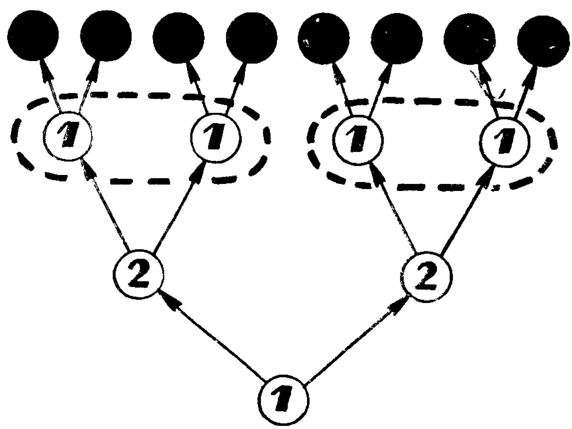
библиотека по автоматике

Д. А. Поспелов

ИГРЫ и АВТОМАТЫ



ИЗДАТЕЛЬСТВО «ЭНЕРГИЯ»

БИБЛИОТЕКА ПО АВТОМАТИКЕ

Выпуск 188

Д. А. ПОСПЕЛОВ

ИГРЫ И АВТОМАТЫ

ИЗДАТЕЛЬСТВО «ЭНЕРГИЯ»

МОСКВА 1966 ЛЕНИНГРАД

УДК 519.95:62-50

Книга посвящена изложению основных идей, лежащих в основе построения логики работы автоматов игрового типа. Снабжена примерами автоматов для широко известных игр.

Рассчитана на читателя, знающего математику в объеме высшего технического учебного заведения.

Поспелов Дмитрий Александрович

Игры и автоматы

М.—Л., издательство „Энергия", 1966, 136 с. с черт.

(Библиотека по автоматике, вып. 188)

$\frac{3\text{-}3\text{-}13}{280\text{-}66}$

Редактор *В. Г. Лазарев* Техн. редактор *Т. Г. Усачева*

Сдано в набор 15/III 1966 г. Подписано к печати 16/VI 1966 г.
Т-07171 Бумага типографская № 1 84×108[1]/[32] Печ. л. 7,14 Уч.-изд. л. 6,52
Тираж 15 000 экз. Цена 35 коп. Заказ 2292

Московская типография № 10 Главполиграфпрома
Комитета по печати при Совете Министров СССР.
Шлюзовая наб., 10.

ПРЕДИСЛОВИЕ

Теория игр и теория статистических решений — два раздела математики, изучающие методы принятия целесообразных решений в конфликтных ситуациях. Не давая точного определения понятия конфликтной ситуации, укажем, что такая ситуация может возникнуть либо при воздействии на один и тот же процесс группы лиц, интересы которых не совпадают между собой, либо в случае необходимости принятия решения о некоторых воздействиях на объект, когда окончательные результаты этих воздействий могут быть оценены только вероятностным образом. Подобные ситуации встречаются, например, при решении задач по управлению сложным процессом, для которого отсутствует его точное математическое описание, а имеющаяся информация о ходе процесса носит статистический характер. В этих случаях поиск оптимального алгоритма управления может осуществляться лишь на основе теории статистических решений и теории игр.

В настоящее время на русском языке появилось значительное число работ, посвященных теории игр и теории статистических решений. Список таких работ приведен в конце книги. В настоящей книге не ставится задача изложения всех (и даже всех основополагающих) результатов, полученных в этих теориях. Автор стремился в первую очередь показать, как могут быть использованы такие результаты при разработке автоматов, предназначенных для управления сложными процессами.

Опыт теоретико-игрового подхода к решению задач теории управления пока еще невелик. Поэтому в большинстве приведенные в книге примеры несложны и относятся не к задаче управления, а к решению обычных игровых задач (карточных игр, игр в домино и т. д.). Однако подходы к решению этих задач можно использовать для задач управления.

В данной книге автомат понимается или как некоторое физическое устройство, реализующее алгоритм решения схемно, или как программа, реализуемая на специализированной или универсальной цифровой вычислительной машине. Понимание автомата как программы не должно смущать читателя, так как основная задача книги состоит не в поиске схемных решений, а в поиске методов решений, которые при желании могут быть воплощены в конструкцию устройства, реализующего данный алгоритм.

В первых двух параграфах излагаются основные элементы теории игр и конечных автоматов в таком объеме, который необходим для дальнейшего изложения. Читатели, знакомые с этими разделами математики, могут без всякого ущерба для понимания дальнейшего материала их опустить. Для чтения § 6—8 необходимо знакомство с основными результатим из теории дискретных цепей Маркова. Минимум необходимых сведений о таких цепях дан в § 6. Читатели, интересующиеся только теоретическими проблемами, связанными с поведением автоматов в случайных средах и играми автоматов, могут читать только три последних параграфа книги.

При работе над книгой большую помощь автору оказали замечания и пожелания, сделанные при знакомстве с рукописью Р. Л. Стратановичем и В. Г. Лазаревым. Автор приносит им за это свою благодарность.

Автор

1. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ИГР И ТЕОРИИ ПРИНЯТИЯ РЕШЕНИЙ

В этом параграфе рассмотрены некоторые общие положения из двух тесно связанных между собой разделов математики: теории игр и теории принятия решений. При этом приведены лишь те положения и результаты этих разделов, которые необходимы для дальнейшего изложения. Все читатели, заинтересованные в более глубоком изучении этих разделов, могут обратиться к специальной литературе по этим вопросам, список которой указан в конце книги.

Прежде всего разъясним то, что мы в дальнейшем будем понимать под термином «игра n лиц». Пусть задано некоторое множество $M=\{m_1, m_2, \ldots, m_k\}$, называемое множеством ходов. Пусть M разбито на n подмножеств $M_1, M_2, \ldots, M_n$, которые в общем случае могут пересекаться. Назовем подмножество M_i множеством возможных ходов для игрока $И_i$ $(i=1, 2, \ldots, n)$. Процесс игры состоит в последовательном выборе игроками $И_i$ каких-либо ходов из своих множеств возможных ходов. Различаются ходы двух типов: личные и случайные. При личном ходе игрок на основании известных ему соображений выбирает один из возможных ходов на данном шаге игры. При случайном ходе выбор хода производит некоторый «случайный механизм». К случайным ходам можно отнести выбор игроком карты из колоды «не глядя», получение тех или иных камней при игре в домино, выпадение числа очков на игральной кости и т. д.

Вся совокупность уже сделанных игроками ходов определяет ситуацию игры b_j. В множестве ситуаций $B=\{b_1, b_2, \ldots, b_l\}$ выделяется подмножество $B'\subseteq B$, элементы которого называются заключительными ситуациями. Если очередной ход некоторого игрока приводит к заключительной ситуации, то игра заканчивается. Всем заключительным ситуациям сопоставлены платежи между игроками, так что полученная заключительная ситуация однозначно определяет выигрыши и проигрыши всех игроков. Это сопоставление осуществляется с помощью платежной функции f, определенной на множестве заключительных ситуаций. В каждом конкретном случае ход игрока регламентируется некоторыми правилами, которые зависят от ситуации игры и возможностей игрока. С помощью этих правил в множесте M_i игрока $И_i$ выделяется подмножество $M'_i \subseteq M_i$ ходов, допустимых для данного конкретного положения в игре.

Игра считается заданной, если заданы множества M_i, функция f и правила выделения подмножеств M'_i для каждой ситуации b_j и возможностей игроков $И_i$.

Существует много различных процессов, которые могут быть сведены к вышеописанной игровой модели. Полезно построить некоторую классификацию игр, используя для нее один из возможных признаков. Можно, например, классифицировать игры по максимальному количеству ходов, сделанных каким-либо из игроков. В наиболее общем случае можно считать, что все игроки делают одинаковое число ходов, рассматривая в качестве хода и отказ игрока от очередного хода. При такой классификации можно разбить все игры на одноходовые и многоходовые. Особенно просто задание одноходовой игры. В этом случае нет необходимости задавать правила выделения подмножеств M'_i и задание игры сводится к заданию множеств M_i и значений функции f.

Другим примером классификации игр может служить классификация по числу игроков. При этом игры разбиваются на игры с одним участником (пасьянс), игры с двумя участниками (шашки) и т. д.

Можно в качестве признака классификации выбрать отношение игроков к исходу игры. По этому признаку все игры можно разделить на два типа: антагонистические и неантагонистические. Антагонистические игры характеризуются тем, что интересы участников игры прямо противоположны. Каждый из игроков стремится обеспечить себе максимальный выигрыш и, следовательно, стремится к максимальному проигрышу своих противников. Теория антагонистических игр, создателем которой является Дж. Нейман, и носит в современной математике название теории игр. При неантагонистических играх один из игроков (или по крайней мере один из игроков) не стремится к максимизации своего выигрыша. Такие игры часто называют играми с природой. Теория неантагонистических игр входит в качестве составной части в теорию принятия решений, активно развивающуюся в настоящее время ветвь математики.

Если в качестве признака классификации игр использовать признак полноты информации, имеющейся у игроков на данном этапе игры, о предшествующих ходах противника и его возможностях, то все игры можно разделить на два типа. В играх с полной информацией игроки при выборе своего очередного хода имеют полную информацию о ситуации b_j и возможностях каждого из игроков (шахматы, шашки, карточные игры «в открытую»). В играх с неполной информацией такого положения нет (домино, карточные игры «в закрытую»).

Наконец, если в качестве признака классификации выбрать наличие или отсутствие детерминизма при выборе игроком своего очередного хода, то игры можно разделить на игры со случайными ходами и игры без случайных ходов. Примером игры со случайными ходами может служить игра в бросание монеты (орлянка).

Кроме вышеперечисленных, возможны другие классификации игр. Так как эти классификации происходят на основе рассмотрения различных признаков, то можно характеризовать игру с помощью совокупности признаков. Например, шахматы являются антагонистической, многошаговой игрой с полной информацией, без случайных ходов и с двумя участниками.

Геометрически игру можно изобразить в виде дерева игры (рис. 1). Узлы этого дерева соответствуют возможным ситуациям в игре, заключительным ситуациям соответствуют закрашенные узлы. Внутри каждого из узлов написана цифра, указывающая номер игрока, делающего при данной ситуации ход (на рисунке рассмотрено дерево игры для двух игроков). Подмножества M'_i изображаются множеством ветвей, выходящих из данного узла. Около заключительных ситуаций написано распределение выигрышей и проигрышей между игроками после окончания игры.

Любой путь на дереве игры соответствует последовательности ходов, сделанных игроками в процессе игры, и определяет партию игры. Число различных партий является важным признаком при исследовании игр

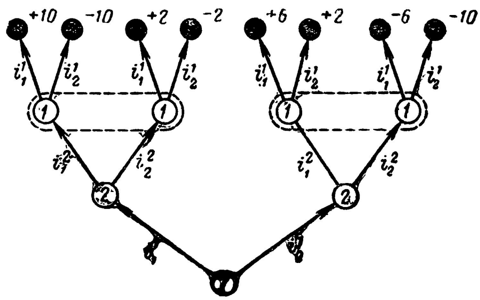


Рис. 1.

теми или иными средствами. Число различных партий игры совпадает с числом заключительных ситуаций. Для игры, показанной на рис. 1, имеется восемь различных партий игры.

Условимся, наконец, о графической интерпретации объема информации, используемом игроком при выборе своего очередного хода. Для этого будем объединять на дереве игры пунктирной замкнутой линией те вершины дерева, которые этот игрок не может различить на данном шаге игры. Если, например, игрок $И_1$ на третьем шаге игры, показанной на рис. 1, не знает хода игрока $И_2$ на предыдущем шаге, то он не может различить, в какой из двух возможных ситуаций, ограниченных на дереве игры пунктиром, находится в данный момент игра. Совокупность ситуаций, попавших внутрь пунктирной линии, называется информационным множеством. Если игра является игрой с полной информацией, то информационные множества состоят в точности из одной ситуации.

Назовем выбор игроком $И_i$ того или иного хода на данном шаге игры ходовой стратегией игрока, а набор указаний, который позволяет игроку в любой ситуации игры, или более точно при любой информации об игре, выбирать ход, полной стратегией игры или просто стратегией игрока $И_i$. Таким образом, стратегии выбираются игроком на основе использования некоторой решающей функции, определенной на множестве стратегий, и использующей для выбора той или иной стратегии ту информацию, которую имеет игрок на данном шаге игры. Для одноходовой игры ходовые стратегии и полные стратегии совпадают между собой.

Для обозначения ходовых стратегий игроков используем букву i с двумя индексами. Верхний индекс указывает номер игрока, которому принадлежит эта стратегия, а нижний — порядковый номер ходовой стратегии в множестве стратегий данного игрока. Например, i_2^1 означает вторую ходовую стратегию игрока $И_1$. Полные стратегии мы будем обозначать аналогичным образом, использовав для их обозначения букву J. Отметим, что общее число ходовых стратегий игрока $И_i$ совпадает с числом элементов подмножества M'_i, а общее число полных стратегий может быть весьма произвольным. Однако не имеет смысла рассматривать полные стратегии, которые хотя и различны по форме, но приводят к реализации одинаковых партий в игре. Поэтому число полных стратегий мы будем считать не превосходящим числа различных последовательностей ходов, которые могут быть сделаны игроками в данной игре.

Для дерева игры, показанного на рис. 1, число ходовых стратегий у каждого из игроков равно двум, независимо от шага игры. Число же полных стратегий для игрока $И_2$ равно четырем, так как игрок $И_2$ знает первый ход игрока $И_1$ и для каждого хода этого игрока может ответить одним из двух своих ходов. Число полных стратегий игрока $И_1$ также равно четырем, так как выбор его хода определяется лишь его первым ходом и не зависит от хода игрока $И_2$, ибо ход этого игрока, по условиям игры, игроку $И_1$ перед его вторым ходом неизвестен. Если бы игра, показанная на рис. 1, была игрой с полной информацией (т. е. на дереве игры все информационные множества состояли только из одной вершины), то число полных стратегий игрока $И_2$ осталось бы равным четырем, а число полных стратегий для игрока $И_1$ увеличилось бы до восьми и стало равным числу партий в игре.

Если через S_j^i обозначить j-ю последовательность ходов игрока $И_i$ в данной партии, то для случая двух игроков можно построить матрицу игры размера $S^1 \times S^2$, где S^1 — число различных последовательностей ходов, которые может сделать в данной игре игрок $И_1$, а S^2 — аналогичное число для игрока $И_2$. Тогда последовательность ходов, которую осуществляет игрок в реализуемой партии игры, полностью определяется номером столбца или строки этой матрицы.

В качестве элементов матрицы возьмем численные значения платежа в конце игры для заключительной ситуации, к которой приводят выбранные игроками последовательности ходов. На рис. 1 показано дерево игры, в которой каждая партия заканчивается после трех ходов. На каждом шаге игрок может реализовать один из двух имеющихся в его распоряжении ходов. Матрица, соответствующая этому дереву игры, имеет четыре строки и два столбца. Это определяется тем, что игрок $И_1$ имеет четыре возможных последовательности ходов

$$S_1^1 = i_1^1,\ i_1^1;\quad S_3^1 = i_2^1,\ i_1^1;$$

$$S_2^1 = i_1^1,\ i_2^1;\quad S_4^1 = i_2^1,\ i_2^1,$$

а игрок $И_2$ — две последовательности ходов $S_1^2 = i_1^2$; $S_2^2 = i_2^2$. Матрица S имеет вид:

$$\begin{array}{c} \\ S_1^1 \\ S_2^1 \\ S_3^1 \\ S_4^1 \end{array}
\begin{array}{c} \begin{array}{cc} S_1^2 & S_2^2 \end{array} \\ \begin{bmatrix} 10 & 2 \\ -10 & -2 \\ 6 & -6 \\ 2 & -10 \end{bmatrix} \end{array}.$$

Выбор той или иной полной стратегии игроками сводится к тому, что на каждом шаге выбирается такая ходовая стратегия, которая является в данных условиях игры наиболее выгодной для игрока. Если оба игрока выбирают некоторые стратегии, то тем самым определяется последовательность ходов, сделанных ими в данной партии, и, следовательно, выбор игроками некоторых стратегий можно рассматривать как выбор первым из игроков строки матрицы S, а вторым — столбца матрицы S.

Если элемент матрицы положителен, то это означает, что при соответствующем выборе игроками полных стратегий игрок $И_1$ получит выигрыш, численно равный значению этого элемента, и этот выигрыш будет выплачен ему игроком $И_2$. Если элемент матрицы отрицателен, то это указывает на то, что при таком выборе игроками полных стратегий игрок $И_1$ выплачивает игроку $И_2$ сумму, численно равную значению этого элемента матрицы игры. Из анализа матрицы игры (платежной матрицы) для нашего примера вытекает, что для игрока $И_1$ стратегия S^1_2 явно невыгодная, ибо при выборе такой стратегии он всегда проигрывает. Наоборот, стратегия S^1_1 для него всегда выигрышная. Для игрока $И_2$ обе его стратегии являются проигрышными, но стратегия S^2_2 для него предпочтительнее, так как при выборе такой стратегии его проигрыш не может превзойти двойки. Если оба игрока произведут анализ платежной матрицы, то игрок $И_1$ выберет стратегию, определяемую первой строкой матрицы, а игрок $И_2$ — стратегию, определяемую вторым столбцом матрицы. После этого выбора сама игра уже не представляет в этом случае никакого интереса, так как выигрыш игрока $И_1$ и проигрыш игрока $И_2$ предопределены.

Таким образом, действия игрока $И_1$ направлены на поиск стратегии, при которой его выигрыш максимален, что соответствует поиску строки в матрице игры, наименьший элемент в которой является наибольшим по сравнению с наименьшими элементами всех других строк матрицы игры. Итак, для игрока $И_1$ полная стратегия будет оптимальной, если будет достигнут

$$\underset{i}{\text{макс}}\left\{\underset{j}{\text{мин}}\, a_{ij}\right\},$$

где a_{ij} — элементы матрицы игры; $i=1, 2, \ldots, l$; $j=1, 2, \ldots, m$, где l — число полных стратегий игрока $И_1$, а m — число полных стратегий игрока $И_2$. Оптимальная стратегия называется поэтому максиминной стратегией.

Игрок $И_2$ тоже хочет выбрать стратегию так, чтобы максимизировать свой выигрыш. Его выигрыши в матрице игры изображаются отрицательными числами, поэтому при выборе оптимальной стратегии игрок $И_2$ ищет такой столбец в матрице игры, в котором максимальный член минимален. Его оптимальная стратегия будет достигнута, если будет достигнут

$$\min_{j}\left\{\max_{i} a_{ij}\right\},$$

где $i=1, 2, \ldots, l$; $j=1, 2, \ldots, m$. Оптимальная стратегия игрока $И_2$ называется минимаксной стратегией.

Если выполняется равенство

$$\max_{i}\left\{\min_{j} a_{ij}\right\}=\min_{j}\left\{\max_{i} a_{ij}\right\}, \tag{1}$$

то говорят, что игра имеет седловую точку и элемент матрицы, определяемый на основании соотношения (1), называется ценой игры. Можно показать, что игры, имеющие седловую точку, являются играми с полной информацией. Для игр с седловой точкой характерно то, что, производя полный анализ матрицы игры, игроки выбирают свои оптимальные стратегии и на этом игра заканчивается.

Однако может случиться, что матрица игры такова, что седловой точки для нее не существует. Примером игры без седловой точки может служить игра, определяемая следующей матрицей:

$$\begin{pmatrix} 10 & -5 \\ -10 & 5 \end{pmatrix}.$$

В этом случае

$$\max_{i}\left\{\min_{j} a_{ij}\right\}=-5;$$

$$\min_{j}\left\{\max_{i} a_{ij}\right\}=5$$

и они не совпадают между собой. Для игроков, играющих в игру без седловой точки, необходимо при выборе своей стратегии избегать какой-либо закономерности выбора, так как на основе анализа его игры противник может разгадать принятую закономерность. Выгодным способом выбора стратегии игры в этом случае является способ случайного выбора. Этот выбор может характеризоваться законом распределения дискретной случайной величины ξ, в качестве значений которой выступают целые числа, соответствующие номерам полных стратегий, имеющихся у данного игрока,

ξ	1	2	....	l
p	p_1	p_2		p_l

В этой таблице p_i означает вероятность, с которой на данном шаге игры игрок выбирает стратегию с номером i. При этом, как всегда, $\sum_{i=1}^{l} p_i = 1$. Стратегия игрока, определяемая последовательным выбором полных стратегий на основании таблицы вышеприведенного типа, называется смешанной стратегией.

Рассмотрим теперь матрицу

$$\begin{pmatrix} 10 & -5 \\ -10 & 5 \end{pmatrix}.$$

Пусть вероятности, с которыми игрок $И_1$ выбирает свои полные стратегии, равны p_1 и p_2, а для игрока $И_2$ эти вероятности равны q_1 и q_2. Тогда математическое ожидание выигрыша для первого игрока при условии, что второй игрок выбрал первую стратегию, определяется как $10p_1 - 10p_2$. Если же второй игрок выбрал свою вторую стратегию, то математическое ожидание выигрыша для $И_1$ есть $-5p_1 + 5p_2$. Окончательно математическое ожидание выигрыша для первого игрока имеет вид $(10p_1 - 10p_2)q_1 + (-5p_1 + 5p_2)q_2$. Аналогично для второго игрока математическое ожидание проигрыша определится как $(10q_1 - 5q_2)p_1 + (-10q_1 + 5q_2)p_2$.

