

УДК 004.896

ПРИМЕНЕНИЕ МЕТОДА ПОРОГОВОЙ ФИЛЬТРАЦИИ ДЛЯ ФОРМИРОВАНИЯ ВИРТУАЛЬНЫХ ЦЕЛЕЙ ПРИ
ОБУЧЕНИИ С ПОДКРЕПЛЕНИЕМ В ЗАДАЧЕ МАНИПУЛИРОВАНИЯ ОБЪЕКТОМ

Труфанова А.А. (Университет ИТМО)

Научный руководитель – к.т.н., доцент Ведяков А.А.

(Университет ИТМО)

В работе рассматривается метод обучения с подкреплением, основанный на ретроспективной переигровке, с применением пороговой фильтрации при формировании виртуальных целей для корректировки смещенного распределения опыта в задаче манипулирования объектом.

Введение. При использовании метода ретроспективной переигровки в обучении с подкреплением, может возникать проблема смещенного распределения опыта. К примеру, в задаче толкания кубика в заданное положение, алгоритм может выбрать «пустое» действие (при котором положение кубика не изменится, хотя звенья манипулятора и его конечный инструмент изменят своё положение). Тогда он получит нулевое (но не отрицательное) вознаграждение и может выбрать такую стратегию как лучшую, что будет не совсем корректно. Также в задаче вставки детали, при выборе контактных сил в качестве вектора состояния и фиксированной силы взаимодействия как цели, может возникать смещение из-за возможности получить фиксированную контактную силу при взаимодействии с любой плоскостью. Одним из вариантов решения проблемы смещения является дополнение вектора состояния новыми измеряемыми величинами. В задаче толкания кубика это может быть вектор сил взаимодействия конечного инструмента и объекта взаимодействия, установленного в некотором пределе. В задаче вставки детали это может быть глубина по оси вставки. Однако, такой подход не всегда применим в силу отсутствия возможности добавления датчика измерения или сложности подбора таких измерений, которые позволяли бы решить описанную проблему.

Основная часть. Одним из методов, описанным в магистерской диссертации Vinuamin Manela, является добавление фильтрации для установления порога повторяемости виртуальной цели из текущего положения. Для этого проверяется буфер виртуальных целей для текущего состояния, и, если таковая уже была достигнута ранее, не добавляется повторно, чтобы избежать смещения. На основании вышесказанного применяется метод пороговой фильтрации при формировании ретроспективных целей. В качестве вектора состояния используются положения и скорости звеньев артикулированного робота-манипулятора UR5e, конечного инструмента, положение объекта манипулирования. В качестве действий рассматриваются желаемые скорости звеньев.

Выводы. В виртуальной среде проведено сравнение применения метода обучения с подкреплением с использованием ретроспективной переигровки в классической реализации и с использованием пороговой фильтрации при формировании виртуальных целей. Было установлено, что решение проблемы смещения распределения опыта добавлением фильтрации позволяет добиться более высокой скорости обучения на меньшем количестве итераций. В дальнейшем планируется апробация полученных результатов на реальном роботе.

Труфанова А.А. (автор)

Подпись

Ведяков А.А. (научный руководитель)

Подпись