УДК 621-501.72

А.А. Стенин, Е.Ю. Мелкумян, П.С. Хоменко

РЕАЛИЗАЦИЯ ЭТАЛОННЫХ СТРАТЕГИЙ ПРОЦЕССА ПРИНЯТИЯ РЕШЕНИЙ В АВТОМАТИЗИРОВАННЫХ ОБУЧАЮЩИХ СИСТЕМАХ

Аннотация: Рассмотрены два пути получения эталонных стратегий принятия решений, которые позволяют использовать их при обучении операторов широкого класса эргатических систем.

Ключевые слова: эталонные стратегии, принятие решений, эргатические системы.

Введение

Как известно [1], автоматизированные обучающие системы (АОС) являются эффективным средством повышения уровня подготовки операторов различных эргатических систем (ЭС). Наиболее сложными для операторов являются интеллектуальные ЭС, в которых оператор зачастую работает в режимах постоянного принятия оптимального или близкого к оптимальному решений. При реализации АОС по схеме “учитель-ученик” [2] для оценки правильности выбора той или иной стратегии управления необходимо иметь эталонные стратегии, относительно которых можно строить систему контроля и оценки деятельности оператора в процессе принятия решений.

Постановка задачи

Принятие человеком решений в общем виде можно рассматривать как ветвящийся процесс образования и преобразования иерархических структур, реализуемый двумя способами, отображающими характер изменения этих структур от исходной к заданной [3]. При первом способе исходная ситуация является начальной $S(\tau_0)$ и выбирается такая цепочка операций $R_{jk}(k = 1, 2, \ldots, T - \tau_0)$, которая ведет к заданной ситуации $S_T = S(T)$. Указанный процесс можно представить последовательностью:

$$S(\tau_0)R_{j1}S(\tau_0 + 1)R_{j2} \ldots S(T - \tau_0)R_{jT-\tau_0}S(T) \qquad (1)$$

Если $R \equiv R_{j1}S(\tau_0 + 1)R_{j2} \ldots S(T - \tau_0)R_{jT-\tau_0}$, то указанная последовательность сократится до триады $S(\tau_0)RS(T)$, т.е. по $S(\tau_0)$ и $S(T)$ отыскивается R. Процесс нахождения R может быть представлен индуктивной цепочкой последовательного нахождения состояний $S(\tau_0 + k)$.

При втором способе используется так называемое попятное движение. В качестве начальной ситуации $S(\tau_0)$ выступает заданная ситуация $S(T)$, а в качестве цели $S(T)$ - исходная ситуация $S(\tau_0)$. Так же как и при первом способе отыскивается цепочка операций R_{jk}, переводящая $S(\tau_0) = S(T)$ в $S(T) = S(\tau_0)$, т.е. реализуется последовательность (1). При

определении общего преобразования R последовательность также как и в первом случае сократится до триады. Здесь также имеется схема индуктивного решения, однако процесс нахождения R_{jk} носит дедуктивный характер, т.е. по $S(T)$ отыскивается $S(T-\tau_0)$ и т.д.

В обоих способах процесс решения предполагает исключения из рассмотрения запрещенных ситуаций. При этом обычно используется прошлый опыт.

Таким образом, задача определения оптимальной стратегии процесса принятия решений заключается в определении таких составляющих r_{ij}^k матрицы выигрышей R^k, которые для некоторых стратегий k принесут максимальный выигрыш.

Реализация эталонных стратегий принятия решений

В обоих способах на практике возможны два пути оптимизации принимаемых оператором решений.

Первый путь заключается в выборе решений из сравнительно небольшого набора возможных вариантов, а критерием служит отклонение от оптимального способа решения, т.е. ошибки. Ограничением здесь выступает время принятия решений.

Второй путь соответствует режиму работы, при котором основное внимание обращается на недопустимость ошибок. Время принятия решения изменяется при этом в широких пределах, а критерием оптимальности принятых решений является время, за которое найдены наилучшие решения.

Рассмотрим первый из упомянутых путей оптимизации.

Для управляемой системы с N возможными состояниями ($i = 1,\ldots,N$). В каждом состоянии i оператор может принять Δ_i возможных решений (стратегий), совокупность которых для всех состояний системы в рассматриваемый интервал времени составляет политику оператора.