Найдем теперь значения p_1, p_2, q_1 и q_2, при которых математическое ожидание выигрыша для игрока $И_1$

максимально. Запишем для нахождения условного экстремума функцию Лагранжа:

$$F(p_1, p_2, q_1, q_2, \lambda_1, \lambda_2) = (10p_1 - 10p_2)q_1 +$$

$$+ (-5p_1 + 5p_2)q_2 + \lambda_1(p_1 + p_2 - 1) +$$

$$+ \lambda_2(q_1 + q_2 - 1).$$

Для поиска максимума составляем систему уравнений:

$$\frac{\partial F}{\partial p_1} = 10q_1 - 5q_2 + \lambda_1 = 0; \quad \frac{\partial F}{\partial q_2} = -5p_1 + 5p_2 + \lambda_2 = 0;$$

$$\frac{\partial F}{\partial p_2} = -10q_1 + 5q_2 + \lambda_1 = 0; \quad \frac{\partial F}{\partial \lambda_1} = p_1 + p_2 - 1 = 0;$$

$$\frac{\partial F}{\partial q_1} = 10p_1 - 10p_2 + \lambda_2 = 0; \quad \frac{\partial F}{\partial \lambda_2} = q_1 + q_2 - 1 = 0.$$

Решая ее, получим:

$$p_1 = \frac{1}{2}; \quad p_2 = \frac{1}{2}; \quad q_1 = \frac{1}{3}; \quad q_2 = \frac{2}{3}.$$

Составляем теперь функцию Лагранжа для игрока $И_2$ и ищем такие значения p_1, p_2, q_1 и q_2, для которых эта функция достигает минимума. Получаем путем дифференцирования и приравнивания частных производных нулю систему уравнений, совпадающую с системой уравнений, которую мы только что рассмотрели. Это следует из совпадения выражений для математического ожидания выигрыша для $И_1$ и проигрыша для $И_2$. Тогда минимальный проигрыш второго игрока будет достигаться при тех же частотах смены стратегий, при которых достигается максимальный выигрыш игрока $И_1$. При подстановке значений вероятностей смены стратегий в выражения для математического ожидания выигрыша игроков мы получим, что для обоих игроков значение математического ожидания равняется нулю. Смешанные стратегии, найденные нами для игроков $И_1$ и $И_2$ называют, обычно, смешанными максиминными и минимаксными стратегиями. В рассматриваемом случае существует цена игры, так как значения математических ожиданий у нас совпали. В теории игр доказывается фундаментальная теорема о возможности нахождения игроками минимаксных и максиминных стратегий, которая формулируется следующим образом.

Теорема о минимаксе

Для игры двух лиц, определяемой матрицей игры, всегда существуют оптимальные смешанные стратегии для игроков $И_1$ и $И_2$ и оптимальная стратегия для $И_1$ является смешанной максиминной его стратегией, а оптимальная стратегия для $И_2$ является его смешанной минимаксной стратегией.

Эта теорема доказывает справедливость утверждения о том, что для любой матричной игры может быть определена цена игры, которая соответствует выигрышу, получаемому игроком при использовании им оптимальной (чистой или смешанной) стратегии.

Отсюда вытекает одна из основных прикладных задач теории матричных игр: задача о нахождении вероятностей p_i, определяющих оптимальные смешанные стратегии игроков. Существует несколько универсальных вычислительных методов, с помощью которых удается решить поставленную задачу. В качестве примера можно указать на хорошо известный метод сведения поставленной задачи к стандартной задаче линейного программирования, для решения которой существуют весьма эффективные методы. Покажем принцип сведения задачи о нахождении оптимальных смешанных стратегий игроков к задаче линейного программирования[1]. Пусть игра задана матрицей S размера $m \times n$. Элементы этой матрицы мы обозначим через f_{ij}. Величины f_{ij} характеризуют платежи в случае, когда игрок $И_1$ выбирает i-ю полную стратегию, а игрок $И_2$ j-ю полную стратегию. Пусть X означает некоторую смешанную стратегию игрока $И_1$. X представляет собой вектор, координатами которого являются вероятности выбора полных страте-

[1] При первом чтении описание процесса сведения задачи нахождения оптимальных смешанных стратегий к задаче линейного программирования можно опустить.

гий игроком $И_1$. Таким образом, $X=\{p_1, p_2, \ldots, p_m\}$. Естественно, что все

$$0 \leqslant p_i \leqslant 1 \text{ и } \sum_{i=1}^{m} p_i = 1.$$

Аналогично $Y=\{q_1, q_2, \ldots, q_n\}$ означает некоторую смешанную стратегию игрока $И_2$. Обозначим через $E(X, Y)$ — математическое ожидание выигрыша для игрока $И_1$ при условии, что игроки $И_1$ и $И_2$ выбирают соответственно смешанные стратегии X и Y. Если через X^* и Y^* обозначить оптимальные смешанные стратегии игроков, то имеют место очевидные неравенства, вытекающие из определения оптимальных смешанных стратегий:

$$E(X^*, Y) \geqslant E(X^*, Y^*);$$

$$E(X, Y^*) \leqslant E(X^*, Y^*).$$

Величина $E(X^*, Y^*)$ есть цена игры. Обозначим ее через v. Если игрок $И_1$ выбирает свою оптимальную смешанную стратегию, а его противник — j-ю чистую полную стратегию, то математическое ожидание выигрыша для игрока $И_1$ удовлетворяет неравенству

$$p^*_1 f_{1j} + p^*_2 f_{2j} + \ldots + p^*_m f_{mj} \geqslant v.$$

Меняя значение $j=1, 2, \ldots, n$, мы получим систему n неравенств указанного вида. Разделим обе части неравенств на v и введем обозначения $\bar{p}_i = p^*_i/v$. Тогда получим систему неравенств:

$$\begin{cases} \bar{p}_1 f_{11} + \bar{p}_2 f_{21} + \ldots + \bar{p}_m f_{m1} \geqslant 1; \\ \bar{p}_1 f_{12} + \bar{p}_2 f_{22} + \ldots + \bar{p}_m f_{m2} \geqslant 1; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \bar{p}_1 f_{1n} + \bar{p}_2 f_{2n} + \ldots + \bar{p}_m f_{mn} \geqslant 1; \end{cases}$$

здесь

$$\bar{p}_i \geqslant 0 \text{ и } \sum_{i=1}^{m} \bar{p}_i = \frac{1}{v} \sum_{i=1}^{m} p^*_i = \frac{1}{v} = \bar{v}.$$

Игрок $И_1$ стремится максимизировать цену игры v (или, что то же самое, минимизировать величину $\bar{v} =$

$=1/v$). Другими словами, он стремится минимизировать линейную форму

$$L=\bar{v}=\bar{p}_1+\bar{p}_2+\ldots+\bar{p}_m.$$

Таким образом, задача поиска оптимальной смешанной стратегии сведена нами к поиску таких значений $\bar{p}_i$, которые удовлетворяют системе вышеприведенных неравенств, и таких, при которых линейная форма L минимальна. Такая постановка задачи есть классическая постановка задачи линейного программирования и для решения ее могут быть использованы любые универсальные методы линейного программирования (например, симплекс-метод).

Сложность вычислений с помощью универсальных методов с ростом числа полных стратегий игроков растет столь быстро, что практически эти методы очень скоро перестают быть эффективными, даже при наличии средств вычислительной техники. Поэтому наряду с поиском универсальных методов решения основной задачи теории игр строятся и эффективные частные методы решения игровых задач определенных классов, которые являются более эффективными при решении задач данного класса по сравнению с универсальными методами.

Одним из методов подобного рода является метод геометрического решения задачи о нахождении оптимальной смешанной стратегии игроков. Этот метод может эффективно применяться в играх, для которых выполнено условие, что у одного из игроков имеется не более двух полных стратегий. Поясним сущность геометрического метода на следующем простом примере. Пусть мы имеем матрицу игры следующего вида:

$$\begin{array}{cc} & \begin{array}{cc} S_1'^2 & S_2'^2 \end{array} \\ \begin{array}{c} S_1^1 \\ S_2^1 \end{array} & \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} \end{array}.$$

Найдем для этой игры значения p_1 и p_2, определяющие оптимальную смешанную стратегию для игрока $И_1$. Так как $p_1+p_2=1$, то все смешанные стратегии игрока $И_1$ можно изобразить геометрически точками на оси абсцисс, лежащими на отрезке [0, 1], следующим образом.

Возьмем отрезок длины единица и отождествим концы этого отрезка с выбором полных стратегий игроком $И_1$. Для левого конца отрезка при этом $p_1=0$, $p_2=1$, а для правого конца отрезка $p_1=1$, а $p_2=0$. Любая внутренняя точка этого отрезка будет характеризоваться двумя координатами, определяющимися как длины отрезков между этой точкой и концевыми точками отрезка (рис. 2). При этом выполняется основное соотноше-

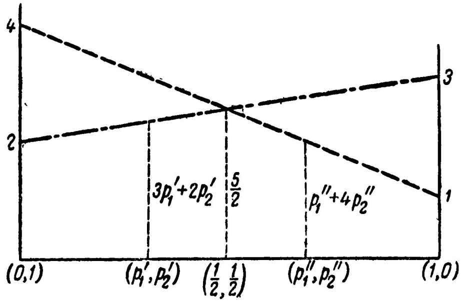


Рис. 2.

ние $p_1+p_2=1$. Очевидно, что множество точек рассматриваемого отрезка совпадает с множеством всех возможных смешанных стратегий для игрока $И_1$. Если игрок $И_2$ выбирает свою первую полную стратегию, то математическое ожидание выигрыша для игрока $И_1$ равно $p_1+4\,p_2$ и геометрически изображается прямой, показанной на рис. 2 пунктиром. Если же игрок $И_2$ выбирает свою вторую полную стратегию, то математическое ожидание выигрыша для первого игрока есть $3\,p_1+2\,p_2$ и геометрически этому математическому ожиданию соответствует на рис. 2 прямая, показанная на этом рисунке как пунктирно-точечная прямая. За счет удачного выбора смешанной стратегии игрок $И_1$ старается увеличить свой выигрыш. Этому соответствует геометрически нахождение такой точки на рассматриваемом отрезке, в которой величина ординаты была бы по возможности наибольшей. Игрок $И_2$, наоборот, стремится уменьшить выигрыш игрока $И_1$ (при рассматриваемой матрице игры игрок $И_2$ всегда проигрывает и его цель состоит в минимизации этого проигрыша). Геометрически действия игрока $И_2$ заключаются в выборе такой прямой, характеризующей математическое ожидание выигрыша для $И_1$, на которой ордината, соответствующая точке смешанной стратегии, используемой $И_1$, имела бы наименьшее значение. Совместные усилия обоих игроков приводят к тому, что геометрическая характеристика их поведения отражается так, как это показано на рис. 3 жирной линией. Как видно из этого рисунка, существует только одна оптимальная для игрока $И_1$ максиминная страте-

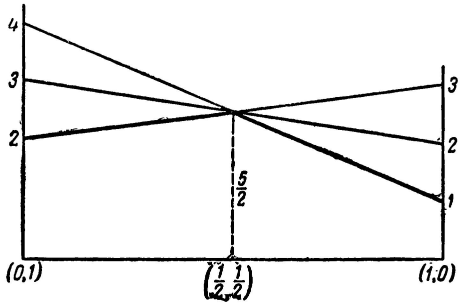


Рис. 3.

гия, которая соответствует точке с координатами ($^1/_2$, $^1/_2$). Цена игры равна значению ординаты в этой точке $v = {^5/_2}$.

Рассмотрим теперь действия игрока $И_2$. Для этого игрока также необходимо найти оптимальную смешанную стратегию, которая не позволила бы игроку $И_1$ получить выигрыш, превышающий цену игры. При этом игрок $И_2$ не может выбирать в качестве оптимальной стратегии какую-либо из своих полных стратегий, ибо это привело бы к увеличению выигрыша игрока $И_1$. Если бы, например, игрок $И_2$ в качестве своей стратегии выбрал бы свою первую полную стратегию, то, узнав об этом, игрок $И_1$ мог бы в качестве стратегии игры выбрать свою вторую полную стратегию и получил бы выигрыш, равный четырем. Поэтому игрок $И_2$ выбирает такие значения q_1 и q_2, при которых математическое ожидание выигрыша для игрока $И_1$ равно $^5/_2$. Математическое ожидание выигрыша первого игрока есть $q_1(p_1+4p_2)+q_2(3p_1+2p_2)$. Тогда смешанная стратегия игрока $И_2$, характеризующаяся значениями q_1 и q_2, геометрически представляется прямой, целиком заключенной между прямыми, показанными на рис. 2 пунктиром и пунктиром с точками. Эта прямая должна проходить через точку, соответствующую максиминной стратегии игрока $И_1$. Нетрудно обнаружить, что оптимальная смешанная стратегия для игрока $И_2$ геометрически определяется прямой, лежащей между прямыми, характеризующими математическое ожидание выигрыша для $И_1$. Отклонение от этой прямой позволяет для $И_1$ увеличить

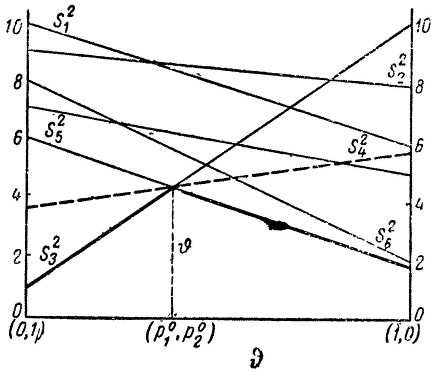


Рис. 4.

свой выигрыш по сравнению с ценой игры за счет выбора удачной полной стратегии. Таким образом, оптимальная минимаксная смешанная стратегия для игрока $И_2$ в рассматриваемом примере определяется значениями $q_1 = {}^1/_4$ и $q_2 = {}^3/_4$.

Рассмотренный нами метод геометрического нахождения оптимальных смешанных стратегий можно обобщить на случай, когда игрок $И_2$ имеет не две полных стратегии, а m таких стратегий. На рис. 4 показана геометрическая интерпретация игры, в которой у игрока $И_2$ имеется шесть полных стратегий. Жирной линией показана ломаная, соответствующая гарантированному выигрышу игрока $И_1$ при выборе им различных смешанных стратегий и выборе игроком $И_2$ наихудшей для игрока $И_1$ полной стратегии. Точка (p_1^0, p_2^0) соответствует максиминной смешанной стратегии для игрока $И_1$. Для игрока $И_2$ оптимальная смешанная стратегия определяется смешением полных стратегий S_3^2 и S_5^2 (и только их), причем q_1 и q_2 определяются из условия того, что математическое ожидание выигрыша для игрока $И_1$ равно v.

В случае, когда число полных стратегий игроков становится больше двух, геометрическое решение становится затруднительным, так как размерность пространства, в котором происходят все построения, определяется минимальным числом полных стратегий, имеющихся в распоряжении игроков.

В задачах теории автоматического регулирования роль противника чаще всего играет то, что в теории игр принято называть природой. Под природой в этой теории понимается игрок, действия которого не направлены на достижение оптимальной стратегии, т. е. такой игрок, которому безразлична величина его проигрыша или выигрыша. Если при этом известно априорное распределение вероятностей, с которыми природа выбирает те или иные полные стратегии, то возникает задача нахождения такой полной стратегии, которая обеспечивала бы оптимальное поведение игрока. Такая задача носит название байесовской задачи принятия решений и изучается в теории принятия решений. Если, например, в рассмотренном нами выше примере игры с матрицей

$$\begin{array}{cc} & \begin{array}{cc} S_1^2 & S_2^2 \end{array} \\ \begin{array}{c} S_1^1 \\ S_2^1 \end{array} & \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} \end{array}$$

игрок $И_2$ всегда придерживается стратегии с $q_1 = {}^1/_3$ и $q_2 = {}^2/_3$, то игрок $И_1$ с помощью стратегии S_2^1 может получить выигрыш, равный ${}^8/_3$, в то время, как при игре с противником, пользующимся оптимальной стратегией, он может получить выигрыш, равный лишь ${}^5/_2$.

Если априорное распределение вероятностей, с которыми противник игрока $И_1$ выбирает свои стратегии, неизвестно, то игроку $И_1$ необходимо выбирать свою стратегию, основываясь на некоторых априорных предположениях о характере игры $И_2$. Эти предположения формулируются обычно в виде некоторого набора правил, содержащих в своей основе некоторый априорный критерий поведения игрока $И_2$, с помощью которого игрок $И_1$ может однозначно выбирать свою смешанную стратегию. В литературе известно несколько таких априорных критериев поведения игрока $И_2$. Одним из наилучших критериев является критерий, известный в литературе под названием критерий Гурвица. Критерий Гурвица устроен следующим образом. Пусть для полной стратегии S^1_i a_i означает наименьший проигрыш для игрока $И_1$, а d_i — наибольший выигрыш для игрока $И_1$, если игрок $И_1$ выберет полную стратегию S^1_i. Тогда построим функцию

$$f(i) = \alpha a_i + (1-\alpha) d_i, \qquad (2)$$

которую будем называть функцией полезности выбора игроком $И_1$ стратегии $S^1{}_i$. Значение α заключено между нулем и единицей. Критерий Гурвица утверждает, что игрок $И_1$ должен выбирать полную стратегию, имеющую наибольшее значение функции полезности. При $\alpha=1$ этот критерий можно назвать критерием максиминной полезности. В случае $\alpha=0$ — критерием максимаксной полезности. В первом случае мы предполагаем, что природа ведет себя как игрок, выбирающий оптимальную стратегию игры, а во втором — предполагаем, что природа играет наилучшим для нас образом. Очевидно, более реально рассматривать α, принимающее промежуточное значение между нулем и единицей. Выбор этого значения носит субъективный характер и зависит от оценки игроком $И_1$ характера игры природы. Отметим, что критерий Гурвица в ряде практически важных случаев может привести к неправильным результатам.

В качестве другого критерия можно рассмотреть критерий, основанный на предположении, что все полные стратегии игроком $И_2$ выбираются с одинаковой вероятностью. Однако и этот критерий может быть подвергнут весьма обоснованной критике. В дальнейшем мы еще вернемся к обсуждению игр с природой и критериев, связанных с выбором стратегии для игрока $И_1$. Более полные сведения об играх с природой в условиях частичного или полного незнания о поведении природы можно найти в книгах, относящихся к теории принятия решений, список которых дан в конце книги.

В заключение этого параграфа дадим описание типов игр, которые нам будет удобно выделить при дальнейшем изложении материала. В силу того, что классификация по типам происходит на основе различных признаков, конкретные игры могут принадлежать одновременно нескольким из нижеперечисленных типов.

1. Игры с полной информацией

Такие игры характеризуются тем, что все игроки имеют информационные множества, состоящие только из единичных вершин дерева игры. Другими словами, в таких играх все игроки знают всю предысторию игры и настоящие возможности всех игроков. Все игры с полной информацией имеют седловую точку, и на основании анализа матрицы игры игроки могут найти свои минимаксные и максиминные стратегии и определить цену игры. После этого игра практически заканчивается. С прикладной точки зрения удобно разделить игры с полной информацией на два подкласса: игры с большим числом полных стратегий и игры с малым числом полных стратегий. Конечно, это деление весьма условно и зависит от мощности вычислительных средств, с помощью которых производится поиск оптимальных стратегий и определение значения цены игры. Таким образом, выделение конкретной игры в один из двух названных подклассов целиком определяется размерами матрицы этой игры и конструкцией устройства, используемого для решения игровой задачи. Игры с полной информацией будут рассмотрены нами в § 3 и 4.

2. Игры с неполной информацией

Для таких игр информационные множества для некоторых (или всех) игроков содержат более одной вершины дерева игры, что вызывает вероятностный характер оценок при выборе очередного хода в игре и, следовательно, стратегии, применяемой игроками в данной партии. Если известно, что противник стремится к максимизации своего выигрыша, то на основании теоремы о минимаксе можно найти оптимальные смешанные стратегии и определить цену игры. Если же один или несколько из игроков не стремятся к максимизации своих выигрышей, то для поиска стратегии каждого из игроков, стремящихся к оптимальному характеру своей игры, необходимо использовать аппарат теории принятия решений. Игры с неполной информацией первого типа рассматриваются в § 5, а пример игры с неполной информацией второго типа рассмотрен в § 6.

3. Игры со случайными ходами

В таких играх в качестве одного из игроков выступает «случай», т. е. природа. Такие игры по существу сводятся к играм с неполной информацией с участием природы. Примеры таких игр даны в § 6 и 7.

4. Коалиционные игры

Коалиционные игры отличаются от прочих игр тем, что в процессе игры некоторые участники могут образовывать временные или постоянные коалиции с договорным распределением выигрыша между участниками коалиции. В результате предварительных переговоров между игроками коалиции игроки могут более точно оценить намерения друг друга и свои возможности и принять для себя более выгодные решения. Примерами таких игр могут служить арбитражные споры, торг, дипломатические переговоры и т. д. Теория таких игр пока еще слабо развита. Примеры таких игр будут рассмотрены в § 8.

2. КОНЕЧНЫЕ АВТОМАТЫ И ИГРЫ

На протяжении всего дальнейшего изложения управляющее устройство будет мыслиться нами как конечный автомат, реализованный схемно или в виде программы для цифровой вычислительной машины.

Рассмотрим устройство, изображенное на рис. 5. Оно имеет n входов и m выходов. На каждый вход устройства может быть подан любой сигнал из входного алфавита

$$X=\{x_0,\ x_1,\ x_2,\ \ldots,\ x_{l-1}\}.$$

Комбинация сигналов входного алфавита, одновременно поданных на все входы устройства, образует входное слово в момент времени t и будет обозначаться как X^t. Сигнал x_0 играет особую роль. Он является сигналом отсутствия подачи сигнала на данный вход (пустой сигнал). Аналогично на каждом выходе рассматриваемого устройства может появиться любой сигнал из алфавита

$$Y=\{y_0,\ y_1,\ y_2,\ \ldots,\ y_{s-1}\},$$

называемого выходным алфавитом. Выходное слово в момент времени t представляет собой набор всех сигналов, появившихся на выходах устройства в этот момент времени. Сигнал y_0 играет роль пустого символа.

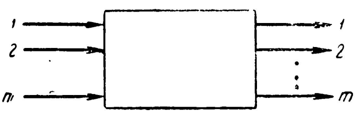


Рис. 5.

В силу конечности алфавитов X и Y существует лишь конечное число входных и выходных слов. Их число соответственно равно l^n и s^m.

Если работа автомата задается таблицей вида

Вход	Выход
X_1^t	$Y_{j_1}^t$
X_2^t	$Y_{j_2}^t$
....	....
$X_{l^n}^t$	$Y_{j_{l^n}}^t$

$$j_i = 1, 2, \ldots, m^s, \qquad (3)$$

то автомат называется конечным автоматом без памяти. Реакция такого автомата на входное слово определяется только видом этого слова и не зависит от предыстории работы автомата.

