
ПРОБЛЕМЫ ПРЕДПРИЯТИЙ

Модель управления человеческим капиталом предприятия на основе методов «машинного обучения с подкреплением»

© 2025 г. Е.В. Орлова

Е.В. Орлова,

Уфимский университет науки и технологий, Уфа, e-mail: ekorl@mail.ru

Поступила в редакцию 01.08.2023

Аннотация. Человеческий капитал является одним из важнейших движущих сил устойчивого экономического роста предприятия, что приобретает еще большую значимость в условиях изменений характера труда в период цифровой трансформации экономики. Портрет работника становится все более многогранным вследствие расширения сфер его активности. Поэтому проблема управления человеческим капиталом на основе формирования индивидуальных траекторий профессионального развития работников представляется актуальной, своевременной, социально и экономически значимой. В работе предлагается модель управления человеческим капиталом, предназначенная для разработки индивидуальных траекторий профессионального развития работников предприятия, формирование которых основано на методах машинного обучения — «машинного обучения с подкреплением». Модель формирует оптимальный режим управления и рассматривается как последовательный набор программных мероприятий, направленных на развитие работника в профессиональной сфере с учетом его изменяющихся в динамике индивидуальных характеристик состояния здоровья, уровня профессиональных и надпрофессиональных компетенций, мотивации, социального капитала. Архитектуру системы управления можно рассматривать как цифровой двойник работника предприятия, который объединяет среду — модель работника как марковского процесса принятия решений и модель управления — агента — центра принятия решений предприятия. Для максимизации функции полезности агента используются алгоритмы «машинного обучения с подкреплением» DQN, DDQN, SARSA, PRO. На основе проведенных экспериментов показано, что наилучшие результаты в смысле достижения максимальной полезности агента обеспечивает алгоритм DDQN. Практическую значимость имеют результаты, сгенерированные предлагаемой моделью, реализация которых позволит в кратчайшие сроки обеспечить рост инновационности и конкурентоспособности предприятия за счет улучшения качества человеческого капитала и роста ресурсной эффективности труда.

Ключевые слова: человеческий капитал, цифровой двойник работника, индивидуальные траектории, машинное обучение, «машинное обучение с подкреплением», *Q*-обучение, оптимальное управление, марковский процесс принятия решений.

Классификация JEL: C61, O15.

УДК: 330.4, 519.865.7.

Для цитирования: **Орлова Е.В.** (2025). Модель управления человеческим капиталом предприятия на основе методов «машинного обучения с подкреплением» // *Экономика и математические методы*. Т. 61. № 1. С. 70–83. DOI: 10.31857/S0424738825010072

1. ВВЕДЕНИЕ

Человеческий капитал является одним из важнейших движущих сил устойчивого экономического роста предприятия, и это приобретает все большую значимость с ходом инновационного развития экономики. Существующие условия технологического развития неизбежно изменяют характер и условия труда. Портрет работника становится все более многогранным вследствие расширения разнообразных сфер активности, в том числе в цифровом пространстве.

Важнейшей чертой четвертой промышленной революции является «персонализация», связанная не только с предоставлением клиентам персонализированных продуктов, но и с обеспечением большей персонализации при построении эффективного взаимодействия с работниками предприятия в трудовых отношениях. Это может усилить рост ценности человека в цифровой экономике.

Целью исследования является разработка методологии цифрового двойника работника, основанной на новой методике оценки человеческого капитала и модели управления для формирования

индивидуальных траекторий профессионального развития, обеспечивающих повышение качества человеческого капитала и рост эффективности предприятия в целом. Для реализации модели управления человеческим капиталом применяется метод «машинного обучения с подкреплением» (reinforcement learning, RL) и семейство алгоритмов, которые используются для вычисления оптимальных стратегий управления.

Управление человеческим капиталом рассматривается как задача последовательного принятия решений в стохастической среде. Для ее решения применяется принцип максимальной ожидаемой полезности. Оптимальный выбор последовательности решений основан на рассуждении о будущих последствиях действий. Для формализации задачи последовательного принятия решений используется марковский процесс принятия решений (markov decision process, MDP). Модель MDP определяет стохастическую динамику системы, а также полезность, связанную с ее эволюцией и со стратегий принятия решений. Обучение агента осуществляется на базе алгоритмов, основанных на полезности, на стратегии, и комбинированных алгоритмов.

2. МЕТОДЫ

2.1. Анализ подходов и методов к управлению человеческим капиталом

Человеческий капитал на уровне конкретного работника представляет собой совокупность знаний, умений и навыков, состояние физического и психического здоровья человека, влияющую на результаты его трудовой деятельности и соответствующие доходы. На уровне предприятия (организации) человеческий капитал формирует экономический ресурс, воспроизводство которого (формирование, накопление и использование) требует, в отличие от физического капитала, постоянной мотивации. На уровне государства и региона человеческий капитал формируется за счет инвестиций в накопление знаниевой, интеллектуальной составляющей, в повышение уровня и качества жизни населения, в том числе и в воспитание, образование, здоровье, безопасность, культуру, искусство (Макаров, Макаров, Клейнер, 2007). Человеческий капитал является сложным фактором социального и экономического развития. Он вносит вклад в повышение качества и производительности труда во всех видах жизнедеятельности общества.

При оценке человеческого капитала на уровне конкретного работника в менеджменте большое распространение получили методы, направленные на получение обратной связи от работников (Church, 2019; Steelman, Williams, 2019) с целью управления эффективностью предприятий. В рамках этой системы для оценки и развития персонала, как правило, выполняются следующие этапы: оценка результатов деятельности работника, оценка потенциала, карьерное планирование, обучение и карьерное наставничество, ротация персонала, изучение компетенций в поведении. Проблема управления карьерой в системе развития человеческих ресурсов также активно обсуждается в научных работах. Например, предлагается использовать различные коэффициенты при факторах для достижения большей эффективности (Zhang et al., 2019; Hernaus, Pavlovic, Klindzic, 2019; Alzyoud, 2018). Эти факторы развития человеческого капитала не являются независимыми, часто имеют тесную корреляцию (Hitka et al., 2019; Stokowski et al., 2018), что усложняет поиск конкретных механизмов управления карьерой. Рекомендуется регулярная и непрерывная оценка, мониторинг сотрудников для оценки степени достижения цели (Jung, Takeuchi, 2018).

Анализ научной литературы об оценке и управлении человеческим капиталом предприятия выявил общую цель — формирование и обеспечение устойчивой системы управления человеческими ресурсами компании (Osranek, Zink, 2014). Существует разрыв между необходимостью учета всех аспектов человеческого капитала и возможностью существующих подходов к такой оценке и управлению. Оценка корпоративного человеческого капитала в ряде работ формируется только с одной стороны — с позиций компетентности работников, или состояния здоровья работников, или мотивационных характеристик и т.п. Однако комплексная оценка человеческого капитала на корпоративном уровне, отражающая множество факторов, формирующих человеческий капитал, и такие особенности, как образование, состояние здоровья, квалификация, вовлеченность и мотивация, коммуникабельность, надпрофессиональные навыки под требования новых профессий и способности быстро адаптироваться, не используются.

