

РАЗРАБОТКА СТРАТЕГИЙ УПРАВЛЕНИЯ СЛОЖНЫМИ ПРОЕКТАМИ НА ОСНОВЕ АЛГОРИТМОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

С.А. Баркалов, bsa610@yandex.ru, <https://orcid.org/0000-0001-6183-3004>
А.В. Белоусов, alexbelousov19@yandex.ru, <https://orcid.org/0000-0003-1464-298X>
Е.А. Серебрякова, sea-parish@mail.ru, <https://orcid.org/0000-0001-5129-246X>
Воронежский государственный технический университет, Воронеж, Россия

Аннотация. В данной работе рассматриваются основные подходы к формированию стратегий управления сложными проектами на основе методов машинного обучения с подкреплением. Масштаб и разнообразие задач сложных проектов, количество исполнителей и используемых ресурсов могут создавать весьма сложные комбинаторные схемы, поэтому эффективное оперативное интерпретирование таких структур и принятие обоснованных решений, гарантирующих соблюдение сроков, представляют собой значительную сложность. Чтобы преодолеть обозначенные сложности, применяются интеллектуальные системы, способные анализировать и предсказывать временные и ресурсные характеристики, относящиеся как к отдельным задачам, так и к их совокупности. Такой процесс интегрирован в марковскую модель принятия решений. В ее состав входит разработка системы показателей как количественных, так и качественных, предназначенных для оценки работы проектов. Все эти аспекты играют существенную роль в определении тех технических и экономических рисков, которые могут возникнуть в процессе реализации проектов путем определения стратегий. **Цель исследования** заключается в формировании интеллектуальной модели и алгоритма принятия решений при выборе стратегий реализации сложных проектов на основе моделей машинного обучения с подкреплением. **Методы исследования.** Для решения задачи выбора стратегий управления сложными проектами использовалась марковская модель принятия решений, используемая для оценки ценности состояний и действий агента при выборе стратегий на основе метода temporal difference. В результате оценка ценности действий производилась на основе алгоритма SARSA, позволяющего получить оптимальные варианты действий на каждой задаче выполнения проекта в зависимости от факторов внутренней и внешней неопределенности. **Результаты.** Представленные в статье методы предоставляют эффективный инструмент для оперативного разрешения широкого спектра задач, неизбежно возникающих при воплощении сложных проектов, учитывая как внутренние, так и внешние факторы неопределенности. Благодаря использованию модели машинного обучения с подкреплением, основанной на марковском процессе, создается основа для системы поддержки принятия решений. Эта система способна динамически оценивать ход выполнения проекта и формировать адаптивные и точные стратегии с низким уровнем погрешности. В ее структуру могут быть интегрированы различные подмодели, такие как регрессии, классификаторы, кластеризаторы и глубокие нейронные сети. **Заключение.** Полученные результаты в полной мере применимы для формирования эффективных стратегий управления сложными проектами. Доказано, что использование марковских моделей принятия решений в полной мере позволяет нивелировать неопределенность при определении характера закона распределения случайной величины генеральной совокупности данных, необходимых для обучения проекта. Кроме того, дискретность в МППР соответствует характеру формирования и управления сложным проектом, осуществляемым варьированием параметров ϵ и γ , а в случае, когда проект длится значительное время, возможно использование параметра β , эффективность применения которого подлежит дальнейшим исследованиям.

Ключевые слова: алгоритм, марковская модель принятия решений, граф, обучение с подкреплением, стратегии, таблица ценности действий

Для цитирования: Баркалов С.А., Белоусов А.В., Серебрякова Е.А. Разработка стратегий управления сложными проектами на основе алгоритмов обучения с подкреплением // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2026. Т. 26, № 1. С. 90–101. DOI: 10.14529/ctcr260108

DEVELOPMENT OF STRATEGY OF MANAGEMENT OF THE DIFFICULT PROJECTS ON THE BASIS OF TUTORING ALGORITHMS WITH THE REINFORCEMENT

S.A. Barkalov, bsa610@yandex.ru, <https://orcid.org/0000-0001-6183-3004>

A.V. Belousov, alexbelousov19@yandex.ru, <https://orcid.org/0000-0003-1464-298X>

E.A. Serebryakova, sea-parish@mail.ru, <https://orcid.org/0000-0001-5129-246X>

Voronezh State Technical University, Voronezh, Russia

Abstract. In this work the main approaches to formation of strategy of management of the difficult projects on the basis of methods of machine learning with a reinforcement are considered. The scale and a variety of tasks of the difficult projects, the number of performers and the used resources can create very the compound combinatory circuits therefore effective expeditious interpretation of such structures and adoption of the justified decisions guaranteeing keeping of terms represent the considerable complexity. To overcome the designated difficulties, the intellectual systems capable to analyze and predict the temporary and resource characteristics relating both to separate tasks, and to their set are applied. Such process is integrated into Markov model of a decision making. Development of the system of indexes as the quantitative, and the qualitative, intended for job evaluation projects is its part. All these aspects play an essential role in determination of those technical and economic risks which can arise during implementation of projects a path of definition of strategy. The research objective consists in formation of intellectual model and algorithm of a decision making when choosing strategy of implementation of the difficult projects on the basis of models of machine learning with a reinforcement. **Research techniques.** For the solution of a problem of the choice of strategy of management of the difficult projects the Markov model of acceptance of decisions used for assessment of value of states and actions of the agent when choosing strategy on the basis of the temporal difference method was used. As a result, assessment of value of actions was made on the basis of the algorithm SARSA allowing to receive optimal variants of actions on each problem of implementation of the project depending on factors of internal and external indeterminacy. **Results.** The methods presented in article provide the effective tool for operational permission of a wide range of the tasks which are inevitably arising at the embodiment of the difficult projects, considering both internal, and external factors of indeterminacy. Thanks to the model of machine learning with a reinforcement based on the Markovian process the basis for the system of support of a decision making is created. This system is capable to estimate dynamically the course of implementation of the project and to form adaptive and exact strategy with low level of an error. Various sub models, such as regressions, qualifiers, clustering and deep neural networks can be integrated into its structure. **Conclusion.** The received results in to the complete measure are applicable for formation of effective strategy of management of the difficult projects. It is proved that use of Markov models of a decision making fully allows to level indeterminacy when determining nature of the distribution law of a random value of a universe of the data necessary for tutoring of the project. Besides the discretization in MPPR corresponds to the nature of formation and management of the difficult project which is carried out by variation of parameters ϵ and γ and in case the project lasts the considerable time perhaps of use of parameter β which effectiveness of application is subject to further researches.

