

ПОДХОД МАШИННОГО ОБУЧЕНИЯ К ФОРМИРОВАНИЮ ФУНКЦИИ УПРАВЛЕНИЯ ГРУППОЙ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

М.С. Товарнов

Московский государственный технический университет им. Н.Э. Баумана
Россия, 105005, Москва, 2-ая Бауманская ул., 5с1
E-mail: mtovarnov@mail.ru

Н.В. Быков

Российский университет транспорта (РУТ-МИИТ)
Россия, 127994, Москва, Образцова ул., 9с9
E-mail: nik.bkv@gmail.com

Ключевые слова: алгоритмы управления, беспилотные летательные аппараты, роевое управление, обучение с подкреплением.

Аннотация: Предложен новый подход к формированию функции управления роем беспилотных летательных аппаратов (БПЛА). Модель полета групп БПЛА составлена в трехмерной постановке, где индивидуальные траектории полета агентов представлены при помощи кривых Безье различных порядков, а фактическая динамика БПЛА рассчитывается с учетом большинства аэродинамических и маневренных характеристик БПЛА. Кривые Безье формируются для каждого БПЛА индивидуально интеллектуальным агентом, который был обучен при помощи методов машинного обучения с подкреплением. Предложенный подход к формированию роевого управления БПЛА может быть применен в более сложных постановках, что позволит использовать полученные функции управления на реальных БПЛА.

1. Введение

Исследования в области формирования роевого и группового управления начали проводиться относительно недавно. И наибольшее развитие таких технологий сейчас происходит в сфере БПЛА. Основные направления исследований направлены на решение задач управления порядком роя [1-3]. Задача управления порядком включает в себя две цели: с одной стороны, все особи в рое должны следовать некоторому общему эталонному пути, а с другой стороны, особи должны сохранять приемлемое положение относительно друг друга. В литературе накопилось некоторое количество эвристических подходов к математическому описанию структуры роя и связей между его элементами [3-6]. Существующие подходы позволяют реализовывать различные системы управления роем, сохраняющие устойчивость в условиях зашумленной и неполной информации о состоянии и положении отдельных особей и препятствий. Они показывают хорошие результаты в задачах, где хорошо определено оптимальное поведение роя в целом. К таким задачам можно отнести: проведения аэрофотосъемки, видеосъемки, мониторинга природных ресурсов и окружающей среды [7], контроля

транспортных магистралей, охраны границ и «поиска и спасания» [8]. Также при решении задач роевого управления используются элементы машинного обучения, но только в качестве замены или дополнения составных элементов существующих алгоритмов [8-11].

Что касается управления БПЛА на индивидуальном уровне, то уже сформировались достаточно эффективные подходы к синтезу систем управления. За непосредственный контроль органов управления БПЛА отвечает контур стабилизации, обеспечивающий заданное положение и ориентацию БПЛА в пространстве для отработки команд управления. Задача стабилизации БПЛА [12, 13] и отработки команд управления может быть решена отдельно от задачи формирования таких команд. Сами команды управления могут формироваться вручную при управляемом полете, или алгоритмом управления при автономном полете. При полете по заранее определенной траектории хорошо зарекомендовал себя подход, где такие траектории представляются в виде полинома – кривой Безье. Такая кривая представляет собой упрощенное представление реальной траектории, однако позволяет учитывать достаточное количество требований и ограничений реальной траектории. Формирование приемлемых траекторий (кривой Безье) является оптимизационной задачей с ограничениями и может вестись различными методами: генетическими и эволюционными алгоритмами [14], алгоритмами машинного обучения [9, 15] и в режиме реального времени [16].

В данной работе предлагается новый подход к формированию функции управления роем БПЛА. Модель полета групп БПЛА составлена в трехмерной постановке, где индивидуальные траектории полета агентов представлены при помощи кривых Безье различных порядков, а фактическая динамика БПЛА рассчитывается с учетом большинства аэродинамических и маневренных характеристик БПЛА. Кривые Безье формируются для каждого БПЛА индивидуально интеллектуальным агентом, который был обучен при помощи методов RL. Предложенный подход к формированию роевого управления БПЛА может быть применен в более сложных постановках, учитывающих ограничения бортовой аппаратуры БПЛА, зашумленность и неполноту информации о положении других БПЛА и препятствий.

2. Сценарии моделирования

В работе моделировался сценарий перехвата группы вражеских БПЛА роем БПЛА-перехватчиков. Количество вражеских БПЛА – 3 шт., перехватчиков – 5 шт. Вражеские БПЛА летят по заранее рассчитанным траекториям и не меняют их на протяжении моделирования. Перехватчиками индивидуально управляет интеллектуальный агент: в определенные моменты времени (один раз в секунду) он задает новую траекторию кривой Безье третьего порядка. Все БПЛА, участвующие в моделировании, имеют одинаковые параметры характерные для FPV-дронов квадрокоптерного типа с размером пропеллера 6 дюймов.