С позиции используемого инструментария управления сложными социально-экономическими системами свою высокую эффективность показали методы агент-ориентированного моделирования (АОМ) (Schelling, 1971; Granovetter, 1973; Rachid, Mohamed, Khouaja, 2018; Макаров и др., 2020, 2022;

Таблица 1. Экономические и социальные эффекты развития человеческого капитала

Уровень работника	Уровень предприятия	Общество
Гарантия сохранения рабочего места	Рост производительности труда, дохода и прибыли	Развитие трудового потенциала общества
Приобретение новых знаний, умений, навыков, раскрытие способностей	Снижение текучести кадров	Рост производительности общественного труда
Рост стоимости на рынке труда	Повышение мотивации труда	
Расширение возможностей профессионального и карьерного роста	Рост вклада каждого сотрудника в достижение целей	
Расширение социальных сетей и связей	Облегчение делегирования полномочий	
Рост самооценки и уверенности в себе	Совершенствование корпоративной культуры	
	Улучшение морально-психологического климата в коллективе и проектных командах	
	Положительное влияние на трудовую дисциплину	

Акопов, 2023), реализующие управленческие схемы ситуационного анализа и моделирования. Вычислительный инструмент, основанный на АОМ, представляет собой систему в виде совокупности агентов (объектов) с заданными свойствами, а концептуальным ядром таких систем является достижение равновесия (или псевдоравновесия). Методология управления человеческим капиталом, предлагаемая в настоящей работе, расширяет возможности имитационного моделирования на основе методов АОМ за счет использования «машинного обучения с подкреплением».

Традиционно управление человеческим капиталом реализуется на основе методов тактического, а не стратегического управления, при котором долгосрочные эффекты от принятых решений не рассматриваются. Также не учитывается динамический характер самих показателей человеческого капитала, связанных с изменением профессиональных качеств работника, его мотивации и социальных характеристик. Стратегическая позиция при построении системы управления предприятием отражает необходимость инвестировать в развитие персонала, что является выгодным для предприятия. Обоснованная политика гарантирует получение компетентной, мотивированной и слаженной команды, которая обеспечит предприятию дополнительную прибыль. В табл. 1 представлены разноуровневые экономические и социальные эффекты, которые возникают в результате воздействия на человеческий капитал предприятия.

На стадии разработки стратегии развития человеческого капитала требуется четко определить, какие результаты предприятие ожидает получить от сотрудников, какие профессиональные и личные качества нужно развивать. Стратегия может быть ситуативной и системной. В первом случае она привязана к конкретной бизнес-задаче (например, обеспечение роста продаж). При системном варианте реализуется постоянное обучение и развитие внутри предприятия. Благодаря этой стратегии сотрудники совершенствуют весь спектр навыков и применяют их на практике без отрыва от работы. Предприятие должно формировать специалистов, которые способны к разработке и внедрению инновационных идей и решений. В этом процессе большое значение имеет подбор и реализация адекватных методов управления человеческим капиталом.

2.2. Приложения «машинного обучения с подкреплением» в организационных системах

Приложения систем «машинного обучения с подкреплением» (reinforcement learning, RL) разнообразны, связаны с задачами оптимизации (динамического программирования, ДП) процессов и систем. Новейшие исследования в области управления организационными системами представлены ниже:

— в области промышленности, менеджмента «машинное обучение с подкреплением» используется по всему спектру задач управления ресурсами (Ding et al., 2022; Li et al., 2023), разработки принципов календарного планирования производства (Wang et al., 2022), разработки планов пополнения запасов, устанавливающих момент и объем пополнения запасов, разработки логистических маршрутов и цепочек поставок (Abideen et al., 2021; Yan et al., 2022);

— в робототехнике «машинное обучение с подкреплением» имеет множество приложений, включая улучшение движения, разработку автономных транспортных средств (Orr, Dutta, 2023).

- «машинное обучение с подкреплением» улучшает управление движением на дорогах и используется в алгоритмах управления умными городами (Mohammadi, 2018);
- множество приложений «машинного обучения с подкреплением» в области здравоохранения используются для формирования схем расчета и дозирования лекарственных средств (Yu, 2019a, 2019b);
- при конструировании систем образования и электронного обучения, которые могут повысить свою эффективность за счет подбора учебных программ на базе «машинного обучения с подкреплением» (Chi et al., 2011);
- для разработки стратегии принятия решений в биржевой торговле (Azhikodan, Bhat, Jadha, 2019; Пономарев, Оселедец, Чихоцкий, 2019);
- для построения рекомендательных систем в приложениях онлайн-торговли (Liu et al., 2019; Zheng et al., 2018).

В задачах управления персоналом, управления человеческими ресурсами на уровне предприятия методы «машинного обучения с подкреплением» до сих пор не использовались. Учитывая потенциальные возможности методов «машинного обучения с подкреплением», их применение для решения задачи управления человеческим капиталом представляется возможным и целесообразным.

2.3. Методология управления человеческим капиталом

Потенциальная эффективность систем управления производственно-экономическими, социально-экономическими системами на основе цифровых двойников, а также методов и инструментов имитационного моделирования, методов и моделей оптимального управления уже демонстрировалась рядом авторов (Макаров, Бахтизин, Бекларян, 2019; Боровков, 2021; Orlova, 2022). Представленная в работе методология управления человеческим капиталом основана на цифровом двойнике работника предприятия и укрупненно состоит из блока оценки человеческого капитала и блока управления им (рис. 1).

Технология построения цифрового двойника работника для решения задачи управления человеческим капиталом основана на предложенной ранее модели его оценки (Орлова, 2021) и модели управления им, отличается от существующих системностью и комплексностью оценки

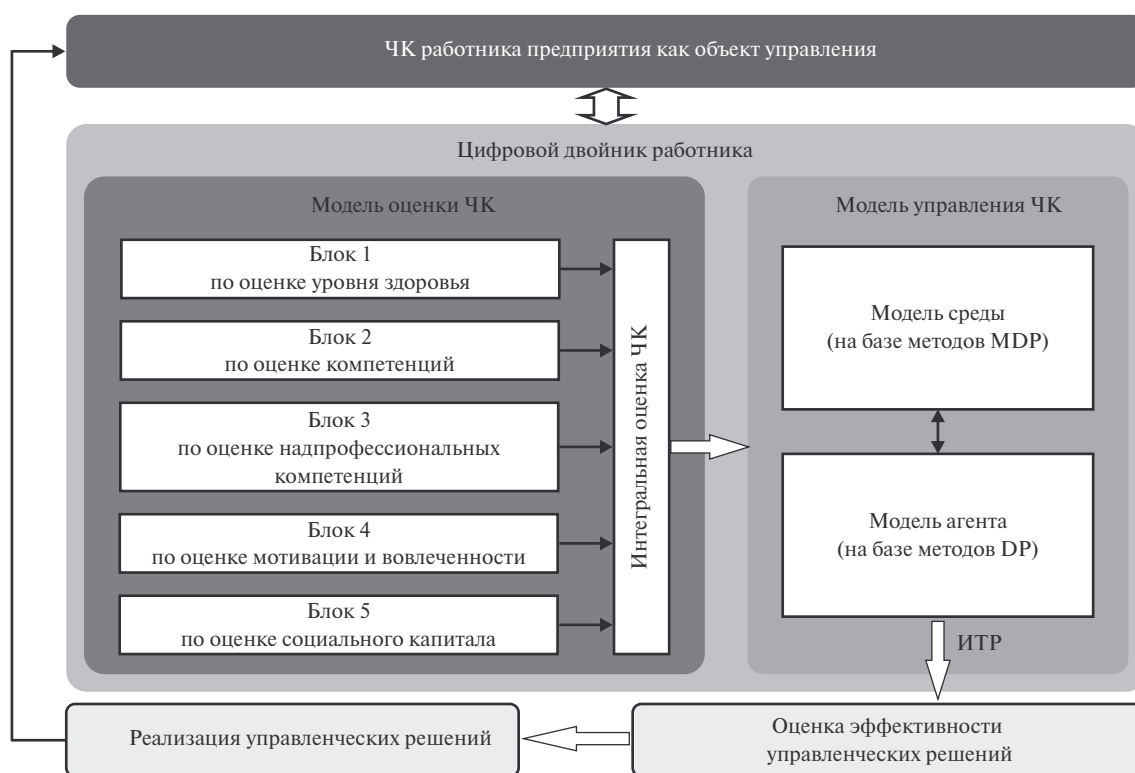


Рис. 1. Концептуальная схема технологии построения цифрового двойника работника

по связанным блокам, отражающим разные стороны и свойства человеческого капитала, а также научно-обоснованной стратегией управления. Модель оценки базируется на всестороннем учете свойств человеческого капитала, проявляющихся в цифровой экономике, и учитывает традиционные характеристики — возраст, образование, профессиональный опыт и компетенции, а также дополнительные характеристики — социальный капитал, состояние здоровья, надпрофессиональные компетенции, мотивацию и вовлеченность.