Keywords: algorithm, Markov model of a decision making, columns, tutoring with a reinforcement, strategy, action value table

For citation: Barkalov S.A., Belousov A.V., Serebryakova E.A. Development of strategy of management of the difficult projects on the basis of tutoring algorithms with the reinforcement. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics.* 2026;26(1):90–101. (In Russ.) DOI: 10.14529/ctcr260108

Введение

В России проектная методология набирает все большую популярность во многих секторах экономики. Однако масштаб и разнообразие задач, исполнителей и используемых ресурсов могут создавать весьма сложные комбинаторные схемы. Эффективное оперативное интерпретирование таких структур и принятие обоснованных решений, гарантирующих соблюдение сроков, представляют собой значительную сложность. Чтобы преодолеть обозначенные сложности, приме-

няются интеллектуальные системы, способные анализировать и предсказывать временные и ресурсные характеристики, относящиеся как к отдельным задачам, так и к их совокупности. Эти системы способны оценивать риски и, в соответствии с текущей внутренней и внешней обстановкой, реализовывать альтернативные задачи. Такой процесс интегрирован в марковскую модель принятия решений [1]. В его состав входит разработка системы показателей как количественных, так и качественных, предназначенных для оценки работы проектов. Эти показатели дают возможность не только измерять эффективность проектов, но и прогнозировать их дальнейшее развитие. Подобные расчеты приобретают особую актуальность в связи с тем, что реализация сложных проектов неизбежно сталкивается с непредвиденными факторами, требующими оперативного анализа и корректировки стратегии выполнения задач. Применение такого подхода значительно повысит конкурентоспособность компаний и даст менеджерам возможность использовать передовые цифровые инструменты для управления. Эти инструменты интегрируют принципы машинного обучения с подкреплением и производственными операциями, что особенно важно в условиях существующего технологического диссонанса между этапами проектирования и воплощения в жизнь. При определении подходящих моделей необходимо принимать во внимание специфику конкретного производственного процесса, где на выбор оказывают влияние разнообразные как внешние, так и внутренние обстоятельства. Внутренние возможности компании формируются под влиянием ряда ключевых элементов. К ним относятся уровень ее технической базы, включающий применяемые технологии и программное обеспечение, отвечающие требованиям современных цифровых решений; финансовое состояние, обеспечивающее стабильность и возможность инвестиций; квалификация персонала и его готовность к освоению новых, конкурентных технологий. Все эти аспекты играют существенную роль в определении тех технических и экономических рисков, которые могут возникнуть в процессе реализации проектов. Выбор методик прогнозирования, планирования и управления напрямую зависит от этих факторов. Однако если при составлении прогнозов не учитывать их воздействие, то результаты могут оказаться неточными [1, 2]. В результате реальные показатели выполнения проектов и работы компаний могут существенно расходиться с прогнозируемыми значениями.

Постановка задачи

Некоторые проекты обладают высокой степенью инновационности, что часто сопровождается дефицитом или полным отсутствием статистических данных. Длительный срок реализации многих проектов также создает трудности в применении регрессионных моделей для прогнозирования показателей. При недостаточном объеме данных для достоверного определения распределения генеральной совокупности применяются методы повторной выборки, известные как ресемплинг [1, 3]. Ресемплинг позволяет минимизировать риск ошибочных выводов, связанных с гипотезой о распределении. В контексте управления сложными проектами эффективным вариантом ресемплинга выступает бутстрэп. Для создания классификаторов и регрессоров в таких проектах хорошо зарекомендовали себя ансамблевые методы, основанные на адаптивном бустинге.

Анализируем подход адаптивного бустинга в контексте бинарной классификации [4]. Предположим, что задан таргет

$$Y = \{-1, +1\},$$

признаковое множество

$$D = \{(x_1, y_1), \dots, (x_d, y_d)\},$$

тогда цель заключается в тренировке ансамбля, состоящего из множества (M) моделей.

Итоговая модель будет выглядеть следующим образом:

$$\hat{f}_M(x) = \text{sign}(\sum_{m=1}^M \rho_m a_m(x)), \quad (1)$$

где ρ_m – поправочные коэффициенты для базовых алгоритмов $a_m(x)$.

Введем следующие весовые коэффициенты:

$$w_{i,M} = \exp(-y_i \sum_{m=1}^M \rho_m a_m(x_i));$$

$$\tilde{w}_{i,M} = \frac{w_{i,M-1}}{\sum_j w_{j,M-1}}.$$

В таком случае алгоритм, основанный на адаптивном бустинге [1, 5], будет функционировать следующим образом:

1. Задаем начальные значения для весов $w_{i,1}$, где $i \in [1, d]$ (при этом значение весов рассчитывается как $\frac{1}{d}$, т. е. $w_{i,1} \leftarrow \frac{1}{d}$).

