Предисловие редактора серии «Adaptive computation and machine learning»

Меня радует, что книга Ричарда Саттона и Эндрю Барто — одна из первых в новой серии «Adaptive computation and machine learning». Данное руководство представляет собой исчерпывающее введение в захватывающую область обучения с подкреплением. В число основателей этой области входят и авторы предлагаемой книги. Она позволяет студентам, специалистам-практикам и исследователям получить наглядное представление о важнейших концепциях обучения с подкреплением, а также познакомиться с математическими основами этой области знаний. В книге говорится также о новейших примерах практических применений обучения с подкреплением, о соотношении обучения с подкреплением с ключевыми проблемами искусственного интеллекта. Обучение с подкреплением обещает стать чрезвычайно важной новой технологией с огромным потенциалом практических применений, а также важным инструментом для проникновения в суть организации интеллектуальных систем.

Задача построения систем, которые могли бы приспосабливаться к окружающей среде, а также обучаться на основе получаемого опыта, давно привлекает внимание исследователей из самых различных областей, включая информатику, технические науки, математику, физику, нейробиологию и когнитологию. Проведенные ими исследования привели к созданию широкого спектра методов обучения, которые могут оказать существенное влияние на многие области науки и техники. В последние годы различные научные сообщества начинают приходить к общему пониманию проблем, связанных с обучением с учителем, обучением без учителя и с обучением с подкреплением. Серия «Adaptive computation and machine learning»[1)], выпускаемая издательством The MIT Press, нацелена на унификацию разнообразных направлений в исследованиях по обучению машин, а также на поощрение высококачественных исследовательских работ и инновационных приложений.

Томас Диттерих

[1)] Русское издание книги выходит в серии «Адаптивные и интеллектуальные системы». — *Прим. ред.*

Посвящается Харри Клопфу

Предисловие

К тому, что сейчас принято называть обучением с подкреплением, мы впервые пришли в 1979 г. Мы оба работали тогда в Массачусетском университете над одним из наиболее ранних проектов, связанных с возвратом к сетям нейроноподобных адаптивных элементов как многообещающему подходу к решению задачи адаптивного искусственного интеллекта. В этом проекте изучалась «гетеростатическая теория адаптивных систем», созданная А. Харри Клопфом. Работа Харри была богатым источником идей, и перед нами стояла задача критически изучить их, а также сопоставить с тем, что было наработано за долгую предшествующую историю исследований в области адаптивных систем. Нам надо было выявить составные элементы этих идей, понять соотношение между ними и их относительную важность. Эта работа продолжается и сейчас, но в 1979 г. мы поняли, что, несмотря на свою простоту, одна из идей, на которых основывается рассматриваемый подход, привлекала удивительно мало внимания с точки зрения вычислительной перспективы. Это была просто-напросто идея обучающейся системы, которая *хочет* чего-то, которая подстраивает свое поведение, чтобы максимизировать значение некоторого особого сигнала из окружающей среды. Это была идея «гедонистической»[1)] обучающейся системы, или, как сказали бы мы сейчас, идея обучения с подкреплением.

Как и другим, нам казалось, что обучение с подкреплением было уже достаточно подробно исследовано еще в ранний период развития кибернетики и искусственного интеллекта. Однако при более глубоком изучении состояния дел в данной области выяснилось, что на самом деле эти вопросы рассматривались весьма поверхностно. Хотя обучение с подкреплением явно служило отправной точкой ранних исследований в области обучения машин, большинство исследователей, которые начинали работать в этой области, переключились на такие вопросы, как

[1)] Гедонизм — направление в этике, в котором наслаждение, удовольствие считается высшей целью и основным мотивом человеческого поведения. Добро при этом определяется как то, что приносит наслаждение, а зло — как то, что влечет за собой страдание. — *Прим. ред.*

классификация образов, обучение с учителем, адаптивное управление, либо просто перестали заниматься исследованиями в области обучения. В результате аспекты обучения, связанные с воздействием среды, пользовались относительно малым вниманием исследователей. Оглядываясь назад, можно утверждать, что эти аспекты были особенно важны, именно работа над ними обеспечила развитие данной научной области. Прогресс в ее компьютерном изучении был невелик до тех пор, пока не удалось осознать, что такая фундаментальная идея все еще слабо изучена.

