М. О. Савоневская, В. В. Столярова, Л. А. Черникова

ИСПОЛЬЗОВАНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ (RL) ДЛЯ ПРЕДОТВРАЩЕНИЯ ТУПИКОВЫХ СИТУАЦИЙ В СИСТЕМАХ С ОБЩИМИ РЕСУРСАМИ

Рассматривается применение обучения с подкреплением (RL) для предотвращения тупиковых ситуаций (deadlock) в системах с общими ресурсами. Предлагается использование Q-обучения в условиях ограниченной наблюдаемости среды, где агенты адаптивно учатся избегать коллизий при доступе к ресурсам.

Введение

Возникновение тупиковых ситуаций в системах с общими ресурсами представляет собой серьезную проблему, которая препятствует их эффективному функционированию. Применение методов обучения с подкреплением (Reinforcement Learning, RL), основным из которых является Q-обучение, позволяет разрабатывать стратегии предотвращения коллизий даже в условиях ограниченной наблюдаемости среды.

I. Тупиковые ситуации в системах с общими ресурсами

Тупиковые ситуации (deadlocks) — это критические состояния, при которых группа процессов взаимно блокируется из-за циклического ожидания недоступных ресурсов, удерживаемых другими процессами [1]. В распределённых системах проблема усугубляется неполной наблюдаемостью и децентрализованным управлением, что требует адаптивных методов (например, Reinforcement Learning).

II. Обучение с подкреплением (RL) как подход для предотвращения тупиков

Обучение с подкреплением (Reinforcement Learning, RL) представляет собой парадигму машинного обучения, в которой агент обучается принимать оптимальные решения через взаимодействие со средой. RL использует систему вознаграждений и штрафов, позволяя агенту самостоятельно находить стратегии, максимизирующие долгосрочную выгоду. Ключевое преимущество RL в данном контексте - способность адаптироваться к изменяющимся условиям без явного перепрограммирования.

III. Особенности избегания коллизий через Q-обучение

Основная задача алгоритма Q-обучения — сформировать оптимальную стратегию поведения, обновляя значения Q-функции на основе опыта, полученного в процессе взаимодействия со средой. Процесс Q-обучения основан на поэтапном пересмотре значений Q-функции. Через итеративное обновление Q-таблицы алгоритм выявляет паттерны, ведущие к deadlock, и вырабатывает стратегию их избегания. В работе он использует только локальную информацию, адаптируясь к изменениям системы без полного перерасчета политики. Каждый раз, когда программа совершает действие и получает обратную связь от среды (в виде награды и перехода в новое состояние), он корректирует свои представления о ценности этого действия [2]. Этот итеративный подход позволяет со временем всё точнее определять, какие решения будут наиболее выгодными в долгосрочной перспективе, при этом избегая тупиковые ситуации.

IV. Концептуальная модель интеграции Q-обучения в системы с общими ресурсами

Предлагается концептуальная модель, в которой агенты, использующие Q-обучение, взаимодействуют с системой с общими ресурсами, получая частичную информацию о ее состоянии и принимая решения, направленные на предотвращение коллизий и тупиков. Модель включает следующие компоненты: агенты (программные сущности, обучающиеся с использованием Q-обучения для принятия решений), среда (все элементы системы, с которыми взаимодействуют агенты), награды (определяются на основе эффективности работы системы, включая показатели предотвращения тупиков).

1. Chen, M. Deadlock-Detection via Reinforcement Learning / M. Chen, L. Rabelo // University of Central Florida – 2017. – Vol. 6. – P. 1-6.
2. Watkins, C. Q-Learning / C. H. Watkins, P. Dayan // Q-Learning – 1992. – P. 281-285.

Савоневская Маргарита Олеговна, студентка ФИТиУ БГУИР, go.to.mychannel745@gmail.com.
Столярова Виолетта Витальевна, студентка ФИТиУ БГУИР, violettastolarova@icloud.com.
Черникова Лолита Александровна, студентка ФИТиУ БГУИР, lolita190511@gmail.com.
Научный руководитель: Хаджинова Наталья Владимировна, старший преподаватель кафедры ИТАС БГУИР, khajynova@bsuir.by.