2. Обучим базовые алгоритмы:

2.1. Пусть $a_m (m \in [1, M])$ – базовый алгоритм с достаточно незначительной ошибкой

$$N(a_m, \tilde{w}_m) \leftarrow \sum_{i=1}^d \tilde{w}_{i,m} 1(y_i a_m(x_i) \leq 0) \quad (2)$$

и коэффициентами

$$a_m: p_m \leftarrow \frac{1}{2} \log \frac{1 - N(a_m, \tilde{w}_m)}{N(a_m, \tilde{w}_m)}.$$

В рамках обработки каждой записи выборки, то есть выполняя итерацию от 1 до d , мы будем обновлять значения коэффициентов $w_{i,m+1}$ и $\tilde{w}_{i,m+1}$:

$$w_{i,m+1} \leftarrow w_{i,m} \exp(-\rho_m y_i a_m(x_i)), \quad (3)$$

$$\tilde{w}_{i,m+1} \leftarrow \frac{w_{i,m+1}}{\sum_{j=1}^d w_{j,m+1}}.$$

2.2. Повысим эффективность модели путем обновления ее состава:

$$f_m \leftarrow \sum_{s=1}^m \rho_s a_m(x_i). \quad (4)$$

3. Тогда результирующая модель будет выглядеть как

$$\hat{f}_m = \text{sign}(f_m). \quad (5)$$

Определившись с ансамблем моделей, необходимо выбрать аппарат для управления сетевым графиком реализации сложного проекта с учетом следующих недостатков, присущих сложным проектам [6, 7]:

– используемые в сложных проектах параметры представляют собой случайные величины, законы распределения которых меняются в процессе работы. Основанные на точечных оценках модели не всегда точны, что может вызвать задержку принятия решений и не позволить добиться желаемых результатов;

– сложные проекты, управляемые по жесткому календарному графику, демонстрируют низкую гибкость. Это часто влечет за собой несвоевременность и неточность действий, которые чаще всего представляют собой реакцию на негативные внешние и внутренние обстоятельства. В таких ситуациях исправление ошибок становится сложным и дорогостоящим;

– не всегда незначительные краткосрочные неудачи приводят к глобальным потерям. Однако постоянная попытка исправить их может негативно повлиять на общий прогресс и эффективность долгосрочных задач.

Среди существующих подходов особенно перспективны модели reinforcement learning (RL), поскольку они предоставляют широкие возможности для разработки гибких стратегий принятия решений (ПП) в сложных проектах. Ключевое отличие RL заключается в отсутствии заранее заданных целевых значений (откликов). Обучение в таких моделях направлено на формирование стратегий агента таким образом, чтобы он максимизировал общую сумму полученных «наград» [1, 7].

Обучение модели учитывает случайную природу всех входящих в нее переменных, что позволяет оперативно корректировать стратегии в краткосрочном плане. При этом долгосрочная стратегическая цель остается неизменной: обеспечение организационно-технологической надежности при выполнении сложных задач. Для достижения этой цели эффективно применяются марковские процессы принятия решений, основанные на определенной четверке параметров (S, A, R, P), описанных в [1, 8]:

S – различные фазы функционирования агента (этапы исполнения расписания сложного проекта);

A – набор операций, которые агент способен совершить (в рамках функционирования сложного проекта);

$R: S \times A \rightarrow \mathbb{R}$ – механизм вознаграждения, активирующийся при переходе между состояниями S и S' , если выполнено конкретное действие a ;

$P: S \times A \rightarrow \pi(S)$ – граф переходов между состояниями описывается набором вероятностных распределений, где каждое распределение относится к множеству S . При этом вероятность перехода в следующее состояние не зависит от истории предыдущих состояний, т. е.

$$P_k(S_{t+1} = s' | S_t, a_t, r_t, S_{t-1}, a_{t-1}, r_{t-1}, \dots, S_0, a_0) = P_k(S_{t+1} = s' | S_t, a_t). \quad (6)$$

Следовательно, переход агента в новое состояние на очередном шаге определяется исключительно его нынешним состоянием и принятым решением. Данная система правил МППР отлично подходит для оценки эффективности исполнения сложного проекта и определения корректирующих действий при отклонениях от плана [8].

В силу предопределенного числа этапов, свойственного сложным проектам, мы имеем дело с конечным марковским процессом принятия решений. Задача завершения проекта соответствует терминальной узловой точке графа МППР. В рамках обучения мы примем во внимание оценку ожидаемой пользы после совершения шага t :

$$G_t = r_{t+1} + r_{t+2} + \dots + r_T, \quad (7)$$

где r_{t+i} – система поощрений для агента, принимающего решение a_k в актуальном состоянии; r_T – награда за совершение шага, который приводит к завершению работы системы.

Ввиду сложности толкования полученных данных о предполагаемой выгоде мы примем ее за основу при определении стратегии реализации сложного проекта:

$$G_t = r_{t+1} + \gamma r_{t+2} + \dots + \gamma^{T-t-1} r_T, \quad (8)$$

где $\gamma \in [0,1]$ – коэффициент дисконтирования, позволяющий управлять выбором стратегий реализации сложного проекта (когда его значение приближается к нулю, для ЛПП приоритетом становится получение максимального результата на данном этапе проекта [9]. В остальных ситуациях ЛПП предпочитает реализовать проект в полном объеме, соблюдая сроки и заданное качество (такой подход носит долгосрочный характер).

Далее необходимо оценить ценность различных состояний и действий, исходя из полученной выгоды, которая будет реализована в рамках выбранной стратегии реализации сложного проекта, учитывая ранее проанализированные случайные факторы.

В этом случае значение функции S , которая отражает ценность состояний (достижимых результатов проекта), определяется при применении данной стратегии π :

$$v_\pi(s) = M_\pi(\sum_{k=0}^T \gamma r_{t+k+1} | s_t = s), \quad (9)$$

где $s \in S$, т. е. оценить текущее состояние графиков проекта можно лишь на основе имеющихся задач.

