитый уровень ассоциативной организации памяти;

– агентно-ориентированная организация обработки знаний, которая управляется самой обрабатываемой БЗ.

Обучаемость OSTIS-систем определяется их способностью обнаруживать противоречия (ошибки), информационные дыры (пропуски) и информационный мусор, которые появляются в текущем

состоянии БЗ как в результате приобретения знаний и навыков извне, так и в процессе решения различных задач. При этом во втором случае каждое обнаруженное противоречие, информационная дыра или информационный мусор явно связываются с породившим их информационным процессом для уточнения причины их возникновения.

Технология OSTIS, как и разработанные на ее основе ИС, также обладает высоким уровнем гибкости (адаптируемости), поскольку реализована в виде интеллектуальной метасистемы, которая сама является OSTIS-системой. Общая архитектура ИС как OSTIS-системы, создаваемой на основе данной технологии, приведена на рисунке 2.

Под самообучением ИС понимается автоматизация различных процессов, направленных на обучение ИС и осуществляемых самой ИС. К таким автоматизируемым процессам относятся следующие:

– перманентный анализ качества БЗ и интегрированного решателя задач обучаемой ИС, результатом которого являются, например, выявление различного рода ошибок (противоречий), информационного мусора, информационных дыр (недостающих знаний и навыков), оценка достоверности (правдоподобия) новых приобретаемых (в том числе вводимых) знаний и навыков, а также формирование спецификации этих знаний и навыков (кто автор, момент появления в системе, тип и т.д.);

– автоматизируемые виды совершенствования текущего состояния БЗ и интегрированного решателя задач, автоматическое исправление некоторых ошибок, ликвидация информационного мусора, спецификация информационных дыр, систематизация приобретенных знаний и навыков, извлечение неявно представленных знаний из данных (индуктивный вывод, самообучение по прецедентам и на основе аналогий и т.д.);

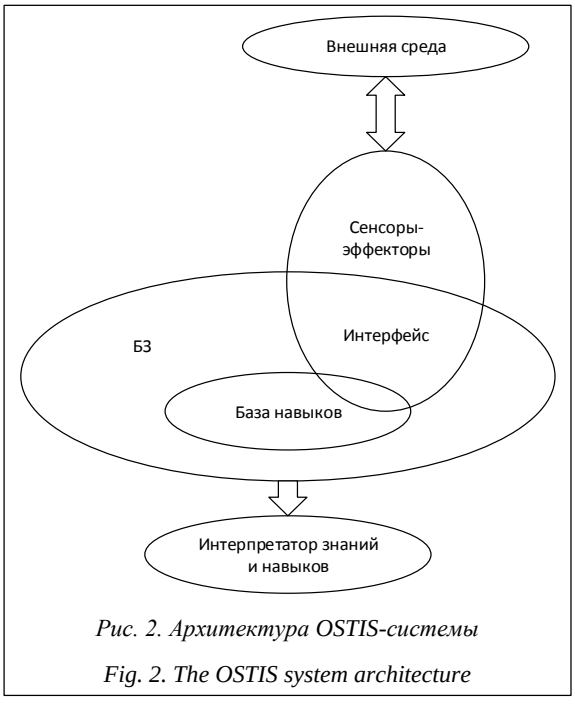

Рис. 2. Архитектура OSTIS-системы

Fig. 2. The OSTIS system architecture

– координация деятельности учителей-разработчиков (экспертов, ЛПР – лиц, принимающих решения) ИС, которые становятся самостоятельными агентами (субъектами) разработки ИС, управляемыми БЗ этой системы.

Отметим, что разработка ИС, обладающей развитыми навыками самообучения, принципиально отличается от разработки ИС, не обладающей такими возможностями. Это обусловлено тем, что самообучаемая ИС сама становится одним из субъектов собственного обучения, то есть одним из своих учителей. И это является существенным фактором повышения эффективности обучения, поскольку никто лучше самой обучаемой ИС не знает о ее внутреннем состоянии (состоянии БЗ, интегрированного решателя и других компонентов).

Методика обучения ИС во многом определяется тем, какими средствами самообучения обладает обучаемая ИС.

Направления самообучения ИС типа OSTIS-систем:

– приобретение новых знаний из разных источников;

– извлечение неявных знаний из приобретенных знаний;

– обнаружение закономерностей;

– структуризация БЗ;

– поддержание целостности БЗ (непротиворечивости, полноты, отсутствие «мусора» и т.д.);

– повышение эффективности решения задач на основе анализа собственного функционирования.