С тех пор рассматриваемая область исследований прошла долгий путь, обретая зрелость и развиваясь по нескольким направлениям. Обучение с подкреплением постепенно стало одним из наиболее активно исследуемых направлений в области обучения машин, искусственного интеллекта, нейросетевых исследований. Создана прочная математическая база обучения с подкреплением, решен ряд интересных прикладных задач. Сейчас компьютерные исследования в области обучения с подкреплением — это обширная область, в которой активно работают сотни исследователей во всем мире, основные интересы которых лежат в таких разнообразных областях, как психология, теория управления, искусственный интеллект, нейробиология. В частности, большое внимание уделяется выявлению и развитию взаимосвязей обучения с подкреплением с теорией оптимального управления и динамическим программированием. В целом проблема обучения на основе взаимодействия системы со средой для достижения некоторых целей все еще весьма далека от окончательного решения, однако наше понимание данной проблемы значительно улучшилось. Уже понятна роль таких элементов, как обучение на основе временны́х различий, динамическое программирование, аппроксимация функций, прояснилась и перспектива решения проблемы в целом.

Наша цель при написании данной книги состояла в том, чтобы дать ясное и простое изложение важнейших идей и алгоритмов обучения с подкреплением. Нам хотелось, чтобы изложенный материал был доступен читателям из всех упоминавшихся выше научных областей, однако задачу показать во всех деталях значение для них обучения с подкреплением мы не ставили. Материал дается почти исключительно с позиций искусственного интеллекта и технических наук. Подробное изложение взаимосвязей с психологией, нейробиологией и прочими областями отложим до другого времени или оставим другим авторам. Мы не стали также давать скрупулезное формальное описание обучения с подкреплением, поскольку не стремились излагать материал на максимально возможном уровне математической абстракции в стиле

«теорема-доказательство». Мы выбрали такой уровень изложения, который позволяет передать основные математические идеи данной области, не затеняя при этом простоты и потенциальной общности ее важнейших положений.

Книга состоит из трех частей. Часть I — вводная и проблемно-ориентированная. Здесь мы акцентируем внимание читателя на простейших аспектах обучения с подкреплением, а также на основных чертах, отличающих его от родственных направлений. Целая глава в этой части посвящена описанию конкретной задачи обучения с подкреплением, решение которой изучается затем на протяжении всей оставшейся части книги. В части II представлены три важнейших базисных метода поиска решений: динамическое программирование, простые методы Монте-Карло, обучение на основе временны́х различий. Первый из перечисленных методов представляет собой одну из разновидностей методов планирования и предполагает, что имеется явным образом сформулированное знание обо всех аспектах решаемой проблемы, тогда как оставшиеся два метода — это методы обучения. Часть III посвящена обобщению этих методов, а также построению ряда комбинированных подходов на их основе. Используя следы приемлемости, можно унифицировать методы Монте-Карло и обучения на основе временны́х различий, а методы аппроксимации функций, в частности на основе искусственных нейронных сетей, развивают все эти методы таким образом, что они становятся применимыми к задачам существенно более высокой размерности. Мы опять возвращаемся к идее объединения методов, основанных на планировании и на обучении, а также соотносим их с эвристическим поиском. Наконец, мы обсуждаем состояние исследований в области обучения с подкреплением, а также рассматриваем ряд конкретных случаев, включая несколько наиболее интересных и впечатляющих примеров применения обучения с подкреплением, существующих на сегодняшний день.

Наша книга задумывалась как основа для односеместрового курса, ее можно дополнить другой литературой, например более математизированным руководством Бертсекаса и Цициклиса (Bertsekas, Tsitsiklis 1996). Данная книга может быть использована также в качестве части курса более широкой направленности, например, машинного обучения, искусственного интеллекта или нейронных сетей. В таком случае может оказаться желательным ограничиться лишь частью материала, содержащегося в книге. Мы рекомендовали бы тогда взять гл. 1 в качестве краткого обзора тематики обучения с подкреплением, гл. 2 по разд. 2.2, гл. 3 без разд. 3.4, 3.5 и 3.9, а затем взять какое-то количество материа-

ла из оставшихся глав соответственно интересам и имеющемуся в распоряжении времени. Главы 4, 5 и 6 взаимосвязаны, и их лучше давать именно в такой последовательности, из них гл. 6 наиболее важна как для всей области обучения с подкреплением, так и для остального материала книги. Курс, основным содержанием которого является обучение машин или нейронные сети, должен включать гл. 8, а курс, связанный с тематикой искусственного интеллекта или планированием, — гл. 9. Материал гл. 10 надо включать практически во всех случаях, поскольку глава короткая и содержит обобщающий материал, дающий единую точку зрения на излагаемые методы обучения с подкреплением. В течение всей книги разделы, более трудные для восприятия и не очень существенные для понимания оставшейся части книги, помечаются символом «⋆». Они могут быть опущены при первом чтении, что не создаст особых затруднений при чтении последующих разделов. Некоторые упражнения также помечены символом «⋆», чтобы показать, что они относятся к повышенному типу сложности и не существенны для понимания основного материала соответствующей главы.

