

Правилами является набор записей, в которых каждое из управляющих воздействий приписывается каждому участнику с одной из трех записей разрешения:

- воздействие разрешено ("Р");
- воздействие запрещено ("З");
- решается голосованием ("Г").

Участники сеанса взаимодействия, к которым управляющие воздействия приписаны с записью разрешения "Р", могут выполнять их независимо от желания остальных. Если запись для участника содержит разрешение "З" – воздействие участником выполняться не может. Для правил с записью разрешения "Г" вопрос о применении управляющего воздействия выносится на голосование. Системой голосований в автоматическом режиме формируется вопрос и варианты ответов, который оперативно запрашивается у участников. Решение выносится по принципу большинства голосов.

Любой участник может инициировать голосование по поводу изменения политики управления. В этом случае им предлагается новое содержание политики, которая выносится на голосование.

Данный подход обладает следующим рядом важных качеств, позволяющих эффективно его применять:

- возможность гибкой настройки и добавления правил управления;
- участие всех пользователей в распределенном управлении сеансом связи;
- система голосований, позволяющая смещать участников, злоупотребляющих своими полномочиями.

Разработанные методики реализованы в программных модулях управления сеансом связи в системе ВКС «Дельта-конференция». Они позволяют эффективно организовать многостороннее взаимодействие между терминалами и могут быть в дальнейшем использованы не только в системах ВКС, но и в других системах многостороннего взаимодействия (например, в системах видеонаблюдения, удаленного сбора данных, дистанционного обучения и т.п.).

В.Н. Вичугов, Г.П. Цапко

МЕТОД ПОДКРЕПЛЯЕМОГО ОБУЧЕНИЯ В ЗАДАЧАХ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

Одной из наиболее актуальных проблем в области автоматического управления является проблема построения адаптивных систем управления, способных приспосабливаться к неизвестным или изменяющимся параметрам объекта управления (ОУ). В данной статье рассмотрен метод построения адаптивных систем управления, в основе функционирования которых лежит метод обучения с подкреплением, также называемый методом подкрепляемого обучения.

Метод подкрепляемого обучения является достаточно новым методом в группе методов машинного обучения и занимает промежуточное положение между методами обучения с учителем и без учителя. В основе метода обучения с подкреплением лежат те основополагающие принципы адаптивного поведения, которые позволяют живым организмам приспосабливаться к изменяющимся или неизвестным условиям обитания. Метод обучения с подкреплением (Reinforcement Learning) был представлен и подробно изложен в книге [1]. В данном методе в обобщенном виде рассматривается взаимодействие агента с внешней средой, в результате которого агент путем проб и ошибок самостоятельно определяет наиболее оптимальное поведение для достижения максимума некоторого критерия. Во время взаимодействия со средой агент получает сигнал подкрепления, который

является скалярной величиной, характеризующей, насколько «хорошо» функционирует агент в данный момент времени. Целью функционирования агента является максимизация суммарного сигнала подкрепления, которое получит агент при взаимодействии с внешней средой. В исходном виде метод обучения с подкреплением предполагает конечное количество состояний внешней среды и возможных воздействий агента на внешнюю среду, а также взаимодействие агента с внешней средой в дискретные моменты времени. Указанные ограничения не позволяют свободно использовать указанный метод в задачах автоматического управления, так как сигналы в системах управления обычно являются непрерывными как по уровню, так и во времени. Тем не менее, метод обучения с подкреплением был успешно применен в некоторых задачах, которые можно рассматривать как задачи автоматического управления: система управления объектом, представляющим собой тележку с шестом [1], система управления роботом, который учится плавать в водной среде [2], система управления перевернутым маятником [3].

На основе метода подкрепляемого обучения была разработана структурная схема обобщенной САУ, функционирующей на основе метода обучения с подкреплением (RL-САУ), и алгоритмы работы структурных блоков. На основе структурной схемы и алгоритмов функционирования RL-САУ в среде программирования Borland Delphi было разработано программное средство «Исследование RL-САУ», предназначенное для моделирования и исследования дискретных RL-САУ. Программное средство позволяет задавать математическую модель ОУ в виде системы дифференциальных уравнений, определять вид и параметры задающего воздействия, задавать параметры настройки управляющего устройства (УУ), управлять процессом моделирования, отображать на экране значения всех моделируемых сигналов и их графики, определять значения показателей качества управления, сохранять результаты моделирования в файлы.

Ниже представлены результаты экспериментальных исследований, осуществленных с помощью разработанного программного средства с дискретной адаптивной системой управления ОУ второго порядка «Колебательное звено» с передаточной функцией

$$W = \frac{1}{0.5 \cdot p^2 + 0.1 \cdot p + 1}.$$

На рис. 1 показана переходная характеристика данного ОУ.

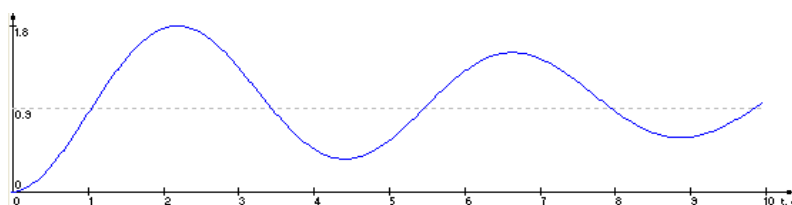


Рис. 1. Переходная характеристика ОУ «Колебательное звено»

Для функционирования RL-САУ необходимо установить значения параметров настройки УУ. Для проведения исследований были выбраны следующие значения параметров настройки: количество уровней квантования выходной величины $y - 100$, переменной состояния $x_2 - 50$, переменной состояния $x_1 - 0$, задающего воздействия $g - 50$, производной задающего воздействия $g' - 10$, возможные значения управляющего воздействия: 5, минус 5, 0, 15, минус 15. В качестве задающего воздействия был выбран импульсный сигнал. В эксперименте величина сиг-