Модель управления человеческим капиталом предназначена для формирования индивидуальных траекторий профессионального роста и развития работников (в частности, ИТР), выработка которых основана на полученной из модели оценки человеческого капитала результатов и базируется на использовании методов обучения с подкреплением. Дизайн модели управления рассматривается как последовательное решение задачи принятия решений, которая хорошо вписывается в структуру «машинного обучения с подкреплением». Правила принятия решений эквивалентны политикам в «машинном обучении с подкреплением», а результаты управления выражаются функциями вознаграждения (награды). Входными данными является совокупность данных о работниках по всему спектру факторов человеческого капитала — знания, умения, состояние здоровья, социальный капитал, инновационность, социально-демографические факторы, а также данные о причинных связях указанных факторов. Выходные данные формируются в виде управленческих решений для каждого этапа (как состояния «машинного обучения с подкреплением»).

Применение методов «машинного обучения с подкреплением» для решения задач управления человеческим капиталом и формирования ИТР дает несколько преимуществ:

1) позволяет принимать наилучшее решение во времени для каждого работника в каждый момент времени с учетом разнообразия его характеристик. Это может достигаться без построения точной математической модели (поведения) работника, а также без наличия данных о причинно-следственных связях между решением (воздействием) и результатом (отдачей);

2) позволяет улучшить долгосрочные результаты с учетом распределения во времени и отсроченного эффекта от воздействия (решения);

3) позволяют проектировать функцию вознаграждения, которая является адаптивной и основывается на экспертных знаниях в данной предметной области;

4) обеспечивает многокритериальную оптимизацию в критериях эффективности и риска (например, для предприятия — потерять конкурентного работника, который может уйти к другому работодателю).

В основе модели управления лежит последовательность правил принятия решений для формирования действий в соответствии с текущими показателями деятельности работника и предшествующими воздействиями на него. Политика разработки индивидуальной *профессиональной траектории работника* становится все более актуальной и может использоваться для выбора эффективных инструментов управления человеческим капиталом на предприятии. Эта политика представляет собой набор правил для определения оптимального состава мероприятий и программ в определенный момент времени в зависимости от характеристик работников, а также показателей эффективности воздействия уже реализованных программ. Оптимальный режим воздействия (управления) — как последовательный набор программ — позволяет максимизировать средний ожидаемый доход за весь период реализации программ (управленческих решений).

Задача, которую необходимо решить, состоит в том, чтобы определить набор мероприятий (программ), которые приводят к росту эффективности всего предприятия. При этом учитывается, что эффективность следующей программы зависит от результатов эффективности предыдущих программ у конкретного работника.

Объектом управления является человеческий капитал работников предприятия, оценка которого по отдельным работникам подробно представлена в (Орлова, 2021б; Orlova, 2021a). На основе разработанной методики каждый работник имеет определенную численную оценку его человеческого капитала. Оценка производится по пяти группам показателей (блокам оценки) — оценка уровня здоровья, оценка компетенций, оценка надпрофессиональных компетенций, оценка мотивации и вовлеченности и оценка социального статуса. В зависимости от итогового значения человеческого капитала работник попадает в одну из пяти групп по каждому блоку оценки (каждый блок имеет пять градаций значений).

Предложенная методика оценки человеческого капитала является основанием для формирования управленческих решений, направленных на развитие потенциала работников и качества человеческого капитала. Управленческие решения имеют персональный характер, зависят от существующего уровня

человеческого капитала, реализованы на основе динамического режима и формируют ИТР. Разработан перечень решений по каждому блоку оценки этого капитала в зависимости от балла, набранного работником (Орлова, 2020а). Эти решения (мероприятия, программы) направлены на улучшение качества человеческого капитала в разрезе отдельных показателей. Общая траектория профессионального развития работника определяется композицией управленческих решений из каждого блока. Формирование такой траектории и является задачей управления на основе методов «машинного обучения с подкреплением».

2.4. Постановка задачи управления

Среда задается как марковский процесс принятия решений (MDP): $M = \{S, A, P, R, \gamma\}$, где S — пространство состояний, в котором определены состояния среды (работников предприятия) $s_t \in S$; A — пространство действий, в котором определены действия агента (центра принятия решений предприятия) $a_t \in A$; P — переходные вероятности марковского процесса, на каждом шаге времени t следующее состояние s_{t+1} берется из распределения вероятностей p : $p(s_{t+1}|s_t, a_t) \in P$; R — функция награды (вознаграждения), на каждом временном шаге t агент получает награду, зависящую от действия a_t , реализованного для состояния s_t и нового состояния s_{t+1} : $r_t(s_t, a_t, s_{t+1}) \in R$; γ — коэффициент дисконтирования, используемый для суммирования наград, $\gamma \in [0, 1]$. Далее будут равнозначно использоваться обозначения для состояний $s_t = s$, $s_{t+1} = s'$, для действий $a_t = a$ и наград $r_t = r$.

Сеть марковского процесса принятия решений состоит из множества вершин, соответствующих разным уровням качества человеческого капитала. Состояния заданы в 5-мерном пространстве. Изменение состояния происходит под действием 25 различных решений — по пять решений для каждого из пяти показателей оценки качества человеческого капитала. Введено предположение, что для каждого работника в единицу времени возможна реализация решения (действия) из одного блока $\{A_1, \dots, A_5\} \in A$. Поэтому движение работника по его траектории при реализации определенного решения возможно на один уровень вперед по определенному показателю (т.е. значение этого показателя человеческого капитала растет), назад по этому показателю (т.е. значение этого показателя снижается), неизменность показателя этого капитала. Если в результате реализации действия a_t у работника повышается показатель его человеческого капитала, на улучшение которого направлено это действие, то агенту возвращается награда $r_t(a_t)$.

Всего имеется пять блоков показателей оценки человеческого капитала и пять градаций внутри каждой группы показателей, таким образом можно сформировать 5^5 состояний графа MDP (вершины соответствует разным уровням качества этого капитала). В каждом блоке показателей оценки ЧК для каждого из пяти уровней интегрального показателя в блоке имеется определенное управленческое решение по наращиванию уровня человеческого капитала. Поэтому имеется 25 различных действий, реализующих марковский процесс принятия решений. Зададим состояния графа как S_{ijkl} , где индексы соответствуют показателям пяти блоков оценки ЧК. В каждом состоянии агент может принимать разные решения, образуя множество решений a_{ij} , где i — блок показателей человеческого капитала, подлежащих управлению; j — соответствующее решение в блоке i . Например, при реализации действия переход из состояния S_{22111} через действие a_{13} означает, что применяется решение 3 в блоке 1 (состояние здоровья), позволяющее увеличить уровень человеческого капитала с уровня S_{22111} до уровня S_{32111} .