Книга большей частью самодостаточна. Не предполагается, что читатель обладает какой-либо специальной математической подготовкой, он должен быть знаком лишь с элементарными сведениями из теории вероятностей, в частности с понятием математического ожидания случайной величины. Материал гл. 8 будет значительно проще понять, если читатель обладает некоторыми знаниями в области искусственных нейронных сетей или знаком с какими-либо другими подходами к организации обучения с учителем. Этот материал, однако, можно читать и без такой предварительной подготовки. Мы настоятельно рекомендуем прорабатывать упражнения, содержащиеся в книге. Преподаватели могут воспользоваться соответствующими руководствами по решению этих задач. Эти руководства и другие регулярно обновляемые материалы по теме доступны через Интернет.

В конце большинства глав содержится раздел, озаглавленный «Библиографические и исторические замечания». В этих разделах указываются источники идей, излагаемых в соответствующей главе, даются ориентиры для дальнейшего изучения предмета и исследований в нем, а также описывается история вопроса. Несмотря на наше стремление сделать эти разделы возможно более корректными и полными, вне всякого сомнения, какие-то из важных работ не попали в поле нашего зрения. Мы приносим свои извинения по этому поводу и будем всячески приветствовать поправки и дополнения, которые могли бы быть включены в последующие издания.

В каком-то смысле мы работаем над данной книгой вот уже двадцать лет, и у нас есть причины высказать свою благодарность очень многим людям. Первым делом, мы благодарим тех, кто очень помог нам выработать общую точку зрения, представленную в этой книге, — это Харри Клопф (Harry Klopf), который помог нам осознать необходимость возобновления исследований в области обучения с подкреплением; Крис Уоткинс (Chris Watkins), Димитрий Бертсекас (Dimitri Bertsekas), Джон Цициклис (John Tsitsiklis) и Пол Вербос (Paul Werbos), которые помогли нам понять значение связей обучения с подкреплением и динамического программирования; мы благодарны Джону Муру (John Moore) и Джиму Кехое (Jim Kehoe) за озарения и вдохновение, связанные с теорией обучения животных; Оливеру Селфриджу (Oliver Selfridge) за подчеркивание широты и важности такой области, как адаптация. В более широком плане мы признательны нашим коллегам и студентам, которые воздействовали на нас очень многими способами — это Ron Williams, Charles Anderson, Satinder Singh, Sridhar Mahadevan, Steve Bradtke, Bob Crites, Peter Dayan и Leemon Baird. Наши идеи в области обучения с подкреплением существенно обогатились в результате дискуссий с такими коллегами, как Paul Cohen, Paul Utgoff, Martha Steenstrup, Gerry Tesauro, Mike Jordan, Leslie Kaelbling, Andrew Moore, Chris Atkeson, Tom Mitchell, Nils Nilsson, Stuart Russell, Tom Dietterich, Tom Dean и Bob Narendra. Мы признательны Michael Littman, Gerry Tesauro, Bob Crites, Satinder Singh и Wei Zhang за помощь в работе со специфическим материалом разд. 4.7, 11.1, 11.4, 11.5 и 11.6 соответственно. Мы признательны также Управлению по научным исследованиям ВВС США, Национальному научному фонду и фирме GTE Laboratories за их долгую и дальновидную поддержку.

Мы хотели бы также поблагодарить многочисленных коллег, которые прочитали черновой вариант данной книги и высказали ценные замечания. Среди этих коллег Tom Kalt, John Tsitsiklis, Pawel Cichosz, Olle Gällmo, Chuck Anderson, Stuart Russell, Ben Van Roy, Paul Steenstrup, Paul Cohen, Sridhar Mahadevan, Jette Randlov, Brian Sheppard, Thomas O'Connell, Richard Coggins, Cristina Versino, John H. Hiett, Andreas Badelt, Jay Ponte, Joe Beck, Justus Piater, Martha Steenstrup, Satinder Singh, Tommi Jaakkola, Dimitri Bertsekas, Ben Van Roy, Sascha Engelbrecht, Torbjörn Ekman, Christina Björkman, Jakob Carlström и Olle Palmgren. В заключение мы признательны Gwyn Mitchell за многообразную помощь, а Harry Stanton и Bob Prior — за отстаивание наших интересов в издательстве MIT Press.