В основе обучаемости ИС лежат способность к рефлексии, то есть способность анализировать и оценивать собственное качество и качество своей деятельности, а также способность оперативно (что особенно важно для ИС РВ типа ИСППР РВ) усваивать новые знания и навыки и совершенствовать уже приобретенные. Так, например, обучаемость искусственных нейронных сетей определяется наличием способа автоматической корректировки параметров искусственной нейронной сети на основе результатов ее тестирования, направленной на построение оптимальной структуры связей, настройку параметров связей (метод стохастического градиента, метод обратного распространения ошибки).

Таким образом, обучаемость ИС обеспечивается:

– систематизацией внутреннего представления знаний и навыков (все накапливаемые знания и навыки должны быть приведены в стройную систему);

– достаточно простой моделью интеграции (погружения) новых знаний и навыков в состав БЗ;

– неограниченными возможностями представлять в БЗ всю необходимую для самоанализа

информацию о себе (содержащую знак собственного «Я», полную документацию, описание своих связей с другими сущностями и в том числе связей собственной точки зрения с точками зрения других субъектов);

– способностью к рефлексии и достаточно простой моделью анализа качества текущего состояния БЗ (качества структуры БЗ, полноты, наличия и локализации обнаруженных противоречий и ошибок);

– уровнем развития средств обнаружения и устранения нештатных (в том числе ошибочных) ситуаций в процессе функционирования ИС;

– уровнем развития средств повышения качества текущего состояния БЗ (совершенствования системы накапливаемых знаний и навыков).

Гибкость (адаптируемость, модифицируемость) ИС является основой ее обучаемости (самообучаемости) и определяется трудоемкостью внесения различных изменений в ИС, осуществляемых на различных уровнях ИС в ходе ее обучения. Гибкость OSTIS-систем обеспечивается предложенными в технологии OSTIS базовыми принципами: кодирования информации в памяти ИС на основе SC-кода (Semantic Computer Code); организации памяти ИС (SC-памяти), обеспечивающей хранение и обработку текстов SC-кода (SC-текстов); обработки информации в SC-памяти ИС.

Отметим, что важным достоинством смыслового представления информации на основе SC-кода является то, что в нем явно и четко задаются связи между описываемыми сущностями в виде связей между знаками этих сущностей, а также указывается семантический тип каждой такой связи. Любую знаковую конструкцию можно представить как множество знаков описываемых сущностей и множество знаков связей, связывающих эти описываемые сущности с другими сущностями. При этом на описываемые сущности и на связи между ними не накладывается никаких ограничений. Знаковая конструкция, являющаяся смысловым представлением информации, в общем случае не может быть линейной, поскольку каждая описываемая сущность, являющаяся денотатом соответствующего знака, может иметь множество связей с другими сущностями, описываемыми в этой же знаковой конструкции. Таким образом, смысловое представление информации есть сетевая (графовая) структура в виде *семантической сети* (СС) или совокупности СС [4], что дает возможность в полной мере использовать теорию графов для исследований и построения алгоритмов обработки СС.

Память OSTIS-системы (SC-память) представляет собой нелинейную (графовую) ассоциативную структурно перестраиваемую (графодинамическую) память, в которой обработка информации сводится к изменению не только состояния элементов памяти, но и конфигурации связей между ними. С формальной точки зрения SC-память является динамическим SC-текстом, в котором могут происходить следующие события: удаление SC-узла вместе с удалением всех инцидентных ему SC-связок (знаков различных связей); удаление SC-связки вместе с удалением всех SC-связей, компонентом которых она является; замена типа элемента SC-текста (например, SC-узел может быть преобразован в SC-коннектор); добавление нового элемента SC-текста с указанием связи нового генерируемого в памяти SC-элемента с каким-либо уже присутствующим (хранимым) в памяти SC-элементом.

В SC-памяти обеспечивается возможность хранения и обработки любых внешних информационных конструкций (терминов, иероглифов, текстов, изображений, видеоинформации, аудиоинформации), представленных в электронной форме в виде соответствующих файлов. Через такие файлы осуществляется, в частности, обмен информацией между OSTIS-системой и внешними по отношению к ней субъектами (пользователями, другими OSTIS-системами).