Следовательно, модификация состояния влечет за собой трансформацию задач и, как следствие, приводит к изменению ее ценности как в краткосрочном, так и в долгосрочном аспектах для проекта [10]. Далее обратим внимание на функцию, определяющую ценность действий, обозначенных как a , для задачи s при применении стратегии π :

$$q_\pi(s, a) = M_\pi(\sum_{k=0}^T \gamma r_{t+k+1} | s_t = s, a_t = a). \quad (10)$$

Формулы (9) и (10) рассчитывают среднее значение полученных вознаграждений для ЛПП, если он примет определенные решения в рамках строительного проекта. В расчетах учитывается дисконтирование вознаграждений, а также внутренние и внешние факторы неопределенности [10, 11].

С учетом понимания ключевых характеристик строительного проекта мы можем использовать рекуррентные формулы для определения функций ценности [11]. Предполагая, что известны вероятностные показатели переходов между задачами (s и s') при осуществлении работ a :

$$P_{ss'}^a = P(s_{t+1} = s' | s_t = s, a_t = a). \quad (11)$$

Система вознаграждений для ЛПП при выборе задач теперь полностью сформирована:

$$R_{ss'}^a = M_\pi(r_{t+1} | s_t = s, a_t = a, s_{t+1} = s'). \quad (12)$$

В настоящее время мы можем сформулировать уравнения Беллмана для определения стоимости состояний и действий в рамках выбранной стратегии:

$$v_\pi(s) = \sum_{a \in A(s_t)} \pi(a|s) \sum_{s' \in S} P_{ss'}^a (R_{ss'}^a + \gamma v_\pi(s')), \quad (13)$$

$$q_\pi(s, a) = \sum_{s' \in S} P_{ss'}^a (R_{ss'}^a + \gamma v_\pi(s')). \quad (14)$$

При определении стоимости задач в строительном проекте уравнение Беллмана отображает связь между ценностью текущей задачи и ценностью задач, которые следуют за ней [12]. Аналогично для функций ценности работ оно показывает, как ценность конкретной пары (задача, работа) определяется ценностью последующих пар подобных задач и работ.

Чтобы определить важность различных состояний и шагов в строительном проекте, сначала создается граф множественных прецедентных [13] отношений (МППР).

Пример такого графа представлен на рис. 1.

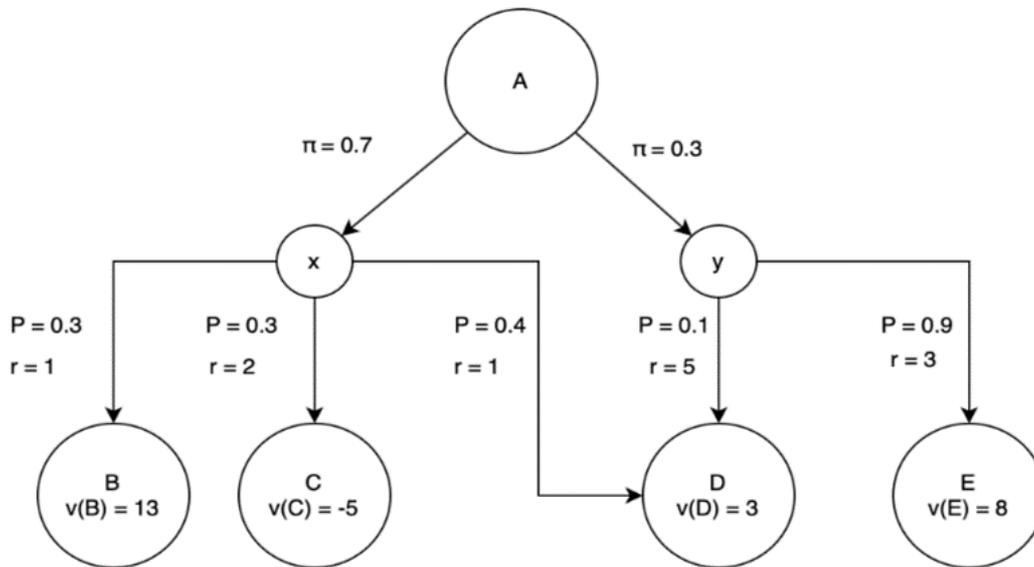


Рис. 1. Граф множественных прецедентных отношений (МППР)
Fig. 1. Count of the multiple case relations (MDP)

Теперь можно приступить к выбору стратегии управления сложным проектом. Для этого применим инструмент ϵ -жадной стратегии:

$$\pi_t(a|s_t) = \begin{cases} \frac{1-\epsilon}{|A_t|} - \frac{\epsilon}{|A(s_t)|}, & a \in A_t \\ \frac{\epsilon}{|A(s_t)|}, & a \notin A_t \end{cases}, \quad (15)$$

которая в зависимости от заданного параметра ϵ позволяет переходить от жадной (с вероятностью $(1 - \epsilon)$)

$$\pi_t(a|s_t) = \frac{1}{|A_t|}$$

к абсолютно исследовательской стратегии (вероятность ϵ)

$$\pi_t(a|s_t) = \frac{1}{|A(s_t)|}$$

К сожалению, в большинстве проектов не известны распределения вероятностей P_{ss}^a , выбора доступных действий $\pi(a|s)$ для всех состояний, распределения вероятностей переходов в следующие состояния и математические ожидания R_{ss}^a , получаемых при этом наград. Мы можем лишь задать коэффициент дисконтирования γ и вероятность ϵ для ϵ -жадной стратегии, а также скорость обучения α – *learning rate*.