I

Постановка задачи и подходы к ее решению

1 Введение

Когда мы размышляем о самой природе обучения, первой, вероятно, возникает идея о том, что мы обучаемся через взаимодействие с окружающей средой. Когда играет младенец, размахивая руками или глядя по сторонам, явно выраженного учителя у него нет, но зато у него есть прямой сенсомоторный контакт со средой. Постоянно повторяясь, контакты такого рода в изобилии дают информацию относительно причин и следствий, относительно последовательностей действий, а также о том, что надо делать, чтобы добиться определенных целей. В течение всей нашей жизни такие взаимодействия представляют собой, несомненно, важный источник знаний об окружающем мире и о нас самих. Учимся мы водить автомобиль или поддерживать разговор, мы четко сознаем, каким образом среда реагирует на наши действия, и стараемся добиться нужных нам результатов, организуя соответствующим образом свое поведение. Обучение через взаимодействие — основополагающая идея, на которой базируются почти все теории обучения и интеллекта.

В этой книге мы принимаем *вычислительный* подход к обучению на основе взаимодействия. Вместо того чтобы напрямую строить теории того, как обучаются люди и животные, мы исследуем идеализированную ситуацию обучения и оцениваем эффективность различных подходов к обучению. Иначе говоря, мы рассматриваем данную проблему с позиций исследователя или специалиста в области искусственного интеллекта. Мы изучаем проекты машин, которые были бы эффективны для решения задач обучения, интересных с научной и практической точек зрения, оценивая эти проекты с привлечением математических методов и компьютерных экспериментов. Изучаемый подход именуется *обучением с подкреплением*. Он, в отличие от других подходов к обучению машин, значительно более ориентирован на обучение, направляемое целями и основанное на взаимодействии со средой.

1.1. Обучение с подкреплением

Обучение с подкреплением — это обучение тому, что надо делать, как следует отображать ситуации в действия, чтобы максимизировать некоторый сигнал поощрения (вознаграждения), принимающий числовые

значения. Обучаемому не говорят, какое действие следует предпринять, как это имеет место в большинстве подходов к обучению машин. Вместо этого он, пробуя выполнять различные действия, должен найти, какие из них принесут ему наибольшее вознаграждение. В наиболее интересных и важных случаях действия могут влиять не только на вознаграждение, получаемое немедленно, но также и на возникающую ситуацию, а через нее — на все последующие поощрения. Эти две характеристики — поиск методом проб и ошибок, а также отсроченные поощрения — представляют собой две наиболее важные отличительные черты обучения с подкреплением.

Определение обучения с подкреплением дается не через описание методов обучения, а путем выявления характерных свойств *задачи* обучения. Любой метод, который хорошо подходит для решения сформулированной задачи, будем рассматривать как метод обучения с подкреплением. Подробное описание задачи обучения с подкреплением в терминах оптимального управления марковскими процессами принятия решений отложим до гл. 3. Но уже сейчас можно сказать, что основная идея состоит просто в том, чтобы уловить и зафиксировать наиболее важные аспекты реальной задачи, имея в виду организацию взаимодействия агента со средой для достижения некоторой цели. Совершенно ясно, что такого рода агент должен иметь возможность в какой-то мере воспринимать состояние среды, а также быть в состоянии предпринимать действия, которые могут повлиять на состояние среды. Агент должен иметь также цель или цели, связанные с состояниями среды. Формулировка задачи обучения с подкреплением должна учитывать все три аспекта (восприятие, действие, цели) в их наиболее простых формах, но без их вульгаризации.

Обучение с подкреплением отличается от *обучения с учителем*, которое рассматривается в большинстве современных исследований по обучению машин, статистическому распознаванию образов, искусственным нейронным сетям. Обучение с учителем — это обучение по примерам, предъявляемым некоторой информированной внешней инстанцией. Это важный вид обучения, однако без привлечения дополнительных средств он не пригоден для обучения через взаимодействие. В задачах, решаемых на основе взаимодействия, зачастую непрактично пытаться получать примеры требуемого поведения, которые были бы одновременно корректными и представительными для всех ситуаций, в которых должен действовать агент. На «ничейных» территориях, в ситуациях, когда обучение более всего и нужно, агент должен быть в состоянии учиться только на основе своего собственного опыта.