Особое значение для обеспечения гибкости OSTIS-системы имеет развитая форма ассоциативного доступа к любым видам знаний и навыков, хранимых в памяти системы, благодаря развитым средствам спецификации искомых знаний. Так, например, существенно упрощается процедура ассоциативного поиска знаков на основе априорной информации о связях между ними. Такой поиск осуществляется с помощью волновой навигации по пространству связанных знаков. Существенно упрощается также процедура ассоциативного поиска фрагментов хранимой БЗ, удовлетворяющих заданным запросам (требованиям), а также существенно расширяется многообразие видов таких запросов.

Интегрированный решатель задач OSTIS-системы представляет собой иерархическую систему агентов, которые осуществляют обработку БЗ, представленную в SC-коде и хранимую в SC-памяти, и взаимодействуют между собой только через указанную память. Таким образом, весь процесс обработки БЗ в OSTIS-системе управляется самой БЗ. Агенты (называемые SC-агентами) подразделяются на неатомарные, представляющие собой коллективы SC-агентов более низкого уровня, и атомарные, не являющиеся коллективами SC-агентов. При этом каждому SC-агенту соответствует свой класс ситуаций или событий в SC-памяти, инициирующих деятельность этого агента, порождая соответствующий информационный процесс в SC-памяти, основные характеристики и текущее состояние которого представляются в памяти и используются при выполнении этого процесса.

В качестве примера рассмотрим решение некоторой вычислительной задачи с использованием параллельного асинхронного многоагентного подхода. Вычисление числовой функции в SC-памя-

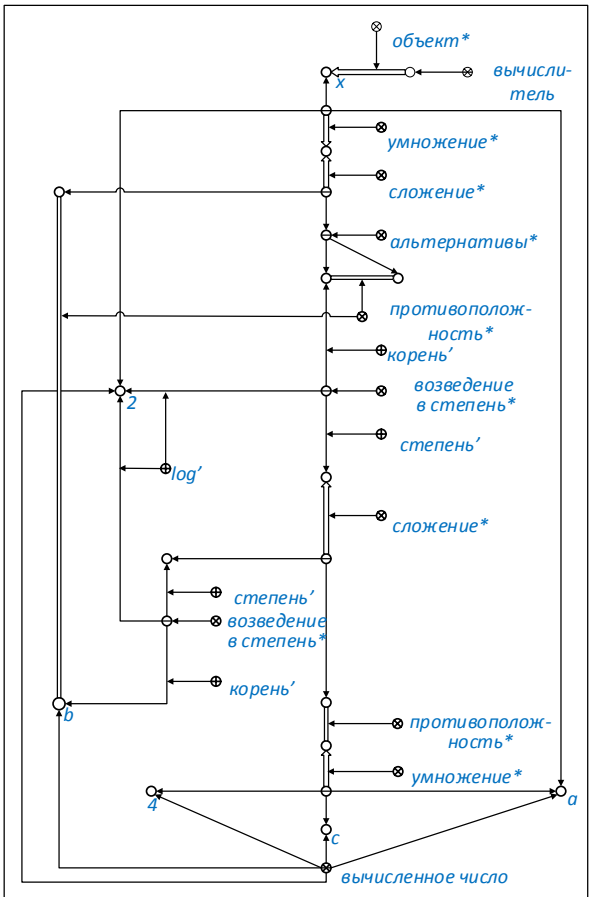

Рис. 3. Формулировка задачи решения квадратного уравнения в SC-коде

Fig. 3. The SC code formulation of the problem of solving a quadratic equation

ти – это построение SC-текста, являющегося представлением числа в заданной системе счисления (например десятичной). Подобная задача формулируется с помощью ключевого понятия «вычислить», обозначающего класс действий, направленных на вычисление чисел, и являющегося условием инициирования арифметических SC-агентов (сложения, вычитания, умножения, деления, возведения в степень, взятия корня, взятия логарифма и т.д.), который при необходимости может пополняться SC-агентами, обеспечивающими вычисление требуемых числовых функций. На рисунке 3 дано представление (формулировка) задачи решения квадратного уравнения в SC-коде.

Заключение

Обучаемость ИС, являясь их важнейшим свойством, создает хорошие предпосылки для существенного расширения жизненного цикла систем и обеспечения высокого уровня конкурентоспособности по сравнению с традиционными компьютерными системами. Но для обеспечения конкурентоспособности ИС, кроме их обучаемости, необходима хорошо продуманная методика обучения (совершенствования). Управление обучением ИС имеет свою достаточно сложную специфику прежде всего в силу специфики объекта обучения. Особенно, если речь идет об ИС РВ.

