

Заключение. Анализ результатов моделирования позволяет выделить схему фазометра на компараторах с транзисторными ключами. Эта схема имеет максимальную чувствительность и линейную прямо пропорциональную зависимость выходного напряжения от разности фаз входных сигналов. Такой сигнал на выходе схемы может использоваться для управления ключами коммутации конденсаторных батарей с минимальной корректировкой.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Железко, Ю.С. Компенсация реактивной мощности в сложных электрических системах. – М.: Энергоиздат, 1981. – 200 с.
2. Сульжиц, А. Тарифные системы на электрическую энергию для населения / А. Сульжиц, А. Сульжиц // Энергетика и ТЭК. – № 12(69). – 2008. – Режим доступа: <http://www.energetika.by/arch>.

3. Ярошевич, А.В. Схема компенсации реактивной мощности в квартирных электрических сетях // Вестник Брестского государственного технического университета – № 5(71): Физика, математика, информатика. – Брест: БрГТУ, 2011. – С. 66–67.
4. Режим доступа: www.ensytech.com.
5. Клаассен, К.Б. Основы измерений. Электронные методы и приборы в измерительной технике. – Москва: Постмаркет, 2000. – 352 с.
6. Бутев, В. Электронный фазометр // Радио – 1990. – № 5. – С. 56.
7. Горошков, Б.И. Радиоэлектронные устройства: справочник. – М.: Радио и связь, 1984. – 400 с.
8. Гончаренко, А. Фазометр на микросхемах // Радио. – 1984. – № 12 – С. 29.

Материал поступил в редакцию 28.06.13

YARASHEVICH A.V. Modeling of phase detector for reactive power regulator

The task of developing a compensator of reactive power for residential consumers of electricity is solved by application of the analog circuit of the transmitter. The first stage of calculation capacity balancing capacitors is to determine the phase difference of voltage and current consumption. Circuit implementation phase detector most simple method based on transforming an interval of time in voltage. Consider three schemes of detectors:

- phase detector on limiters;
- phase meter on comparators with transistor keys;
- phase meter on comparators with *D* – trigger.

Circuit simulation with service *Electronics Workbench* was made at various values of input voltage in the range $\Delta\phi$ from 0 to 100 angular degrees. Analysis of results of modeling allowed us to highlight the scheme of *phase meter on comparators with transistor keys*. This scheme has a maximum sensitivity and linear directly proportionate the dependence of the output voltage from the phase difference of the input signals.

УДК 004.853

Дёмин В.В., Кабыш А.С., Головки В.А., Stetter R.

АЛГОРИТМЫ ПОДКРЕПЛЯЮЩЕГО ОБУЧЕНИЯ ДЛЯ ЭНЕРГОЭФФЕКТИВНОГО УПРАВЛЕНИЯ МНОГОКОЛЕСНЫМИ ПРОИЗВОДСТВЕННЫМИ РОБОТАМИ

Введение. Эффективное управление мобильным роботом на производстве позволяет экономить множество ресурсов: время автономной работы, возможность перевозки более тяжелых грузов на более длинные расстояния, маневренность при перевозке габаритных грузов в ограниченном пространстве, прокладывание оптимальной траектории. Важными задачами являются **оптимизация энергопотребления и оптимальное планирование траектории**. Задача энергосбережения в общем случае должна обеспечиваться подсистемами управления. Например, проблема энергопотребления моторов решается при их проектировании [1]. Подсистема управления не сможет влиять на КПД моторов, но должна обладать стратегией эффективного управления (оптимальная скорость мотора, оптимальный разгон, плавная функция торможения).

Оптимальное планирование траектории, как правило, реализуется на уровне подсистемы планирования [2]. Такая подсистема строит траекторию до цели и разбивает ее на части, которые могут быть представлены в виде кривых определенного радиуса и прямолинейных промежутков. Система управления роботом позволяет передвигаться (по возможности без остановок) по этой траектории, затрачивая как можно меньше энергии батарей.