На рис. 2 схематично представлен граф переходов из состояния s при реализации действия a . Каждая стрелка на рисунке соответствует тройке (s', s, a) , на стрелках отражены вероятности перехода $p(s'|s, a)$ из текущего состояния s в новое состояние s' на следующем шаге при реализации

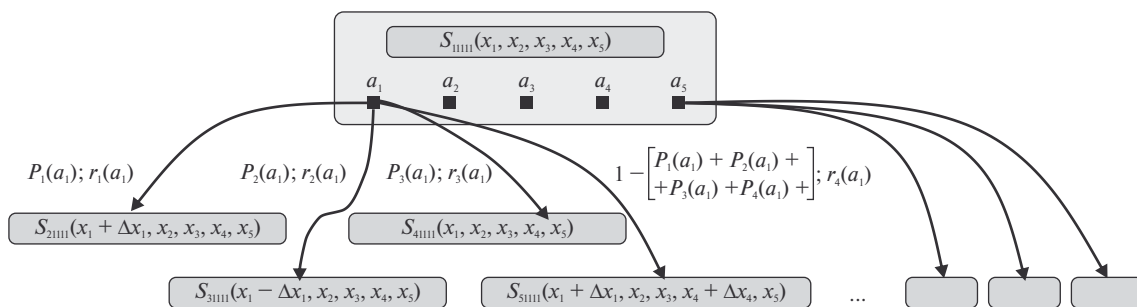


Рис. 2. Граф переходов из состояния s в состояние s' при реализации действия a

действия a , а также соответствующая награда (ожидаемое вознаграждение) $r(s, a, s')$ за этот переход. Сумма вероятностей перехода на стрелках, исходящих из вершины, равна 1. Вероятности переходов $p(s'|s, a)$ определяются исходя из результатов рандомизированных испытаний (Orlova, 2023).

Целью задачи является выработка такой последовательности управленческих решений, направленных на формирование ИТР, которая обеспечивает рост человеческого капитала, т.е. максимизацию награды на протяжении моделируемого периода.

2.5. Алгоритмы обучения агента

Алгоритм «обучение с подкреплением» представляет собой последовательность адаптированных процедур, соответствующих динамическому изменению состояния системы. Таким образом, стратегия построения ИТР, разработанная на основе алгоритмов «обучение с подкреплением», будет динамически меняться с течением времени по мере накопления наблюдений. Для обучения агента тестируются алгоритмы следующих классов — алгоритмы, основанные на полезности, алгоритмы, основанные на стратегии, и комбинированные алгоритмы.

3. АЛГОРИТМЫ, ОСНОВАННЫЕ НА ПОЛЕЗНОСТИ (DQN-АЛГОРИТМЫ)

Алгоритм DQN (Deep Q-Networks) — как алгоритм обучения глубоких нейронных сетей — аппроксимирует функцию Q . Настроенная функция Q используется агентом для выбора действий. Для обновления текущей стратегии используется опыт, накопленный при реализации разных стратегий (не только текущей). В Q -обучении две стратегии: целевая (постоянно улучшается) и поведенческая ϵ -жадная, используемая для взаимодействия со средой. Агент на основе сведений о состоянии объекта управления s_t и полученном из среды вознаграждении r_t за действие a_t , переведшее состояние объекта в следующее состояние, вычисляет значение функции $Q(s, a)$, оценивающее ценность действия a_t в состоянии s_t . Настройка функции Q осуществляется с помощью метода TD-обучения (метода временных различий), значение функции обновляется на накопленные дисконтированные будущие вознаграждения и определяет принцип оптимальности Беллмана:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right], \quad (1)$$

где α — скорость обучения функции ценности (при $\alpha < 1$ осуществляется приближение старого состояния к новому, при $\alpha = 1$ осуществляется замена старого состояния новым); r_{t+1} — награды, полученные из среды, за действия a_t из состояния s_t ; γ — коэффициент дисконтирования; $\max_a Q(s_{t+1}, a)$ — максимальное ожидаемое значение из состояния s_{t+1} (новое значение); $Q(s_t, a_t)$ — предыдущая оценка Q -значения (старое значение).

Полученные Q -значения используются для обучения агента и для определения следующего действия. Для этого используется нейронная сеть (сеть полезности, value networks), которая оценивает Q -значения пар (s, a) и выбирает действия с максимальным Q -значением (максимальной полезностью):

$$Q_{\text{target}}^\pi(s, a) = r + \gamma \max_{a'} Q^\pi(s', a'). \quad (2)$$

Алгоритм SARSA (State–Action–Reward–State–Action) является одним из фундаментальных «алгоритмов обучения с подкреплением». Идея данного алгоритма такая же, как и у алгоритма DQN, за исключением того, что алгоритм DQN рассчитывает функцию полезности за несколько временных шагов, т.е. буферизует опыт. Алгоритм DQN реализует вычисления по множеству пакетов данных, это увеличивает вычислительную нагрузку на вычислительную систему, но при этом может значительно ускорить обучение.

Для предотвращения ошибок в максимизации в Q -обучении и повышения устойчивости обучения используется алгоритм двойного Q -обучения. В алгоритме DQN для выбора действия и получения оценки функции Q используется одна и та же нейронная сеть. В алгоритме двойной DQN (double DQN, DDQN) применяются две нейронные сети. Первая — обучаемая θ -сеть используется для выбора действия a ; вторая — прогнозная ϕ -сеть используется для расчета Q -значения для пар (s, a) , т.е. для оценки этого действия a . На рис. 3 представлена архитектура модели управления, в основе которой лежит алгоритм DDQN.

Алгоритм DDQN уменьшает завышение оценок Q -значений путем настройки оценок функций Q . Применение двух сетей в данном алгоритме могут замедлять процесс обучения, если параметры

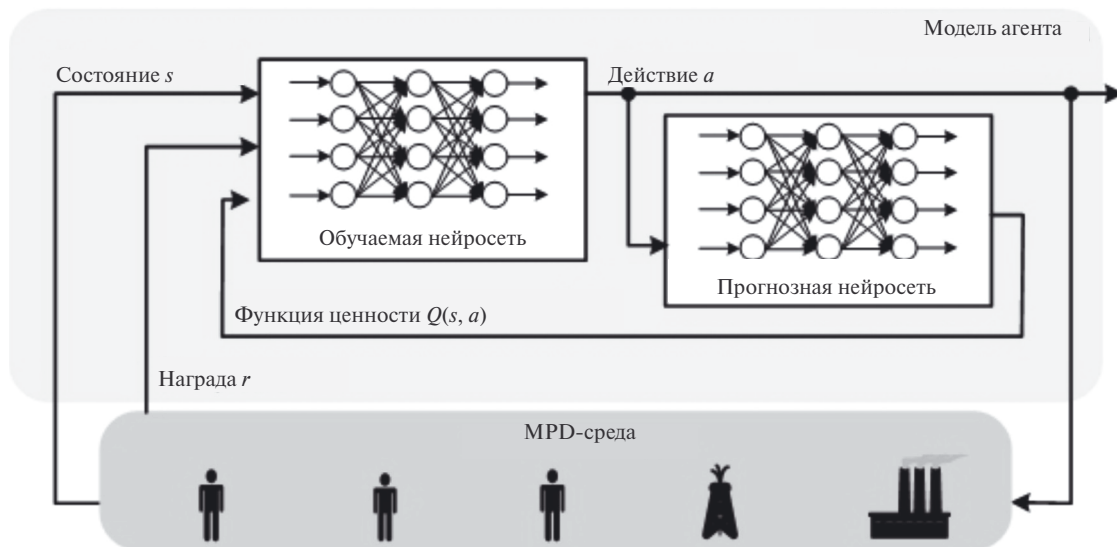


Рис. 3. Архитектура проектируемой модели управления человеческим капиталом



Рис. 4. Укрупненный алгоритм обучения DDQN-агента

θ и ϕ имеют очень близкие значения; в этом случае обучение может быть неустойчивым. Для поиска разумного соотношения между устойчивостью и скоростью обучения нужно настраивать гиперпараметр «частота обновления прогнозной сети», управляющий скоростью изменения ϕ . Адаптированный к данной задаче алгоритм обучения DDQN-агента представлен на рис. 4.

