

УДК 004.896

РАЗРАБОТКА ДЕЦЕНТРАЛИЗОВАННОГО АЛГОРИТМА КООПЕРАТИВНОЙ СБОРКИ ДЕТАЛЕЙ  
ДВУМЯ МАНИПУЛЯЦИОННЫМИ РОБОТАМИ НА ОСНОВЕ МЕТОДА MADDPG

Антипов В.А. (Университет ИТМО)

Научный руководитель – к.т.н. Ведяков А.А.  
(Университет ИТМО)

В докладе рассматривается система из двух артикулированных манипуляторов, осуществляющих сборку детали. На основе метода мультиагентного глубокого градиентного обучения (MADDPG) для рассматриваемой системы синтезирован алгоритм вставки детали в отверстие. Предложенный алгоритм позволяет после централизованного обучения осуществлять децентрализованное управление манипуляторами, увеличивая отказоустойчивость системы.

**Введение.** Кооперативные манипуляционные системы начали получать все большую популярность, в связи с развитием антропоморфной робототехники. Также востребованность обусловлена развитием концепции умных фабрик. Кооперативные робототехнические системы в отличие от одиночных роботов способны выполнять больший спектр задач в виду повышенного числа степеней свободы и увеличенной общей грузоподъемности. Однако, для управления такими системами требуются новые алгоритмы, позволяющие роботам кооперативно и децентрализованно принимать решения, направленные на достижение общих целей. В данной работе рассматривается задача сборки детали, состоящей из двух частей, посредством вставки одной в другую с помощью двух манипуляционных роботов.

**Основная часть.** Алгоритм управления кооперативной системой из двух манипуляционных роботов основывается на методе мультиагентного глубокого градиентного обучения. Используемый метод позволяет после централизованного обучения выполнять задачи децентрализованно. Каждый манипулятор (агент) при расчете сигнала управления (стратегии) основывается на собственных наблюдениях и получает индивидуальную награду. Целью каждого агента является максимизация индивидуального вознаграждения, при этом каждый агент не знает ни реального состояния, ни стратегии других агентов.

В основе метода MADDPG лежит несколько базовых подходов к обучению. Функция ценности состояния оценивается критиком, а сигнал управления рассчитывается актором. При обучении информация о состоянии среды есть у каждого критика, а траектория системы записывается в буфер опыта. Также для случая, когда манипуляторы одинаковы (агенты однородны) можно использовать одного критика, что повышает скорость обучения.

Разрабатываемый алгоритм в качестве входных данных использует положение одной детали относительно другой, оцениваемое на основе информации о положении рабочих инструментов и габаритах деталей, а также силы и моменты, прикладываемые к детали. Рассчитываемый сигнал управления представляет собой желаемую скорость рабочего инструмента для каждого манипулятора. Обучение происходит с применением подхода *sim2real* с последующим дообучением на реальных данных.

**Выводы.** Предложенный алгоритм ставит перед собой цель максимизировать число успешных вставок детали в отверстие и позволяет решать задачу децентрализованной кооперативной сборки детали из нескольких частей. Предложенный алгоритм устойчив к отключению или отказу одного из агентов при условии невырожденности матрицы Якоби манипуляторов.

Антипов В.А. (автор)

Подпись

Ведяков А.А. (научный руководитель)

Подпись