Ключевым вкладом этой статьи является предложенная модель координации колесных модулей на основе виртуального лидера и обучения с подкреплением для эффективного управления многоколесным роботом. По сравнению с аналогичными подходами кругового движения [3], предложенный позволяет повысить эффективность потребления энергии роботом. Представленная модель решает проблему кругового движения платформы относительно центра разворота, даже если он динамически меняет свое положение. Подход требует лишь информации о положении агентов относительно центра платформы.

Многоколесная производственная платформа. Задача перевозки тяжелых грузов до сих пор остается актуальной на современных производствах. Для перевозки больших грузов все чаще приме-

няются автономные мобильные грузовые платформы. Одна из таких платформ – производственный грузовой робот, разработанный в лаборатории университета Равенсбург-Вайнгартена [4]. Фотография робота представлена на рис. 1а. Основные характеристики платформы: размер платформы 1200 см на 800 см, максимальная грузоподъемность 500 кг при комплектации 4-мя модулями, ёмкость аккумуляторов 52Ah, минимальная скорость 1 м/с, независимое управление каждым модулем.

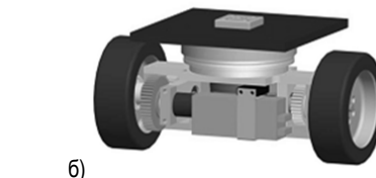


Рис. 1. а) производственная грузовая мобильная платформа; б) инновационный модуль

Платформа построена на базе инновационных модулей [4]. Такой модуль (рис. 1б) состоит из двух колес, приводимых в движение двумя независимыми моторами, и имеет дифференциальную схему управления. К платформе такие модули подсоединены подшипником (рис. 2), что позволяет им поворачиваться относительно платформы на любой угол.

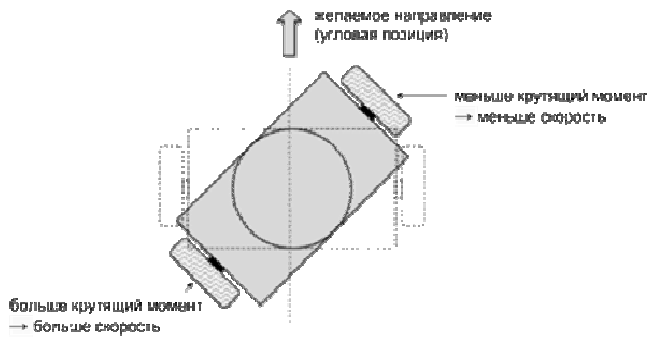


Рис. 2. Маневренность моторного модуля

Представленная платформа использует четыре модуля, но так же возможно собрать платформу и с большим количеством модулей.

Проблема управления многоколесным роботом. Классическим подходом для управления колесными платформами является решение обратной кинематической задачи. Такой подход хорошо работает для распространенных схем управления роботом: дифференциальная схема, автомобильная схема, велосипедная схема. Но в представленной производственной платформе все четыре модуля управляются независимо, что требует обобщения всех схем управления. В реальности же схема управления рассчитана только для симметричного управления [4]. Другой проблемой является необходимость повторных расчетов при изменении положения или количества модулей платформы.

Исходя из вышеизложенного, отметим недостатки существующей системы управления:

- ограниченное управление, и как следствие, уменьшение маневренности платформы;
- необходимость перерасчета схемы управления при реконфигурации модулей платформы.

Для решения задачи планирования подсистема управления должна быть способна повторить траектории, сгенерированные в рамках этой задачи. Для построения траектории платформы планировщик разбивает кратчайший путь на отрезки кривых с разным радиусом и прямым. На рисунке 3 изображен пример разбиения траектории на 4 отрезка:

1. Радиус разворота R_1 , центр разворота (X_1, Y_1) .
2. Прямолинейный отрезок.
3. Радиус разворота R_2 , центр разворота (X_2, Y_2) .
4. Прямолинейный отрезок.