Тогда можно использовать алгоритм обучения с подкреплением SARSA, в котором для обновления ценности действий $q(s_t, a_t)$ используется значение $q(s_{t+1}, a_{t+1})$, т. е. ценность действия a_{t+1} , выбранного в состоянии s_{t+1} согласно некоторой установленной стратегии $\pi_{t+1}(a|s_{t+1})$. В этом случае применяем принцип **Temporal Difference (TD)** для оценки ценности состояния (задачи проекта):

$$v(s_t) \leftarrow v(s_t) + \alpha(r_{t+1} + \gamma v(s_{t+1}) - v(s_t)). \quad (16)$$

Интерпретируется данное выражение следующим образом: находясь в состоянии s_t (задачи проекта), агент, следуя какой-то стратегии, выполняет действие a_t (выбирает возможные операции для выполнения задачи проекта), получает за это награду r_{t+1} и переходит в состояние s_{t+1} . Ценность $v(s_{t+1})$ нового состояния s_{t+1} также известна, значит, можно обновить s_t .

В процессе обучения необходимо совершенствовать подход к определению значимости различных действий:

$$q(s_t, a_t) \leftarrow q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a \in A(s_{t+1})} q(s_{t+1}, a_{t+1}) - q(s_t, a_t)). \quad (17)$$

Вначале происходит присвоение значений всем действиям $a \in A(s)$, которые могут быть выполнены из каждого нетерминального состояния.

Затем агент проходит ряд игр. Начало каждой игры характеризуется случайным начальным нетерминальным состоянием агента.

Далее в рамках каждой игры на каждом шаге t агент, находясь в s_t (задаче проекта), выбирает действие a_t (из списка возможных действий) согласно выбранной стратегии (определяемой значением ϵ), переходит в s_{t+1} (следующую задачу) и получает r_{t+1} . Если задача проекта s_{t+1} , то выбирается действие a_{t+1} , а значение $q(s_t, a_t)$ пересчитывается методом TD:

$$q(s_t, a_t) \leftarrow q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a \in A(s_{t+1})} q(s_{t+1}, a_{t+1}) - q(s_t, a_t)). \quad (18)$$

Рассмотрим пример расчёта для графа МППР, изображенного на рис. 1.

Зададим $\gamma = 0,8$, $\alpha = 0,1$. Стратегия ϵ -жадная со значением $\epsilon = 0,1$. Неизвестны ни вероятности переходов из состояния s в s' в результате выполнения действия a , ни то, какие награды получим в этом случае.

Все значения придется оценивать по ходу игры. Зададим матрицу ценности действий для инициализации алгоритма SARSA (табл. 1). Крестиками обозначены недоступные для данного состояния действия. Начальные ценности зададим равными нулю.

Таблица 1
Матрица ценности действий для инициализации
алгоритма SARSA

Table 1
Matrix of value of actions for initialization of an algorithm SARSA

	a_1	a_2	a_3	a_4
s_1	0	0	×	×
s_2	0	×	×	0
s_3	×	0	0	
s_4	×	0		0
s_5	×	×	×	×

После выбора a_2 агент переходит в позицию s_2 и получает отрицательное вознаграждение в размере 2. В состоянии s_2 ему предстоит принять решение: выполнить действие a_1 или a_4 . Ввиду равных значений, присвоенных обоим действиям (ноль), агент решает действовать по варианту a_4 . Теперь можно обновить значение $q(s_1, a_2)$:

$$q(s_1, a_2) \leftarrow q(s_1, a_2) + \alpha(r + \gamma q(s_2, a_4) - q(s_1, a_2)) = 0 + 0,1(-2 + 0,8 \cdot 0 - 0) = -0,2.$$

Получим матрицу ценности действий на шаге 1 (табл. 2).

Таблица 2
Матрица ценности действий на шаге 1

Table 2
Matrix of value of actions on a step 1

	a_1	a_2	a_3	a_4
s_1	0	-0,2	×	×
s_2	0	×	×	0
s_3	×	0	0	
s_4	×	0		0
s_5	×	×	×	×

Продолжим работу алгоритма:

Текущее состояние – s_2

Выбранное действие – a_4

Следующее состояние – s_5

Награда – 7

Игра заканчивается.

$$q(s_2, a_4) \leftarrow 0 + 0,1(7 - 0) = 0,7.$$

Теперь построим матрицу для оценки ценности действий для терминального состояния финального марковского процесса принятия решений (табл. 3).

Таблица 3
Матрица ценности действий для терминального состояния
Table 3
Matrix of value of actions for a terminal state

	a_1	a_2	a_3	a_4
s_1	0	-0,2	×	×
s_2	0	×	×	0,7
s_3	×	0	0	
s_4	×	0		0
s_5	×	×	×	×

Проведя серию игр, получим результирующую табл. 4.

Таблица 4
Результирующая матрица
Table 4
The resulting matrix

	a_1	a_2	a_3	a_4
s_1	-0,32	0,88	×	×
s_2	-0,02	×	×	6,5
s_3	×	-0,09	-0,19	
s_4	×	0,99		6,14
s_5	×	×	×	×

Наша оценка совершенного действия тесно связана с будущими действиями: мы предвосхищаем их, что, в свою очередь, влияет на наше восприятие настоящего.

Пример выбора стратегии управления сложным проектом

В качестве примера разработанной стратегии выберем сложные строительные проекты на стадии возведения объекта с использованием информационной поддержки всех этапов реализации. Ввиду многообразия критериев, определяющих эффективность управления строительством, требуется разработка методики оценки достижения целей строительных проектов. Влияние погодных условий, выполнение обязательств со стороны субподрядчиков, стабильность поставок материалов и техники – лишь некоторые из факторов, которые могут вывести проект с намеченного курса, нарушить график и подрывать организационно-технологическую надежность на всех этапах жизненного цикла объекта. Таким образом, требуется разработка системы, которая будет предоставлять интеллектуальную помощь в процессе управления ИСП с применением алгоритмов машинного обучения для принятия решений.

