

Планирование задач и оптимизация логистических операций в облачных производственных системах с помощью обучения с подкреплением

Д.И.Журавлёв¹, В.В.Поздняков^{2,3}

¹Московский физико-технический институт (национальный исследовательский университет)

²Институт AIRI

³НИУ ВШЭ

В настоящий момент всё большую популярность набирает использование облачного производства для моделирования производственных и бизнес процессов. Его идея заключается в объединении ресурсов и последующем централизованном их распределении для достижения оптимального значения целевой функции затрат [1].

Планирование облачного производства формально можно представить, как решение задачи планирования работы цеха (JSSP), которая является классической задачей оптимизации в области исследования операций и управления производством. Она включает планирование набора заданий, каждое из которых состоит из последовательности операций, на множестве машин с учетом различных ограничений и целей. Более того, она естественно обобщается на задачу гибкого планирования (FJSP), когда существует множество возможностей для выполнения операций на различных станках в различных городах [2][3]. Сейчас для решения задач этого типа разработаны точные методы, но они работают долго, так как класс задач относится к NP-трудным.

Отличительной чертой данной работы является процесс планирования задач в облачном производстве с учетом затрат на логистику, которые не учитывались при решении задачи гибкого планирования. Для достижения этой цели мы предлагаем использовать обучение с подкреплением на основе алгоритма Q-learning. Результаты работы мы сравниваем с жадным решением, точным решением и решением на основе обучения с учителем модели графовой нейронной сети. Во всех методах по набору данных предлагается матрица планирования, в каком городе какую задачу производить. Набор данных включает в себя матрицы допустимых операций, расстояние между городами, время и стоимость выполнения каждой операции в каждом городе.

Для решения поставленной задачи нами была разработана модель обучения с подкреплением на основе «эпсилон-жадной» стратегии и уравнения Беллмана:

$$Q^{new}(S_t, A_t) = (1 - \alpha) \cdot Q(S_t, A_t) + \alpha \cdot (R_{t+1} + \gamma \cdot \min_a Q(S_t, a)), \text{ где}$$

Q – функция ожидаемых вознаграждений, α – скорость обучения ($0 < \alpha \leq 1$), R_{t+1} – стоимость производства при переходе из состояния S_t в S_{t+1} , γ – коэффициент дисконтирования, S_t – состояние, определяемое текущей операцией, субоперацией и городом, A_t – возможные города производства следующей за текущей субоперации.

Были проведены эксперименты на данных Фатахи [4]. Результат работы данной модели при разных гиперпараметрах представлен на рис. 1. Итоговая стоимость запланированного маршрута, полученная разными методами, для сравнения представлена в табл. 1.

Операции, субоперации, города	Точное решение	Жадный алгоритм	Q-learning	GNN
5, 5, 10	7652	12522	12522	12369
10, 10, 10	14200	23082	23082	22574
5, 10, 20	5866	9414	9414	9625