Рассмотрены методы машинного обучения с подкреплением (методы RL-обучения) на основе временных различий (TD-методы), ориентированные на перспективные ИС РВ, а также интеграция методов обучения с подкреплением и гибких алгоритмов, ориентированная на использование в составе ИС РВ типа ИСППР РВ. Предложена функциональная схема подсистемы прогнозирования для ИС РВ, интегрирующая методы машинного RL-обучения на основе темпоральных различий с гибким алгоритмом поиска решения. В настоящее время данная подсистема внедряется в разрабатываемый прототип ИСППР РВ для управления сложным технологическим объектом (одной из подсистем энергоблока АЭС) и его мониторинга.

Отмечено, что высокий уровень обучаемости OSTIS-систем является основой для обеспечения высоких темпов эволюции самой технологии OSTIS, поскольку эта технология реализуется также в виде OSTIS-системы (метасистемы IMS.ostis). Важнейшими направлениями эволюции технологии OSTIS являются расширение и совершенствование структуры библиотеки многократно используемых компонентов OSTIS-систем, что обеспечивает существенное снижение трудоемкости разработки ИС и ИС РВ. Высокие темпы эволюции технологии OSTIS, открытый характер этой технологии, а также открытый характер участия в ее развитии (в рамках open-source проекта IMS.ostis) обеспечивают высокую конкурентоспособность и перспективность технологии OSTIS.

Работа выполнена при финансовой поддержке РФФИ (проекты №№ 17-07-00553, 18-51-00007) и БРФФИ (проект № Ф16Р-102).

Литература

1. Саттон Р.С., Барто Э.Г. Обучение с подкреплением; [пер. с англ.]. М.: БИНОМ. Лаборатория знаний, 2011. 400 с.

2. Еремеев А.П., Кожухов А.А. Реализация методов обучения с подкреплением на основе темпоральных различий и мультиагентного подхода для интеллектуальных систем реального времени // Программные продукты и системы. 2017. № 1. С. 28–33.

3. Башлыков А.А., Еремеев А.П. Методы и программные средства конструирования интеллектуальных систем поддержки принятия решений для объектов энергетики // Вестн. МЭИ. 2018. № 1. С. 72–85.

4. Осипов Г.С. Методы искусственного интеллекта. М.: Физматлит, 2015. 296 с.

5. Busoniu L., Babuska R., and De Schutter B. Multi-agent reinforcement learning: An overview. Berlin, Germany, Springer, 2010, vol. 310, pp. 183–221.

6. Еремеев А.П., Кожухов А.А., Гулякина Н.А. Исследование и реализация методов обучения с подкреплением для интеллектуальных систем реального времени // Нечеткие системы, мягкие вычисления и интеллектуальные технологии (НСМВИТ-2017): тр. VII Всерос. науч.-практич. конф. СПб: Политехника-сервис, 2017. С. 50–62.

7. Hansen E.A., Zilberstein S. Monitoring and control of anytime algorithms: A dynamic programming approach. J. of Artificial Intelligence, 2001, vol. 126, pp. 139–157.

8. Golenkov V.V., Gulyakina N.A., Grakova N.V., Nikulenka V.Y., Eremeev A.P., Tarasov V.B. From training intelligent systems to training their development means. Proc. Intern. Conf. OSTIS-2018. Minsk, 2018, pp. 81–98.

9. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям. М.: Эдиториал УРСС, 2002. 352 с.

10. Никуленков С.И., Тулупьев А.Л. Самообучающиеся системы. М.: Изд-во МЦНМО, 2009. 288 с.

11. Sutton R.S., Barto A.G. Reinforcement learning. London, The MII Press, 2012, 320 p.

12. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. М.: ДМК Пресс, 2015. 402 с.

Software & Systems
DOI: 10.15827/0236-235X.122.239-245

Received 06.04.18
2018, vol. 31, no. 2, pp. 239–245

ON THE IMPLEMENTATION OF MACHINE LEARNING TOOLS IN REAL-TIME INTELLIGENT SYSTEMS

A.P. Eremeev [1], *Dr.Sc. (Engineering), Professor, eremeev@appmat.ru*
A.A. Kozhukhov [1], *Postgraduate Student, saaanchezzz@yandex.ru*
V.V. Golenkov [2], *Dr.Sc. (Engineering), Professor, golen@bsuir.by*
N.A. Gulyakina [2], *Ph.D. (Physics and Mathematics), Associate Professor, guliakina@bsuir.by*

[1] *National Research University "Moscow Power Engineering Institute", Krasnokazarmennaya St. 14, Moscow, 111250, Russian Federation*
[2] *Belarusian State University of Informatics and Radioelectronics (BSUIR), P. Brovki St. 6, Minsk, 220013, Belarus*

Abstract. The paper analyzes the methods of reinforcement learning in terms of using them in real-time intelligent systems (RT IS) on the example of the real-time intelligent decision support systems (RT IDSS).