Для определения оптимальных стратегий управления строительными проектами были задействованы данные из специального датасета – **Construction_Data_PM_Tasks_All_Projects.csv** [14]. Данный набор данных объединяет информацию о 12 424 проектах, в которых оценивается качество выполненных работ и уровень соблюдения правил безопасности на стройплощадке. Пример, иллюстрирующий структуру данного набора данных, представлен на рис. 2.

Данный набор данных охватывает сведения о таких параметрах, как идентификатор проекта, его текущий статус, географическое положение, название, дата запуска, категория, история изменений статуса, список выполняемых задач и общее количество задач.

Осталось определить стоимость каждого возможного состояния, исходя из выбранной стратегии. Эта стратегия, в свою очередь, будет сформирована на основе функции, учитывающей как внутренние, так и внешние факторы неопределенности, на которые мы можем влиять, меняя стратегии.

Для решения задачи использовалась игра Franka Kitchen из фреймворка Gymnasium-Robotics [15]. В качестве стратегии была использована ϵ -жадная стратегия с изменяемыми параметрами $\epsilon = (0,1-0,6)$, коэффициентами дисконтирования $\gamma = (0,6-0,9)$, параметр $\text{random_seed} = 42$, количество игр – 10 000.

```
pm_forms_data=pd.read_csv('/content/Construction_Data_PM_Tasks_All_Projects.csv')
pm_forms_data
```

0	T1.23963030	Open	JPC Project Management>EHS Management>01 Inspe...	task raised in incorrect location of this form...	14/09/2020	NaN	Safety Notice (Amber) - General Issue	Main Contractor	14/09/2020	FormAnswer	False	
1	T116412.200	Closed	QC & BC(A)R>ITP 02 Architectural & M&E Service...	Metsec	14/09/2020	NaN	JPC - Progress Photo	Ceilings & Partitions	14/09/2020	NaN	False	T
2	T141663.27	EHS Good Observation	JPC Project Management>EHS Management>01 Inspe...	Good clear exclusion zones and access through ...	14/09/2020	NaN	Safety Notice (Green) - Good Observation	Main Contractor	14/09/2020	FormAnswer	False	T
3	T116412.100	Closed	QC & BC(A)R>ITP 02 Architectural &	RC walls	14/09/2020	NaN	JPC - Progress	Precast	14/09/2020	NaN	False	T

Рис. 2. Фрагмент набора данных Construction_Data_Tasks_Forms_All_Projects.csv
Fig. 2. Data set fragment Construction_Data_Tasks_Forms_All_Projects.csv

На рис. 3 приведен сравнительный анализ оценки ценности задач.



Рис. 3. Сравнительный анализ оценки ценности задач строительного проекта
Fig. 3. Comparative analysis of assessment of value of tasks of the structural project

Изучение рис. 3 показывает, что задачи с номерами 1328 и 1330 оказались самыми приоритетными, получив оценки 30,2 и 29,7 % соответственно. Все остальные задачи получили существенно более низкие оценки. Таким образом, при применении ϵ -жадной стратегии принимающий решения (ЛПР) будет первоначально сосредотачиваться на задачах, входящих в приоритетный план, а затем, по мере необходимости, обращаться к остальным задачам, следуя их ранжированию.

В последующем мы проанализируем полученные данные с точки зрения качества выполненных задач (рис. 4).

Изучение данных, представленных на рис. 4, выявило существенные различия между гистограммами. Помимо лидирующих показателей были получены неудовлетворительные результаты, что указывает на необходимость их корректировки путем изменения стратегий обучения.

В ходе обучения получен список приоритетных задач для реализуемых проектов в зависимости от факторов внешней неопределённости (рис. 5).

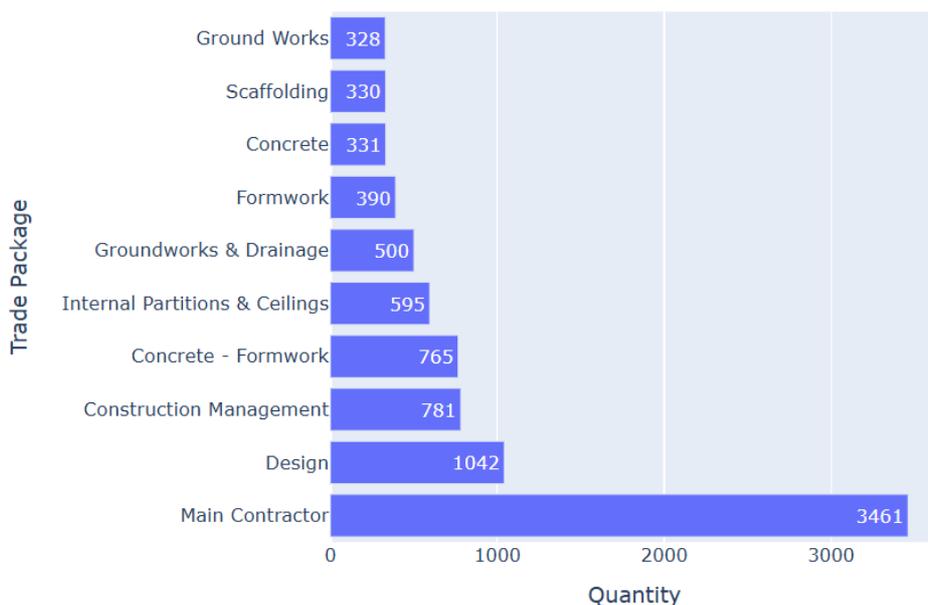


Рис. 4. Оценка ценности задач проекта по критерию качества
Fig. 4. Assessment of value of tasks of the project by criterion of quality