Одна из наиболее серьезных проблем, возникающих в обучении с подкреплением и отсутствующих в других видах обучения, — это проблема поиска компромисса между *изучением* и *применением*[1]. Чтобы получить большее вознаграждение, агент, обучающийся с подкреплением, должен предпочитать действия, которые он уже проверил в прошлой своей деятельности и обнаружил, что они эффективны с точки зрения получения поощрения. Однако, чтобы обнаруживать их, надо пробовать выполнять такие действия, которые еще не выполнялись ранее. Агент должен *применять* те действия, про которые уже известно, что они позволяют получить вознаграждение, но он должен также и *изучать* новые действия, чтобы иметь возможность делать лучший выбор в будущем. Проблема состоит в том, что нельзя только использовать уже проверенные действия или только искать новые эффективные действия, поскольку это ведет к провалу попыток решения задачи. Агент должен пытаться предпринимать разнообразные действия *и* благоприятствовать тем из них, которые окажутся лучшими. В задачах стохастического характера каждое из действий должно быть повторено многократно, чтобы добиться получения надежной оценки ожидаемого вознаграждения. Дилемма изучения–применения интенсивно исследуется математиками вот уже в течение нескольких десятилетий (см. гл. 2). Пока же просто отметим, что в области обучения с учителем, в том виде, как эту область принято обычно определять, проблема равновесия между изучением и применением в полном объеме даже и не возникает.

Еще одна характерная черта обучения с подкреплением состоит в том, что в явном виде рассматривается *целостная* проблема целенаправленного агента, взаимодействующего с неопределенной средой[2]. Это существенно отличается от многих других подходов, где рассмотрение проводится на уровне подзадач без всяких указаний на то, как эти подзадачи можно увязать в общую картину. Например, уже упоминалось, что в большинстве исследований в области обучения машин внимание фокусируется на обучении с учителем, причем явно никак не указывается, как такого рода возможность можно использовать на практике для получения каких-либо полезных результатов. В других исследованиях разрабатываются теории планирования, оперирующего с общими целями, но без рассмотрения роли планирования в процессах принятия решения, идущих в реальном масштабе времени, или же во-

[1] То есть между поведением, направленным на получение знания, и поведением, основанным на использовании уже имеющегося знания. — *Прим. ред.*

[2] То есть со средой, которая содержит неопределенность, неопределенные факторы. — *Прим. ред.*

проса о том, откуда взять предсказывающие модели, которые необходимы для планирования. Такой подход позволил получить много важных результатов, однако концентрация внимания в нем на изолированных подзадачах существенно ограничивает возможности его применения.

Обучение с подкреплением ставит прямо противоположную задачу и исходит из целостного, взаимодействующего со средой, целенаправленного агента. Все агенты, обучающиеся с подкреплением, имеют явным образом выраженные цели, могут воспринимать особенности среды, а также выбирать действия, влияющие на эту среду. Более того, обычно с самого начала предполагается, что агент должен действовать, несмотря на существенную неопределенность в среде. Когда в обучение с подкреплением вовлекается планирование, приходится предпринимать усилия для обеспечения взаимодействия между планированием и выбором действий в реальном масштабе времени, а также для поиска способов получения и улучшения моделей среды. Рассмотрение специфичных для данной задачи соображений относительно того, какие характеристики критичны, а какие нет, может привести к включению в обучение с подкреплением средств обучения с учителем. Чтобы добиться успеха в области обучения, надо выделить и изучить важные подзадачи общей задачи обучения, но это должны быть такие подзадачи, роль которых ясна с точки зрения целостного, взаимодействующего со средой, целенаправленного агента, даже если нет возможности во всех подробностях описать этого агента.

Исследования в области обучения с подкреплением можно считать частью общего процесса, развивающегося в последние годы. Он состоит во все расширяющихся контактах и взаимодействиях между искусственным интеллектом и другими инженерными дисциплинами. Еще совсем недавно искусственный интеллект рассматривался как область исследований, никак не связанная с такими областями, как теория управления и статистика. Искусственный интеллект имел дело с логикой и символами, не с числами. Его моделями были большие программы на языке LISP, а не дифференциальные уравнения, соотношения линейной алгебры и статистики. В последние десятилетия эта точка зрения постепенно размывалась. Современные исследователи в области искусственного интеллекта допускают использование моделей статистики и алгоритмов управления, например, в качестве вполне подходящих конкурирующих методов или же просто как часть арсенала методов анализа. Области, которым ранее не придавали значения, находящиеся на стыке искусственного интеллекта и традиционных инженерных методов, сейчас развиваются наиболее активно. В их число входят такие новые на-

правления, как нейронные сети, интеллектуальное управление, а также предмет нашего рассмотрения — обучение с подкреплением. В обучении с подкреплением развиваются идеи, почерпнутые в теории оптимального управления и в стохастической аппроксимации, следуя более общим и более амбициозным целям искусственного интеллекта.