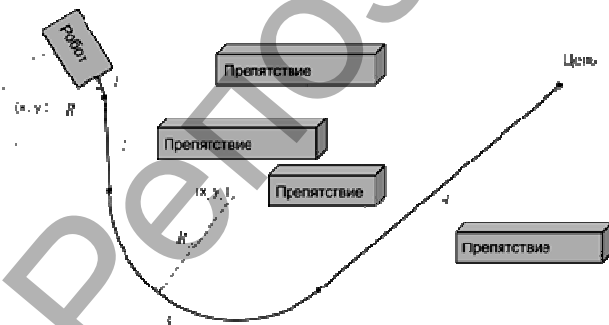


Рис. 3. Пример разбиения траектории на отрезки

В рамках данной задачи будем называть центр разворота робота маяком, представляющим точку в пространстве с двумерными координатами. Движение робота заключается в движении вокруг заданного маяка. Прямолинейный отрезок можно представить в виде маяка с радиусом ∞ .

Декомпозиция робота на колесные модули-агенты. Используя утверждение о том, что каждый модуль платформы управляется независимо, произведем декомпозицию платформы на модули-агенты, где каждый модуль платформы будет представлен агентом управления. Агенты располагаются в двумерной среде с привязкой к маяку, как показано на рис. 4. Местоположение маяка определяется координатами (X_b, Y_b) . Радиус разворота ρ – это расстояние от центра модуля до маяка. Ошибка угла поворота вычисляется как разность между направлением робота и желаемым направлением.

В модели окружающая среда предоставляет всю необходимую информацию об агенте и соответствующего ему центре разворота. В реальной платформе подсистема навигации для определения ориентации модуля использует данные с одометров и датчика угла поворота модуля.

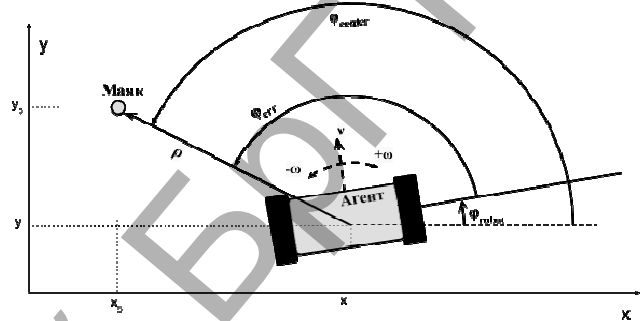


Рис. 4. Состояние агента по отношению к маяку

Состояние модуля-агента представлено в табл. 1.

Таблица 1. Состояние модуля-агента

№	Состояние	Значение
1	Положение робота по X, x	Координата, м
2	Положение робота по Y, y	Координата, м
3	Центр маяка по X, X _b	Координата, м
4	Центр маяка по Y, Y _b	Координата, м
5	Ориентация робота, φ_{robot}	Радiani, $-\pi < \varphi_{robot} \leq \pi$
6	Расположение маяка, φ_{center}	Радiani, $-\pi < \varphi_{center} \leq \pi$
7	Радиус разворота, ρ	Вещественное число, м

Множества действий агента-модуля $A = \{A_w, A_v\}$ связаны с управлением угловой $A_w = \{\omega_+, \omega_-, \emptyset\}$ и линейной скоростью $A_v = \{v_+, v_-, \emptyset\}$. Значения действий приведены в табл. 2.

Таблица 2. Действия модуля-агента

№	Действия робота	Value
1	Увеличить линейную скорость, v+	+0,1 м/с
2	Уменьшить линейную скорость, v-	-0,1 м/с
3	Увеличить скорость поворота, $\omega+$	+0,1 рад/с
4	Уменьшить скорость поворота, $\omega-$	-0,1 рад/с
5	Ничего не делать, \emptyset	+0 м/с, +0 рад/с

Многоагентная система колесных модулей-агентов. Т.к. модули идентичны между собой и способ крепления к платформе одинаков, то после декомпозиции можно считать таких агентов голономными. Голономные агенты это агенты, у которых действия и состояния совпадают. Агенты с таким свойством обладают двумя важными особенностями:

1. Знания одного агента можно полностью передать второму агенту, при этом второй агент становится точной копией первого.
2. Множество голономных агентов можно обучать взаимодействию одновременно, используя общую базу знаний.