Рассмотрим теперь алфавит

$$Z=\{z_1,\ z_2,\ \ldots,\ z_r\},$$

который будем называть алфавитом внутренних состояний автомата. Символы z_i^t характеризуют состояние автомата в момент времени t. Если работа автомата определяется не только входным словом, поданным на его вход в момент времени t, но и тем состоянием, в котором находится в это время автомат, то такой автомат называется конечным автоматом с памятью. Его работа задается с помощью таблицы следующего вида:

$$
\begin{array}{c|l}
X_1^t, z_1^t & Y_{j_{11}}^t, z_{q_{11}}^t, \\
X_1^t, z_2^t & Y_{j_{12}}^t, z_{q_{12}}^t \\
\cdot\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
X_{l^n}^t, z_r^t & Y_{j_{l^n r}}, z_{q_{l^n r}}
\end{array}
\qquad
\begin{array}{l}
j_{ij} = 1, 2, \ldots, m^s, \\
q_{ij} = 1, 2, \ldots, r.
\end{array}
\tag{4}
$$

Число строк этой таблицы равно $l^n r$.

Оба типа рассмотренных конечных автоматов характеризуются тем, что задание входного слова или входного слова и состояния автомата однозначно определяют его выходное слово или выходное слово и новое состояние. Такие автоматы мы будем называть детерминированными автоматами. Детерминированные автоматы тесно связаны с игровыми задачами, характеризующими игры с полной информацией. Для игр с неполной информацией более естественно рассмотреть автоматы, для которых принимаемые решения (выходные слова) носят вероятностный характер. Работа такого автомата в случае автомата с памятью определяется двумя матрицами следующего вида:

$$
A = \begin{array}{c}
\\
X_1^t, z_1^t \\
X_1^t, z_2^t \\
\cdot \\
\cdot \\
\cdot \\
X_{l^n}^t, z_r^t
\end{array}
\begin{array}{c}
\begin{array}{ccc} Y_1^t & Y_2^t \ldots & Y_{m^s}^t \end{array} \\
\begin{bmatrix}
p_{11}^1 & p_{11}^2 \ldots & p_{11}^{m^s} \\
p_{12}^1 & p_{12}^2 \ldots & p_{12}^{m^s} \\
\cdot & \cdot\ \cdot\ \cdot & \cdot \\
\cdot & \cdot\ \cdot\ \cdot & \cdot \\
\cdot & \cdot\ \cdot\ \cdot & \cdot \\
p_{l^n r}^1 & p_{l^n r}^2 \ldots & p_{l^n r}^{m^s}
\end{bmatrix}
\end{array}
\tag{5}
$$

и

$$B = \begin{array}{c} \\ X_1^t,\ z_1^t \\ X_1^t,\ z_2^t \\ \cdot \\ \cdot \\ \cdot \\ X_{l^n}^t,\ z_r^t \end{array} \begin{array}{c} \begin{array}{cccc} z_1^{t+1} & z_2^{t+1} & \ldots & z_r^{t+1} \end{array} \\ \begin{bmatrix} q_{11}^1 & q_{11}^2 & \ldots & q_{11}^r \\ q_{12}^1 & q_{12}^2 & \ldots & q_{12}^r \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ q_{l^n r}^1 & q_{l^n r}^2 & \ldots & q_{l^n r}^r \end{bmatrix} \end{array}. \tag{6}$$

Первая матрица определяет вероятности появления тех или иных выходных слов при подаче на вход автомата данного входного слова и при наличии определенного внутреннего состояния автомата. Вторая матрица для тех же исходных условий определяет вероятности перехода автомата в новое внутреннее состояние. При этом, конечно, $0 \leqslant p_{ij},\ q_{ij}^k \leqslant 1$ и выполнены равенства:

$$\sum_{k=1}^{m^s} p_{ij}^k = 1 \quad \text{и} \quad \sum_{k=1}^{r} q_{ij}^k = 1 \quad \begin{pmatrix} i = 1, 2, \ldots, l^n \\ j = 1, 2, \ldots, r \end{pmatrix}.$$

Автоматы такого типа мы будем называть конечными автоматами марковского типа или просто марковскими автоматами. В частном случае, когда в матрицах A и B в каждой строке имеется только одна единица, а остальные элементы строки равны нулю, марковский автомат превращается в конечный автомат с памятью. Если множество внутренних состояний автомата состоит из одного элемента, то работа марковского автомата определяется одной матрицей следующего вида:

$$C = \begin{array}{c} \\ X_1^t \\ X_2^t \\ \cdot \\ \cdot \\ \cdot \\ X_{l^n}^t \end{array} \begin{array}{c} \begin{array}{cccc} Y_1^t & Y_2^t & & Y_{m^s}^t \end{array} \\ \begin{bmatrix} p_1^1 & p_1^2 & \ldots & p_1^{m^s} \\ p_2^1 & p_2^2 & \ldots & p_2^{m^s} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ p_{l^n}^1 & p_{l^n}^2 & \ldots & p_{l^n}^{m^s} \end{bmatrix} \end{array}. \tag{7}$$

Такой автомат обычно называют преобразователем законов распределения. Мы будем называть его марковским автоматом без памяти.

Все рассмотренные автоматы характеризуются тем, что в процессе их работы таблицы, определяющие его выходное слово и внутреннее состояние или матрицы A и B, не меняют своего вида. Поэтому все такие автоматы называются автоматами с постоянной структурой. Кроме автоматов с постояной структурой, для теории игр представляют интерес автоматы с переменной структурой, которые характеризуются тем, что в зависимости от предшествующего опыта реакция автомата на входное слово меняется. Автомат с переменной структурой можно рассматривать как марковский автомат, у которого после очередного такта работы происходит перерасчет значений p_{ij}^{k} и q_{ij}^{k} из матриц A и B с помощью соотношений:

$$\left.\begin{aligned} \tilde{p}_{ij}^{k} &= p_{ij}^{k} + \Delta p_{ij}^{k}; \\ \tilde{q}_{ij}^{k} &= q_{ij}^{k} + \Delta q_{ij}^{k}. \end{aligned}\right\} \tag{8}$$

Добавки Δp_{ij}^{k} и Δq_{ij}^{k} выбираются таким образом, чтобы выполнялись равенства:

$$\left.\begin{aligned} \sum_{k=1}^{m^s} \Delta p_{ij}^{k} &= 0; \\ \sum_{k=1}^{r} \Delta q_{ij}^{k} &= 0, \end{aligned}\right\} \tag{9}$$

а новые элементы матриц A и B $\tilde{p}_{ij}^{k}$ и $\tilde{q}_{ij}^{k}$ должны лежать в пределах от нуля до единицы. Конкретные значения Δp_{ij}^{k} и Δq_{ij}^{k} выбираются на основе некоторого знания о характере задачи или подбираются эмпирическим путем. Автоматы с переменной структурой можно отождествить с линейными моделями обучения, рассматривавшимися рядом авторов в связи с общей теорией обучения[1].

---

[1] Э. Буш, Мостеллер, Стохастические модели обучения, Изд-во иностранной литературы, 1962.

Все рассмотренные автоматы являются табличными автоматами, так как их работа может определяться таблицами. При программной реализации таких автоматов в памяти машины необходимо хранить таблицы выходов и смены состояний для детерминированных автоматов или значения элементов матриц A и B для недетерминированных автоматов. При больших значениях l^n, m^s и r необходимый объем памяти может оказаться слишком большим. Однако на основании результатов, полученных в теории конечных автоматов, всегда возможен переход от табличного задания логики работы детерминированного автомата к функциональному представлению значения выхода через значения входа и внутреннего состояния[1]. Такие способы задания автоматов более удобны для программной реализации. При дальнейшем изложении материала мы не будем акцентировать внимание на форме здания работы автомата, а будем интересоваться главным образом возможностью построения автомата детерминированного или недетерминированного типа с постоянной или переменной структурой для решения игровых задач того или иного типа.

Рассмотрим более подробно, как связаны между собой структура конечного автомата и тип игры и как с помощью конечного автомата можно имитировать игру. Для этого мы будем изображать работу автомата графически с помощью диаграмм, построенных по следующему принципу. Состояниям автомата мы сопоставим вершины в диаграмме, а возможные переходы из состояния в состояние под воздействием входных слов — изобразим стрелками, направленными из данной вершины диаграммы в вершину, соответствующую нужному состоянию. Над этой стрелкой мы будем писать входное слово, соответствующее такому переходу, а рядом в скобках указывать выходное слово, которое при этом получается. На рис. 6 показан пример такой диаграммы. Если внутреннее состояние автомата в момент времени t есть состояние, обозначенное номером 3, то при подаче на вход автомата слова X_3^t возникнет переход автомата во

---

[1] См., например, изложение метода такого перехода в книге Д. А. Поспелова, Логические методы анализа и синтеза схем, изд-во «Энергия», 1964.

внутреннее состояние с номером 1 и на выходе автомата появится выходное слово Y^t_2. Если автомат недетерминирован, то, кроме отметки на стрелках переходов о входном и выходном словах, соответствующих этому переходу, необходимо еще отмечать значение вероятности

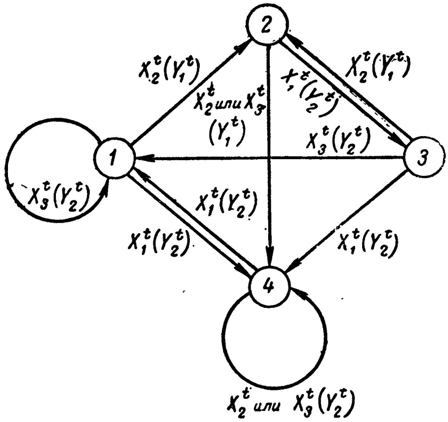


Рис. 6.

такого перехода. Для автоматов с переменной структурой диаграмма переходов меняется во времени.

Будем предполагать, что на вход автомата подается входное слово, несущее информацию об очередном ходе противника, а в памяти автомата хранятся все возможные чистые стратегии, которые могут быть выбраны автоматом на всех этапах игры.

Рассмотрим вначале игру с полной информацией в предположении, что противник, играющий с автоматом, может допускать ошибки в игре (т. е. придерживаться неоптимальной стратегии). В этом случае автомат должен использовать допускаемые противником ошибки для увеличения своего выигрыша.

Рассмотрим игру, дерево которой показано на рис. 1. Это дерево нетрудно отобразить как диаграмму переходов автомата, который играет оптимальным образом. Такая диаграмма переходов показана на рис. 7,*а*. Как следует из этой диаграммы, при безошибочной игре противника автомат получает выигрыш, равный цене игры, а при ошибке, допущенной противником, автомат увеличивает свой выигрыш до максимально возможного. Первому случаю на диаграмме переходов соответствует заключительное состояние *3*, а второму — заключительное состояние *4*. Состояние *1* соответствует началу игры. Так, в начале игры автомат не имеет никакой внешней информации о ходах противника.

Рис. 7.

Это происходит потому, что автомат «ходит» первым, а его противник еще не «ходил», то есть на его вход ничего не подается. Положение, когда автомат игры не имеет внешней информации на диаграмме переходов условно обозначено подачей на вход автомата пустого слова λ. Диаграмма, показанная на рис. 7,*а*, отображает только ту часть дерева игры, показанного на рис. 1, которая приведена на рис. 7,*б*.

Диаграмма переходов может быть реализована либо схемно, либо в виде программы для универсальной вычислительной машины. Эта программа будет представлять собой таблицу с двумя выходами, характеризующими ситуацию на данном этапе игры перед очередным ходом автомата, и выбранный ход автомата. На пересечении соответствующих входов в таблице стоит оценка ситуации, получающейся после выбранного хода машины. Для игр с полной информацией выбор очередного хода машины производится при неизменных оценках выбора, которые в любом случае обеспечивают машине выигрыш, равный цене игры, а при ошибках противника позволяют увеличить его. Заметим, что при большом числе полных стратегий такой способ решения игр может оказаться весьма плохим и более эффективными являются способы решения, учитывающие специфику данной конкретной игры (например, наличие эквивалентных стратегий, наличие симметрии в матрице игры и т. д.). При рассмотрении конкретных примеров игр с полной информацией мы в дальнейшем будем везде, где это возможно, использовать частные особенности игры для уменьшения сложности автомата, дающего решение игры.

Описанный нами табличный автомат для игр с полной информацией является конечным автоматом без памяти и его работа определяется таблицей вида (3).

Если автомат выбирает не полную оптимальную стратегию, а лишь ходовую оптимальную стратегию (в предположении, что на основании предварительного анализа игры удалось найти последовательность ходовых оптимальных стратегий, дающих в совокупности полную оптимальную стратегию), то при одной и той же ситуации в игре действия автомата могут быть различными, так как выбор хода определяется не только данной ситуацией, сложившейся перед очередным ходом автомата, но и предысторией игры. В этом случае реакция автомата определяется двумя параметрами: входом — ситуацией в настоящее время и состоянием памяти автомата, определяемым предысторией его работы. Такой автомат является конечным автоматом с памятью, а общий характер его функционирования определяется с помощью задания таблицы вида (4).

Рассмотрим теперь случай игры с неполной информацией. В этом случае у автомата отсутствует информация, которая позволила бы ему однозначно выбрать свой ход или свою полную стратегию в игре.

Рассмотрим несколько случаев, которые сводятся к играм с неполной информацией. Если в игре интересы игроков противоположны и есть основание считать, что противник автомата играет оптимальным образом, то для автомата оптимальным является такой выбор своих стратегий, который соответствует его оптимальной смешанной стратегии, определяемой на основе теоремы об оптимальных смешанных стратегиях. Таким образом, автомат в этом случае должен выдавать на выходе оптимальную смешанную стратегию. Если отождествить выходной сигнал автомата с номером выбираемой на данном этапе игры полной стратегии, то такой автомат будет автоматом марковского типа без памяти, работа которого определяется матрицей C (соотношение (7)), имеющей одну строку и такое число столбцов, которое соответствует числу полных стратегий, имеющихся у автомата. Если выход автомата отождествляется с выбором хода на данном этапе игры, то в матрице C число строк соответствует числу ситуаций в игре (ситуаций в данный момент в совокупности с известной автомату предысторией игры), а число столбцов матрицы C соответствует максимальному числу всех возможных выборов хода.

В последнем случае, естественно, перейти от марковского автомата без памяти к марковскому автомату с памятью, работа которого определяется с помощью матриц A и B (соотношения (5) и (6)). В этом случае предыстория игры, известная автомату, будет соответствовать определенному состоянию автомата, а значение входа будет соответствовать ситуации в данный момент игры.

Если допустить, что противник может ошибаться или что противником автомата является природа, то автомат должен использовать эти ошибки или неоптимальную игру природы в целях увеличения своего выигрыша. Если противником автомата является природа, применяющая известную смешанную стратегию, то автомат должен отвечать на нее полной стратегией, обеспечивающей ему наибольший выигрыш. Если эта смешанная стратегия остается неизменной, то автомат может быть выполнен в виде конечного немарковского автомата без памяти (выдающего номера полных стратегий) или с памятью (выдающего последовательность ходов). Если же априорные вероятности смешивания заранее неизвестны, то в конструкции автомата должен быть предусмотрен блок, с помощью которого можно оценивать относительные частоты применения противником своих полных стратегий (в стационарном случае) или оценивать вероятности выбора тех или иных стратегий в нестационарном случае. На рис. 8 показана упрощенная структура автомата, работающая в условиях, когда априорные сведения о выборе противником той или иной стратегии игры отсутствуют. Блок условных вероятностей определяет смешанную (или полную в частном случае) стратегию, которую использует противник на данном этапе игры. По выбранным значениям вероятностей смешивания определяется: использует ли противник оптимальную смешанную стратегию или не

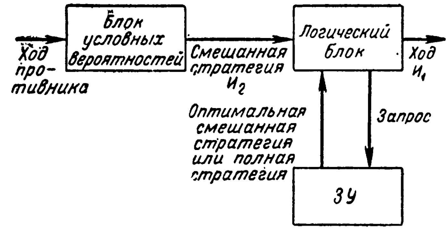


Рис. 8.

использует. В первом случае автомат отвечает своей оптимальной смешанной стратегией (работает как марковский автомат, определяемый матрицей C или матрицами A и B), а во втором случае автомат отвечает полной стратегией, обеспечивающей ему максимально возможный выигрыш. В стационарном случае блок условных вероятностей после достаточного объема наблюдений может прекратить работу и в дальнейшем работа автомата будет происходить как работа автомата с заранее известными априорными вероятностями выбора противником тех или иных стратегий.

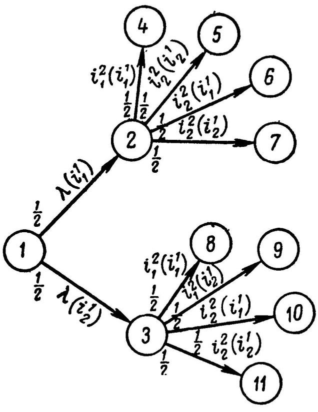


Рис. 9.

Если противник автомата является природой и априорные вероятности выбора им своих стратегий неизвестны, то автомат может использовать для своего функционирования один из возможных критериев, рассмотренных нами в § 1. Если, например, игра определяется деревом, показанным на рис. 1, но значения платежей заранее неизвестны, то автомат, используя критерий одинаковой вероятности может реализовать диаграмму переходов, показанную на рис. 9. В этом случае автомат является марковским автоматом и не содержит блока условных вероятностей.

В целом автомат с блоком условных вероятностей можно рассматривать как автомат с переменной структурой, работа которого определяется матрицами A и B и соотношениями (8). В случае стационарной игры противника перестройка структуры автомата должна привести в пределе к автомату, реализующему оптимальную смешанную стратегию или оптимальную полную стратегию и имеющему постоянную структуру. В нестационарном случае автомат является по существу автоматом с переменной структурой.

Автоматы с переменной структурой будут рассмотрены нами более подробно в § 7.

3. ИГРЫ С ПОЛНОЙ ИНФОРМАЦИЕЙ И МАЛЫМ ЧИСЛОМ СТРАТЕГИЙ

Игры такого типа не представляют большого теоретического интереса, ибо при небольших размерах матрицы игры и наличии в матрице игры седловой точки легко осуществить полный анализ игры и найти оптимальные стратегии для игроков. Поэтому в случае игры, когда оба противника выбирают свои оптимальные стратегии, игра по существу полностью заканчивается и сам процесс игры (реализация ходов в соответствии с выбранными максиминной и минимаксной стратегиями) никакой новой информации не несет.

Однако в системах автоматического регулирования роль одного из игроков часто играет природа, не стремящаяся к ведению игры оптимальным образом. Поэтому при конструировании автомата, играющего в такую игру, возникает проблема создания конструкции автомата, которая позволила бы ему получить возможно больший выигрыш при любом выборе природой своей полной стратегии.

Пусть, например, игра задается с помощью матрицы следующего вида:

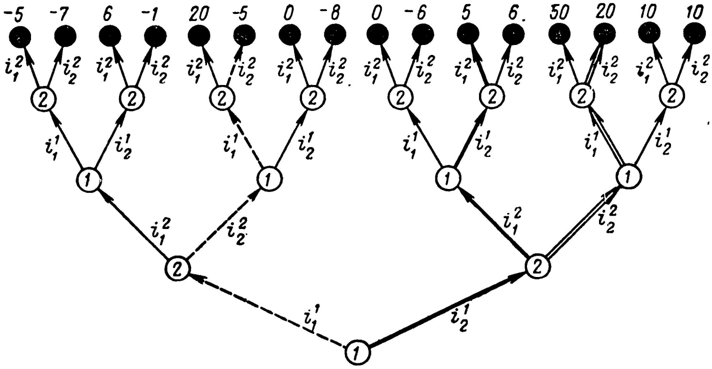
$$\begin{pmatrix} -5 & -7 & 20 & -5 \\ 6 & -1 & 0 & -8 \\ 0 & -6 & 50 & 20 \\ 5 & 8 & 10 & 10 \end{pmatrix}$$

и пусть роль игрока $И_1$ выполняет автомат, а в качестве игрока $И_2$ выступает природа, которая выбирает свою полную стратегию, не сообразуясь с правилом выбора оптимальных стратегий. Если бы $И_2$ был заинтересованным игроком, то на основе анализа матрицы игры он бы обнаружил, что у него существует оптимальная минимаксная стратегия S_1^2, на которую игрок $И_1$ мог бы ответить наилучшим для него образом, используя свою максиминную стратегию S_4^1. При этом $И_1$ получал бы в каждой партии игры выигрыш, равный пяти единицам. Учитывая, что игрок $И_2$ является незаинтересованным игроком, игрок $И_1$ (автомат) должен использовать ошибки, допускаемые природой, для максимального увеличения своего выигрыша.

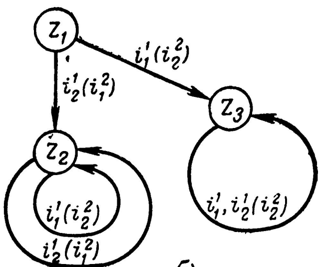
Если игроку $И_1$ известен выбор полной стратегии игроком $И_2$, то для этой полной стратегии $И_1$ выбирает такую ответную полную стратегию, которая максимизирует его выигрыш. Для возможности организации такого ответа в памяти автомата должна храниться таблица, состоящая из двух строк. В верхней строке перечислены все полные стратегии природы, а в нижней — те полные стратегии автомата, которые обеспечивают ему наибольший выигрыш при данной полной стратегии природы. Для вышеприведенной матрицы игры такая таблица имеет следующий вид:

Стратегии природы	S_1^2	S_2^2	S_3^2	S_4^2
Стратегии автомата	S_2^1	S_4^1	S_3^1	S_3^1

Если оценка полной стратегии, выбранной природой, до начала игры автоматом невозможна до появления некоторой ходовой стратегии, однозначно определяющей выбранную природой полную стратегию, то для правильной работы автомата необходимо построить таблицу, которая была бы устроена так же, как и вышеприведенная таблица, но в строках этой таблицы были бы



а)



б)

Рис. 10.

перечислены не полные стратегии игроков, а ходовые стратегии, которые они могут применять в процессе игры.

Для игры, дерево которой показано на рис. 10,*а*, такая стратегия имеет следующий вид:

Ход $И_2$	λ	i_1^2	i_2^2
Ход $И_1$	i_2^1	i_2^1	i_1^1

Здесь λ обозначает сигнал начала игры, в соответствии с которым автомат делает свой первый ход.