3.1. Алгоритмы, основанные на стратегии (REINFORCE)

Данный класс алгоритмов предназначен для настраивания стратегии. Хорошие состояния должны порождать действия, обеспечивающие траектории, которые максимизируют целевую функцию агента как сумму дисконтированных наград, усредненных по нескольким эпизодам. Агенту нужно действовать в среде, а действия, которые будут оптимальными в данный момент, зависят от состояния. Функция стратегии π принимает на входе состояние, а на выходе выдает действие $a \sim \pi(s)$. То есть агент может принимать эффективные решения в разных ситуациях.

3.2. Комбинированные алгоритмы (PRO)

Алгоритм PRO (Proximal Policy Optimization) представляет собой метод градиента стратегии с преобразованием целевой функции, что может повысить устойчивость и эффективность выборов в процессе обучения за счет меньших затрат вычислительных ресурсов и более высокой производительности. Однако у этого алгоритма есть недостатки, например низкая чувствительность к гиперпараметру ϵ , что дает близкие значения производительности при разных значениях этого параметра.

4. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Система моделирования реализована в программе Matlab2021b¹ и состоит из трех процессов — создания агента Q -обучения, обучение агента и тестирование. Обучение нейросети проводится каждые 200 эпизодов, каждый из которых включает 50 испытаний (шагов). Эксперименты проводятся на данных крупного нефтедобывающего и нефтеперерабатывающего предприятия. Имеются данные об оценке человеческого капитала по пяти блокам оценки, производительности труда работников. Обобщенные значения количественных показателей работников по выборке (без выделения по отдельным категориям) представлены в табл. 2. Сформирован также комплекс управленческих решений по воздействию на элементы ЧК.

Здесь рассматриваются два показателя, формирующие человеческий капитал, — показатель уровня здоровья работника и показатель уровня его компетенций. Каждый из приведенных показателей имеет пять уровней. Задача решается на двумерной сетке 5×5 , где по строкам отражаются уровни первого показателя — уровня здоровья работника, по столбцам — уровни второго показателя — компетенций.

По результатам оценки человеческого капитала, определенной по методике, предложенной автором в (Орлова, 2021; Orlova, 2021b), каждый работник получает по каждому из этих двух показателей определенный балл от 1 до 1000. В соответствии с этим он попадает в одну из пяти групп по каждому показателю и на сетке 5×5 представляется в виде кружка, рис. 5. На рисунке представлен работник, попавший в ячейку (3, 2), это — его начальный уровень человеческого капитала.

При определении цели, т.е. уровня потенциально возможного человеческого капитала для конкретного работника, учитывается ряд ограничений, связанных с его возможностями. На рисунке показана область ограничений в виде ячеек, залитых черным цветом. Здесь показано, что для этого работника существует возможность движения в сторону улучшения человеческого капитала — цели, ячейка (5, 5). Если работник с учетом его текущего уровня образования и объективных

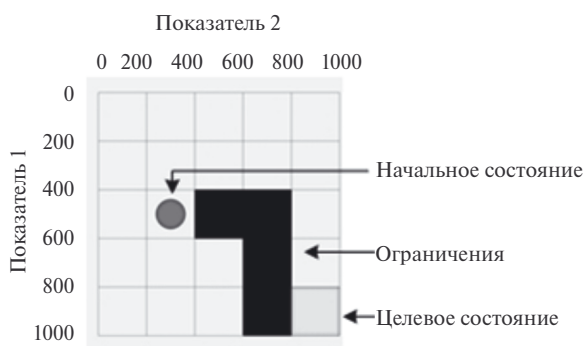


Рис. 5. Представление задачи в виде сетки 5×5

характеристик состояния здоровья не может достичь максимально возможного уровня (ячейка (5, 5)), то этому работнику определяется другая цель. Также по совокупности факторов здоровья ему не удастся добиться повышения уровня профессионализма в данной области с учетом низких значений состояния здоровья. То есть условия профессии таковы, что достижение уровней 3 и 4 профессионального развития невозможно для работника с низкими показателями здоровья, ячейки (3, 3), (3, 4). Также при высоких показателях качества здоровья работника ему доступен переход на высший уровень компетентности, т.е. быстрый переход на уровень 5 по показателю 2, ячейки (4, 5), (5, 5).

Таблица 2. Статистические характеристики количественных показателей в выборке 70 работников, принявших участие в эксперименте

Показатель	Среднее значение	Среднеквадратическое отклонение
Возраст работников, лет	41	5
Состояние здоровья, баллы. Максимально возможный балл 1000	564	61
Уровень профессиональных компетенций, баллы. Максимально возможный балл 1000	625	43
Уровень надпрофессиональных компетенций, баллы. Максимально возможный балл 1000	469	34
Уровень мотивации и вовлеченности, баллы. Максимально возможный балл 1000	692	51
Уровень социального капитала. Максимально возможный балл 1000	721	20
Индивидуальная производительность труда, руб./человек	543	62