Для получения эффекта от второго свойства будем использовать архитектуру многоагентной системы с использованием вирту-

ального лидера [5]. Это позволит накапливать базу знаний только у одного агента.

Основная идея – определить *виртуального лидера*, который расположен в центре формации относительно всей группы, и, соответственно, его *виртуальные координаты*. Состояние каждого агента будет определяться относительно виртуального лидера или виртуального центра координат. После того, как определена динамика виртуального лидера, агенты начинают свое движение. Таким образом, задача планирования пути и построения траекторий может быть реализована только для виртуального агента. Тогда как для модулей-агентов будет решаться задача удержания формации во время движения. Будем называть N колесных модулей-агентов с виртуальным лидером, которые образуют многоагентную систему, **платформой**.

На рис. 5 изображен пример такой структуры, состоящей из четырех модулей, где (X_b, Y_b) – это координаты маяка, C представляет собой координаты виртуального лидера (X_c, Y_c) , Φ_c – угол ориентации лидера относительно маяка, и ρ_c – центр разворота. Ошибка d_i^{err} определяет отклонение *i*-го модуля от правильного положения в платформе и вычисляется как разница между желаемым положением агента и текущим.

Положение виртуального центра определяется как центр тяжести площади платформы. Виртуальному агенту-лидеру задана информация об оптимальной линейной скорости $V_{opt} \in [V_{opt}^{min}, V_{opt}^{max}]$, где V_{opt}^{min} и V_{opt}^{max} являются соответственно минимальным и максимальным значением оптимальной скорости. Для реального робота учитывается функция энергопотребления, построенная на основе технических данных из документации моторов.

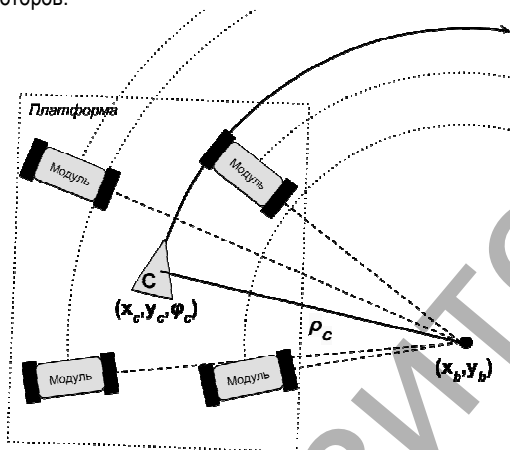


Рис. 5. Структура платформы с четырьмя модулями и виртуальным лидером

Обучение модулей-агентов. Процесс обучения агентов движению вокруг маяка состоит из двух этапов: (а) обучение агента повороту на заданный угол ориентации и (б) обучение всех агентов платформы движению по кругу за виртуальным лидером. Обучение одиночного агента повороту на угол нужной ориентации проводилось с использованием стандартного *Q-learning* алгоритма без следов преемственности [6].

При обучении состояние агента характеризуется парой значений $s_t = [\Phi_{err}^t, \omega_t]$. Множество действий $A_\omega = \{\emptyset, \omega_+, \omega_-\}$, где действие агента $a_t \in A_\omega$ – это изменение угловой скорости $\Delta\omega_t$ для момента времени *t*.

Система обучения сообщает положительное подкрепление, если ориентация робота ближе к целевой ($\Phi_{err}^t \rightarrow 0$) используя оптимальную скорость $\omega_t \rightarrow \omega_{orb}$ и штраф, если ориентация агента отклоняется от целевой или выбранное действие не оптимально для текущего положения (агент не начал во время тормозить).

Значение награды определяется как:

$$r^t = R(\Phi_{err}^{t-1}, \omega^{t-1}), \quad (1)$$

где *R* – функция награды, которая представлена деревом принятия решений, изображенным на рис. 6. Оптимальная скорость w_{opt} представляет собой константное значение скорости и используется для того, чтобы показать, что функция ценности способна находить эффективную скорость. Для реального робота используется функция энергопотребления моторов, рассчитанная по их документации.