В начальный момент времени БПЛА перехватчики находятся на уровне земли в кругу с центром в начале координат и радиусом 50 м и имеют нулевые скорости. вражеские БПЛА появляются на удалении от 1000 до 1100 м на высоте от 90 до 100 м и двигаются по случайным траекториям со скоростями от 60 до 90 км/ч, проходящим на удалении до 50 м от центра координат. Если вражеский БПЛА долетает до конца своей траектории, то считается, что он выполнил свою задачу и больше не участвует в текущем моделировании. Если расстояние между двумя любыми БПЛА меньше 3 м, то считается, что произошло столкновение, и оба БПЛА больше не участвуют в текущем моделировании. Если перехватчик приблизился к вражескому БПЛА на это расстояние,

то засчитывается удачный перехват. Моделирование продолжается, пока существуют БПЛА перехватчики.

3. Результаты обучения

Обучение происходило в два этапа. На первом этапе оценивались эффективности алгоритмов, расчеты проводилось по 2 млн. шагов для каждого алгоритма, все параметры обучения задавались стандартными значениями, которые определялись во фреймворке. На втором проводился подбор оптимальных параметров для самого эффективного алгоритма – TQC и проводилось более длительное обучение (450 млн. шагов).

На рис. 1 показан прогресс обучения на первом этапе: отслеживался показатель средней награды за эпизод. График на рисунке сглажен, чтобы нагляднее показать линии трендов. Сглаживание при помощи функции экспоненциальной скользящей средней (ЕМА) с коэффициентом сглаживания 0.99.

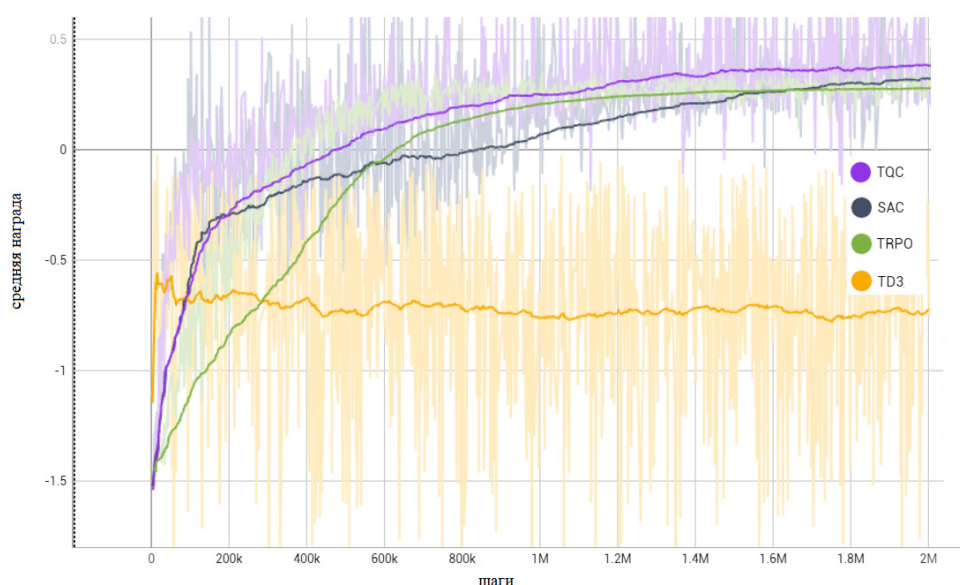


Рис. 1. Прогресс обучения на первом этапе: средняя награда за эпизод.

Из графиков видно, что алгоритмы TQC, SAC и TRPO в принципе подходят для использования, т.к. средняя награда за эпизод становится больше нуля, что свидетельствует о том, что агент научился строить траектории, которые удовлетворяют ограничениям маневренности.

Для второго этапа был выбран алгоритм, показавший лучшие предварительные результаты – алгоритм TQC. Была проведена оптимизация архитектуры нейронной сети агента и его «критика» (нейронная сеть, используемая алгоритмом при обучении). Нейронная сеть представляет собой двухслойный персептрон с количеством нейронов на первом слое 2048, на втором – 1024. Функция активации ELU.

Обучение длилось 450 млн. шагов. Результаты приведены на рис. 2. Как видно из графиков, в процессе обучения агенты научились перехватывать 2/3 вражеских БПЛА. Это связано с тем, что во время обучения интеллектуальный агент имеет некоторое стохастическое поведение.