¹ <https://www.mathworks.com/>

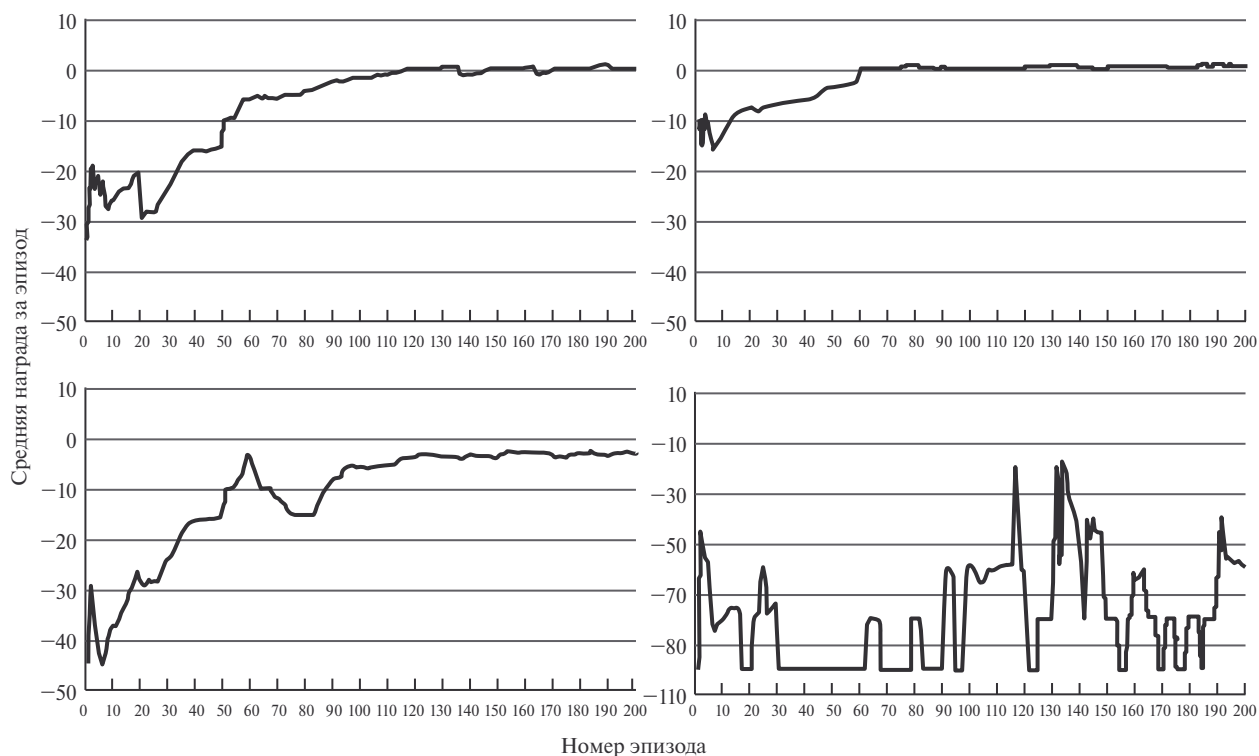


Рис. 6. Результаты обучения агента, коэффициент дисконтирования $\gamma = 0,99$, вероятность случайного действия $\epsilon = 0,04$: а) эксперимент 1 — обучение на основе алгоритма DQN; б) эксперимент 2 — обучение на основе алгоритма DDQN; в) эксперимент 3 — обучение на основе алгоритма SARSA; г) эксперимент 4 — обучение на основе алгоритма PRO

Работник представлен как среда, в которой заданы ограничения достижимости цели, начальное и конечное состояние, функции переходов состояний, награды за эти переходы. Возможны переходы вправо и вниз, что соответствует перемещению работника на следующий уровень показателя при реализации управленческих решений в соответствующей группе. Действия дискретны и отражают одно из управленческих решений, предназначенных для данной категории работников. Под решением понимается реализация определенного мероприятия (например, повышение квалификации, обеспечение занятий спортом и др.), направленного на рост человеческого капитала. Полный перечень таких мероприятий можно найти в (Orlova, 2021a). Можно привести в исполнение 24 различных решения при движении работника из ячейки (1, 1) в ячейку (5, 5). Каждый шаг имеет штраф -1 балл, достижение цели имеет награду в 3 балла, переход через область ограничений, например перемещение из (4, 3) в (4, 5), награждается в размере 1 балла. Цель алгоритма — научить агента формировать последовательность управленческих решений, обеспечивающих за минимальное время достижение максимальной награды.

Проведена серия обучающих экспериментов для ряда работников с разными характеристиками человеческого капитала. Для обучения агента использовались различные алгоритмы — DQN, DDQN, SARSA и PRO. На рис. 6 представлены результаты экспериментов в виде динамики средней награды в разрезе эпизодов (всего 200 эпизодов моделирования, в каждом из которых реализовывалось 50 испытаний).

Производительность алгоритмов оценивается по двум критериям — эффективности политики (средней награды) и эффективности обучения агента (скорости сходимости). Значения средних наград рассчитаны за каждый из 200 эпизодов моделирования, усредненных по 50 испытаниям. Показано, что наилучший результат обеспечивает алгоритм DDQN, дающий сравнительно

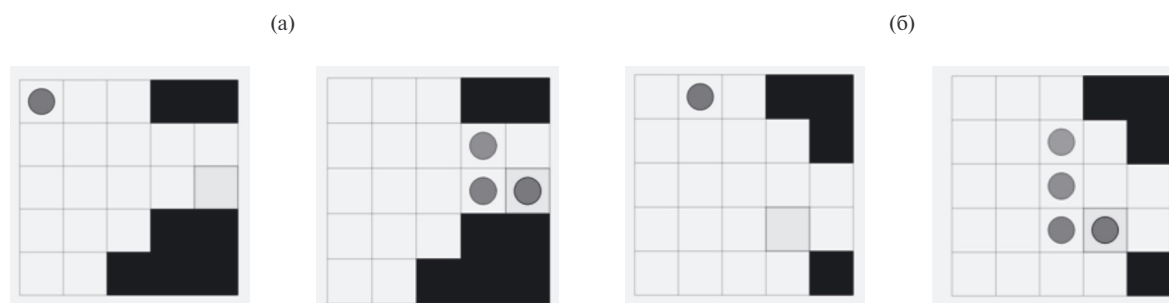


Рис. 7. Оптимальные политики (ИТР): а) эксперимент 1, начальное состояние (1, 1), целевое состояние (3, 5), возможен переход (прыжок) из состояния (4, 3) в состояние (3, 4); б) эксперимент 2, начальное состояние (2, 1), целевое состояние (4, 4), возможен переход (прыжок) из (1, 3) в (2, 4)

Таблица 3. Результаты оценки эффективности политик и эффективности обучения

Алгоритм	Эффективность политики — средняя награда	Эффективность обучения — скорость сходимости (число эпизодов)
DQN	−0,15	0,27 (53)
DDQN	0,2	0,29 (58)
SARSA	−3,1	0,59 (117)
PRO	−52	—

быстрое обучение и положительную награду (табл. 3). Поэтому выработка ИТР должна быть основана на DDQN-агенте.

На основе модельных экспериментов для каждого работника с учетом начального и конечного (целевого) состояния были сформированы оптимальные стратегии в виде ИТР (см. рис. 6). Они отражают последовательность принятия управленческих решений, обеспечивающих рост качества человеческого капитала. На рис. 7 по мере продвижения по сформированной траектории текущее состояние (кружок) все более затемняется.

Результаты моделирования демонстрируют, что для первого работника для достижения целевого состояния наилучшим решением будет следующая последовательность действий. Сначала необходимо реализовать программные мероприятия, направленные на улучшения состояния здоровья работника до второго уровня, а затем постепенно выполнять программы обеспечения роста основных компетенций до уровня 4, после этого применить меры, направленные на оздоровление, и только после этого снова обеспечить повышение профессиональных компетенций. Для второго работника траектория его развития следующая: сначала реализация программы, обеспечивающей рост профессионализма до уровня 3, затем проведение поэтапных мероприятий, направленных на улучшение состояния здоровья до уровня 4, затем применение мер, направленных на рост компетентности. Реализация сконструированных ИТР на основе обученного агента позволит повысить качество человеческого капитала и обеспечит рост эффективности труда.

5. ЗАКЛЮЧЕНИЕ

В работе показано, что методы «машинного обучения с подкреплением» имеют доказанную эффективность, когда особенности решаемых задач управления следующие: 1) объект управления характеризуется стохастической динамикой своих показателей, а управленческие решения недетерминированы; 2) задачи управления имеют стратегический характер; 3) решение задачи управления представляется в виде последовательного принятия решений. Задача управления человеческим капиталом удовлетворяет всем перечисленным свойствам. Поэтому применение методов и инструментов «машинного обучения с подкреплением» для ее решения теоретически обосновано.

В работе проведена систематизация областей применения моделей и методов «машинного обучения с подкреплением». Показано, что для таких сложных организационных систем, как системы управления человеческим капиталом предприятия, обладающих свойствами самостоятельного целеполагания,

рефлексии и ограниченной рациональности при принятии решений, не существует комплексного методологического подхода выработки оптимального режима управления качеством человеческого капитала.