Стратегия оператора может быть выражена как матрица переходных вероятностей $p^k = \left\|p_{ij}^k\right\|$ от состояния i к состоянию j. Приведем ей в соответствие матрицу выигрышей $R^k = \left\|r_{ij}^k\right\|$, получаемых от реализации каждого решения оператора.

Если проводится анализ поиска оптимальной стратегии на конечном интервале времени, то может быть оценен общий выигрыш от той или иной политики оператора.

При этом очевидно, что система управления переходит из одного состояния в другое через равные интервалы времени. Считая, что последействие отсутствует, процесс принятия решений можно описать аппаратом теории Марковских процессов.

Кроме того, для анализа стратегий работы оператора на конечном интервале времени, когда имеется набор возможных вариантов решений, для получения эталонной оптимальной стратегии предлагается использовать стандартную процедуру метода динамического программирования [4].

В этом случае полный ожидаемый выигрыш за n шагов будет равен:

$$\nu_i(n) = \sum_{j=1}^{N} p_{ij}[r_{ij} + \nu_j(n-1)] = \sum_{j=1}^{N} p_{ij} r_{ij} + \sum_{j=1}^{N} p_{ij}\nu_j(n-1) \qquad (2)$$

$$\text{или} \quad \nu_i(n) = q_i + \sum_{j=1}^{N} p_{ij}\nu_j(n-1),$$

где $q_i = \sum\limits_{j=1}^{N} p_{ij} r_{ij}$ – минимальный ожидаемый выигрыш, а оптимальная стратегия будет соответствовать критерию оптимальности $(n+1)$–го решения:

$$\nu_i(n+1) = \max_k \{q_i + \sum_{j=1}^{N} p_{ij}\nu_j(n)\}, k = 1, \ldots, \Delta_i, i = 1, \ldots, N. \qquad (3)$$

При неограниченном времени протекания оцениваемой деятельности оператора суммарный выигрыш системы также растет неограниченно, поэтому политику оператора можно оценивать по среднему ожидаемому доходу от реализации одного решения.

В этом случае задачу поиска оптимальной стратегии можно рассматривать по отдельным актам принятия решений вне зависимости от интервалов времени, в которые они реализуются.

Для нахождения такой политики можно воспользоваться итеративным процессом. Если политика оператора фиксирована, то поиск оптимальной стратегии сводится к решению системы уравнений:

$$q + \nu_i = g_i + \sum_{j=1}^{N} p_{ij}\nu_j \qquad (4)$$

с использованием критерия

$$q^k = \max_k \{g_i^k + \sum_{j=1}^{N} p_{ij}^k \nu_j - \nu_i\} \qquad (5)$$

Общая процедура поиска оптимальной стратеги принятия решений при фиксированных ν_i сводится к следующему. При фиксированной политике A решается система уравнений (4), причем в ней ν_j при $j = N$ полагается равным нулю (поскольку нам важны разности $\nu_i - \nu_j$ относительных весов, а не их абсолютные значения). Затем найденные веса ν_i^A подставляются в формулу (5), и для каждого i находится максимальное значение критерия, и набор стратегий, на котором достигается максимальное среднее значение критерия, принимается за новую политику A_1.

При втором пути реализации процесса принятия решений время принятия отдельных решений колеблется в широких пределах за счет достижения точного решения. В этом случае для реализации эталонных

стратегий принятия решений предлагается использовать аппарат теории полумарковских процессов. При этом система управления перед тем как перейти из состояния i в состояние j, находится в состоянии i случайное время $\tau_{ij} = m$ с плотностью распределения $f_{ij}(\tau_{ij} = m) = N(\tau_{ij} = m)$. Значением $f_{ij}(m)$ определяется вероятность того, что система в состоянии i проводит ровно m единиц времени, прежде чем перейти в i. Если политика решений оператором описывается как полумарковский процесс, то оптимальная стратегия i в состоянии i может быть выбрана на основании анализа переходных вероятностей p_{ij}^k, плотности распределения $h_{ij}^k(m)$, а также системных выигрышей типа $y_{ij}^k(l)$ и $b_{ij}^k(m)$. Здесь $b_{ij}^k(m)$ – системный выигрыш от перехода системы, находившейся в течение времени m в состоянии i из состояния i в состояние i при i-й оптимальной стратегии оператора, $y_{ij}^k(l)$ - выигрыш от пребывания системы в данном состоянии i в течении временного интервала $(l-1, l)$, пропорциональный некоторой норме выигрыша за единицу времени при i-й оптимальной стратегии.