The authors describe implementation of reinforcement learning algorithms based on temporal differences and consider the main advantages of using flexible (anytime) algorithms that can have a significant impact on RT IS efficiency and productivity. Flexible algorithms can be crucial to RT IDSS, as they are able to find acceptable solutions under tight time constraints and improve them (up to optimal ones) while increasing available resources (especially temporary ones). The proposed flexible algorithm includes a statistical forecasting module and a module of multi-agent reinforcement learning.

The paper considers the possibilities of implementation of the developed flexible algorithm into a RT IS forecasting subsystem of RT IDSS type to control and monitor a complex technological object.

The paper considers the approach to implementation of a transition from knowledge-based intelligent systems training to training of their development tools. At the same time, the architecture of such intelligent systems is considered as the basis for its flexibility and learning capacity. The paper also examines the areas of intelligent system learning and self-learning, as well as their ability to acquire knowledge and skills from different sources.

The authors provide rationalization for the application the developed OSTIS technology to develop knowledge-based intelligent systems, including the RT IS.

Keywords: artificial intelligence, intelligent system, real time, decision support, machine learning, self-learning, reinforcement learning, flexible (anytime) algorithm, development technology of intelligent systems, software tool.

Acknowledgements. *The work has been supported by the Russian Foundation for Basic Research, projects no. 17-07-00553, no.18-51-00007, by the Belarusian Foundation for Basic Research, projects no. Ф16Р-102.*

References

1. Sutton R.S., Barto A.G. *Reinforcement Learning*. Moscow, BINOM Publ., 2011, 400 p.

2. Eremeev A.P., Kozhukhov A.A. Implementation of reinforcement learning methods based on temporal differences and multi-agent approach for real-time intelligent systems. *Programmnye produkty i sistemy* [Software and Systems] 2017, no. 1, pp. 28–33 (in Russ.).

3. Bashlykov A.A., Eremeev A.P. Methods and software tools of designing of intelligentl decision support systems for power objects. *Vestn. MEI* [Vestn. MPEI]. 2018, no. 1, pp. 72–85 (in Russ.).

4. Osipov G.S. *Metody iskusstvennogo intellekta* [Methods of artificial intelligence]. 2nd ed. Moscow, Fizmatlit Publ., 2015, 296 p.

5. Busoniu L., Babuska R., De Schutter B. *Multi-agent reinforcement learning: An overview*. Berlin, Germany, Springer Publ., 2010, pp. 183–221.

6. Eremeev A.P., Kozhukhov A.A., Gulyakina N.A. Rresearch and implementation of reinforcement learning methods for real time intelligent systems. *Nechetkie sistemy, myagkie vychisleniya i intellektualnye tekhnologii (NSMVIT-2017): tr. VII Vseros. nauch.-praktich. konf.* [Fuzzy Systems, Soft Computing and Intelligent Technology (NSMVIT-2017): Proc. 7th Sci. and Pract. Conf.] St. Petersburg, 2017, pp. 50–62 (in Russ.).

7. Hansen E.A., Zilberstein S. Monitoring and control of anytime algorithms: A dynamic programming approach. *J. Artificial Intelligence*. 2001, vol. 126, pp. 139–157.

8. Golenkov V.V., Gulyakina N.A., Grakova N.V., Nikulenka V.Y., Eremeev A.P., Tarasov V.B. From training intelligent systems to training their development means. *Proc. Int. Conf. "Open Semantic Technologies for Intelligent Systems" (OSTIS-2018)*. Minsk, 2018, pp. 81–98.

9. Tarasov V.B. *Ot mnogoagentnykh sistem k intellektualnym organizatsiyam* [From multi-Agent Systems to Intellectual Organizations]. Moscow, Editorial URSS Publ., 2002, 352 p.

10. Nikulenkov S.I., Tulupev A.L. *Samoobuchayushchiesya sistemy* [Self-Learning Systems]. Moscow, MTSNMO Publ., 2009, 288 p.

11. Sutton R.S., Barto A.G. *Reinforcement Learning*. London, MIT Press, 2012, 320 p.

12. Flach P. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge Univ. Press, 409 p. (Russ. ed.: Moscow, DMK Press, 2015, 402 p.).