Эта таблица позволяет автомату либо реализовать оптимальную максиминную стратегию, либо увеличить свой выигрыш при ошибках в игре, допущенных противником. Оптимальная максиминная стратегия автомата показана на рис. 10,*а* в виде пути по дереву, обозначенного жирной линией. В начале игры автомат всегда выбирает ход i_2^1, так как при выборе хода i_1^1 он в худшем для него случае может проиграть игроку $И_2$, выплатив ему сумму, равную 5. Путь, отражающий такой процесс игры, показан на рис. 10,*а* пунктиром. После выбора ходовой стратегии игроком $И_2$ результат этого выбора подается на вход автомата. Если на вход автомата поступает сигнал о том, что $И_2$ выбрал стратегию i_1^2, то автомат выбирает i_2^1, обеспечивая себе по крайней мере выигрыш, гарантированный оптимальной максиминной стратегией этого варианта. В случае, если стратегия, выбранная $И_2$, есть i_2^2, автомат выбирает стратегию i_1^1, обеспечивающую ему выигрыш, равный 20 единицам. На рис. 10,*а* этот случай игры показан двойной линией.

Если теперь отождествить автомат с игроком $И_2$, а природу — с игроком $И_1$ и попытаться составить таблицу работы автомата на основе анализа очередного хода противника, то нетрудно убедиться, что составление такой таблицы окажется невозможным. Связано это с тем, что ответ автомата на ходы природы зависит не только от выбранной природой ходовой стратегии, но и от предыстории игры. Это соответсвует тому, что автомат, моделирующий поведение игрока $И_2$, должен быть автоматом с памятью, в отличие от автомата, моделирующего действия игрока $И_1$, который является автоматом без памяти.

Для составления таблицы ответов автомата на ходовые стратегии природы сопоставим каждому ярусу дерева игры, кроме яруса заключительных позиций, одно из внутренних состояний автомата. Для рассматриваемого дерева игры число внутренних состояний автомата будет равно трем. Обозначим эти состояния символами z_1, z_2, z_3. В начале игры автомат находится в состоянии z_1.

В зависимости от хода природы автомат переходит в состояние z_2 или z_3, выдавая на выходе сигнал, соответствующий своему ходу. Свой очередной ход автомат выбирает, находясь в состоянии z_2 или z_3. Во время второго хода состояние автомата не изменяется, так как после этого хода игра заканчивается. Учитывая ход игрока $И_1$ и внутренние состояния автомата на данном шаге игры, можно построить следующую таблицу, определяющую поведение автомата в процессе игры.

Ход природы	i_1^1			i_2^1		
Состояние автомата	z_1	z_2	z_3	z_1	z_2	z_3
Ход автомата	i_2^2	i_2^2	i_2^2	i_1^2	i_1^2	i_2^2

На основании этой таблицы может быть построена диаграмма работы автомата, показанная на рис. 10,*б*.

По этой диаграмме может быть синтезирована логическая схема автомата на основании методов, хорошо известных в теории конечных автоматов[1].

Для программной реализации автомата на цифровой вычислительной машине необходимо переписать таблицу игры в следующей форме

Состояние автомата	Ход противника	
	i_1^1	i_2^1
z_1	i_2^2; z_3	i_1^2; z_2
z_2	i_2^2; z_2	i_1^2; z_2
z_3	i_2^2; z_3	i_2^2; z_3

[1] См., например, книгу В. М. Глушкова «Синтез цифровых автоматов», Физматгиз, 1962.

Такая таблица с двумя входами хранится в памяти машины. После очередного хода противника на основании информации о его ходе и состоянии автомата в данный момент игры на пересечении соответствующих столбца и строки находится ответный ход автомата, который выдается машиной на печать, и новое внутреннее состояние автомата, которое запоминается в отведенной для этого ячейке памяти внутри машины.

Из разобранного нами примера игры с небольшим числом стратегий и полной информацией следует, что на основе анализа игры всегда можно построить схемную или программную реализацию автомата, наилучшим образом ведущего себя в данной игре. Сложность автомата определяется числом ярусов дерева игры, относящихся к ходам противника, и числом различных выборов, которые может сделать противник в результате своего хода. Если при каждом своем ходе противник делает один из r ходов, а число ярусов дерева игры, соответствующих его ходу, равно l, то число внутренних состояний автомата в наихудшем случае равно r^{l-1}, если автомат начинает игру, или $r^{l-1}+1$, если игру начинает противник.

4. ИГРЫ С ПОЛНОЙ ИНФОРМАЦИЕЙ, НО С БОЛЬШИМ ЧИСЛОМ СТРАТЕГИЙ

Хотя с теоретической точки зрения все игры с полной информацией одинаковы, но при попытке на практике использовать метод построения матрицы игры и определения на основе анализа этой матрицы оптимальных стратегий игроков оказывается в целом ряде случаев невозможным из-за больших размеров этой матрицы. Типичным примером такого случая является игра в шахматы. При игре в шахматы на каждом шаге игры оба игрока имеют полную информацию о ходе игры. Существует лишь конечное число стратегий, которые могут быть использованы игроками. Казалось бы, в силу того, что для игры в шахматы существуют оптимальные стратегии, игра в шахматы не представляет никакого интереса, ибо, если оба игрока выбирают оптимальные стратегии, результат игры не зависит от игроков. Однако, несмотря на свое многовековое существование, шахматы до сих пор остаются одной из самых увлекательных и сложных игр, в которых максимально проявляется умение каждого игрока выбрать в данной позиции наиболее сильное продолжение игры. Вся трудность «исчерпания» возможностей шахматной игры заключается в огромном числе возможных стратегий для каждого игрока. Если допустить, что партия заканчивается вничью после пятидесяти ходов, сделанных одним из игроков подряд без движения пешек и взятия фигур, то число ходов партии ограничено числом 6 300, а число выборов на каждом этапе игры не превышает 2 732. Практически невозможно построить матрицу шахматной игры и найти оптимальные полные стратегии.

Аналогичное положение имеет место и для игр, существенно более простых, чем шахматы. Даже для совсем простой игры в крестики — нолики на поле из девяти клеток число различных партий настолько велико, что полный анализ матрицы игры хотя и возможен (матрица имеет размер $9!! \times 8!!$ *), но затруднителен из-за большого размера этой матрицы.

Поэтому для реализации в автоматах таких игр и решения их на машине требуется принципиально иной подход к нахождению оптимальных стратегий. В этом параграфе на ряде примеров мы рассмотрим, какого рода подходы можно использовать при решении игр с большим числом стратегий. В качестве первого примера рассмотрим игру в крестики — нолики на поле, состоящем из девяти клеток, показанном на рис. 11,*а*. В игре участвуют два игрока $И_1$ и $И_2$. Ход игрока $И_1$ заключается в том, что он в любую свободную клетку игрового поля ставит крестик, а ход игрока $И_2$ состоит в заполнении свободной клетки игрового поля значком нолик. Выигрывает игрок, сумевший заполнить своими значками одну из двух больших диагоналей, столбик или строку игрового поля.

В этой игре у игрока, начинающего игру, имеется 9!! полных стратегий, а у его противника 8!! полных стратегий. Однако в силу симметрии игрового поля число различных полных стратегий на самом деле существенно меньше. Так, например, в начале игры у иг-

---

* $(2k)!! = 2k \cdot (2k-2) \ldots 4 \cdot 2$; $(2k+1)!! = (2k+1)(2k-1) \ldots 3 \cdot 1$.

рока, делающего первый ход, имеется девять возможных ходов, но с точки зрения продолжения игры число различных ходов равно всего трем: ход в центральную клетку игрового поля, ход в угловую клетку и ход в одну из клеток, расположенных на стороне квадрата между угловыми клетками. Если первый игрок поставил

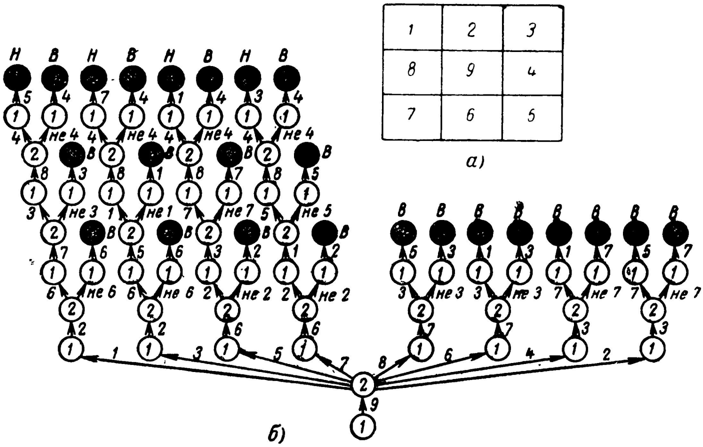


Рис. 11.

крестик в центральную клетку игрового поля, то у второго игрока на самом деле есть только два различных хода: поставить нолик в угловую клетку поля или поставить его в клетку, расположенную между угловыми клетками. Однако если даже учесть все такие симметрии в игре, то общее число полных стратегий все же остается весьма большим. Если считать, что игрок $И_1$ не делает ошибок во время игры, а игрок $И_2$ играет не обязательно оптимально, то, например, для того случая, когда в начале игры игрок $И_1$ делает ход в центральную клетку поля, что представляется наиболее выгодным, дерево игры все же остается довольно большим (см. рис. 11,*б*).

Поэтому желательно найти метод решения игры, который не сводился бы к полному анализу всего дерева игры, но позволял бы получать оптимальную полную стратегию.

Идея такого метода сводится к тому, что вместо определения оптимальной полной стратегии на каждом шаге игры выбирается на основе некоторого критерия оптимальная ходовая стратегия Эти оптимальные ходовые стратегии выбираются так, чтобы их совокупность всегда приводила к оптимальной полной стратегии.

Критерий выбора ходовых стратегий обычно строится следующим образом. Каждому возможному выбору приписывается некоторая числовая оценка, зависящая от ситуации, в которой делается данный выбор. На данном шаге игры выбирается та ходовая стратегия, оценка для которой максимальна (или минимальна). Сами оценки носят субъективный характер и отражают то знание об игре, которым обладает данный игрок.

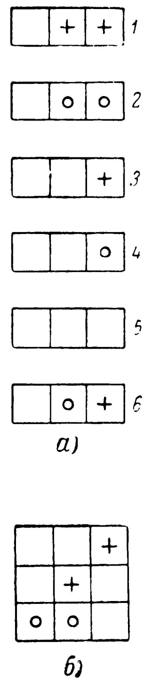


Рис. 12.

Для игры в крестики — нолики введем оценки игровых полей и ситуаций следующим образом. Если через δ_i обозначить оценку игрового поля с номером i ($i=1, 2, \ldots, 9$), то в начале игры все поля, лежащие по сторонам игрового поля и отличные от угловых, получают оценку, равную 1. угловые поля — оценку, равную 2, а центральное поле — оценку, равную 3. С каждой клеткой игрового поля связывается совокупность направлений. Для клеток, лежащих на сторонах игрового поля и не являющихся угловыми, имеется два направления: горизонтальное и вертикальное, которые мы будем обозначать символами g и w. Для угловых клеток имеется еще диагональное направление, которое мы будем обозначать символом d. Наконец, для центральной клетки имеется четыре направления: вертикальное, горизонтальное и два диагональных. Ситуация, которая складывается в ходе игры на любом из направлений, может быть расклассифицирована на шесть типов, как это показано на рис. 12,a. Эти типы ситуаций относятся друг к другу в порядке важности так, что тип ситуации с меньшим номером является менее важным, чем тип ситуации с большим номером. Типом клетки игрового поля называется наиболее важный из типов ситуаций, которыми обладают направления, связанные с этой клеткой игрового поля. Типам ситуаций на основе эмпирического опыта игры условно сопоставляются некоторые числовые оценки. Будем, например, считать, что оценки типов имеют следующий вид:

Тип ситуации	1	2	3	4	5	6
Оценка	40	20	10	5	1	0

Тогда оценка каждой клетки игрового поля в данный момент игры определяется следующим образом. Суммируются оценки ситуаций по всем направлениям, соответствующим данной клетке игрового поля, и полученная сумма отождествляется с δ_i. Для занятых полей $\delta_i=0$. Если, например, в некоторый момент игры перед ходом игрока $И_1$, начинавшего игру, складывается ситуация, показанная на рис. 12,*б*, то оценки для свободных игровых полей подсчитываются следующим образом:

$$\delta_1=g+w+d=10+5+10=25;$$

$$\delta_2=g+w=10+0=10;$$

$$\delta_4=g+w=10+10=20;$$

$$\delta_5=g+w+d=20+10+10=40;$$

$$\delta_8=g+w=10+5=15.$$

Выбор клетки игрового поля, в которую должна сделать ход машина, определяется максимальной оценкой, которую получают клетки игрового поля в результате этой оценки. В рассматриваемом случае ход будет осуществлен в клетку с номером 5. Ясно, что это единственно разумный ход в сложившейся ситуации игры. Всякий другой ход привел бы машину к проигрышу. Кроме того, этот ход обеспечивает машине выигрыш при следующем ходе, независимо от хода противника. Способ введения оценок на игровые поля применяется весьма часто. Его удобство заключается в том, что выбор машиной очередной чистой стратегии сводится к ряду проверок, арифметических подсчетов и определению на основании этих подсчетов наиболее выгодной чистой стратегии. На принципе введения оценок можно построить метод решения игры, являющейся обобщением рассмотренной нами игры в крестики — нолики. Игра проводится на игровом поле произвольного размера и конфигурации. Выигрывает игрок, который сумеет подставить подряд по горизонтали, вертикали или диагонали четыре однотипных значка (крестика или нолика). Программа, дающая решение такой игры, была составлена Н. И. Гусаковым.

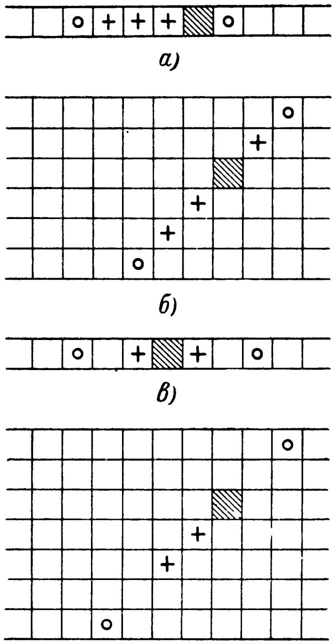


Рис. 13.

Пусть для определенности машина делает первый ход в игре. Первый ход машины произволен (если игровое поле не имеет границ, в противном случае машина выбирает любую неграничную клетку игрового поля). Второй ход машины состоит в приписывании крестика к ранее поставленному машиной. При реализации последующих ходов машина выбирает тот или иной ход в зависимости от ситуации, сложившейся на игровом поле. Если в эту ситуацию включены типовые комбинации *а* или *б*, показанные на рис. 13, то машина заканчивает игру и выигрывает. Если у машины нет выигрышной ситуации типа *а* или *б*, то она проверяет их наличие у противника. Если у противника две или больше выигрышных комбинации, то машина проигрывает. Если выигрышная комбинация у противника единственная, то машина ее блокирует. Если у противника также нет комбинаций типа *а* или *б*, то рассматривается возможность создания таких комбинаций для машины и блокирование комбинаций типа *в* и *г* для противника (если это возможно). В остальных случаях ход машины определяется на основании вычисления функции оценок игровых ситуаций. Автором программ выбраны следующие оценки. Для комбинаций типа *а* и *б* оценка равна ± 8, для комбинации типа *в*—± 2, а для комбинации типа *г* — ± 1. Знак плюс берется, если в результате очередного хода машины соответствующие комбанации появляются у машины, а знак минус берется в том случае, когда соответствующая комбинация появляется у противника. В качестве очередного хода машина выбирает тот, для которого суммарная оценка максимальна.

На аналогичном принципе подсчета оценок Стречи была построена машинная программа, полностью решающая проблему игры в шашки на доске 8×8. Однако основым недостатком метода оценок является необходимость предварительного полного анализа игры с целью установления разумных оценок ситуаций, дающих при суммировании оценку данного хода. Этот анализ может быть настолько сложным, что практически сам уже требует многих часов работы машины. В силу этого основной интерес приобретает не игра машины по уже найденной системе оценок, а принципы нахождения таких оценок в процессе работы машины. К этому вопросу мы еще вернемся после рассмотрения еще одного примера игры с полной информацией.

Другим примером игр с полной информацией может служить игра, известная в США под названием «Ним». В этой игре участвуют два игрока $И_1$ и $И_2$. В начале игры имеется *n* кучек однородных предметов. По правилам игры ход игрока заключается в том, что он берет из любой кучки любое число предметов. Выигрывает тот игрок, который возьмет последние предметы.

Эта игра является игрой с полной информацией, но ее особенностью является то, что начальное положение носит неопределенный характер, так как ни число кучек, ни число предметов, находящихся в каждой кучке, никак не регламентировано. В связи с этим нет возможности произвести полный анализ игры на основании дерева игры. Такой анализ возможен только в том случае, когда начальная позиция игры задана. Если, например, в начальной позиции имеются две кучки, в которых соответственно находится по два предмета, то дерево игры имеет вид, показанный на рис. 14. На этом

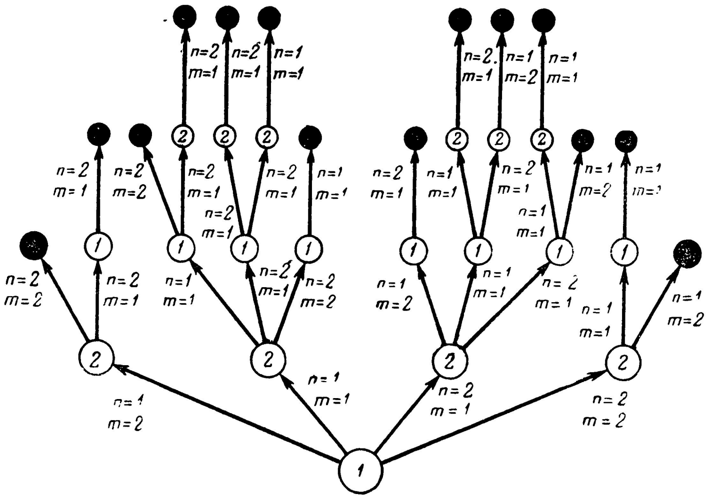


Рис. 14.

рисунке n означает номер кучки, из которой игрок берет предметы, а m — число взятых предметов. Общее число различных полных стратегий равно 14. Как можно заметить, даже при весьма небольших размерах начальной позиции дерево игры получается громоздким, а нахождение оптимальной стратегии — затруднительным.

На примере игры в «Ним» мы покажем, что весьма желательным является нахождение общей теории игры данного вида, учитывающей специфику игры. Такая теория, хотя и не будет иметь универсального характера, как методы, описанные в § 1, будет все же гораздо удобнее универсальных методов с точки зрения практического поиска оптимальной стратегии игроками.

Пусть для определенности $n=4$ и после очередного хода противника в кучках осталось соответственно 9, 16, 35 и 7 предметов. Запишем количество оставшихся предметов в двоичной системе счисления:

$$
\begin{array}{r}
9 - 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 1 \\
16 - 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0 \\
35 - 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 1 \\
7 - 0\ \ 0\ \ 0\ \ 1\ \ 1\ \ 1
\end{array}
$$

Найдем теперь сумму по модулю двух чисел, записанных двоичным кодом. Это суммирование происходит поразрядно по следующему правилу: если число единиц у суммируемых чисел в данном разряде четно, то в результирующем числе в этом разряде пишется ноль, в противном случае в данном разряде результирующего числа пишется единица. Для рассматриваемого примера результат описанной операции дает число 111101. Обозначим это число буквой R. Определяем самый левый разряд в R, в котором содержится единица. В нашем случае таким разрядом является крайний левый разряд. После этого среди чисел, из которых мы получим число R, ищем то, у которого содержится единица в том же разряде, что и разряд в R, содержащий самую левую единицу. В нашем примере таким числом является число 35. Это число мы обозначим буквой Q. Находим теперь сумму по модулю два чисел R и Q:

$$
\begin{array}{l}
R - 1\ 1\ 1\ 1\ 0\ 1 \\
\underline{Q - 1\ 0\ 0\ 0\ 1\ 1} \oplus \\
T - 0\ 1\ 1\ 1\ 1\ 0
\end{array}
$$

Полученное нами число T, записанное в десятичной системе, соответствует числу 14. Утверждается, что это число определяет количество предметов, которое должно остаться в кучке, содержащей 35 предметов. Покажем, что такой выбор хода обеспечивает игроку оптимальную полную стратегию.

Доказательство этого факта состоит из обоснования справедливости следующих четырех утверждений: 1) $M \oplus T = 0$, где M — сумма по модулю два количества предметов во всех кучках, кроме кучки, соответствующей Q; 2) всегда можно найти такую кучку, характеризуемую числом Q, что $M \oplus T = 0$; 3) если после хода игрока $И_1$ $R = 0$, то после очередного хода игрока

$И_2$ $R \neq 0$; 4) если после каждого хода $И_1$ $R=0$, то $И_1$ выигрывает игру.

Справедливость первого утверждения вытекает из следующего: $M \oplus Q = R$. Но из $R \oplus Q = T$ следует, что $T = M \oplus Q \oplus Q = M$. Отсюда $R = M \oplus T = M \oplus M = 0$.

Для справедливости второго утверждения надо показать, что при $R \neq 0$ перед ходом $И_1$ всегда можно найти T такое, что $M \oplus T = 0$ после хода машины. Появление единицы в самом левом разряде R возможно лишь при условии, что по крайней мере при записи количества предметов в одной из кучек в этом разряде есть единица. Но тогда $T = R \oplus Q$ в этом разряде будет иметь обязательно ноль, а левее этого разряда T и Q совпадают. Это означает, что $T < Q$ и ход игрока $И_1$ реализуем.

Справедливость третьего утверждения основывается на правилах игры. Так как после хода машины $R=0$, а игрок и $И_2$ может взять предметы только из одной кучки, то после его хода изменится лишь одно из слагаемых, с помощью которых будет подсчитано новое R. Отсюда вытекает, что новое R уже не будет нулем.

