ЛОКАЛЬНОЕ ПЛАНИРОВАНИЕ НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В СРЕДЕ С ДИНАМИЧЕСКИМИ ПРЕПЯТСТВИЯМИ: ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ

Б. Ангуло (*brian.angulo@phystech.edu*)
Московский физико-технический институт, Москва

Аннотация. Локальное планирование траектории для автомобилей в среде со статическими и динамическими препятствиями до сих поре остается нетривиальной задачей. С одной стороны при планировании необходимо обеспечивать учет динамики автомобиля, что сама по себе уже является трудоемкой задачей, а с другой стороны необходимо избежать столкновения автомобиля со статическими и динамическими препятствиями. Для решения данной задачи в работе предлагается использовать метод на основе обучения с подкреплением.

Ключевые слова: планирование траектории, обучение с подкреплением, объезд препятствий, автономные автомобили.

Введение

Кинодинамическое планирование движения в среде с динамическими препятствиями широко применяется в системах автономных автомобилей. Учет динамику автомобиля при планировании является нетривиальной задачей. Однако, динамические препятствия накладывают еще больше ограничений на задачу локального планирования из-за непредсказуемых будущих их траекторий движений.

Существуют два общих подхода к кинодинамическому планированию: методы планирования на основе примитивов движения [Likhachev, M. 2009] и на основе сэмплинга [Karaman, S. 2010]. Данные методы растут некоторое дерево, где от родителя строятся динамически выполнимые траектории, которые затем проверяются на столкновение и добавляются в дерево только те траектории, которые свободны от столкновения. В отличие от этих методов, методы на основе обучения с подкреплениям в используют сенсорную информацию автомобиля, позволяющую агенту напрямую избежать столкновения [Faust, A. 2018]. Это в свою очередь позволяет строить более безопасные и приближенные к реальности траектории. Предварительные результаты показывают превосходство данного метода перед классическим методом (ExpStab) [Astolfi, A. 1999].

1 Постановка задачи

В качестве упрощенной модели рассмотрим велосипедную модель автомобиля. Состояние автомобиля в каждый момент времени определяется тройкой $\boldsymbol{s}(t) = (x, y, \theta)$, где x, y – координаты передней оси колес автомобиля, ϑ - угол поворота главной оси автомобиля. Управление задается двойкой $u(t) = (v, \gamma)$: линейная скорость и угол поворота колес автомобиля. Обозначим через $X_{free}(t)$ свободное множество состояний в момент t, а через $D_{obs}(t)$ – множество динамических препятствий. Пусть даны два состояния: начальное *start* и конечное *goal*. Требуется найти управления $u(t)$ для перехода из *start* в *goal* так, что каждое промежуточное состояние $s(t) \in X_{free}(t)$ и $s(t) \notin D_{obs}(t)$, т.е. состояние автомобиля должно принадлежать свободному пространству, а также, избежать столкновения с динамическими препятствиями.

2 Метод решения

Предлагаемый метод состоит в применении метода обучения с подкреплением для генерации управлений автомобиля. Агент обучается предсказывать действия $a(t)$ на основе его наблюдении $o(t)$, которое включает в себя сенсорную информацию автомобиля. На каждом временном шаге обучения агент получает наблюдение $o(t)$, затем выполняет действие$a(t)$ и получает некоторую награду $r(t)$.

Конфигурационное пространство агента $\mathrm{C} = (x, y, \theta, v, \gamma)$ - это набор всевозможных его состояний. В качестве наблюдения агента $o(t)$ используется кортеж, состоящий из элементов из конфигурационного пространства $o(t)=(\Delta x, \Delta y, \Delta\vartheta, \Delta\upsilon, \Delta\gamma, \vartheta, \upsilon, \gamma)$ вместе с показаниями псевдолидара, которые излучаются из передней оси автомобиля и покрывают 180° относительно его главной оси с шагом 10°. Символ Δ обозначает разность между значением соответствующего параметра в конечном состоянии и параметра в текущем состоянии.

Действия агента описываются двойкой $a(t) = (a, \omega)$, гдеa- линейное ускорение автомобиля, а ω - угловая скорость. Управления автомобиля $u(t)$ можем получить от действий нашего агента $a(t)$ с помощью следующих преобразований: $v = v_0 + a \cdot t, \gamma = \gamma_0 + \omega \cdot t$.

Награда $r = \left(r_{goal}, r_{timeStep}, r_{backward}, r_{field}, r_{collision}\right)$ состоит из следующих слагаемых, где r_{goal} – награда за достижение цели, $r_{timeStep}$ – штраф за каждый шаг обучения, $r_{backward}$ – штраф за движение задним ходом, r_{field} – плотная награда за приближение к цели, $r_{collision}$ – штраф за столкновение с препятствием.

Предлагаемый подход состоит из трех этапов обучения агента с возрастающей сложностью (curriculum learning): обучение агента в среде без

препятствий, в среде со статическими препятствиями, и в среде со статическими и динамическими препятствиями. На каждом этапе создаются обучающая и валидационная выборки в рассматриваемой среде. Каждая выборка состоит из сгенерированных случайных заданий, где каждое задание представляет с собой кортеж из двух элементов из конфигурационного пространства *start* и *goal*. Переход на следующий этап обучения определяется 90% порогом успешности на валидационной выборке.

Итак, агент генерирует последовательность действий $\big(a(t), \dots, a(t+T)\big)$ для перехода автомобиля из одного состояния в другое, обеспечивая кинодинамическую выполнимость траектории и избежание столкновения со статическими и динамическими препятствиями.

3 Результаты экспериментальных исследований

Исследование проводилось на основе алгоритма сэмплирования RRT как и в предлагаемом методе [Chiang, H. 2019], где в качестве алгоритма локального планирования был взят наш обучаемый алгоритм. Эксперименты проводились на двух картах парковки размера 60x100 м. Были определены десять заданий разной сложности, где начальных состояний всего десять и одно конечное состояние. Применение нашего агента в среднем уменьшает время достижения цели автомобиля примерно в 30% по сравнению с классическим методом (ExpStab).

Список литературы

[Likhachev, M. 2009] Planning long dynamically feasible maneuvers for autonomous vehicles. The International Journal of Robotics Research, (2009): 933-945.

[Karaman, S. 2010] Optimal kinodynamic motion planning using incremental sampling-based methods. IEEE Conference on Decision and Control, (2010): 7681-7687.

[Astolfi, A. 1999] Exponential Stabilization of a Wheeled Mobile Robot Via Discontinuous Control, Journal of Dynamic Systems, Measurement, and Control, 121(1): 121–126.

[Faust, A. 2018] PRM-RL: PRM-RL: Long-range Robotic Navigation Tasks by Combining Reinforcement Learning and Sampling-Based Planning. IEEE International Conference on Robotics and Automation (ICRA), 2018: 5113-5120.

[Chiang, H. 2019] RL-RRT: Kinodynamic motion planning via learning reachability estimators from RL policies. IEEE Robotics and Automation Letters (2019): 4298-4305.