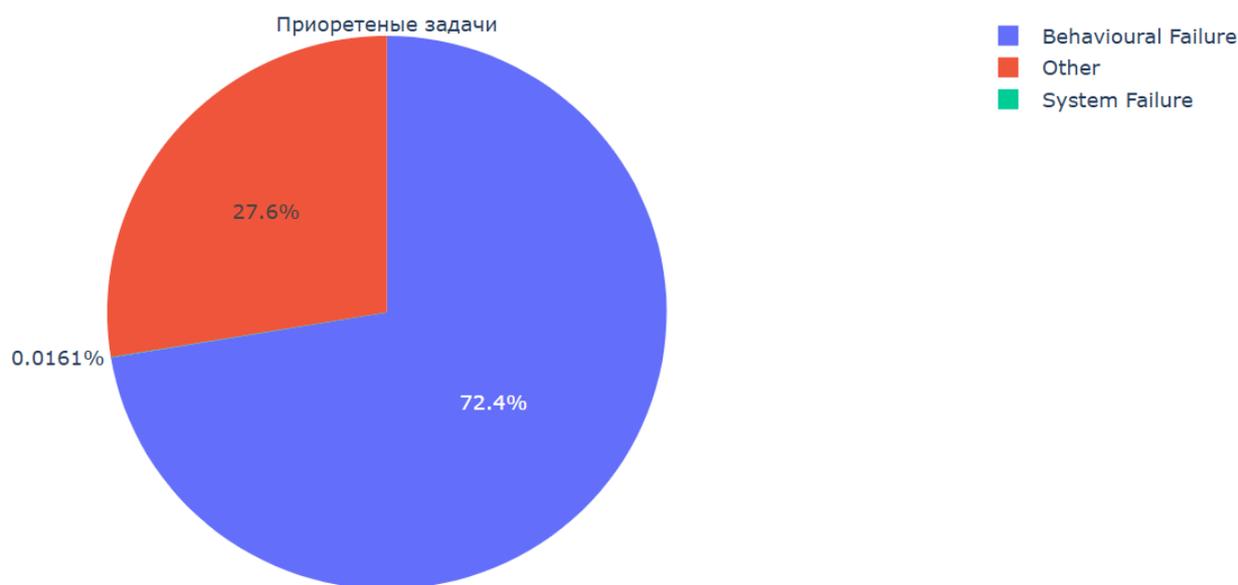


Рис. 5. Сравнительный анализ приоритетных задач проекта в зависимости от ценности состояний
Fig. 5. The comparative analysis of priority tasks of the project depending on the value of states

Только около 27,6 % задач идеально подходят для проекта. Остальные показывают значительную вариацию, что указывает на несовершенство выбранной стратегии и необходимость корректировки параметров вероятностных переходов между задачами в календарном графике.

Заключение

Представленные в статье методы предоставляют эффективный инструмент для оперативного разрешения широкого спектра задач, неизбежно возникающих при воплощении сложных проектов, учитывая как внутренние, так и внешние факторы неопределенности. Благодаря использованию модели машинного обучения с подкреплением, основанной на марковском процессе, создается основа для системы поддержки принятия решений. Эта система способна динамически оценивать ход выполнения проекта и формировать адаптивные и точные стратегии с низким уровнем погрешности. В ее структуру могут быть интегрированы различные подмодели, такие как регрессии, классификаторы, кластеризаторы и глубокие нейронные сети.

Полученные результаты в полной мере применимы для формирования эффективных стратегий управления сложными проектами. Доказано, что использование марковских моделей принятия решений в полной мере позволяет нивелировать неопределённость при определении характера закона распределения случайной величины генеральной совокупности данных, необходимых для обучения проекта. Кроме того, дискретность в МППР соответствует характеру формирования и управления сложным проектом, осуществляемым варьированием параметров ε и γ , а в случае, когда проект длится значительное время, возможно использование параметра β , эффективность применения которого подлежит дальнейшим исследованиям.

Список литературы

1. Грессер Лаура, Кенг Ван Лун. Глубокое обучение с подкреплением: теория и практика на языке Python. СПб.: Питер, 2022. 416 с.
2. Аверина Т.А., Половинкина А.И., Шумарин В.В. Управление рисками в условиях инновационного развития организации // Вестник Воронежского государственного технического университета. 2009. Т. 5, № 5. С. 87–89.
3. Баркалов С.А., Бурков В.Н., Порядина В.Л. Механизмы активной экспертизы в задачах комплексного оценивания // Вестник Воронежского государственного технического университета. 2009. Т. 5, № 6. С. 64–66.
4. Белоусов В.Е., Баркалов С.А., Нижегородов К.А. Ресурсно-временной анализ в задачах календарного планирования строительных предприятий // Материалы XVI Всероссийской школы-конференции молодых ученых «Управление большими системами». Тамбов: Изд-во ТГТУ, 2019. Т. 1. С. 98–101.
5. Горелик А.Л., Скрипкин В.А. Методы распознавания. М.: Высшая школа, 2004. 341 с.
6. Галинская А.А. Модульные нейронные сети: обзор современного состояния разработок // Математические машины и системы. 2003. № 3-4. С. 87–102.
7. Эффективное управление организационными и производственными структурами: моногр. / О.В. Логиновский, А.В. Голлай, О.И. Дранко и др. М.: Инфра-М, 2020. 450 с. (Научная мысль). ISBN 978-5-16-016217-1. DOI: 10/12737/1087996
8. Вапник В.Н. Восстановление зависимости по эмпирическим данным. М.: Наука, 1979. 295 с.
9. Алгоритмы: построение и анализ: пер. с англ. / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. 2-е изд. М.: Вильямс, 2005. 1296 с.
10. Белоусов В.Е., Абросимов И.П., Губина О.В. Алгоритм идентификации состояний многоуровневой технической системы с использованием расплывчатых категорий модели представления знаний // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2017. № 3. С. 124–129.
11. Белоусов В.Е., Нижегородов К.И., Соха И.С. Алгоритмы получения упорядоченных правил предпочтения в задачах принятия решений при планировании производственных программ // Управление строительством. 2019. № 1 (14). С. 105–110.
12. Jordan M.I. Attractor dynamics and parallelism in a connectionist sequential machine // The Eighth Annual Conference of the Cognitive Science Society. Amherst, MA, 1986. P. 531–546.
13. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование: учеб. М.: Финансы и статистика, 2001. С. 203–211.
14. Construction project data analysis. URL: <https://www.kaggle.com/code/vadbel66/construction-project-data-analysis> (дата обращения: 09.11.2025).
15. Gymnasium-Robotics is a collection of robotics simulation environments for Reinforcement Learning. URL: <https://robotics.farama.org> (дата обращения: 09.11.2025).