Впервые представлено, что одним из эффективных инструментов повышения качества человеческого капитала является индивидуализация решений при разработке стратегии управления. На базе методологии цифрового двойника работника разработана модель управления человеческим капиталом, основанная на использовании методов «машинного обучения с подкреплением», обеспечивающая формирования индивидуальных траекторий профессионального развития работника с учетом его индивидуальных характеристик (состояния здоровья, уровня профессиональных и надпрофессиональных компетенций, мотивации, социального капитала) и текущих показателей деятельности.

Новизна модели управления человеческим капиталом состоит в том, что, во-первых, в ней отражены существенные свойства человеческого капитала, которые формируют условия для его управления; во-вторых, она основана на новой схеме поддержки принятия управленческих решений для формирования индивидуальных траекторий профессионального развития, позволяющих выработать комплекс решений для развития каждого работника, адекватный его потенциалу здоровья, интеллекта, социальных и карьерных возможностей. Модель управления человеческим капиталом базируется на концепции MDP, методах динамического программирования и алгоритмах «машинного обучения с подкреплением». Выработанная на основе модели политика представляет собой набор правил, сформированный на основе алгоритма «машинного обучения с подкреплением», для определения оптимального состава программ в определенный момент в зависимости от характеристик работника в текущий момент и результатов действия программ в прошедшие моменты времени. Отличие предложенной модели управления человеческим капиталом от других состоит в том, что политика управления вырабатывается без наличия точной математической модели (поведения) работника, а также без наличия данных о причинно-следственных связях между решением (программой) и результатом (эффективностью программы).

Проведены экспериментальные исследования использования предлагаемой модели управления человеческим капиталом на базе данных крупного предприятия. Для работников с разным профилем человеческого капитала сформированы оптимальные политики, которые являются реализацией ИТР. Реализация выработанных политик позволит повысить качество человеческого капитала предприятия, повысит его ресурсную эффективность и обеспечит рост интегральных показателей деятельности.

Теоретическая значимость результатов обусловлена тем, что процесс организации, структурирования и поиска эффективных стратегий управления человеческим капиталом рассматривается как задача последовательного принятия решений в стохастической среде. Оптимальные решения формируются в среде, динамика которой описывается с помощью MDP, по критерию максимизации ожидаемой полезности, на основе алгоритмов Q-обучения. Практическую значимость представляет система поддержки принятия решений, которая может быть сконструирована на базе разработанных моделей и являться основой для управления человеческими ресурсами предприятия, обеспечивая в минимальные сроки рост инновационности и конкурентоспособности предприятия за счет улучшения качества человеческого капитала.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Акопов А.С.** (2023). Моделирование и оптимизация стратегий принятия индивидуальных решений в многоагентных социально-экономических системах с использованием машинного обучения // *Бизнес-информатика*. Т. 17. № 2. С. 7–19. DOI: 10.17323/2587-814X.2023.2.7.19 [**Akopov A.S.** (2023). Modeling and optimization of strategies for making individual decisions in multi-agent socio-economic systems with the use of machine learning. *Business Informatics*, 17, 2, 7–19. DOI: 10.17323/2587-814X.2023.2.7.19 (in Russian).]
- Боровков А.И.** (2021) Цифровые двойники в условиях четвертой промышленной революции // *CONNECT. Мир информационных технологий*. № 1–2. С. 50–53. [**Borovkov A.I.** (2021). Digital twins in the fourth industrial revolution. *CONNECT. The World of Information Technologies*, 1–2, 50–53 (in Russian).]
- Макаров В.Л., Бахтизин А.Р., Бекларян Г.Л.** (2019). Разработка цифровых двойников для производственных предприятий // *Бизнес-информатика*. Т. 13. № 4. С. 7–16. DOI 10.17323/1998-0663.2019.4.7.16 [**Makarov V.L., Bakhtizin A.R., Beklaryan G.L.** (2019). Developing digital twins for production enterprises. *Business Informatics*, 13, 4, 7–16. DOI: 10.17323/1998-0663.2019.4.7.16 (in Russian).]
- Макаров В.Л., Бахтизин А.Р., Бекларян Г.Л., Акопов А.С., Ровенская Е.А., Стрелковский Н.В.** (2022). Агентное моделирование социально-экономических последствий миграции при государственном регулировании занятости // *Экономика и математические методы*. Т. 58. № 1. С. 113–130. DOI: 10.31857/S042473880018960-5 [**Makarov V.L., Bakhtizin A.R., Beklaryan G.L., Akopov A.S., Rovenskaya E.A., Strelkovsky N.V.** (2022). Agent-based