нала подкрепления равна $1 - \varepsilon^2$, где ε – ошибка управления. Такое выражение было выбрано в связи с тем, что максимизация суммарной величины подкрепления приводит к минимизации среднеквадратической ошибки управления. Графики, описывающие функционирование системы в начале периода обучения, показаны на рис. 2. В УУ отсутствует априорная информация о математической модели ОУ. На рисунке видно, что выходное воздействие в начале функционирования формируется в основном случайным образом. По мере обучения в УУ определяются точные значения функции оценки воздействий, что позволяет УУ формировать такие воздействия на ОУ, которые приведут к максимизации суммарной величины подкрепления, что приведет к минимизации среднеквадратической ошибки управления.

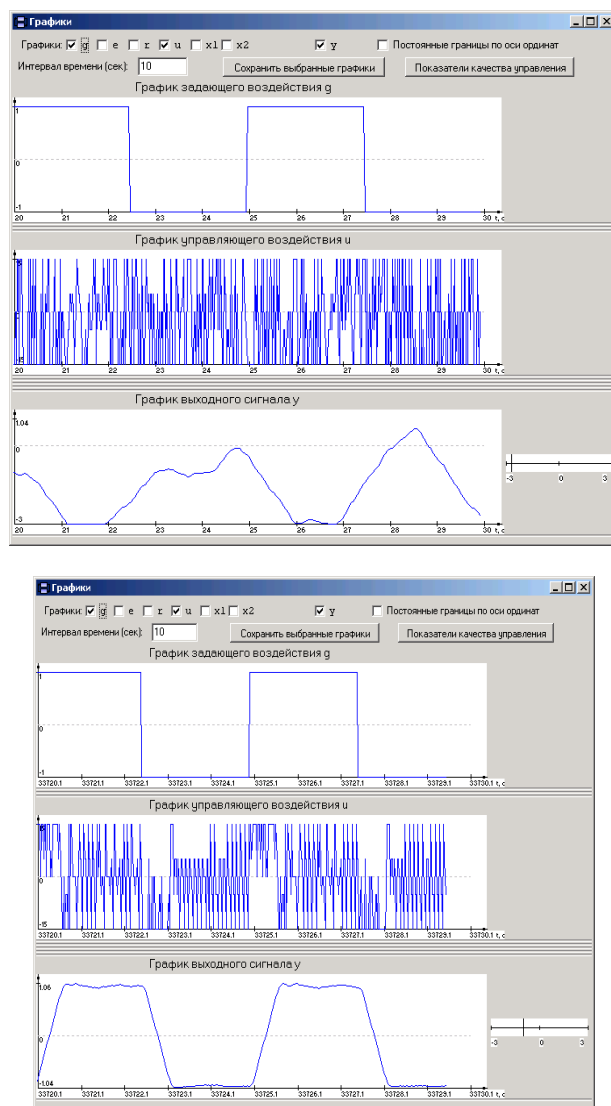


Рис. 2. Графики задающего воздействия, управляющего воздействия и выходной величины для ОУ «Колесательное звено» в начале периода обучения и через 10 часов модельного времени

На рис. 2 также показаны графики, которые характеризуют поведение системы после 10 часов модельного времени. При компьютерном моделировании 10 часов модельного времени соответствуют одной минуте реального времени. На рисунке видно, что УУ научилось формировать такие воздействия, которые приводят к соответствию выходного сигнала задающему сигналу. Здесь следует учесть, что максимальная амплитуда управляющего сигнала ограничена и не позволяет добиться идеального соответствия выходного и задающего сигналов.

Результаты экспериментальных исследований RL-CAУ с линейными и нелинейными ОУ второго порядка показали приемлемое качество управления и способность RL-CAУ адаптироваться к изменяющимся параметрам ОУ. Недостатком предложенного способа построения RL-CAУ является экспоненциальная зависимость объема требуемой памяти от порядка ОУ и от количества уровней квантования сигналов. Эту особенность исследователи в области подкрепляемого обучения называют «проклятием размерности» (англ. «dimensional curse»). В дальнейших исследованиях планируется устранить этот недостаток за счет применения искусственных нейронных сетей для представления функции оценки воздействий.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Sutton R.S., Barto A.G. Reinforcement learning: An introduction. MIT Press, Cambridge, MA, 1998.
2. Coulom R. Reinforcement Learning Using Neural Networks, with Applications to Motor Control. Institut National Polytechnique de Grenoble, 2002.
3. Aamodt T. Intelligent Control via Reinforcement Learning. Bachelors Thesis, University of Toronto, April 1997.

Л.В. Зюзина

ОБОБЩЕННАЯ МОДЕЛЬ УПРАВЛЕНЧЕСКОЙ ПОДГОТОВКИ ИНЖЕНЕРА-ПРОЕКТИРОВЩИКА И ЕЕ ФУНКЦИОНИРОВАНИЕ

Цель управленческого образования инженеров в вузе – обеспечение будущего специалиста системным и целостным содержанием и объемом знаний, достаточных в рамках его полномочий для успешного осуществления оперативного управления. Существующая сегодня система управленческой подготовки инженера, как отмечается в научной литературе, функционирует, но установленные цели не достигаются.

Формой представления целей системы является модель, поэтому для решения проблемы следует сформировать адекватную современным требованиям модель управленческих знаний.

Система управленческой компетенции инженера является сложной системой, для которой характерно наличие элементов, связей, структуры и т.д. Ее можно представить в виде основных элементов макромоделей, перечень и взаимосвязь которых представлены на рис. 1