References

1. Gresser Laura, Keng Wang Loong. *Deep learning with a reinforcement: the theory and practice are always on the lips Python*. St. Petersburg: Piter, 2022. 416 p. (In Russ.)
2. Averina T.A., Polovinkina A.I., Shymarin V.V. Management of risks in the conditions of innovative development of organization. *Bulletin of Voronezh state technical university*. 2009;5(5):87–89. (In Russ.)
3. Barkalov S.A., Burkov V.N., Porjadina V.L. Mechanisms of active examination in problems complex estimation. *Bulletin of Voronezh state technical university*. 2009;5(6):64–66. (In Russ.)

4. Belousov V.E., Barkalov S.A., Nizhegorodov K.A. [Resource timing analysis in problems of scheduling of the construction enterprises]. In: *Materials of the XVI All-Russian school conference of young scientists "Management of big systems"*. Tambov: Tambov State Technical University Publ. 2019. Vol. 1. P. 98–101. (In Russ.)
5. Gorelik A.L., Skripkin V.A. *Metody raspoznavaniya* [Recognition methods]. Moscow: Vysshaya shkola, 2004. 341 p. (In Russ.)
6. Galinskaya A.A. [Modular neural networks: review of the current state of developments]. *Mathematical machines and systems*. 2003;(3-4):87–102. (In Russ.)
7. Loginovskiy O.V., Gollay A.V., Dranko O.I., Shestakov A.L., Shinkarev A.A. *The effective management of organizational and production structures. Monograph*. Moscow: Infra-M Publ., 2020. 456 p. (In Russ.) ISBN 978-5-16-016217-1. DOI: 10.12737/1087996
8. Vapnik V.N. *Vosstanovlenie zavisimosti po empiricheskim dannym* [Recovery of dependence according to empirical data]. Moscow: Nauka, 1979. 295 p. (In Russ.)
9. Cormen T., Leiserson Ch., Rivest R., Stein C. *Introduction to Algorithms*. Transl. from Engl. 2nd ed. Moscow: Williams, 2005. 1296 p. (In Russ.)
10. Belousov V.E., Abrosimov I.P., Gubina O.V. [An algorithm of identification of conditions of a multilevel technical system with use of indistinct categories of model of representation of knowledge]. *Proceedings of Voronezh state university. Series: Systems analysis and information technologies*. 2017;(3):124–129. (In Russ.)
11. Belousov V.E., Nizhegorodov K.I., Soha I.S. Algorithms of obtaining the ordered rules of preference in problems of decision-making when planning production programs. *Upravleniye stroitel'stvom*. 2019;1(14):105–110. (In Russ.)
12. Jordan M.I. Attractor dynamics and parallelism in a connectionist sequential machine. In: *The Eighth Annual Conference of the Cognitive Science Society*. Amherst, MA, 1986. P. 531–546.
13. Afanas'ev V.N., Yuzbashev M.M. *Analiz vremennykh ryadov i prognozirovaniye: uchebnyk* [Analysis of time series and forecasting. Textbook]. Moscow: Finansy i statistika, 2001. P. 203–211. (In Russ.)
14. Construction project data analysis. Available at: <https://www.kaggle.com/code/vadbel66/construction-project-data-analysis>. (accessed 09.11.2025).
15. Gymnasium-Robotics is a collection of robotics simulation environments for Reinforcement Learning. Available at: <https://robotics.farama.org>. (accessed 09.11.2025).

Информация об авторах

Баркалов Сергей Алексеевич, д-р техн. наук, проф., заведующий кафедрой управления, Воронежский государственный технический университет, Воронеж, Россия; bsa610@yandex.ru.

Белоусов Алексей Вадимович, аспирант кафедры управления, Воронежский государственный технический университет, Воронеж, Россия; alexbelousov19@yandex.ru.

Серебрякова Елена Анатольевна, канд. экон. наук, доц., доц. кафедры цифровой и отраслевой экономики, Воронежский государственный технический университет, Воронеж, Россия; sea-parish@mail.ru.

Information about the authors

Sergey A. Barkalov, Dr. Sci. (Eng.), Prof., Head of the Department of Management, Voronezh State Technical University, Voronezh, Russia; bsa610@yandex.ru.

Alexey V. Belousov, Postgraduate student of the Department of Management, Voronezh State Technical University, Voronezh, Russia; alexbelousov19@yandex.ru.

Elena A. Serebryakova, Cand. Sci. (Econ.), Ass. Prof., Ass. Prof. of the Department of Digital and Branch Economy, Voronezh State Technical University, Voronezh, Russia; sea-parish@mail.ru.

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article.

The authors declare no conflicts of interests.

Статья поступила в редакцию 11.11.2025

The article was submitted 11.11.2025