- modeling of the socio-economic consequences of migration under state regulation of employment. *Economics and Mathematical Methods*, 58, 1, 113–130. DOI: 10.31857/S042473880018960-5 (in Russian).]
- Макаров В.Л., Бахтизин А.Р., Бекларян Г.Л., Акопов А.С., Стрелковский Н.В., Ровенская Е.А.** (2020). Агентное моделирование популяционной динамики двух взаимодействующих сообществ: мигрантов и коренных жителей // *Экономика и математические методы*. Т. 56. № 2. С. 5–19. DOI 10.31857/S042473880009217-7 [Makarov V.L., Bakhtizin A.R., Beklaryan G.L., Akopov A.S., Strelkovsky N.V., Rovenskaya E.A. (2020). Agent-based modeling of population dynamics of two interacting communities: Migrants and indigenous residents. *Economics and Mathematical Methods*, 56, 2, 5–19. DOI: 10.31857/S042473880009217-7 (in Russian).]
- Макаров В.Л., Клейнер Г.Б.** (2007). *Микроэкономика знаний*. М.: Экономика. 300 с. [Makarov V.L., Kleiner G.B. (2007). *Microeconomics of knowledge*. Moscow: Economics. 300 p. (in Russian).]
- Орлова Е.В.** (2020а). Методы и модели анализа данных и машинного обучения в задаче управления производительностью труда // *Программная инженерия*. № 4. С. 219–229. DOI: 10.17587/prin.11.219-229 [Orlova E.V. (2020a). Methods and models of data analysis and machine learning in the problem of labor productivity management. *Programmnyaya Ingeneria (Software Engineering)*, 11, 4, 219–229. DOI: 10.17587/prin.11.219-229 (in Russian).]
- Орлова Е.В.** (2020б). Управление производительностью труда с учетом факторов здоровья: технология и модели // *Управленец*. № 6. С. 57–69. DOI: 10.29141/2218-5003-2020-11-6-5 [Orlova E.V. (2020b). Labour productivity management using health factors: Technique and models. *The Manager (Upravlenets)*, 11, 6, 57–69. DOI: 10.29141/2218-5003-2020-11-6-5 (in Russian).]
- Орлова Е.В.** (2021). Оценка человеческого капитала предприятия и управление им в условиях цифровой трансформации экономики // *Journal of Applied Economic Research*. Т. 20. № 4. С. 666–700. DOI: 10.15826/vestnik.2021.20.4.026 [Orlova E.V. (2021). Assessment of the human capital of an enterprise and its management in the context of the digital transformation of the economy. *Journal of Applied Economic Research*, 20, 4, 666–700. DOI: 10.15826/vestnik.2021.20.4.026 (in Russian).]
- Пономарев Е.С., Оселедец И.В., Чихоцкий А.С.** (2019). Использование обучения с подкреплением в задаче алгоритмической торговли // *Информационные процессы*. Т. 19. № 2. С. 122–131. [Ponomarev E.S., Oseledets I.V., Chihotsky A.S. (2019). Using reinforcement learning in algorithmic trading. *Information Processes*, 19, 2, 122–131 (in Russian).]
- Abideen A.Z., Sundram V.P.K., Pyeman J., Othman A.K., Sorooshian S.** (2021). Digital twin integrated reinforced learning in supply chain and logistics. *Logistics*, 5, 84. DOI: 10.3390/logistics5040084
- Alzyoud A.** (2018). The influence of human resource management practices on employee work engagement. *Foundations of Management*, 10, 251–256. DOI: 10.2478/fman-2018-0019
- Azhikodan A.R., Bhat A.G., Jadhav M.V.** (2019). Stock trading bot using deep reinforcement learning. In: *Innovations in computer science and engineering*. Springer: Berlin/Heidelberg, Germany, 41–49.
- Chi M., VanLehn K., Litman D.** et al. (2011). Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. *User Model User-Adapted Interaction*, 21, 137–180. DOI: 10.1007/s11257-010-9093-1
- Church A.H., Bracken D.W., Fleeno J.W., Rose D.S.** (2019). *Handbook of strategic 360 feedback*. New York: Oxford University Press. 637 p.
- Ding Q., Jahanshahi H., Wang Y., Bekiros S., Alassafi M.O.** (2022). Optimal reinforcement learning-based control algorithm for a class of nonlinear macroeconomic systems. *Mathematics*, 10, 499. DOI: 10.3390/math10030499
- Granovetter M.S.** (1973). The strength of weak ties. *American Journal of Psychology*, 78 (6), 1360–1380.
- Hernaus T., Pavlovic D., Klindzic M.** (2019). Organizational career management practices: The role of the relationship between HRM and trade unions. *Employee Relations*, 41, 84–100. DOI: 10.1108/ER-02-2018-0035
- Hitka M., Kucharčíková A., Štarchoň P., Balážová Ž., Lukáč M., Stacho Z.** (2019). Knowledge and human capital as sustainable competitive advantage in human resource management. *Sustainability*, 11, 4985. DOI: 10.3390/su11184985
- Jung Y., Takeuchi N.** (2018). A lifespan perspective for understanding career self-management and satisfaction: The role of developmental human resource practices and organizational support. *Human Relations*, 7, 73–102.
- Li Q., Lin T., Yu Q., Du H., Li J., Fu X.** (2023). Review of deep reinforcement learning and its application in modern renewable power system control. *Energies*, 16, 4143. DOI: 10.3390/en16104143
- Liu J., Zhang Y., Wang X., Deng Y., Wu X.** (2019). *Dynamic pricing on e-commerce platform with deep reinforcement learning*. arXiv:1912.02572.
- Mohammadi M., Al-Fuqaha A., Guizani M., Oh J.** (2018). Semisupervised deep reinforcement learning in support of IoT and smart city services. *IEEE Internet of Things Journal*, 5, 2, 624–635.
- Orlova E.V.** (2021a). Innovation in company labor productivity management: Data science methods application. *Applied System Innovation*, 4, 3, 68. DOI: 10.3390/asi4030068
- Orlova E.V.** (2021b). Design of personal trajectories for employees' professional development in the knowledge society under industry 5.0. *Social Sciences*, 10, 11, 427. DOI: 10.3390/socsci10110427
- Orlova E.V.** (2022). Design technology and ai-based decision making model for digital twin engineering. *Future Internet*, 14, 9, 248. DOI: 10.3390/fi14090248

- Orlova E.V. (2023). Inference of factors for labor productivity growth used randomized experiment and statistical causality. *Mathematics*, 11, 4, 863. DOI: 10.3390/math11040863
- Orr J., Dutta A. (2023). Multi-agent deep reinforcement learning for multi-robot applications: A survey. *Sensors*, 23, 3625. DOI: 10.3390/s23073625
- Osranek R., Zink K.J. (2014). Corporate human capital and social sustainability of human resources. In: I. Ehnert, W. Harry, K. Zink. *Sustainability and human resource management*. CSR, Sustainability, Ethics & Governance. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-37524-8_5
- Rachid B., Mohamed T., Khouaja M.A. (2018). An agent based modeling approach in the strategic human resource. Management, including endogenous and exogenous factors. *Simulation Modelling Practice and Theory*, 88, 32–47.
- Schelling T.C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, (Informa UK Limited), 1 (2), 143–186. DOI:10.1080/0022250x.1971.9989794
- Stelman L.A., Williams J.R. (2019). *Feedback at work*. Switzerland AG: Springer Nature. 280 p.
- Stokowski S., Li B., Goss B.D., Hutchens S., Turk M. (2018). Work motivation and job satisfaction of sport management faculty members. *Sport Management Education Journal*, 12, 80–89. DOI: 10.1123/smej.2017-0011
- Wang R., Chen Z., Xing Q., Zhang Z., Zhang T. (2022). A modified rainbow-based deep reinforcement learning method for optimal scheduling of charging station. *Sustainability*, 14, 1884. DOI: 10.3390/su14031884
- Yan Y., Chow A.H., Ho C.P., Kuo Y.H., Wu Q., Ying C. (2022). Reinforcement learning for logistics and supply chain management: Methodologies, state of the art, and future opportunities. *Transportation Research Part E: Logistics and Transportation Review*, 162, 102712.
- Yu C., Liu J., Nemati S. (2019a). *Reinforcement learning in healthcare: A survey*. arXiv:1908.08796.
- Yu P., Lee J.S., Kulyatin I., Shi Z., Dasgupta S. (2019b). *Model-based deep reinforcement learning for dynamic portfolio optimization*. arXiv:1901.08740.
- Zhang L., Guo X., Lei Z., Lim M.K. (2019). Social network analysis of sustainable human resource management from the employee training's perspective. *Sustainability*, 11, 380. DOI: 10.3390/su11020380
- Zheng G., Zhang F., Zheng Z., Xiang Y., Yuan N.J., Xie X., Li Z. (2018). DRN: A deep reinforcement learning framework for news recommendation. In: *Proceedings of the 2018 World Wide Web Conference*. Lyon, France, 167–176.

Model for human capital management of an enterprise based on reinforcement learning methods

© 2025 E.V. Orlova

E.V. Orlova,

Ufa University of Science and Technology, Ufa, Russia; e-mail: ekorl@mail.ru

Received 01.08.2023

Abstract. Human capital is an important driver for sustainable enterprise's economic growth and becomes more important under digital transformation. The employee profile appears multifaceted due to the expansion of activities. Therefore, the problem of human capital management based on the design of employees' individual trajectories of professional development is relevant, timely, socially and economically significant. The paper proposes a model for employees' individual trajectories of the professional development, which is based on reinforcement learning methods. The model forms an optimal management regime and is considered as a consistent set of program activities aimed at the employee's development in his professional sphere. It considers employee's individual characteristics (health, competencies, motivation and social capital). The total control system is considered as a digital twin of an employee, and creates the environment — the model of an employee as a Markov decision process and the control model — the agent — a center of enterprise's decision-making. We use reinforcement learning algorithms DDQN, SARSA, PRO to maximize the agent's utility function. Based on the experiments, it is shown that the best results are provided by the DDQN algorithm. The results generated by the proposed model are of practical importance, which would contribute to the growth of an enterprise's innovativeness and competitiveness by improving the human capital quality and increasing the labor resource efficiency.

Keywords: human capital, employee's digital twin, individual trajectories, machine learning, reinforcement learning, Q-learning, optimal control, Markov decision process

JEL Classification: C61, O15.

UDC: 330.4, 519.865.7.

For reference: Orlova E.V. (2025). Model for human capital management of an enterprise based on reinforcement learning methods. *Economics and Mathematical Methods*, 61, 1, 70–83. DOI: 10.31857/S0424738825010072 (in Russian).