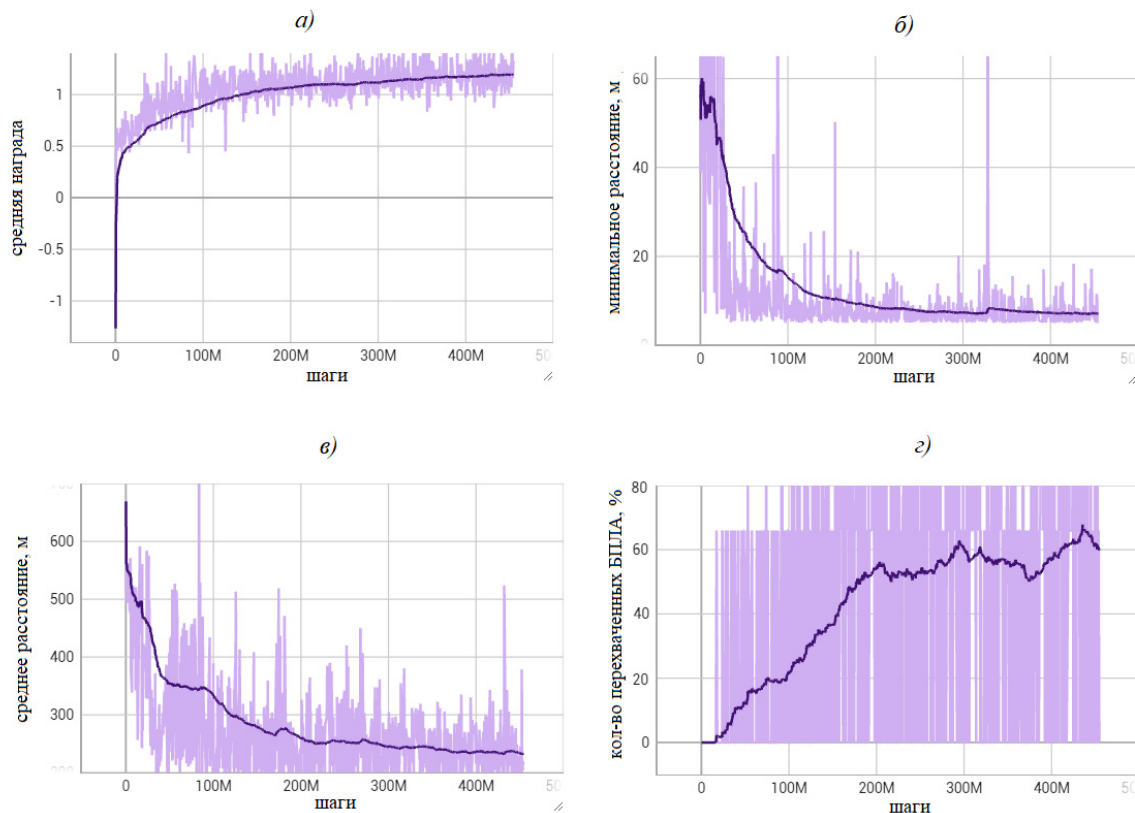


Рис. 2. Прогресс обучения с использованием алгоритма TQC: графики средней награды за эпизод (а), минимального расстояния между перехватчиками и вражескими БПЛА (б), рассчитанного в п.7 алгоритма функции награды, среднего фактического расстояния между перехватчиками и вражескими БПЛА (в) и процент перехваченных БПЛА (г).

Если в течение обучения отдельно проводить по несколько эпизодов в детерминированной постановке, то можно оценить реальную эффективность обучения. На рис. 3 представлен именно такой график.

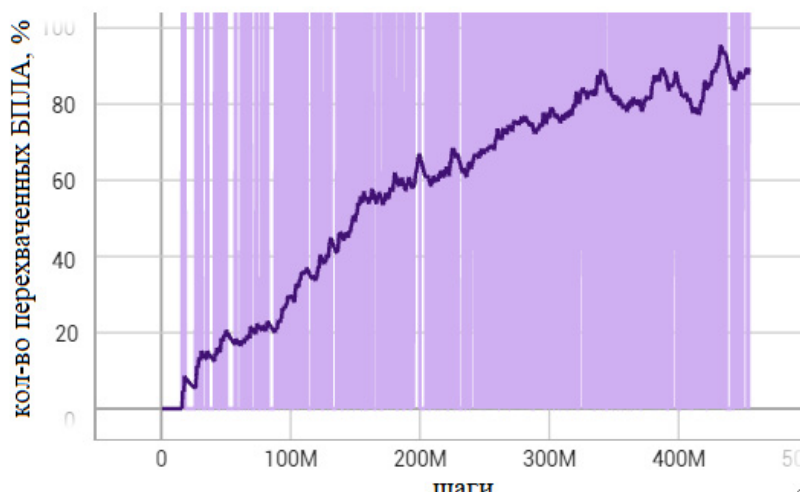


Рис. 3. Доля перехваченных БПЛА при детерминированном агенте TQC.