Четвертое утверждение связано с тем, что когда игрок $И_1$ придерживается правила, то после его хода $R=0$, и игрок $И_2$ не может получить $R=0$. Но при взятии последних предметов всегда $R=0$. Следовательно, игрок $И_1$ выигрывает. В рассмотренной игре оптимальная стратегия игрока $И_1$ состоит из последовательности ходовых стратегий, выбираемых из условия того, что после хода игрока $И_1$ величина $R=0$. Таким образом, при безошибочной игре всегда выигрывает тот игрок, который начинает игру. В этом случае все стратегии другого игрока одинаково оптимальны (все приводят к проигрышу). Однако если игрок, начинающий игру, может ошибаться, то, вычисляя каждый раз значение R после его хода и обнаружив, что после некоторого хода это значение не равно нулю, игрок, играющий вторым, перехватывает инициативу в свои руки и при безошибочной игре выигрывает партию.

Описанная выше теория игры в Ним носит несколько завуалированный характер. Сам метод решения игры основан на довольно глубоких результатах теории графов, которые в настоящей книге не излагаются[1].

В качестве третьего примера мы рассмотрим игру в шахматы. В отличие от двух предыдущих игр игра в шахматы характеризуется отсутствием теории, позволяющей заменить нахождение оптимальной полной стратегии нахождением на каждом шаге игры оптимальной ходовой стратегии. Связано это с тем, что пока не найдено способов выбора среди множества ходовых стратегий, имеющихся у игрока на данном шаге игры, оптимальных стратегий. Эта задача решена только для весьма простых эндшпильных задач, в которых число ходовых стратегий у игроков невелико и все результаты применения той или иной стратегии легко оцениваются с точки зрения оптимальности полной стратегии.

Поэтому для решения таких задач приходится использовать приближенный способ решения задачи, в котором задача нахождения оптимальной полной стратегии заменяется задачей о нахождении полной стратегии, близкой с точки зрения тех или иных критериев к оптимальной.

Общая идея такого способа решения игр основана на методах решения задач на вычислительных машинах, которые в последнее время получили название эвристических методов. Сущность эвристических методов может быть пояснена следующим образом. Пусть мы имеем конечное множество элементов $A=\{a_1, a_2, \ldots, a_n\}$, которое мы будем называть множеством ситуаций. В этом множестве выделяется некоторое подмножество $B \subset A$, элементы которого называются заключительными ситуациями. Кроме того, имеется конечное множество операторов $F=\{f_1, f_2, \ldots, f_m\}$. Элементы этого множества играют роль преобразующих операторов, определенных на множестве A. Наконец, на множестве A для любой пары элементов этого множества определено понятие различия между этими элементами: $\rho(a_i, a_j)$.

Функция $\rho(a_i, a_j)$ обладает следующими свойствами:

1) $\rho(a_i, a_j)$ есть всегда натуральное число;

2) $\rho(a_i, a_j)=0$;

[1] Интересующиеся играми на графах могут прочитать гл. 6 из книги К. Берж, «Теория графов и ее применения», Издательство иностранной литературы, 1962.

3) $\rho(a_i, a_j)=0$ для любых a_i и a_j, одновременно принадлежащих B;

4) $\rho(a_i, a_j)=\rho(a_j, a_i)$;

5) $\rho(a_i, a_j)\neq 0$, если a_i принадлежит B, а a_j не принадлежит B.

Назовем отличием элемента a_i от заключительной ситуации функцию $\mu(a_i)=\underset{j}{\text{мин}}\{\rho(a_i, a_j)\}$, где a_j — любой элемент множества B.

Разобьем теперь множество A на сумму непересекающихся подмножеств

$$A=B\cup A_1\cup\ldots\cup A_r,$$

где A_i — это подмножество, элементами которого являются те элементы a_j, для которых отличие от заключительной ситуации равно i. Подмножества A_i мы будем называть множествами i-го уровня.

Если $\overset{0}{a}_i$ есть начальная ситуация, то решением задачи мы будем называть поиск такой последовательности операторов $f_{i1}, f_{i2}, \ldots, f_{il}$, с помощью которой можно получить переход от элемента $\overset{0}{a}_i$ к элементу из B, при условии, что длина этой последовательности не превосходит числа l. Общее число последовательностей операторов, не превосходящих l, равно

$$m^0+m^1+m^2+\ldots+m^l=\frac{m^{l+1}-1}{m-1}. \qquad (10)$$

Соотношение (10) определяет число различных полных стратегий, которые могут быть использованы при решении задачи. При этом не все эти последовательности дают нужный результат. Оптимальной полной стратегией будет стратегия, в результате применения которой получается элемент множества нулевого уровня за наименьшее число шагов (т. е. последовательность из наименьшего числа операторов, приводящая при поочередном применении этих операторов от исходной ситуации к заключительной).

При большом числе m или l поиск оптимальной стратегии становится затруднительным, так как число различных полных стратегий может быть весьма большим. Для преодоления этой трудности метод эвристического решения можно строить на следующей основе. К элементу a_i^0 (исходному) применяются поочередно все f_j. Пусть в результате этого мы получим следующие m элементов $f_1(a_i^0)$, $f_2(a_i^0)$, . . . , $f_m(a_i^0)$. Для каждого из этих элементов определяем их отличие от заключительной ситуации. Среди $\mu(f_q(a_i^0))$, $q=1, 2, \ldots, m$ находим минимальное значение и сравниваем его с $\mu(a_i^0)$. Если $\mu(\alpha_i^0)$ больше этого значения, то в качестве первого оператора в искомой последовательности выбирается оператор, преобразующий a_i^0 в элемент с наименьшим $\mu(f_q(a_i^0))$.

После этого процесс повторяется заново, причем за исходный элемент принимается вновь найденный элемент. Процесс поиска обрывается в следующих трех случаях.

1. Наименьшее $\mu(f_q(a^0{}_i))$, полученное на последнем шаге, есть ноль. В этом случае решение найдено, что соответствует выигрышу. Подчеркнем, что при этом построена некоторая последовательность ходовых стратегий, дающая в совокупности полную стратегию, приводящую к выигрышу. Но так как выигрыш определяется не только достижением элементов множества нулевого уровня, но и учитывает также число шагов, сделанных для этого, то он может быть не максимален и, следовательно, найденная полная стратегия может оказаться неоптимальной.

2. Число использованных операторов равно l, а наименьшее значение $\mu(f_q(a^0{}_i))$, полученное после l шагов, отлично от нуля. В этом случае имеется чистый проигрыш и найденная полная стратегия является неоптимальной (в предположении, что существуют полные стратегии, с помощью которых за l шагов можно получить для данной исходной ситуации заключительную ситуацию).

3. На очередном шаге все $\mu(f_q(a^0{}_i))$ больше $\mu(a^0{}_i)$. В этом случае процесс построения последовательности

операторов обрывается, а решение задачи на базе используемого критерия выбора ходовых стратегий становится невозможным.

Критерий использования того преобразования, который дает при своем применении к исходному элементу элемент с минимальным отличием от B не является, конечно, единственно возможным. Можно вместо этого критерия использовать критерий, связанный со случайным выбором элемента, получаемого с помощью преобразований из исходного, при единственном условии, что его отличие от B меньше, чем отличие от B исходного элемента. Можно допускать некоторое увеличение отличия на данном шаге, не превосходящее определенного значения и т. д.

Отметим одну из особенностей эвристического метода решения. С помощью такого метока не гарантируется нахождение решения задачи, а сам метод носит характер метода проб и ошибок. Второй особенностью этого метода является наличие разбиения множества ситуаций A на множества уровней и возможность для каждой ситуации определения множества уровня, к которому она принадлежит. В частности, выделение из множества A множества нулевого уровня свидетельствует о том, что у играющего есть знание о том, как выглядят заключительные выигрышные ситуации.

При вышеприведенных рассуждениях мы считали, что в игре фактически участвует лишь один игрок, так что, строго говоря, мы не могли называть описываемый процесс игрой и использовать для его описания терминологию из теории игр. Теперь мы предположим, что имеется игрок $И_2$, интересы которого противоположны интересам игрока $И_1$. А именно, для игрока $И_2$ заключительной ситуацией будет ситуация, принадлежащая к множеству наибольшего уровня. Игроки поочередно выбирают одно из преобразований (что соответствует выбору чистой стратегии) и игра заканчивается либо тогда, когда один из игроков получает заключительную ситуацию, либо после l парных ходов. Если игра закончена, то определяется величина разности $i—j$, где i — уровень множества, из которого была взята начальная ситуация, а j — уровень множества, в котором находится окончательная ситуация. Если эта разность положительна, то выиграл игрок $И_1$, в противном случае — выиграл игрок $И_2$. Величина выигрыша определяется модулем этой разности (или однозначно связана с этой величиной).

Прежде чем описывать эвристический метод игры в шахматы, мы рассмотрим использование такого метода на более простом примере, взяв в качестве игры уже рассмотренную нами игру в крестики — нолики на поле 3×3.

Будем рассматривать вектор, имеющий девять координат, который будет характеризовать ситуацию, складывающуюся в игре. Например, для положения игры, показанного на рис. 14, такой вектор имеет вид $\{—, —, +, —, —, 0, 0, —, +\}$. Этот вектор мы представим в виде «суммы» двух векторов: вектора-конфигурации для игрока $И_1$, ходы которого отмечаются крестиками, и вектора-конфигурации для игрока $И_2$, ходы которого отмечаются ноликами. Эти векторы-конфигурации учитывают только ходы данного игрока, сделанные им в процессе игры. Векторы-конфигурации «в сумме» дают вектор-ситуацию. Для рассмотренного случая вектора-ситуации векторы-конфигурации выглядят следующим образом:

$$K^1 = \{—, —, +, —, —, —, —, —, +\};$$

$$K^2 = \{—, —, —, —, —, 0, 0, —, —,\}.$$

В множестве всех векторов K^1 выделено подмножество K^1_0 выигрышных конфигураций, в множестве векторов K^2 выделено подмножество K^2_0 проигрышных (для игрока $И_1$) конфигураций. Цель игрока $И_1$ заключается в преобразовании данной ситуации в ситуацию, которая содержала бы в себе выигрышную конфигурацию. Цель игрока $И_2$ противоположна: он стремится так преобразовать данную ситуацию, чтобы получить ситуацию, включающую в себя конфигурацию, проигрышную для игрока $И_1$. Игра всегда начинается с начальной ситуации $a_0 = \{—, —, —, —, —, —, —, —, —,\}$. Игра заканчивается, если в результате игры ситуация преобразуется в ситуацию, содержащую выигрышную или проигрышную для $И_1$ конфигурацию, или в том случае, когда вектор-ситуация уже не содержит координат, соответствующих незаполненным полям игрового поля.

Подмножества K_0^1 и K_0^2 состоят из векторов следующего вида:

$$k_1 = \{0, 0, 0, x, x, x, x, x, x\};$$
$$k_2 = \{x, x, x, 0, x, x, x, 0, 0\};$$
$$k_3 = \{x, x, x, x, 0, 0, 0, x, x\};$$
$$k_4 = \{0, x, x, x, x, x, 0, 0, x\};$$
$$k_5 = \{x, 0, x, x, x, 0, x, x, 0\};$$
$$k_6 = \{x, x, 0, 0, 0, x, x, x, x\};$$
$$k_7 = \{0, x, x, x, 0, x, x, x, 0\};$$
$$k_8 = \{x, x, 0, x, x, x, 0, x, 0\},$$

где значок «0» соответствует либо крестику, либо нолику, а значок «×» соответствует крестику, нолику или незаполненной клетке игрового поля.

Множество операторов F состоит для каждого из игроков из девяти операторов $\{f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9\}$, с помощью которых происходит отметка соответствующей клетки игрового поля значком крестик или значком нолик. Ход, осуществляемый с помощью оператора f_i, называется допустимым, если клетка игрового поля с номером i свободна.

Обращением конфигурации мы будем называть переход от вектора-конфигурации для $И_1$ к вектору-конфигурации для $И_2$, соответствующих одному и тому же вектору-ситуации.

Вектор-ситуацию мы будем называть выигрышной, если можно найти такой ход, что, независимо от очередного хода противника, при следующем ходе игрока может быть получена заключительная ситуация, в которую входит конфигурация из подмножества K_0^i, для данного игрока. Вектор-ситуация будет называться проигрышной, если она является выигрышной для противника.

В качестве расстояния в множестве ситуаций (или конфигураций) можно использовать величину их отличия от выигрышных ситуаций. При таком определении расстояния ситуация вида $\{0, 0, —, \times, \times, \times, \times, \times, \times\}$ отличается от первой ситуации из K_i на одну единицу, а ситуация $\{—, —, 0, —, —, —, —, —, 0\}$ находится на расстоянии два от последней ситуации из K_i. При очередном ходе машины ее работа определяется следующей последовательностью шагов: 1. Проверяется, нет ли для машины такого оператора (хода), который приводил бы к выигрышу (т. е. переводил бы данную ситуацию из K_0^1). Если такой ход имеется, то машина выигрывает партию. 2. Происходит обращение данной конфигурации и выявляется, нет ли у противника возможности при своем ходе получить выигрышную ситуацию. Если такая возможность имеется, то ход машины выбирается так, чтобы расстояние от ситуации, полученной после хода машины, до ситуаций из K_0^2 для противника стало бы больше, чем в данной ситуации. 3. Происходит просмотр всех возможных ходов и выбирается ход, дающий наименьшее расстояние полученной после этого ситуации от выигрышной. 4. Проверяется путем обращения конфигурации, которая будет получена после хода машины, не будет ли она выигрышной для противника. Если это не так, то делается выбранный ход. 5. Если новая конфигурация является проигрышной, то необходимо выбрать другой ход, дающий меньшее приближение к K_0^1, и повторить для этого хода пункт 4. 6. Если все ходы машины приводят к проигрышным ситуациям, то машина сдается.

Так как машина играет без ошибок, то случай проигрыша возможен только в процессе обучения машины, если вид ситуаций, которые являются проигрышными, машине заранее неизвестен. В этом случае на основе анализа уже сыгранных машиной партий происходит выделение множества проигрышных и выигрышных ситуаций. После обнаружения всех таких ситуаций машина начинает играть безошибочно.

Рассмотрим теперь игру в шахматы, применяя для машинного решения этой игры эвристические соображения. В отличие от такой простой игры, как крестики — нолики, число различных позиций на шахматной доске (ситуаций) огромно. Оно оценивается приблизительно числом 10^{43}. Конечно, эти позиции можно расклассифицировать на типовые множества A_i, $i=0, 1, \ldots, r$, и определить расстояние между ситуациями с помощью числа пар ходов, необходимых для достижения из данной ситуации предполагаемой ситуации. Однако практически произвести такую классификацию невозможно и такой подход имеется только для эндшпилей, когда число различных типовых ситуаций становится невелико. Поэтому вместо настоящей функции расстояния при машинном решении шахматных задач вводится функция расстояния, носящая эвристический характер, основанная на имеющемся опыте игры. Оценочная функция для каждой ситуации состоит из оценки наличия фигур, их взаимного расположения, очередности хода, наличия или отсутствия возможности проведения рокировки и т. д. Для оценки наличия фигур, как правило, используется оценочная функция Шеннона. Эта функция имеет вид $f(S)=\Sigma p_i(n_i-n'_i)$, где p_i — вес фигуры, а n_i и n'_i — число фигур i-го типа, имеющихся у машины и противника соответственно. Задание весов носит субъективный характер и подбирается экспериментальным путем. В программе для игры в шахматы, составленной немецким математиком Шлибсом, приняты следующие веса фигур: король — 200, ферзь — 9, ладья — 5, слон — 3, конь—3, пешка 1, отсталая пешка—0,5, изолированная—0,4, сдвоенная — 0,3. Кроме оценки материального соотношения сил, в этой функции необходимо учесть еще ряд факторов стратегического и тактического плана. В качестве примера факторов, которые необходимо учесть в функции $f(S)$, укажем на оценку наличия пешек на игровых полях *e*4, *d*4 и *c*4, слабой позиции пешек около короля, наличия пешек на полях, цвет которых отличен от цвета, имеющегося у игрока слона, проходных пешек и т. д., расположение фигур в позиции (продвинутый на поля *e*5, *d*5, *c*5, *f*5, *e*6, *d*6, *c*6, *f*6 конь, наличие ладьи на открытой или полуоткрытой линии, сдвоенные ладьи, связанные фигуры и т. д.), наличие вилок, связок и т. д. Чем больше факторов учитывается в оценочной функции, тем более точна игра машины.

Игра машины происходит по следующим правилам. Проанализировав заданную позицию и все возможные в этой позиции ходы, машина перебирает все возможные варианты продолжения на n ходов вперед. При этом после своего хода машина рассматривает игру с точки зрения своего противника и выбирает оптимальный ответный ход. Для всех вариантов продолжения вычисляется оценочная функция после n ходов машины и выбирается тот вариант игры, при котором эта оценочная функция максимальна.

Несмотря на простоту идеи игры, ее реализация сопряжена с большими трудностями из-за огромного перебора, который должна осуществлять машина даже при небольших значениях n. Число возможных ходов, особенно в начале игры, слишком велико и необходимо принять какие-то меры, сокращающие полный перебор всех возможных вариантов хода в данной ситуации. Из различных способов такого сокращения перебора для случая игр типа шахмат весьма удобен метод оценок и граней, предложенный А. Л. Брудно. Этот метод позволяет на много порядков снизить величину перебора.

Рассмотрим дерево игры для двух игроков. Через H_i будем обозначать вершины этого дерева, а через l_i — его ребра. Как известно, вершины дерева характеризуют позиции, появляющиеся в игре, а ребра — выбранные игроками ходы. Каждому ребру дерева игры сопоставим некоторую величину $a(l_i)$, называемую ценой ребра (ценой хода). Если последовательность ребер образует путь, то ценой пути будем называть сумму цен ребер, образующих этот путь:

$$a(L)=\sum_{i=1}^{n} a(l_i). \tag{11}$$

Ценой позиции $b(H_i)$ назовем тот выигрыш, который получит игрок $И_1$, если после данной позиции оба игрока будут играть оптимальным образом. Наконец, ценой вершины H_i мы назовем величину, определяемую соотношением

$$c(H_i)=a(L)+b(H_i), \tag{12}$$

где L — путь, ведущий в вершину H_i из начальной вершины дерева игры.

Для игрока $И_1$ цель игры состоит в таком течении игры, при котором игра заканчивается в заключительной позиции с максимальной ценой. Цель игрока $И_2$ противоположна и он стремится закончить игру в вершине с минимальной ценой.

Оценкой позиции, характеризуемой вершиной H_i, называется пара чисел $r_1(H_i)$, $r_2(H_i)$, определяющих отрезок $[r_1, r_2]$, таких, что

$$r_1(H_i)\leqslant c(H_i)\leqslant r_2(H_i). \tag{13}$$

Точные значения цен вершин известны, как правило, лишь для вершин, соответствующих заключительным позициям в игре. Цены всех остальных вершин известны приблизительно лишь на основе некоторого опыта игры или анализа дерева игры. Поэтмоу оценки можно рассматривать как интуитивное знание игрока о характере данной позиции. Поэтому эти оценки могут быть весьма широкими, как, например $[-\infty, r_2]$, или $[-\infty, \infty]$.

Для данной вершины H рассмотрим все вершины H_i, лежащие на пути L, идущем из начальной позиции дерева игры к вершине H. Найдем пересечение оценок для всех этих вершин. Точки, принадлежащие к этому пересечению (которое может быть и пустым), одновременно принадлежит всем отрезкам $[r_{i1}, r_{i2}]$ для всех вершин H_i, лежащих на L. Это пересечение называется гранью для вершины H.

Если на вершине H надписана оценка $[r'_1, r'_2]$ и H_i есть множество вершин дерева, непосредственно следующих за вершиной H (т. е. вершин, в которые можно попасть только пройдя через вершину H), то пересечение оценок для H и всех H_i задается соотношением

$$[r_1, r_2] = [r', r'_2] \cap \begin{bmatrix} \text{макс} & \text{макс} \\ r_{i_1} & r_{i_2} \\ \text{мин} & \text{мин} \end{bmatrix}; \qquad (14)$$

здесь верхняя строка во втором члене выбирается в том случае, когда данная вершина дерева соответствует ходу $И_1$, а нижняя строка берется при условии, что H соответствует ходу игрока $И_2$.

Вершина дерева называется удовлетворительно оцененной, если грань, соответствующая этой вершине, не содержит более одной точки (т. е. грань либо пуста, либо содержит одну точку). Дерево игры называется оцененным удовлетворительно, если существует рекуррентное отношение, которое позволяет по удовлетворительно оцененным конечным вершинам дерева удовлетворительно оценивать все вершины дерева. В качестве примера такого рекуррентного отношения можно рассматривать отношение следующего вида:

$$[r''_1, r''_2] = [r'_1, r'_2] \cap \begin{bmatrix} \text{макс} & \text{макс} \\ r''_{i_1} & r''_{i_2} \\ \text{мин} & \text{мин} \end{bmatrix}, \qquad (15)$$

где выбор верхней и нижней строк во втором члене правой части происходит так же, как это делалось для соотношения (14). $[r'_1, r'_2]$ означает надписанную оценку данной вершины дерева, а $[r''_{i1}, r''_{i2}]$ — самые узкие оценки для вершин, следующих за данной вершиной. Наконец, $[r''_1, r''_2]$ означает самую узкую оценку для данной вершины.

А. Л. Брудно доказана теорема о необходимом и достаточном условии удовлетворительного задания дерева игры. Таким условием является удовлетворительная оценка конечных вершин дерева. Как следствие, из этой теоремы можно сформулировать следующее важное для нас положение. Удовлетворительно заданное дерево игры останется удовлетворительно заданным, если выкинуть все ребра, выходящие из любого набора удовлетворительно заданных вершин.

Смысл этого следствия состоит в том, что при организации перебора вариантов на дереве игры при получении для некоторой вершины дерева удовлетворительной оценки можно прекращать просмотр остальных вариантов, начинающихся в этом узле. Частным случаем такой организации перебора является одноветочная схема перебора с использованием оценок и граней, которая была успешно использована А. Л. Брудно и И. Я. Ландау при программировании карточной игры одномастка на вычислительной машине. Применение этой схемы позволило авторам «ускорить» перебор в 10^{10} раз.

