

ГАРАНТИРОВАННАЯ БЕЗОПАСНОСТЬ В ГИБРИДНЫХ СИСТЕМАХ УПРАВЛЕНИЯ: СОЧЕТАНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ И БАРЬЕРНЫХ МЕТОДОВ

Исхаков К. Ф.¹, Свердлов Д. М.¹

Научный руководитель – канд. физ-мат. наук, доцент Бойцев А. А.¹

¹Университет ИТМО

Isxakov2@yandex.ru

Введение

Современные методы обучения с подкреплением демонстрируют высокую эффективность при управлении динамическими системами в непрерывных пространствах состояний и управлений, однако не обеспечивают формальных гарантий безопасности [1]. В связи с этим активно развиваются гибридные подходы, сочетающие обученные политики с методами оптимального управления, в частности с Control Barrier Functions (CBF), предназначенными для обеспечения инвариантности безопасного множества [2]. Большинство существующих работ опирается на локальные дискретные условия безопасности, гарантирующие отсутствие столкновений лишь на одном временном шаге [2,4,5]. Целью данного исследования является анализ достаточности таких условий в контексте онлайн-управления автономной гоночной машинкой формата F1Tenth [3].

Основная часть

В работе рассмотрен гибридный алгоритм, включающий предобученную политику Soft Actor-Critic и онлайн CBF-фильтр, построенный для нелинейной дискретной динамики автомобиля. Проведён анализ строгой дискретной формулировки CBF-условия, обеспечивающего безопасность перехода от шага t к шагу $t+1$ при существовании допустимого управления. На основе вычислительных экспериментов показано, что выполнение данного условия не гарантирует существование безопасного управления на следующем шаге времени. Установлено, что система может войти в состояние, из которого дальнейшее сохранение безопасности невозможно, несмотря на наличие альтернативных управлений, обеспечивающих безопасность при многошаговом рассмотрении.

Выводы

Полученные результаты демонстрируют принципиальное ограничение локальных CBF-гарантий и показывают, что одношаговая инвариантность безопасного множества не эквивалентна многошаговой безопасности. Работа указывает на необходимость разработки методов, учитывающих многократный горизонт управления при интеграции обучения с подкреплением и оптимального управления.

Литература

1. Haarnoja T., Zhou A., Abbeel P., Levine S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor // Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden, 2018. Vol. 80. P. 1861–1870.
2. Ames A. D., Xu X., Grizzle J. W., Tabuada P. Control barrier function based quadratic programs for safety-critical systems // IEEE Transactions on Automatic Control. 2017. Vol. 62, no. 8. P. 3861–3876.

3. Berducci L., Turrini A., Giunchiglia F. Hierarchical potential-based reward shaping from task specifications [Электронный ресурс] // arXiv. 2021. Режим доступа: <https://arxiv.org/abs/2110.02792> (дата обращения: 19.02.2026).
4. Bertsekas D. P. Model predictive control and reinforcement learning: A unified framework based on dynamic programming // IFAC-PapersOnLine. 2024. Vol. 58, no. 18. P. 363–383.
5. Yang L., Chen X., Zhang Y. CBF-RL: Safety filtering reinforcement learning in training with control barrier functions [Электронный ресурс] // arXiv. 2025. Режим доступа: <https://arxiv.org/abs/2510.14959> (дата обращения: 19.02.2026).