При таких допущениях задача может быть сформулирована как задача максимизации одношагового выигрыша, либо как задача максимизации полного ожидаемого дохода.

Как и при марковских процессах будем рассматривать полумарковские процессы конечной и бесконечной длительности. В первом случае задача состоит в определении политики оператора, максимизирующей полный ожидаемый выигрыш от процесса, до конца функционирования которого осталось n единиц времени. В случае процессов бесконечной длительности задача может быть сформулирована либо как задача максимизации одношагового выигрыша (аналогично задаче для марковских процессов бесконечной длительности), либо как задача максимизации полного ожидаемого дохода. В общем виде можно оперировать критерием, получаемым в соответствие с принципом оптимальности Беллмана [4]:

$$\nu_i(n) = \max_k \{ \sum_{j=1}^{N} p_{ij}^k \sum_{m=n+1}^{\infty} f_{ij}^k(m) \\ [\sum_{l=1}^{n-1} y_{ij}^k(l) + \nu_i(0)] + \sum_{j=1}^{N} p_{ij}^k \sum_{m=0}^{n} f_{ij}(m)[\sum_{l=0}^{m-1} y_{ij}^k(l) + r_{ij}^k(m) + \nu_j(n-m)]\} \tag{6}$$

где $i = \overline{1, N}, n = 1, 2, 3 \ldots$.

Средний ожидаемый доход r_i от пребывания системы в состоянии i и ухода из этого состояния при длительном функционировании процесса будет:

$$r_i = \sum_{j=1}^{N} p_{ij} \sum_{m=0}^{\infty} h_{ij}(m) \left[\sum_{l=0}^{m-1} y_{ij(l)} + b_{ij}(m) \right]. \tag{7}$$

 ISSN 1560-8956

Следует заметить, что для процессов с одним эргодическим классом, когда предельная вероятность i-го состояния не зависит от начального состояния, прибыль одинакова для всех состояний процесса.

Политику, оптимальную в смысле максимизации одношаговой прибыли процесса, т. е. среднего ожидаемого дохода за единицу времени, можно найти с помощью итеративного процесса, описанного выше и использующего критерий

$$q^k = q_i^k + \frac{1}{\tau_i^k}\left[\sum_{j=1}^{N} p_{ij}\nu_j - \nu_i\right] \to \max_k \qquad (8)$$

Заключение

Для построения в АОС объективной автоматизированной системы контроля и оценки деятельности операторов эргатических систем требуется наличие эталонных реализаций тестовых задач обучения, т.е. в данном случае эталонных стратегий принятия решений.

В работе предложены два пути оптимизации принимаемых оператором решений, основанных на математическом аппарате теории марковских и полумарковских процессов с использованием принципа оптимальности Беллмана.

Эти пути касаются обучения операторов на непродолжительном конечном интервале времени обучения, что наиболее часто встречается на практике, и длительном (бесконечным с математической точки зрения) интервале времени обучения. Кроме того в работе рассматриваются пути нахождения оптимальных стратегий, когда время принятия единичных решений фиксировано и одинаково и когда время перевода реализуется оператором случайным образом.

Полученные таким образом в статье эталонные стратегии принятия решений позволяют использовать их при обучении операторов такого класса эргатических систем, как диспетчера авиалиний, диспетчера энергоблоков и др., где необходимо принимать решения в штатных и нештатных ситуациях.

Литература

1. Автоматизированные обучающие системы профессиональной подготовки операторов ЛА / Под ред. Шукшунова В.Е. – М.: Машиностроение. – 1986. – 240с.
2. Технические эргатические системы/ Под. Ред. Павлова В.В. – К: Вища школа. - 1977. –344с.
3. Шибанов Г.П. Количественная оценка деятельности человека в системах “человек-техника” – М.: Машиностроение. - 1983. – 224с.
4. Беллман Р., Калаба Р. Динамическое программирование и современная теория управления – М.: Наука. – 1969. - 119с.

Отримано 04.11.2011 р.