При одноветочной схеме перебора в памяти автомата в данный момент времени хранится начальная вершина дерева игры и вершины, лежащие на пути, идущем от начальной вершины до некоторой вершины H, для которой мы хотим найти удовлетворительную оценку. От вершины H происходит движение вперед по дереву игры, с помощью которого мы переходим от H к следующей за ней вершине дерева H'. Если следующих за H вершин в дереве несколько, то выбираем среди них H' по некоторому жесткому правилу (например, берем всегда самую левую от H вершину, которая еще не рассматривалась). Если H' не оценена удовлетворительно, то H' принимается за H и движение вперед повторяется. В силу конечности дерева игры и того, что концевые вершины дерева удовлетворительно оценены, всегда наступит такой момент, что при движении вперед очередная вершина дерева H' окажется удовлетворительно оцененной. Как только такая вершина найдена, то из этой вершины происходит движение назад по дереву игры в вершину, из которой мы попали в данную вершину. Если возвращение в предшествующую вершину H уже не первое, то для H уже есть некоторая (неудовлетворительная) оценка $[q_1, q_2]$, которую можно назвать оценкой вершины H от возвратов. Пользуясь оценкой $[r'_1, r'_2]$, можно улучшить для H оценку от возврата, надписанную оценку $[r_1, r_2]$ и грани в соответствии со следующими очевидными соотношениями:

$$\left.\begin{aligned} \tilde{r}_1 = \tilde{q}_1 &= \text{макс}\,(r'_1,\ r_1); \\ \tilde{q}_2 &= \text{мин}\,\{r_2,\ \text{макс}\,(r'_2,\ q_2)\}; \\ r_2 = r_2;\quad \tilde{\lrcorner} &= \text{макс}\,(\llcorner\,,\ r_1); \end{aligned}\right\} \qquad (16)$$

здесь $\lrcorner$ и $\llcorner$ означают величины граней, а значок тильда означает новые значения оценок и граней для вершины H. Если эта оценка удовлетворительна, то рассмотрение всех путей, идущих из H, прекращается и делается новое движение назад к вершине, предшествующей H. Если H оценена все еще неудовлетворительно, то делается новое движение вперед из вершины H в очередную вершину, следующую за H, которая еще не рассматривалась при оценке вершины H. Если все пути, ведущие из H, уже просмотрены, а оценка H еще неудовлетворительна, то согласно теореме А. Л. Брудно за удовлетворительную оценку H принимается оценка $[\tilde{q}_1, \tilde{q}_2]$ и от H делается движение назад.

Такой процесс приводит в конце концов к начальной вершине дерева игры, определяя, тем самым, оптимальный ход из начального положения для игрока, начинающего ход. После реализации этого хода рассматривается новое дерево игры с новой начальной вершиной и процесс оценки вариантов очередного хода повторяется.

На рис. 15 показан пример использования описанного метода перебора вариантов продолжения игры. Здесь H_0 — начальная вершина. Пусть ее надписанная оценка есть $[-\infty, \infty]$. Движемся из этой вершины вперед в вершину H_1. У этой вершины оценка удовлетворительная. Возвращаемся назад к H_0 и, используя соотношения (16), получаем новую оценку и грань для H_0 [10, ∞]. Эта оценка еще не удовлетворительная. Поэтому осуществляем движение вперед к вершине H_2. Для H_2 грань имеет вид [10, ∞] (т. е. на H_2 переносится уже полученная нами грань для H_0). От этой вершины совершаем

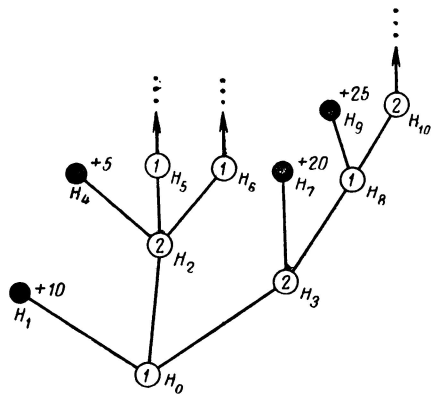


Рис. 15.

движение вперед к вершине H_4, которая оценена удовлетворительно. Следовательно, H_2, имеет оценку [0, 5], лежащую левее ее же грани [10, ∞]. Поэтому рассматривать пути, идущие из H_2, не имеет смысла, так как достаточно лишь запомнить оценку для H_2. Теперь рассматривается вершина H_3, которая не имеет удовлетворительной оценки. Переходим от нее к H_7, для которой имеется удовлетворительная оценка. Отсюда оценки [0, 20], а грани — пересечение граней H_0 [10, ∞] и граней H_3 [0, 20]. Тогда оценка грани H_3 есть [10, 20]. Начинаем рассматривать вершину H_8, обрезая все ребра с оценками, выходящими из граней [10, 20] и т. д.

Метод оценок и граней можно эффективно применять для всех игр, требующих для своей оптимальной реализации большого перебора вариантов продолжения. В частности, в настоящее время методы, подобные описанному нами, начинают широко использоваться при программировании на вычислительных машинах задач, связанных с проведением большого перебора для нахождения оптимального варианта решения задачи.

5. ИГРЫ С НЕПОЛНОЙ ИНФОРМАЦИЕЙ

Игры с неполной информацией представляют собой весьма интересный объект для исследования методами, развитыми нами в § 1, так как идеи, связанные с построением автоматов для подобных игр, наиболее часто могут быть использованы при проектировании систем автоматического управления.

Неполнота информации, имеющаяся у игрока, заключается в том, что он не может однозначно определить ситуацию, которая имеется в игре на данном ее шаге. Чаще всего это происходит из-за того, что в результате наличия случайных ходов формируются подмножества M_i возможных ходов у игроков. Таким образом, неполнота информации практически всегда связана с тем, что игрокам неизвестны множества стратегий, которыми может воспользоваться их противник. Типичными примерами игр с неполной информацией является большинство азартных карточных игр или домино.

Неполнота информации не обязательно связана с наличием в данной игре случайных ходов. Можно представить себе, например, что очередные ходы остаются для противника неизвестными до самого конца партии, хотя каждый из игроков имеет только личные ходы.

Рассмотрим сначала пример игры с неполной информацией, который достаточно прост, но отражает по существу основные особенности решения игр такого типа. Этот пример представляет собой сильно упрощенный вариант популярной карточной игры «веришь — не веришь».

Игра является игрой двух лиц. Колода состоит всего из двух карт: например, туза и десятки. Игрок $И_1$ берет наугад одну из двух карт (делает случайный ход с вероятностью ходов 0,5). После этого игрок $И_1$ смотрит взятую карту. Если эта карта является тузом, то, не показывая ее игроку $И_2$, игрок $И_1$ говорит: «у меня туз» и требует у игрока $И_2$ выигрыш, равный a. Если карта, полученная $И_1$, десятка, то игрок $И_1$ имеет две ходовые стратегии. Он может сказать «у меня туз» и потребовать у игрока $И_2$ выигрыш, равный a, или показать $И_2$ свою карту и заплатить ему проигрыш, равный a. У игрока $И_2$ также две стратегии: верить или не верить, когда игрок $И_2$ говорит: «у меня туз». Игрок $И_2$ может поверить этому и заплатить проигрыш a или не поверить игроку $И_1$ и потребовать от него, чтобы он открыл свою карту. Если в результате открытия карты $И_1$ открывает туза, то игрок $И_2$ проигрывает $2a$, в противном случае $И_2$ выигрывает $2a$.

Итак, $И_1$ имеет две стратегии: J^1_1 — обманывать и J^1_2 — не обманывать, а игрок $И_2$ также две стратегии: J^2_1 — верить и J^2_2 — не верить. Элементы матрицы игры

$$\begin{array}{cc} & \begin{matrix} J^2_1 & J^2_2 \end{matrix} \\ \begin{matrix} J^1_1 \\ J^1_2 \end{matrix} & \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \end{array}$$

вычисляются следующим образом:

$$a_{11}=0{,}5a+0{,}5a=a;$$

$$a_{12}=0{,}5\cdot 2a-0{,}5\cdot 2a=0;$$

$$a_{21}=0{,}5a-0{,}5a=0;$$

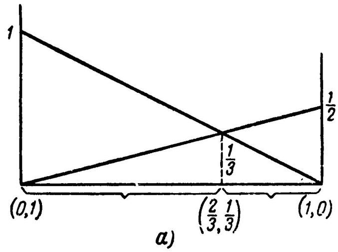
$$a_{22}=0{,}5\cdot 2a-0{,}5a=0{,}5a$$

и, следовательно, матрица игры имеет вид:

$$\begin{pmatrix} a & 0 \\ 0 & 0{,}5a \end{pmatrix}.$$

На рис. 16,*а* показано геометрическое решение данной игры для $a=1$ ($^1/_3$, $^2/_3$), что соответствует смешанной стратегии игрока $И_1$, при которой он обманывает в одном случае из трех. При этом цена игры равна $^1/_3$ и, пользуясь найденной смешанной оптимальной стратегией, игрок $И_1$ всегда может обеспечить себе средний выигрыш $^1/_3$. Для игрока $И_2$ оптимальная смешанная стратегия имеет вид ($^1/_3$, $^2/_3$), т. е. он должен в одном случае из трех обходиться без проверки утверждения игрока $И_1$, а в двух других случаях проверять его утверждение.

На рис. 16,*б* показана схема автомата, реализующего оптимальную стратегию игрока $И_1$. Автомат состоит из трех основных блоков. В блоке подсчета частот подсчитывается смешанная стратегия игрока $И_2$, т. е. относительные частоты выбора этим игроком стратегий «верить» и «не верить». Если эти относительные частоты лежат достаточно близко к оптимальной смешанной стратегии $И_2$, то логический блок выдает сигнал на смеситель стратегий и автомат отвечает на действия игрока $И_2$ своей смешанной стратегией. Если же блок подсчета частот обнаруживает, что частоты использова-



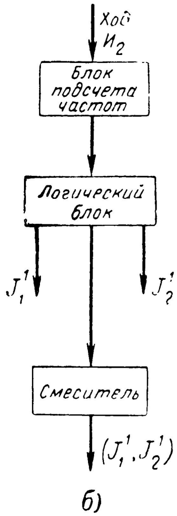


Рис. 16.

ния игроком $И_2$ стратегий «верить» и «не верить» сильно отличается от оптимальных, то логический блок выбирает ту чистую стратегию автомата, которая дает автомату наибольший выигрыш. Этот выбор происходит на основе следующего правила: если игрок $И_2$ выбирает стратегию «верить» с частотой, большей чем $^1/_3$, то автомат начинает использовать чистую стратегию «обманывать», если же игрок $И_2$ использует стратегию «верить» с частотой, меньшей $^1/_3$, то автомат использует чистую стратегию «не обманывать».

В обоих случаях автомат при этом получает выигрыш больший, чем цена игры.

Рассмотрим еще два примера игр с неполной информацией.

В качестве первого примера мы рассмотрим карточную игру, известную под названием «железная дорога» (или «железка»). «Железная дорога» принадлежит к азартным карточным играм и содержит случайные ходы, которые превращают ее в игру с неполной информацией.

В игру играют два игрока $И_1$ и $И_2$. Игрок $И_2$ называется банкометом. Колода состоит из шести колод по 52 карты. При каждой сдаче (случайный ход) каждому из игроков сдается по две карты рубашкой вверх. Каждая карта оценивается числом очков от нуля до десяти, а суммарный счет очков проводится по модулю десять (т. е. сумма, равная десяти очкам, считается равной нулю, одиннадцати очкам — единице и т. д.). Король, дама, валет и туз дают ноль очков. Получив карты, игрок $И_1$ суммирует очки. В зависимости от этой суммы $И_1$ должен поступить следующим образом: если сумма очков меньше пяти, то он требует еще одну карту из колоды, которая сдается рубашкой вниз; если сумма равна шести или семи, то он передает ход банкомету. При сумме, равной пяти, игрок может либо требовать новую карту из колоды, либо передать ход банкомету. Наконец, при сумме восемь или девять игрок и банкомет открывают свои карты и игра заканчивается. Если число очков у $И_1$ больше, чем у банкомета, то он выигрывает, в случае равенства очков исход считается ничейным, если число очков $И_1$ меньше, чем у банкомета, то выигрывает банкомет.

Банкомет играет следующим образом: если сумма очков на его картах меньше трех, то он берет еще одну карту из колоды; при сумме, равной семи, — передает ход игроку $И_1$; при сумме от трех до шести включительно банкомет может либо брать еще одну карту из колоды, либо передать ход игроку $И_1$; наконец, при сумме восемь или девять игроки открывают карты и происходит определение выигравшего по числу набранных очков.

В каждой ситуации, в которой имеется выбор у игроков $И_1$ и $И_2$, имеется некоторый набор ходовых стратегий. У игрока $И_1$ стратегий всего две: брать карту при сумме очков пять или передавать ход банкомету. У банкомета число стратегий существенно больше. Он видит карту, сданную игроку $И_1$ (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), или отмечает, что игрок $И_1$ передал ход. Кроме того, у банкомета может быть сумма очков от трех до шести, при которых он может принимать одно из двух возможных решений: брать еще одну карту из колоды или передавать ход. Совокупность карты, сданной игроку $И_1$, или передачи им хода и суммы у банкомета образуют сорок четыре различные ситуации, в которых банкомет может принять одно из двух возможных решений. Таким образом, число ходовых стратегий банкомета равно 2^{44}.

Матрица игры имеет две строки и 2^{44} столбца. Построение такой матрицы практически невозможно. Тем не менее мы сейчас найдем способ определения оптимальной стратегии игроков в описываемой игре.

Обозначим через p_i вероятность того, что игрок имеет начальную сумму i очков ($i=0$, 1, 2, 3, 4, 5, 6, 7, 8, 9). Все вероятности p_i легко подсчитываются с помощью выявления всех возможных комбинаций двух карт, дающих при суммировании по модулю десять значение i. Например, при подсчете p_0 надо составить следующую таблицу:

Первая карта	Вторая карта	Число комбинаций
0	0	96×95
1	1	24×24
2	2	24×24
3	3	24×24
4	4	24×24
5	5	24×23
6	6	24×24
7	7	24×24
8	8	24×24
9	9	24×24

Учитывая, что всех возможных комбинаций по две карты существует 312×311, можно подсчитать величину p_0:

$$p_0=\frac{96\times95+24^2\times8+24\times23}{312\times311}=\frac{595}{4\,043}.$$

На самом деле подсчет вероятности p_0 произведен не совсем точно. При этом подсчете предполагалось, что вынутая из колоды карта не возвращается обратно. Это было бы верно при колоде, в которой нет одинаковых карт. В нашем же случае в колоде каждая карта встречается 6 раз. Однако проведение точного подсчета вероятностей вызывает значительные технические трудности и ненамного улучшает точность окончательных выводов. Поэтому можно воспользоваться предложенной выше схемой приблизительного вычисления вероятностей значений начальной суммы у игроков.

Буквами l и m обозначим значения карт, добавляемых игроками $И_1$ и $И_2$ к своим картам из колоды. Значения l и m равны 0, 1, 2, ..., 9. Условно будем считать, что если игрок $И_1$ или игрок $И_2$ отказываются от третьей карты, то соответствующее значение l или m равно десяти. Через q_l ($l=0, 1, \ldots, 10$) обозначим вероятность того, что очередная карта из колоды имеет сумму l. При этом $q_{10}=0$. Все значения q_l можно легко вычислить. Например, q_0 получается из тех соображений, что из 52 карт, имеющихся в колоде, лишь 16 дают число очков, равное нулю. Поэтому $q_0={}^{16}/_{52}$. Для всех $1 \leqslant l \leqslant 9$ $q_l={}^{4}/_{52}$.

Перейдем теперь к анализу поведения игрока $И_1$. У этого игрока есть две стратегии: брать или не брать очередную карту из колоды при сумме в пять очков. Пусть игрок $И_1$ мешает эти две своих чистых стратегии с частотой h. Пусть P_{il} означает вероятность того, что когда игрок $И_1$ при сумме двух своих карт, равной i ($i \leqslant 5$), берет третью карту, то сумма очков этой карты равна l. Введем характеристическую функцию $\delta(l)$, определяемую соотношением

$$\delta(l)=\begin{cases} 0 \text{ для } l=0,1,\ldots,9; \\ 1 \text{ для } l=10, \end{cases} \tag{17}$$

тогда

$$P_{il}=\left\{\begin{array}{ll} q_l & \text{для } i=0, 1, 2, 3, 4 \\ hq_l+(1-h)\delta_l & \text{для } i=5 \\ \delta_l & \text{для } i=6, 7, 8, 9 \end{array}\right\}. \tag{18}$$

Теперь рассмотрим игру банкомета. Обозначим через s одну из возможных чистых стратегий банкомета, а через $Q^{(s)}_{jlm}$ — вероятность того, что следующая карта банкомета есть m при условии, что сумма очков на начальных двух картах его равна j и он использует чистую стратегию s, видя, что очередная карта у игрока $И_1$ дает l очков. Выигрыш банкомета на каждом этапе игры определяется значениями четырех параметров i, j, l и m. Обозначим этот выигрыш как $f(i, j, l, m)$. Начальные ставки в игре любые и зависят от договоренности игроков. Все суммирование происходит по модулю десять. Значение $f(i, j, l, m)$ для $j=8{,}9$ определяется из соотношения

$$f(i, j, l, m)=\left\{\begin{array}{r} -1 \text{ для } i>j \\ 0 \text{ для } i=j \\ 1 \text{ для } i<j \end{array}\right\}. \qquad (19)$$

Для прочих j:

$$f(i, j, l, m)=\begin{cases} -1 \text{ для } i=8{,}9 \text{ или для } i+l>j+m; \\ 0 \text{ для } i+l=j+m \text{ и } i\neq 8{,}9; \\ 1 \text{ для } i+l<j+m \text{ и } i\neq 8{,}9. \end{cases} \qquad (20)$$

Тогда ожидаемый выигрыш банкомета выглядит как

$$G(h, s)=\sum_{i=0}^{9}\sum_{j=0}^{9}\sum_{l=0}^{10}\sum_{m=0}^{10} p_i P_{il} q_j Q^{(s)}_{jlm} f(i, j, l, m). \qquad (21)$$

Это выражение является линейной функцией от h, так как единственный член, содержащий h, это P_{5l}.

Теоретически с помощью соотношения (21) можно найти оптимальные стратегии для банкомета и игрока $И_1$. Для этого необходимо перебирать все возможные стратегии банкомета и вычислить все элементы платежной матрицы размера $2^{44}\times 2$. Однако практически такой способ явно непригоден из-за большого числа вычислений, которые необходимо проделать. Для уменьшения числа необходимых вычислений можно воспользоваться графическим методом, описанным в § 1. Для любого h общий ожидаемый выигрыш есть сумма по всем парам (j, l). Для каждой пары (j, l) можно рассмотреть последствия того, что банкомет берет еще одну карту из колоды или не берет. Так как при этом Q^{s}_{jlm} равно q_m или δ_m, то подставляем в (21) вместо Q^{s}_{jlm} q_m или δ_m и выбираем из двух получающихся величин большую. Эту большую величину добавляем в общую сумму выигрыша. Проделав эту операцию для всех возможных пар (j, l), можно определить максимальный выигрыш банкомета и его оптимальную стратегию. Если изменять величину h с достаточно малым шагом, то, повторяя описанную процедуру для каждого значения h, можно получить кусочно-линейную аппроксимацию функции $G(h)$. Минимальная точка на полученном графике будет соответствовать минимаксной стратегии игрока $И_1$, а на основе знания этой стратегии легко вычислить максиминную стратегию для банкомета. На рис. 17 показан результат такого графического решения при аппроксимации с шагом 0, 1. Значения $G(h)$ характеризуются следующей таблицей:

h	0	0,25	0,5	0,75	0,775
$G(h)$	1,539	1,453	1,373	1,302	1,294

На рис. 17 значения $G(h)$ даны в процентах, т. е. они равны ожидаемому выигрышу банкомета при ставке в 100 денежных единиц. Минимальная точка находится около $h=0{,}8$. Частота смешивания для банкомета определяется из построенного графика и равна приблизительно 0,6. Соответствующая оптимальная стратегия для $И_1$ при числе очков, равном 5, брать очередную карту из колоды с частотой 0,8.

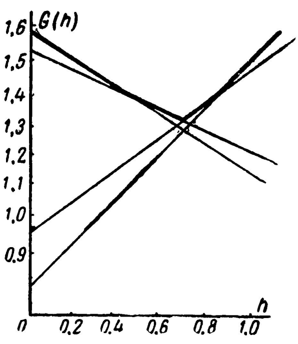


Рис. 17.

В качестве второго примера игры с неполной информацией мы рассмотрим игру в домино. Игра происходит по обычным правилам, двое на двое. Размещение игроков показано на рис. 18,*а*. На рис. 18,*б* показана общая блок-схема программы, составленной Ю. А. Первиным. В зависимости от того, кто на данном этапе игры является заходчиком, машина выбирает одну из четырех возможных тактик ведения игры. Для выбора тактики служит блок переключения тактик (БПТ). При тактике T_1, соответствующей тому, что заходчиком является игрок $И_1$, машина в качестве основной задачи своей игры рассматривает задачу обеспечения передачи первого хода игроку $И_2$. Для этого она выбирает такие ходы, которые с наибольшей вероятностью могут привести к пропуску игроком $И_1$ очередного хода. В качестве второстепенной задачи при тактике T_1 рассматривается задача подыгрывания партнеру машины (игроку $И_2$). При тактике T_2 основная задача машины — подыгрывание игроку $И_2$, а второстепенная — обеспече-

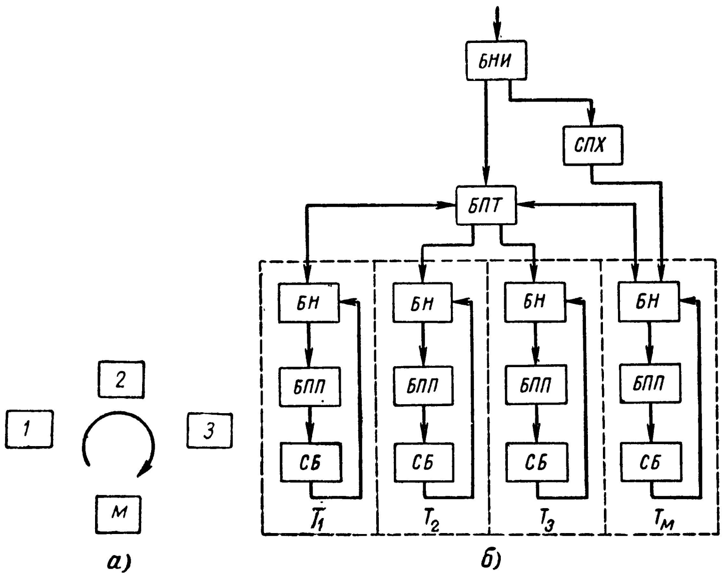


Рис. 18.