Таблица 1. Сравнение результата работы разных алгоритмов

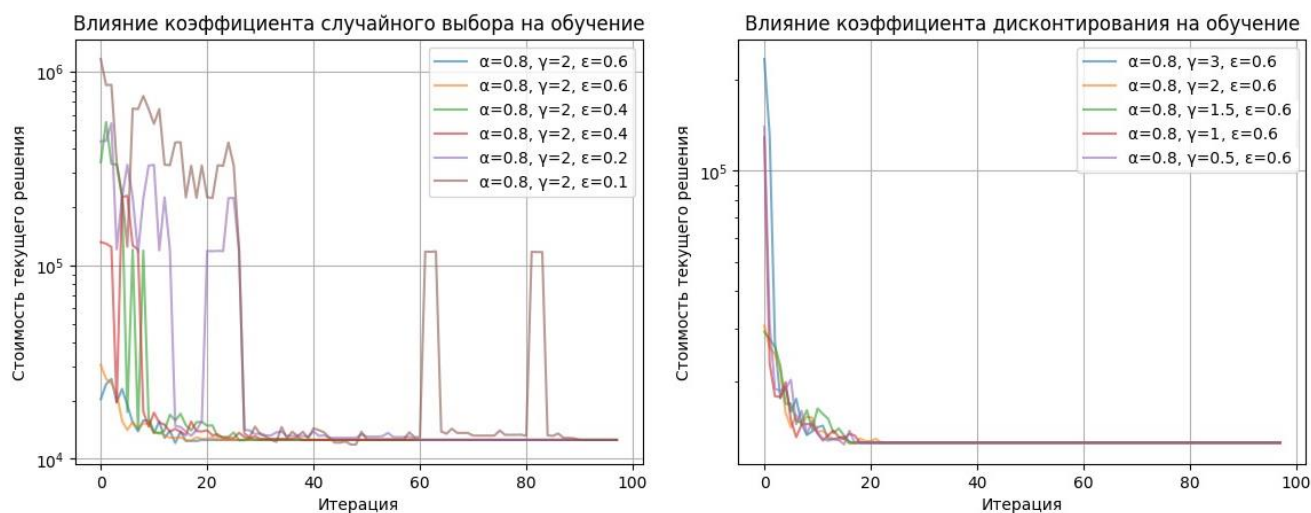


Рисунок 1. Зависимость обучения от гиперпараметров

Для объяснения результата, полученного алгоритмом Q-learning исходя из специфики алгоритма выбирать случайные маршруты, было исследовано пространство возможных решений. Распределение стоимости возможных решений одной из задач планирования представлено на рис. 2. Мы можем заметить, что случайный даже эвристически оптимальный выбор решения с очень низкой вероятностью приводит к субоптимальной стоимости.

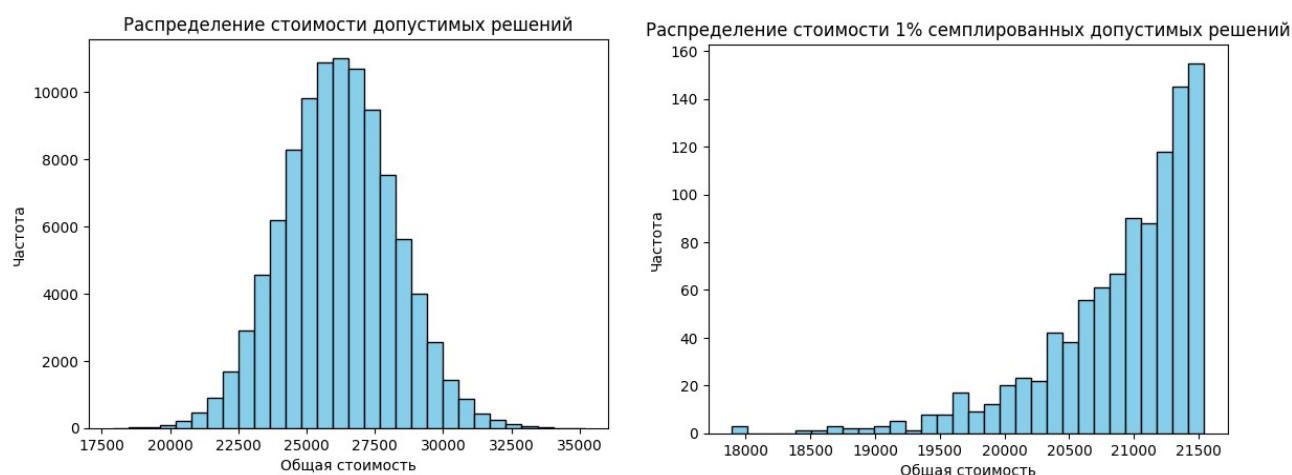


Рисунок 2. Распределение стоимости допустимых решений.

В результате экспериментально мы обнаружили, что Q-learning не дает значительного улучшения по сравнению с жадным алгоритмом из-за экспоненциально большого количества неоптимальных решений, лежащих рядом с найденным решением. Дальнейшим направлением данного исследования может быть применение методов обучения с подкреплением, основанных на алгоритме Actor-critic.

Литература

1. Tarchinskaya, Ekaterina. (2013). Cloud-Based Engineering Design and Manufacturing: State of the Art. IFAC Proceedings Volumes. 46. 335-340. 10.3182/20130619-3-RU-3018.00632.
2. Sadeghi Aghili SA, Fatahi Valilai O, Haji A, Khalilzadeh M. 2021. Dynamic mutual manufacturing and transportation routing service selection for cloud manufacturing with multi-period service-demand matching.
3. Song, Wen & Chen, Xinyang & Li, Qiqiang & Cao, Zhiguang. (2022). Flexible Job Shop Scheduling via Graph Neural Network and Deep Reinforcement Learning. IEEE Transactions on Industrial Informatics. PP. 10.1109/TII.2022.3189725.
4. Delaram, Jalal & Valilai, Omid. (2018). A Mathematical Model for Task Scheduling in Cloud Manufacturing Systems focusing on Global Logistics. Procedia Manufacturing. 17. 387-394. 10.1016/j.promfg.2018.10.061.