Структура взаимодействия агентов на основе обучения с подкреплением для решения задачи кооперативного движения изображена на рис. 7.

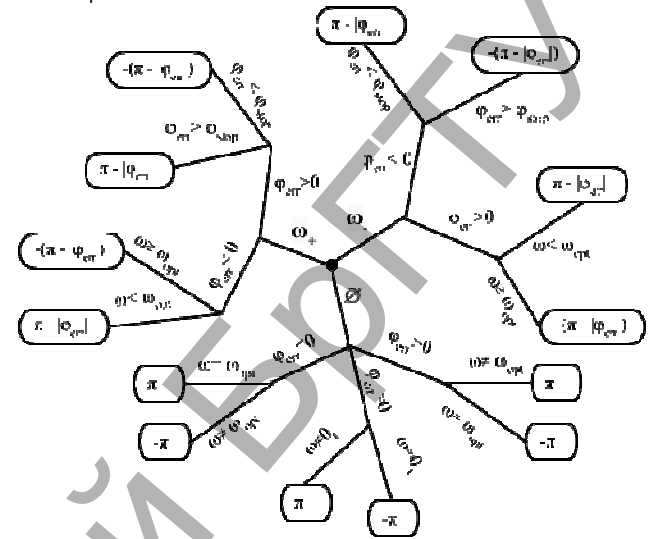


Рис. 6. Функция награды в виде дерева решений

Модуль *i*, находясь в состоянии s_i^t , выбирает действие a_i^t , используя текущую стратегию выбора действий, и переходит в следующее состояние s_i^{t+1} . Виртуальный лидер получает данные об изменениях после выполнения действия, вычисляет и присваивает награду r_i^{t+1} каждому агенту, которая корректность его действий с точки зрения всей платформы.

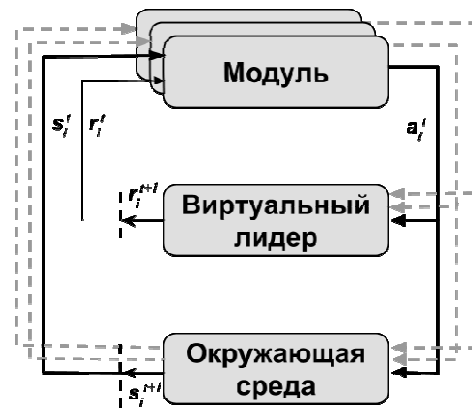


Рис. 7. Архитектура многоагентной системы колесных модулей на основе виртуального лидера и обучения с подкреплением

Для обновления стратегии модулей используется аналог многоагентного *Q-learning* алгоритма [7], в котором виртуальный лидер начисляет награду каждому агенту как оценку состояния в платформе:

$$\Delta Q_i(s_i^t, a_i^t) = \alpha [r_i^{t+1} + \gamma \max_{a \in A(s_i^{t+1})} Q_i(s_i^{t+1}, a) - Q_i(s_i^t, a_i^t)]. \quad (2)$$

Результаты экспериментов с производственным роботом. Для скоростей колеса свыше 0,5 м/с, как и для низких скоростей, необходимо точное позиционирование модуля относительно маяка. Для минимизации погрешности агент обучается управлению скоростями. Первоначально агент обучается разогнаться до заданной системой управления скорости и поддерживать ее. Вторым этапом

агент обучается уменьшать скорость вплоть до остановки, чтобы угловая ошибка была максимально приближена к нулю. Топология Q-функции, которая обучалась в течение 1400 эпох, показана на рисунке 8. Обучение происходило на реальном роботе.