ние пропуска очередного хода игроком $И_3$. При тактике T_3 основная задача машины состоит в отборе у игрока $И_3$ права первого хода, второстепенная задача — игра «на себя». Наконец, при тактике T_M основная задача машины — игра «на себя» и второстепенная — обеспечение пропуска очередного хода игроком $И_1$.

Игра начинается игроком, у которого на руках имеется дупль 1 : 1. Для проверки наличия этого дупля у машины служит блок начала игры (БНИ). Если камень 1 : 1 у машины обнаружен, то включается схема первого хода (СПХ), выдающая указание на выставление этого камня на кон. Если камня 1 : 1 у машины нет, то с помощью блока переключения тактик (БПТ) выбирается соответствующая группа блоков, относящихся к данной тактике. Каждая такая группа состоит из трех блоков: блока наблюдения (БН), блока предтактиковых проверок (БПП) и стратегического блока (СБ). Опишем работу этих основных блоков программы.

Каждому камню $i:j\,(i,\ j=0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6)$ сопоставляются три числа $P_k^{(ij)}$ — вероятности того, что камень находится у игрока $И_k\,(k=1,\ 2,\ 3)$. В начале игры для всех камней, не имеющихся у машины, $P_1^{(ij)}=P_2^{(ij)}=$ $=p_3^{ij}=\frac{1}{3}$. В процессе игры эти вероятности пересчитываются на основе поведения игроков. Этот пересчет вероятностей происходит в блоке наблюдения. Если, например, игрок $И_k$ пропустил ход при значениях кона $(a,\ b)$, то все вероятности $P_k^{(aj)}$ и $P_k^{(bj)}$ должны быть положены равными нулю, а вероятности $P_i^{(aj)}$ и $P_i^{(bj)}\,(i\neq k)$ изменены для выполнения условия равенства единице суммы всех $P_k^{(ij)}$.

Блок наблюдений учитывает также тенденции того или иного игрока к выставлению определенных значений камней и «забиванию» некоторых значений. Практически блок наблюдений осуществляется в виде таблицы, содержащей две колонки. В левой колонке перечислены возможные наблюдения за действиями игроков, а в правом — указаны значения ΔP, служащих для увеличения или уменьшения значений $P_k^{(ij)}$. Таким образом, блок наблюдений является памятью программы и играет центральную роль в обучении машины игре в домино.

Блок предтактиковых проверок выясняет на основе анализа кона и камней машины необходимость хода, для которого выбор однозначен. Такая необходимость может встретиться при опасности «засушивания» дупля, имеющегося у машины, при необходимости сделать «рыбу» и т. д. Этот блок представляет из себя чисто логический блок, последовательно проверяющий все предусмотренные однозначные ситуации.

Если выбор хода машины неоднозначен, то в работу включается стратегический блок. Логика работы этого блока зависит от той тактики, к которой этот блок относится. Для примера ниже рассматривается работа стратегического блока для случая выбора тактики T_M.

Пусть a и b означают число очков на концах кона. При реализации хода машине необходимо решить два вопроса: на какой конец кона выгоднее выставить камень и какой камень выгоднее выставить на выбранный конец кона. Введем следующие числовые характеристики, подсчитываемые машиной при выборе очередного хода:

K_x — сумма числа камней, содержащих x, которые уже выставлены на кон, и имеющихся у машины;

L_x — число камней, содержащих x, которое было у машины в начале игры;

M_x^1 — математическое ожидание числа камней, содержащих „x“ у игрока $И_1$

$$\left(M_x^1 = \sum_{i=0}^{6} P_1^{ix}\right);$$

M_x^2 — математическое ожидание числа камней, содержащих „x“ у игрока $И_2$

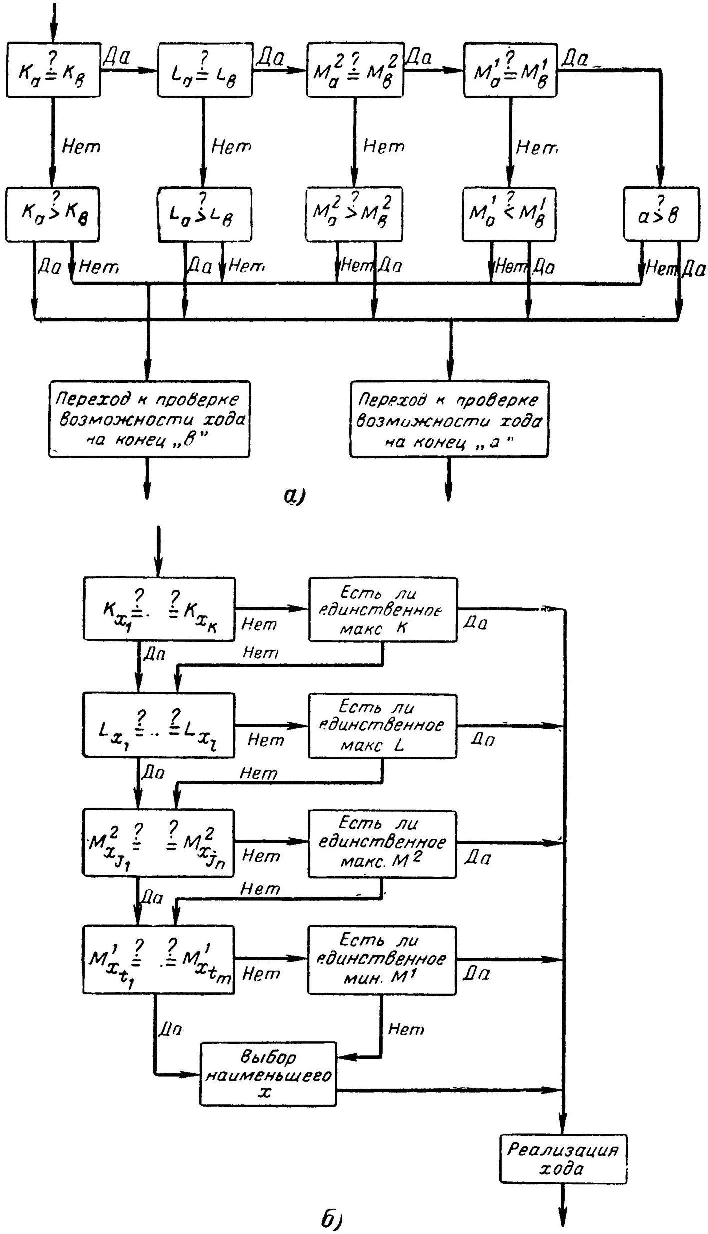
$$\left(M_x^2 = \sum_{i=0}^{6} P_2^{ix}\right).$$

На рис. 19,*а* показана блок-схема выбора машиной конца кона, на который нужно выставлять очередной камень.

Так как по правилам игры допускается в начале игры не более пяти камней, содержащих x, в руках у одного игрока, то число проверок на втором этапе не более пяти. На рис. 19,*б* показана блок-схема реализации второго этапа работы стратегического блока. При этом $k \geqslant l \geqslant n \geqslant m$. Если ни одна из проверок не дает результата, то выставляется камень, для которого имеет место $\underset{x}{\text{мин}} \{(a, x), (b, x)\}$. Этот ход предотвращает выставление игроком $И_1$ камня с большим числом очков.

Остальные стратегические блоки работают аналогичным образом.

Описанный подход к нахождению ходовой стратегии в игре домино не обязательно приводит к нахождению



$K_a \stackrel{?}{=} K_в$
Да
$L_a \stackrel{?}{=} L_в$
$M_a^2 \stackrel{?}{=} M_в^2$
$M_a^1 \stackrel{?}{=} M_в^1$
Нет
$K_a \stackrel{?}{>} K_в$
$L_a \stackrel{?}{>} L_в$
$M_a^2 \stackrel{?}{>} M_в^2$
$M_a^1 \stackrel{?}{<} M_в^1$
$a \stackrel{?}{>} в$
Переход к проверке возможности хода на конец „в"
Переход к проверке возможности хода на конец „а"
а)
$K_{x_1} \stackrel{?}{=} \dots \stackrel{?}{=} K_{x_k}$
Есть ли единственное макс K
$L_{x_1} \stackrel{?}{=} \dots \stackrel{?}{=} L_{x_l}$
Есть ли единственное макс L
$M^2_{x_{j_1}} \stackrel{?}{=} \dots \stackrel{?}{=} M^2_{x_{j_n}}$
Есть ли единственное макс. M^2
$M^1_{x_{t_1}} \stackrel{?}{=} \dots \stackrel{?}{=} M^1_{x_{t_m}}$
Есть ли единственное мин. M^1
Выбор наименьшего x
Реализация хода
б)

Рис. 19.

оптимальной полной стратегии. Этот подход весьма напоминает метод оценки очередного хода при игре в шахматы, рассмотренный нами в предыдущем параграфе. Оба эти метода являются некоторым видоизменением общей схемы эвристического способа решения сложной игровой задачи. Поэтому структура автомата, решающего игровую задачу при большом числе полных стратегий, исключающем возможность практического анализа матрицы игры, одинакова как для автомата, играющего в игру с полной информацией, так и в игру с неполной информацией. При этом игра с полной информацией просто сводится к игре с неполной информацией за счет исключения из рассмотрения части имеющейся информации.

Однако при не слишком большом числе полных стратегий автомат, играющий в игру с неполной информацией, отличается по своей структуре от автомата, решающего задачу о нахождении оптимальной стратегии в случае игры с полной информацией. Это отличие проявляется в том, что такой автомат должен уметь находить оптимальную смешанную стратегию, для чего необходимо иметь в автомате блок выработки смешанных стратегий, построенный по принципу датчика случайных чисел. Для оценки действий противника необходимо иметь блок подсчета относительных частот использования противником своих стратегий. Эти два блока усложняют структуру автомата, решающего игру с неполной информацией, по сравнению с автоматом, играющим в игру с полной информацией, так как этот последний состоит только из блока памяти, представляющего собой таблицу с одним или двумя входами.

6. ИГРЫ ПРОТИВ ПРИРОДЫ

В этом и следующем параграфах мы рассмотрим игры, в которых противником игрока $И_1$ является игрок, незаинтересованный в получении максимально возможного выигрыша. Такой игрок был назван нами в §1 природой и класс подобных игр принято называть играми против природы. Игры против природы можно представить следующим образом. Игрок $И_2$ имеет только случайные ходы и не имеет личных ходов, а игрок $И_1$

имеет некоторое множество личных ходов, с помощью которых он стремится максимизировать свой выигрыш.

В теории игр, развитой Нейманом, модель игры такого типа не рассматривается. В модели Неймана необходимо, чтобы интересы игроков были бы противоположны и каждый из игроков имел бы в своем распоряжении по крайней мере один личный ход. Кроме того, в модели Неймана вероятности исходов всех случайных ходов считаются заранее известными.

Мы в дальнейшем не будем предполагать, что известны априорные вероятности исходов случайных ходов.

Поэтому перед игроком $И_1$ в таких условиях игры стоят две задачи. Во-первых, он должен на основе наблюдений за игроком $И_2$ найти вероятности выбора этим игроком своих стратегий, а, во-вторых, на основе знания об этих вероятностях выбрать свою стратегию оптимальным образом.

В данном параграфе мы рассмотрим решение задачи, стоящей перед игроком $И_1$, на простом примере игры в монетку.

Два игрока $И_1$ и $И_2$ одновременно выбирают одну из двух возможностей: «герб» или «решетку». Их совместный выбор определяет платежи игроков в соответствии со следующей матрицей игры.

Игрок $И_2$

		герб	решетка
Игрок $И_1$	герб	a	$-b$
	решетка	$-c$	d

Игрок $И_1$ делает свой выбор на основе личного хода, а выбор игрока $И_2$ происходит с помощью случайного механизма, дающего выбор герба с вероятностью $P_г$ и выбор решетки с вероятностью $P_г = 1 - P_г$.

Для игрока $И_1$ всегда имеется возможность на основе анализа матрицы игры и теоремы о минимаксе выбрать такую оптимальную минимаксную стратегию, которая обеспечивает ему выигрыш, равный цене игры. Но в силу поведения противника игрок $И_1$ может получить и больший выигрыш, чем это гарантировано теоремой об оптимальных смешанных стратегиях. Для этого необходимо учесть вероятности выбора игроком $И_2$ своих стратегий и на основе этого выбрать ответную стратегию.

Мы будем предполагать, что игрок $И_2$ является игроком стационарного типа. Игрок более общего типа будет рассмотрен нами в последующих параграфах.

Модель игры такого типа может быть использована при решении задач автоматического управления объектами сложной природы. Как правило, описания поведения таких объектов обычно неполны, так как в этих описаниях отсутствует влияние на характер процесса целого ряда факторов, учет которых затруднителен, а иногда и невозможен, так как сами эти факторы остаются неизвестными. Тогда такой объект ведет себя как игрок, выбирающий свою стратегию на основании некоторого закона распределения, который может либо оставаться неизменным, либо меняться в зависимости от изменения некоторых факторов, влияющих на регулируемый объект.

Для иллюстрации поведения игрока, обладающего личными ходами и стремящегося максимизировать свой выигрыш, в играх такого типа рассмотрим описанную выше игру в монетку.

Будем предполагать, что числа a, b, c и d положительны.

Рассмотрим два типа игроков $И_2$: игрока бернуллиевского типа и игрока марковского типа. Игрок бернуллиевского типа характеризуется тем, что, выбрав до начала игры $P_г$ из некоторого распределения, он не изменяет значение вероятности выбора герба на протяжении всей игры. Другими словами, игрок бернуллиевского типа не использует никакой информации о результатах игры. При игре против игрока такого типа игрок $И_1$ знает распределение, из которого выбирается игроком $И_2$ значение $P_г$, но само значение $P_г$ игроку $И_1$ заранее неизвестно.

Игрок марковского типа характеризуется тем, что он выбирает два значения $P_г$ из различных распределений. Эти два значения $P_г$ мы обозначим как $P_{гг}$ и $P_{гр}$. $P_{гг}$ представляет собой вероятность выбора герба игроком $И_2$ при условии, что при своем предыдущем ходе он выбрал герб, а $P_{гр}$ — вероятность того, что игрок $И_2$ выберет на данном ходе герб при условии, что на предыдущем шаге игры он выбрал решетку. Значения $P_{гг}$ и $P_{гр}$ выбираются игроком $И_2$ до начала игры и в дальнейшем не меняются. Игроку $И_1$ эти значения заранее неизвестны, но известны распределения, из которых они могут быть выбраны.

Поведение игрока $И_1$ в игре с подобными противниками определяется следующим образом. В начале игры он задается некоторыми априорными вероятностями выбора герба противником и играет против него так, как будто априорные вероятности соответствуют действительным значениям $P_г$ или $P_{гг}$ и $P_{гр}$. В процессе игры происходит уточнение этих вероятностей. На каждом шаге игры игрок $И_1$ пользуется апостериорными вероятностями выбора герба противником и отождествляет их с действительными вероятностями такого выбора.

Вначале рассмотрим проблему нахождения наилучшей стратегии для игрока $И_1$ при игре против противника бернуллиевского типа. Для такого случая математическое ожидание выигрыша игрока $И_1$ имеет следующий вид:

$$M(И_1)=Q_гP_гa-Q_гP_рb-Q_рP_гc+Q_рP_рd, \qquad (22)$$

где $Q_г$ — вероятность того, что $И_1$ выбирает герб; $P_г$ — вероятность того, что противник выбирает герб; $Q_р$ — вероятность выбора игроком $И_1$ решетки; $P_р$ — вероятность выбора противником решетки. При этом, конечно, $Q_р=1-Q_г$ и $P_р=1-P_г$. Соотношение (22) путем несложных преобразований может быть приведено к следующему виду:

$$M(И_1)=\alpha Q_г+\beta, \qquad (23)$$

где

$$\alpha=(a-b-c+d)P_г+b-d$$

и

$$\beta=(c-d)P_г+d.$$

Из соотношения (23) следует, что игрок $И_1$ при выборе очередного хода должен исходить из знака величины α. Если α положительно, то игрок $И_1$ должен выбирать герб, если α отрицательно, то — решетку. При нулевом значении α выбор игрока $И_1$ безразличен и он может равновероятно выбрать герб или решетку.

Противник может выбирать величину $P_г$ некоторым вероятностным образом, но если он является игроком бернуллиевского типа, то, произведя в начале партии такой выбор, он должен придерживаться его на протяжении всей игры. Это дает возможность оценить значение $P_г$ на основе анализа некоторого количества этапов игры.

Обозначим через $P(P_г)$ априорную вероятность того, что противник на данном этапе игры имеет вероятность выбора герба, равную $P_г$. Пусть после n этапов противник выбрал r раз герб и s раз решетку. Оценим величину условной вероятности выбора вероятности герба $P_г$ при условии, что на прошедших n этапах были выбраны r гербов и s решеток. Для этой оценки используем известную теорему Бейеса:

$$P(P_г/r, s) = \frac{P(r, s/P_г)\,P(P_г)}{\int\limits_0^1 P(r, s/P_г)\,P(P_г)\,dP_г};$$

здесь $P(r, s/P_г)$ — условная вероятность появления r гербов и s решеток при выборе вероятности выбора гербов, равной $P_г$.

Для независимых испытаний

$$P(r, s/P_г) = \frac{n!}{r!\,s!} P_г^r P_р^s$$

и

$$P(P_г/r, s) = \frac{P_г^r P_р^s P(P_г)}{\int\limits_0^1 P_г^r P_р^s P(P_г)\,dP_г}. \qquad (24)$$

Предположим теперь, что $P_г$ имеет равномерное распределение ($P(P_г) = 1$). В этом случае

$$\int\limits_0^1 P_г^r P_р^s\,dP_г = \frac{r!\,s!}{(n+1)!}$$

или

$$P(P_г/r, s) = \frac{(n+1)!}{r!\,s!} P_г^r P_р^s.$$

Тогда можно подсчитать ожидаемое значение $P_\text{г}$ на $(n+1)$-м этапе игры

$$M(P_\text{г})=\int\limits_0^1 \frac{(n+1)!}{r!\,s!}P_\text{г}^{r+1}P_\text{р}^{s}\,dP_\text{г}=\frac{r+1}{n+2}. \qquad (25)$$

Полученное соотношение дает игроку $И_1$ возможность выбрать наилучшее решение на данном этапе игры на основе анализа предыдущих этапов.

Зная на каждом этапе значение $M(P_\text{г})$, игрок $И_1$ может путем сравнения отношения величин γ и δ с величиной $M(P_\text{г})$ определять наилучший выбор. Величины γ и δ определяются соотношениями $\gamma=d-b$ и $\delta=a+$ $+d-b-c$. Если $M(P_\text{г})>\frac{\gamma}{\delta}$, то игрок $И_1$ на данном этапе игры выбирает герб, в противном случае он выбирает решетку. Для случая равномерного распределения $P_\text{г}$ и матрицы игры, для которой $a=b=c=d=1$, выбор герба игроком $И_1$ определяется на основании следующего соотношения

$$Q_\text{г}=\begin{cases}0, \text{ если } \dfrac{r+1}{n+2}>\dfrac{1}{2};\\[2ex] 1, \text{ если } \dfrac{r+1}{n+2}<\dfrac{1}{2}.\end{cases} \qquad (26)$$

Теперь мы рассмотрим игру против противника марковского типа. Прежде чем перейти к анализу поведения в игре против такого противника, приведем без всякого доказательства основные определения и свойства стационарных дискретных марковских цепей, аппарат которых будет нами использоваться в данном примере и в § 7 и 8. Для читателей, не интересующихся обоснованием утверждений, формулируемых ниже, а интересующихся лишь практическими результатами использования этого аппарата, приводимых нами сведений будет достаточно. Лиц, желающих более глубоко изучить марковские цепи, можно отослать к книге В. И. Романовского «Дискретные цепи Маркова», Гостехиздат, 1949 г., в которой дано наиболее полное и строгое изложение этого аппарата.

Рассмотрим конечное множество $S=\{s_1, s_2, \ldots, s_n\}$, которое назовем множеством состояний, и матрицу порядка $n \times n$:

$$P=\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}. \tag{27}$$

Для элементов этой матрицы выполнены соотношения:

$$\left.\begin{aligned} &0 \leqslant p_{ij} \leqslant 1 \\ &\sum_{j=1}^{n} p_{ij}=1 \end{aligned}\right\}, i, j=1,2, \ldots, n, \tag{28}$$

тогда p_{ij} можно рассматривать как вероятность перехода системы из состояния с номером i в состояние с номером j. Зададим еще случайную величину ξ, закон распределения которой имеет вид:

ξ	1	2	...	n
p^0	p_1^0	p_2^0	...	p_n^0

(29)

Величины p_j^0 характеризуют вероятность нахождения системы в начале процесса в состоянии с номером j.

Совокупность (27) и (29) называется дискретной цепью Маркова. При сравнении (27) с матрицей B, определяемой соотношением (6), видно, что матрица (27) может рассматриваться как матрица смены состояний марковского автомата.

. Подчеркнем, что процесс, описываемый с помощью дискретной марковской цепи, характеризуется тем, что состояние цепи в данный момент зависит непосредственно только от состояния цепи в предшествующий момент времени.