1.2. Примеры

Хороший способ понять суть обучения с подкреплением — рассмотреть для него несколько примеров и возможные варианты применений, которые подсказывают, в каком направлении должна идти разработка данного метода.

- Опытный шахматист делает ход. Выбор этого хода обусловлен как планированием — ожиданием ответной реакции противника и реакции на эту реакцию, так и текущими интуитивными соображениями относительно желательности конкретных позиций и ходов.
- Адаптивный регулятор подстраивает параметры процесса перегонки нефти в реальном масштабе времени. Данный регулятор оптимизирует соотношение между выходом продукции, ее стоимостью и качеством на основе заданных граничных значений стоимости без строгой привязки к соображениям, которые закладывались изначально инженерами.
- Детеныш газели старается встать на ноги сразу после рождения. Получасом позже он уже бежит со скоростью более 30 км/ч.
- Мобильный робот решает, должен ли он войти в очередную комнату при сборе мусора или же ему уже пора начинать искать дорогу назад, к месту, где он сможет зарядить свои аккумуляторы. Он принимает соответствующее решение на основе данных о том, насколько быстро и просто удавалось найти зарядную станцию в прошлом.
- Фил готовит себе завтрак. При более подробном рассмотрении оказывается, что даже такая сугубо утилитарная деятельность представляет собой сложный комплекс действий, подчиненных определенным условиям, взаимосвязанных целей и подцелей: подойти к шкафу, открыть его, выбрать коробку с крупой, протянуть к ней руку, взять коробку, вытащить ее. Другие сложные последовательности поведенческих актов, корректируемые по ходу действия, требуются для того, чтобы взять чашку, ложку, кувшин с молоком. Каждый из перечисленных шагов требует выполнения последовательности движений глаз для получения информации, а также для

управления процессами перемещения руки к требуемому предмету и перемещения этого предмета. Постоянно принимаются решения относительно того, что следует делать с тем или иным объектом, стоит ли его перенести на обеденный стол, прежде чем браться за остальные. Выполнение каждого шага направляется целями, такими как «взять ложку» или «подойти к холодильнику». Эти цели могут быть подчинены другим целям, например «обеспечить наличие ложки для еды», когда завтрак готов к употреблению.

Эти примеры имеют ряд общих фундаментальных черт, которые нельзя упускать из виду. Все они включают *взаимодействие* между активным агентом, принимающим решения, и окружающей его средой, в которой агент осуществляет поиск для достижения своей *цели*, несмотря на наличие *неопределенности* в среде. Действия агента направлены на то, чтобы повлиять на будущее состояние среды (им может быть, например, следующая шахматная позиция, уровень жидкости в резервуаре при перегонке нефти, следующее местоположение робота) и через него повлиять на те возможности, которые окажутся в распоряжении агента в следующие моменты времени. Для правильного выбора необходимо принимать во внимание непрямые, отложенные последовательности действий, что может потребовать использования прогнозирования или планирования.

Вместе с тем во всех приведенных примерах нет возможности полностью предсказать влияние предпринимаемых действий. По этой причине агент должен достаточно часто контролировать состояние среды и соответственным образом реагировать на изменения. Например, Фил должен контролировать количество молока, которое он наливает в чашку, чтобы молоко не перелилось через край. Все эти примеры включают явные цели в том смысле, что агент может оценивать характер продвижения к требуемой цели, основываясь на том, что он может воспринимать непосредственно. Шахматист знает, выигрывает он или нет, регулятор нефтеперегонной установки имеет данные о том, сколько нефти должно быть обработано, мобильный робот обнаруживает, что его батарея заряжена недостаточно, а Фил знает, нравится ему завтрак или нет.

Во всех этих примерах агент может использовать накопленный опыт, чтобы со временем улучшать свои характеристики. Шахматист совершенствует свою интуицию, помогающую в оценке позиций, что позволяет ему более успешно играть; детеныш газели развивает свою способность бегать; Фил учится более рационально готовить завтрак. Знание, которое привносится агентом в задачу в начале процесса ее ре-

[. . .]