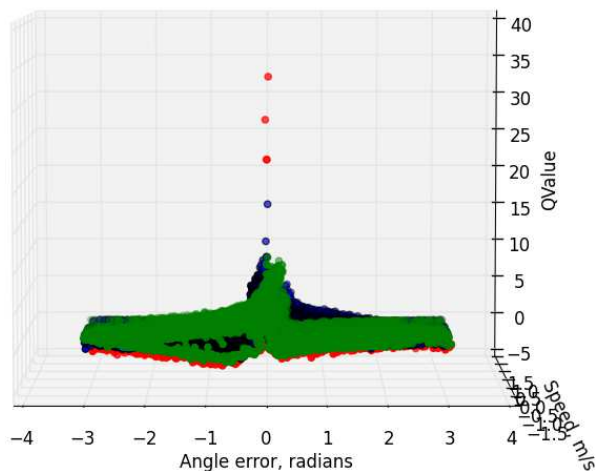


Рис. 8. Топология Q-функции после онлайн-обучения одного модуля с управлением скорости

Результат управления поворотом обученных модулей с центром разворота впереди-справа на рисунке 9.

На рисунке 10 показано, что при передвижении платформы радиусы модулей не изменяются и модули платформы находятся в стабильном положении относительно платформы.

Заключение. Экспериментальная часть демонстрирует успешное применение многоагентного подхода на основе виртуального лидера с использованием обучения с подкреплением для задачи эффективного управления реальной многоколесной роботизированной платформой. Предлагаемый подход включает множество Q-learning агентов, которые определяют оптимальное управление

модулями относительно виртуального лидера. Достоинства разработанного подхода:

- *Декомпозиция* обозначает, что вместо построения глобальной Q-функции мы строим множество локальных.
- *Адаптивность* – платформа адаптирует свое поведение для динамически изменяемого маяка и перенастраивает свою траекторию.
- *Масштабируемость и обобщающая способность* – один метод обучения используется для множества агентов, для любой позиции маяка и для любой позиции агента на платформе.

СПИСОК ЦИТИРОВАННЫХ ИСТОНИКОВ

1. Walters, D.G. The Whole Life Efficiency of Electric Motors // Energy Efficiency Improvements in Electric Motors and Drives. Springer. – 1997. – P. 81–94.
2. Mei, Y. Energy-Efficient Motion Planning for Mobile Robots / Y. Mei, Y.-H. Lu, Y.C. Hu, G. Lee // Robotics and Automation. Proceedings of IEEE International conference on ICRA'04. IEEE. – 2004. – Vol. 5. – P. 4344–4349.
3. Benedettelli, D. Experimental validation of collective circular motion for nonholonomic multi-vehicle systems / D. Benedettelli, A. Garulli, and A. Giannitrapani // Robotics and Autonomous Systems. 2010. – No 58. – P. 1028–1036.
4. Stetter, R. Realization and Control of a Mobile Robot / R. Stetter, P. Ziemniak, A. Pachinski // Research and Education in Robotics-EUROBOT 2010, Communication in Computer and Information Science. Springer. – 2011. – Vol. 156. – P. 130–140.
5. Ren, W. Distributed coordination architecture for multi-robot formation control / W. Ren, N. Sorensen // Robotics and Autonomous Systems. – 2008. – Vol. 56. – № 4. – P. 324–333.
6. Richard, S. // Reinforcement Learning: An Introduction. MIT Press / S. Richard, A. Sutton, G. Barto. – 1998.
7. Kabysh, A. Influence model and reinforcement learning for multi agent coordination // A. Kabysh, V. Golovko, K. Madani / Journal of Qafqaz University, Mathematics and Computer Science. – 2012. – № 33. – P. 58–64.

Материал поступил в редакцию 07.02.14



Рис. 9. Эксперимент поворота модулей к центру разворота впереди-справа



Рис. 10. Эксперимент подтверждения сохранности радиусов во время движения

DZIOMIN U., KABYSH A., GOLOVKO V., STETTER R. Reinforcement learning algorithm for power efficient control of multi-wheeled mobile robot

This paper presents an application of the multi-agent reinforcement learning approach for the efficient control of a real production mobile robot. This approach is based on a multi-agent decomposition applied to multi-wheel control. System efficiency achieved through proper compilation of the value function. Experiments shows that developed system provide stable robust control for complex kinematics.

Репозиторий БрГТУ