4. Заключение

Задача перехвата группы БПЛА роем БПЛА-перехватчиков была решена впервые в трехмерной постановке. Была сделана новая имитационная модель полета групп БПЛА в трехмерной постановке, где индивидуальные траектории полета агентов планируются при помощи кривых Безье различных порядков, а фактическая динамика БПЛА рассчитывается с учетом большинства аэродинамических и маневренных характеристик БПЛА. Также была предложена новая функция оценки планируемой траектории, которая учитывает информацию о положениях соседних агентов в рое и их текущих планируемых траекториях. Было показано, что имитационная модель и функция оценки могут быть использованы в качестве виртуальной для использования современных алгоритмов обучения с подкреплением. Было проведено обучение агента-нейросети. способного асинхронно и строить допустимые траектории индивидуально для каждого БПЛА в рое, цель которого является перехват группы БПЛА. Было показано, что для решения задачи в такой постановке подходят современные алгоритмы обучения с. Предложенный подход к формированию роевого управления БПЛА может быть применен в более сложных постановках, учитывающих ограничения бортовой аппаратуры БПЛА, зашумленность и неполноту информации о положении других БПЛА и препятствий.

Исследование поддержано грантом РФФИ, проект № 19-29-06090 мк.

Список литературы

1. Oh K.K., Park M.C., Ahn H.S. A survey of multi-agent formation control // *Automatica*. 2015. Vol. 53. P. 424-440.
2. Duan H., et al. Hybrid particle swarm optimization and genetic algorithm for multi-UAV formation reconfiguration // *IEEE Comput. Intell. Mag. IEEE*. 2013. Vol. 8, No. 3. P. 16-27.
3. Bandala A.A., Vicerra R.R.P., Dadios E.P. Formation stabilization algorithm for swarm tracking in unmanned aerial vehicle (UAV) quadrotors // *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*. 2015.
4. Guo J., Wang L., Wang X. A Group Maintenance Method of Drone Swarm Considering System Mission Reliability // *Drones*. 2022. Vol. 6, № 10.
5. Parhi D.R., Pothal J.K., Singh M.K. Navigation of multiple mobile robots using swarm intelligence // 2009 World Congr. Nat. Biol. Inspired Comput. NABIC 2009 – Proc. IEEE, 2009. P. 1145-1149.
6. Alvissalim M.S., et al. Swarm quadrotor robots for telecommunication network coverage area expansion in disaster area // *Proc. SICE Annu. Conf. IEEE*, 2012. P. 2256-2261.
7. Albani D., et al. Monitoring and mapping with robot swarms for agricultural applications // 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). 2017. P. 1-6.
8. Khalil H., et al. A UAV-Swarm-Communication Model Using a Machine-Learning Approach for Search-and-Rescue Applications // *Drones*. 2022. Vol. 6, No. 12.
9. Yu X., Luo W. Reinforcement learning-based multi-strategy cuckoo search algorithm for 3D UAV path planning // *Expert Syst. Appl.* 2023. Vol. 223. P. 119910.
10. Hodge V.J., Hawkins R., Alexander R. Deep reinforcement learning for drone navigation using sensor data // *Neural Comput. Appl.* 2021. Vol. 33, No. 6. P. 2015-2033.
11. Azar A.T., et al. Drone deep reinforcement learning: A review // *Electron*. 2021. Vol. 10, No. 9. P. 1-30.
12. Lysukho G.V., Maslennikov A.L. Quadcopter: dynamics and control // *Politech. student J.* 2020.No. 46.
13. Shahida Khatoon, Mohammad Shahid I. Quadrotor Control Using PID Controller // *Emerg. Trends Electr. Electron. Eng.* 2014. July.
14. MacHaret D.G., Neto A.A., Campos M.F.M. Feasible UAV path planning using genetic algorithms and Bézier curves // *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*. 2010. Vol. 6404. LNAI. P. 223-232.
15. De Souza C., et al. Decentralized Multi-Agent Pursuit using Deep Reinforcement Learning // *IEEE Robot. Autom. Lett.* 2021. Vol. 6, No. 3. P. 4552-4559.
16. Sabetghadam B., Cunha R., Pascoal A. Real-time trajectory generation for multiple drones using bézier curves // *IFAC-PapersOnLine*. 2020. Vol. 53. P. 9276-9281.