Рассмотрим k-ю степень матрицы P:

$$P^k=P^{k-1}P=PP^{k-1}$$

или

$$p_{ij}^{k}=\sum_{l=1}^{n} p_{il}^{k-1} p_{lj}. \tag{30}$$

Зафиксируем некоторое состояние цепи с номером i. Вероятность того, что в начале процесса система находится в состоянии s_i согласно (29) равна p_i^0. Вероятность того, что система будет находиться в этом состоянии после первого шага процесса, очевидно, определится следующим образом: на нулевом шаге система может находиться в некотором состоянии s_j, а при переходе к первому шагу она должна перейти из этого состояния в состояние s_i. Но вероятность такого события в силу независимости вероятностей перехода на каждом шаге цепи Маркова есть произведение вероятности того, что система в начальный момент находилась в состоянии s_j, на вероятность перехода из состояния s_j в состояние s_i, которая определяется элементом p_{ji} матрицы P. Беря совокупность всех j, из которых возможен переход в состояние s_i, мы суммируем получаемые произведения, так как события, определяемые найденными вероятностями, несовместны. Итак,

$$p_i^1=\sum_{j=1}^{n} p_j^0 p_{ji}.$$

Проиллюстрируем приведенное выше рассуждение на конкретном примере. Пусть цепь Маркова задана следующим образом:

$$P=\begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$$

ξ	1	2	3
p^0	$\frac{3}{4}$	0	$\frac{1}{4}$

На рис. 20 показана цепь Маркова, соответствующая нашему примеру. Слева от вертикальной пунктирной линии показаны возможные начальные состояния этой цепи и около них написаны вероятности пребывания системы в этих состояниях. Справа от вертикального пунктира показаны состояния, в которые система может перей-

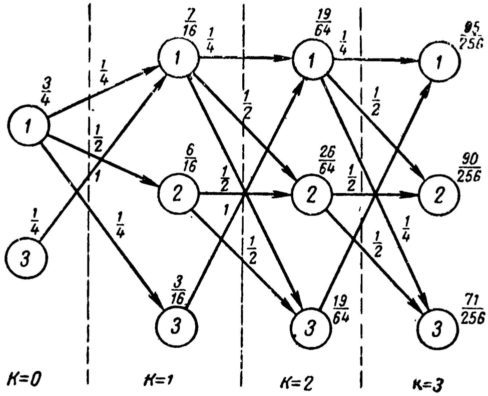


Рис. 20.

ти из начальных состояний, а на соответствующих этим переходам ребрах написаны вероятности переходов, взятые из матрицы P. Тогда вероятность того, что после первого шага система будет находиться в состоянии p_i^1, определяется следующим образом:

$$p_1^1 = p_1^0 p_{11} + p_2^0 p_{21} + p_3^0 p_{31} =$$

$$= \frac{3}{4} \cdot \frac{1}{4} + 0 \cdot 0 + \frac{1}{4} \cdot 1 = \frac{7}{16};$$

$$p_2^1 = p_1^0 p_{12} + p_2^0 p_{22} + p_3^0 p_{32} =$$

$$= \frac{3}{4} \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 = \frac{6}{16};$$

$$p_3^1 = p_1^0 p_{13} + p_2^0 p_{23} + p_3^0 p_{33} =$$

$$= \frac{3}{4} \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 = \frac{3}{16}.$$

Эти значения написаны около соответствующих состояний. Но теперь можно рассматривать величину p_1^1, p_2^1, p_3^1 как начальные для второго шага в цепи Маркова. Тогда

$$p_1^2 = p_1^1 p_{11} + p_2^1 p_{21} + p_3^1 p_{31} =$$

$$= \frac{7}{16} \cdot \frac{1}{4} + \frac{6}{16} \cdot 0 + \frac{3}{16} \cdot 1 = \frac{19}{64};$$

$$p_2^2 = p_1^1 p_{12} + p_2^1 p_{22} + p_3^1 p_{32} =$$

$$= \frac{7}{16} \cdot \frac{1}{2} + \frac{6}{16} \cdot \frac{1}{2} + \frac{3}{16} \cdot 0 = \frac{26}{64};$$

$$p_3^2 = p_1^1 p_{13} + p_2^1 p_{23} + p_3^1 p_{33} =$$

$$= \frac{7}{16} \cdot \frac{1}{4} + \frac{6}{16} \cdot \frac{1}{2} + \frac{3}{16} \cdot 0 = \frac{19}{64}.$$

Продолжая вычисления таким же образом, мы получим:

$$p_1^3 = p_1^2 p_{11} + p_2^2 p_{21} + p_3^2 p_{31} =$$

$$= \frac{19}{64} \cdot \frac{1}{4} + \frac{26}{64} \cdot 0 + \frac{19}{64} \, 1 = \frac{95}{256};$$

$$p_2^3 = p_1^2 p_{12} + p_2^2 p_{22} + p_3^2 p_{32} =$$

$$= \frac{19}{64} \cdot \frac{1}{2} + \frac{26}{64} \cdot \frac{1}{2} + \frac{19}{64} \cdot 0 = \frac{90}{256};$$

$$p_3^3 = p_1^2 p_{13} + p_2^2 p_{23} + p_3^2 p_{33} =$$

$$= \frac{19}{64} \cdot \frac{1}{4} + \frac{26}{64} \cdot \frac{1}{2} + \frac{19}{64} \cdot 0 = \frac{71}{256}.$$

Отсюда следует, что для получения вероятности нахождения системы в состоянии s_i после k шагов необходимо вычислить элементы P^k, определяемые соотношением (30), и умножить вектор начального распределения на P^k. Искомая вероятность есть i-я компонента результирующего вектора.

Если матрица P имеет большие размеры, то вычисления, которые необходимо проделать при вычислении величин p_j^k, могут быть слишком громоздкими. В теории марковских цепей имеется способ, позволяющий находить величины p_j^k без вычисления степеней матрицы P. Этот способ основан на использовании формулы Перрона. Для объяснения этой формулы мы введем понятие характеристического уравнения для матрицы P. Рассмотрим следующую матрицу:

$$P(\lambda) = \begin{pmatrix} p_{11} - \lambda & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} - \lambda & \cdots & p_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ p_{n1} & p_{n2} & \cdots & p_{nn} - \lambda \end{pmatrix},$$

которая получается из матрицы P путем вычитания из всех элементов P, лежащих на главной диагонали, некоторого параметра λ. Рассмотрим уравнение $|P(\lambda)| = 0$, которое будем называть характеристическим уравнением матрицы P, а его корни — собственными значениями P.

Так как матрица P есть матрица довольно специального вида [для нее всегда выполняются соотношения (28)], то можно сформулировать следующее утверждение.

Всякая матрица P, удовлетворяющая соотношениям (28), имеет максимальное собственное значение $\lambda_{\text{макс}} = 1$.

Пусть P имеет собственные значения $\lambda_1, \lambda_2, \ldots, \lambda_q$ с кратностями $m_1, m_2, \ldots, m_q$. Определим функции $\psi_l(\lambda)$ на основании соотношений:

$$|P(\lambda)| = (\lambda - \lambda_l)^{m_l} \psi_l(\lambda), \quad l = 1, 2, \ldots, q.$$

Формула Перрона утверждает следующее. Элемент p_{ij}^k, стоящий в i-й строке и j-м столбце матрицы P^k, может быть вычислен из соотношения

$$p_{ij}^k = \sum_{l=1}^{q} \frac{1}{(m_l - 1)!} D_\lambda^{m_l - 1} \left\{ \frac{\lambda^k M_{ji}(\lambda)}{\psi_l(\lambda)} \right\}_{\lambda = \lambda_l}, \tag{31}$$

где $M_{ji}(\lambda)$ означает алгебраическое дополнение элемента, стоящего в j-й строке и i-м столбце матрицы $P(\lambda)$, а $D_\lambda^{m_l-1}$ означает производную (m_l-1)-го порядка по λ. Подстановка $\lambda=\lambda_l$ выполняется после дифференцирования.

Проиллюстрируем применение формулы Перрона для марковской цепи, определяемой следующей матрицей

$$P=\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Соответствующее характеристическое уравнение для этой матрицы имеет вид:

$$\begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & \frac{1}{2}-\lambda & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2}-\lambda \end{vmatrix}=0$$

или

$$(1-\lambda)\left(\frac{1}{2}-\lambda\right)^2=0,$$

отсюда $\lambda_1=1$; $\lambda_2=\frac{1}{2}$; $m_1=1$; $m_2=2$.

Определим функции $\psi_l(\lambda)$:

$$(1-\lambda)\left(\frac{1}{2}-\lambda\right)^2=(\lambda-1)\,\psi_1(\lambda); \quad \psi_1(\lambda)=-\left(\frac{1}{2}-\lambda\right)^2;$$

$$(1-\lambda)\left(\frac{1}{2}-\lambda\right)^2=\left(\lambda-\frac{1}{2}\right)^2\psi_2(\lambda); \quad \psi_2(\lambda)=1-\lambda.$$

Найдем на основании (35) элемент p_{23}^4 матрицы P^4:

$$p_{23}^4 = \left.\frac{\lambda^4 M_{32}}{-\left(\frac{1}{2}-\lambda\right)^2}\right|_{\lambda=1} + D_\lambda^1\left\{\frac{\lambda^4 M_{32}}{1-\lambda}\right\}_{\lambda=\frac{1}{2}} =$$

$$= \left.\frac{\lambda^4\begin{vmatrix}1-\lambda & 0\\ 0 & \frac{1}{2}\end{vmatrix}}{-\left(\frac{1}{2}-\lambda\right)^2}\right|_{\lambda=1} + D_\lambda^1\left\{\frac{\lambda^4\begin{vmatrix}1-\lambda & 0\\ 0 & \frac{1}{2}\end{vmatrix}}{1-\lambda}\right\}_{\lambda=\frac{1}{2}} =$$

$$= 0 + \left.\frac{4\lambda^3}{2}\right|_{\lambda=\frac{1}{2}} = \frac{1}{4}.$$

Аналогичным образом могут быть найдены и остальные элементы этой матрицы.

Предположим, что матрица P такова, что корень $\lambda=1$ является корнем кратности единица. Тогда запишем формулу Перрона в следующем виде:

$$p_{ij}^k = p_j + \sum_{l=2}^{q}\frac{1}{(m_l-1)!}D_\lambda^{m_l-1}\left\{\frac{\lambda^k M_{ji}}{\psi_l(\lambda)}\right\}_{\lambda=\lambda_l}, \tag{32}$$

где

$$p_j = \frac{M_{jj}(1)}{\sum\limits_{i=1}^{n} M_{ii}(1)}. \tag{33}$$

В силу того, что все остальные собственные значения не превосходят единицы, а величина

$$Q_{ijl}(k) = \frac{1}{(m_l-1)!}D_\lambda^{m_l-1}\left\{\frac{\lambda^k M_{ji}(\lambda)}{\psi_l(\lambda)}\right\}_{\lambda=\lambda_l}$$

представляет собой многочлен от k, степень которого не выше, чем m_l-1, то при переходе к пределу при $k\to\infty$ в (32), мы получим:

$$\lim_{k\to\infty} p_{ij}^k = p_j. \tag{34}$$

Отметим, что результат этого предельного перехода не зависит от i. Вероятности p_j, определяемые соотношением (33), называются **финальными** или **предельными вероятностями** цепи Маркова.

Смысл финальных вероятностей состоит в том, что для достаточно большого числа шагов можно считать, что поведение системы, описываемой цепью Маркова, становится практически независимым от состояния системы в начальный момент времени, а определяется лишь значениями финальных вероятностей, дающих вероятность пребывания системы в состоянии с номером j.

Этот факт можно сформулировать следующим образом. Если вектор $\overline{P}=(p_1, p_2, \ldots, p_n)$ есть вектор, компонентами которого являются финальные вероятности, а P есть матрица, определяющая марковскую цепь, то должно иметь место соотношение

$$\overline{P}=\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \times \overline{P}. \tag{35}$$

Это соотношение может рассматриваться как система уравнений, с помощью которых можно определить значения финальных вероятностей. При этом необходимо еще использовать соотношение нормировки $p_1+p_2+ + \ldots +p_n=1$.

Отметим один достаточный признак существования у цепи Маркова финальных вероятностей. Если все элементы матрицы P положительны (среди элементов этой матрицы нет нулевых), то цепь Маркова, определяемая этой матрицей, всегда имеет финальные вероятности.

Вернемся теперь к рассмотрению игры двух игроков в бросание монеты и будем теперь считать, что игрок $И_2$ является игроком марковского типа. Другими словами, вероятность выбора этим игроком герба или решетки зависит от того выбора, который был сделан этим игроком на предыдущем шаге игры. Обозначим через $p_{гг}$ и $p_{гр}$ вероятность выбора на данном шаге игры герба игроком $И_2$ при условии, что на предыдущем шаге им был выбран соответственно герб и решетка. Марковский процесс является цепью с двумя состояниями и матрица P для него имеет вид:

$$P=\begin{pmatrix} p_{гг} & p_{гр} \\ p_{рг} & p_{рр} \end{pmatrix}=\begin{pmatrix} p_{гг} & 1-p_{гг} \\ p_{рг} & 1-p_{рг} \end{pmatrix}.$$

Условно принято, что первое состояние соответствует выбору герба на предыдущем шаге, а второе состояние — выбору решетки на предыдущем шаге. Обозначим через $p_г$ и $p_р$ финальные вероятности выбора герба и решетки. На основании (35) получаем:

$$\begin{cases} p_г = p_{гг}p_г + p_{гр}p_р; \\ p_р = p_{рг}p_г + p_{рр}p_р; \\ p_г + p_р = 1, \end{cases}$$

отсюда

$$p_г = \frac{p_{рг}}{1 + p_{рг} - p_{гг}}.$$

Теперь можно определить условные математические ожидания выигрыша

$$M_г = p_г(г)\,[\gamma p_{гг} - \delta] + [(-c-d)\,p_{гг} + d];$$

$$M_р = p_р(г)\,[\gamma p_{рг} - \delta] + [(-c-d)\,p_{рг} + d]; \qquad (36)$$

здесь $M_г$ — математическое ожидание выигрыша, если игрок $И_2$ перед этим поставил герб;

$M_р$ — математическое ожидание выигрыша, если игрок $И_2$ перед этим поставил решетку;

$p_г(г)$ — условная вероятность того, что $И_1$ поставит герб после того, как $И_2$ поставит герб;

$p_р(г)$ — условная вероятность того, что $И_1$ поставит герб после того, как $И_2$ поставит решетку.

Из (36) видно, что максимальное значение выигрыша достигается в том случае, когда параметры $p_г(г)$ или $p_р(г)$ принимают одно из двух крайних значений. В нижеследующей таблице приведен выбор этих значений, дающий оптимальную стратегию поведения игрока $И_1$ против игрока $И_2$ марковского типа.

Рядом авторов (Д. Хагельбергером, К. Шенноном, Дж. Уайтом и др.) были сконструированы автоматы, играющие в монетку. Все эти автоматы основаны на

Область	Условные вероятности		Оптимальное поведение		Математическое ожидание выигрыша	Ограничение на $\frac{\delta}{\gamma}$
	$p_{гг}$	$p_{рг}$	$p_г(г)$	$p_р(г)$		
1	$>\frac{\delta}{\gamma}$	$>\frac{\delta}{\gamma}$	1	1	$\gamma p_г - \delta + cp_г + dp_р$	$\frac{\delta}{\gamma} \leqslant 1$
2	$>\frac{\delta}{\gamma}$	$<\frac{\delta}{\gamma}$	1	0	$p_г[\gamma p_{гг} - \delta] + cp_г + dp_р$	$0 \leqslant \frac{\delta}{\gamma} \leqslant 1$
3	$<\frac{\delta}{\gamma}$	$>\frac{\delta}{\gamma}$	0	1	$p_р[\gamma p_{рг} - \delta] + cp_г + dp_р$	$0 \leqslant \frac{\delta}{\gamma} \leqslant 1$
4	$<\frac{\delta}{\gamma}$	$<\frac{\delta}{\gamma}$	0	0	$cp_г + dp_р$	$\frac{\delta}{\gamma} \geqslant 0$
5	$=\frac{\delta}{\gamma}$	$=\frac{\delta}{\gamma}$	Любое значение	Любое значение	$cp_г + dp_р$	$0 \leqslant \frac{\delta}{\gamma} \leqslant 1$

принципе, сущность которого сводится к тому, что человек, играя в подобную игру, как правило, действует не случайным, а закономерным образом и эта закономерность может быть обнаружена автоматом по небольшому ряду последовательных наблюдений. Закономерность игры человека в большинстве случаев им не осознается и он уверен, что его выбор носит случайный характер. Автоматы такого типа можно назвать автоматами, играющими против противника, ходы которого псевдослучайны. Идею подобных автоматов рассмотрим на примере автомата Хагельбергера. На примере этого автомата видны преимущества подобных устройств и недостатки, присущие всем автоматам такого типа.

Автомат и его противник одновременно называют герб или решетку (ту или другую сторону монеты). При совпадении названных сторон выигрывает машина, в противном случае она остается в проигрыше. Автомат

Хагельбергера обладает некоторой памятью. Он помнит результаты двух последних партий игры и информацию о своих действиях в последней партии. Условимся считать, что входной сигнал автомата, равный нулю, свидетельствует о том, что автомат выиграл данную партию игры. Подача же на вход автомата сигнала, равного единице, свидетельствует о проигрыше автомата в данной партии. Состоянием автомата назовем тройку двоичных разрядов. Первые два разряда тройки характеризуют проигрыши и выигрыши в двух последних сыгранных автоматом партиях, а в третьем разряде состояния пишется 0 или 1, в зависимости от того, герб или решетку выбирал автомат при реализации последней партии игры. Таким образом, состояние 100, например, означает, что автомат проиграл предпоследнюю партию, выиграл последнюю и в последней партии выбрал герб. Наконец, выход автомата также является двоичным числом, определяемым следующим образом: ноль на выходе автомата означает, что автомат выбирает в данной партии игры герб, а единица соответствует выбору автоматом в данной партии решетки.

Как следует из вышесказанного, автомат Хагельбаргера является конечным марковским автоматом с памятью. Соответствующие этому автомату матрицы A и B, построенные по типу матриц (5) и (6), имеют вид:

Вход автомата	Состояние автомата	Выход автомата 0	Выход автомата 1
0	000	1	0
1	000	0,75	0,25
0	001	0	1
1	001	0,25	0,75
0	010	0,75	0,25
1	010	0,5	0,5
0	011	0,25	0,75
1	011	0,5	0,5
0	100	1	0
1	100	0,75	0,25
0	101	0	1
1	101	0,25	0,75
0	110	0,75	0,75
1	110	0,5	0,5
0	111	0,25	0,75
1	111	0,5	0,5

Как следует из этих матриц, автомат выбирает свой ход на основе анализа исходов двух последних партий игры. При выигрыше в двух последних партиях автомат сохраняет выбор последней партии; при проигрыше в двух последних партиях автомат с равной вероятностью выбирает герб или решетку; при наличии в двух последних партиях одного выигрыша и проигрыша автомат с вероятностью 0,75 сохраняет выбор, имевший место в последней партии.

Вход автомата	Состояние автомата	Новое состояние автомата							
		000	001	010	011	100	101	110	111
0	000	1	0	0	0	0	0	0	0
1	000	0	0	0,75	0,25	0	0	0	0
0	001	0	1	0	0	0	0	0	0
1	001	0	0	0,25	0,75	0	0	0	0
0	010	0	0	0	0	0,75	0,25	0	0
1	010	0	0	0	0	0	0	0,5	0,5
0	011	0	0	0	0	0,25	0,75	0	0
1	011	0	0	0	0	0	0	0,5	0,5
0	100	1	0	0	0	0	0	0	0
1	100	0	0	0,75	0,25	0	0	0	0
0	101	0	1	0	0	0	0	0	0
1	101	0	0	0,25	0,75	0	0	0	0
0	110	0	0	0	0	0,75	0,25	0	0
1	110	0	0	0	0	0	0	0,5	0,5
0	111	0	0	0	0	0,25	0,75	0	0
1	111	0	0	0	0	0	0	0,5	0,5

На рис. 21 показана диаграмма переходов для автомата Хагельбергера. На дугах надписаны значения входов, при которых реализуется этот переход, а в скобках указана вероятность этой реализации. Выход автомата (т. е. выбор в данной партии герба или решетки) определяется последним разрядом того состояния, в которое переходит на данном шаге автомат.

Автомат Хагельбергера помнит результаты лишь двух последних партий, поэтому его возможности невелики. Он всегда выигрывает у игрока, придерживающегося следующих трех периодических последовательностей выбора хода: *гггг...*, *рррр...*, *гргргр*... В остальных случаях выигрыш машины не является стопроцентным.

Существует оптимальная смешанная стратегия для игрока $И_1$, которая дает ему вероятность выигрыша 0,6. Однако при практических испытаниях автомата Хагельбергера выяснилось, что интуитивно ни один из противников автомата не смог найти оптимальную смешанную стратегию, а все игроки $И_1$ придерживались некоторой более и менее очевидной закономерности в своих выборах. Это позволяло машине получить вероятность выигрыша большей, чем 0,4.

Автомат Хагельбергера содержит около 100 реле. Его схема приведена в работе Хагельбергера, указанной в списке литературы.

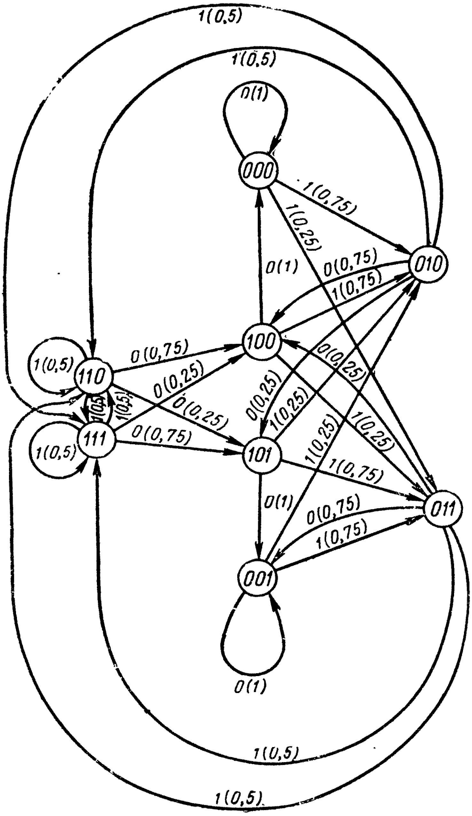


Рис. 21.

7. АВТОМАТЫ В СЛУЧАЙНЫХ СРЕДАХ

В этом параграфе мы продолжим рассмотрение игр автоматов против природы. На рис. 22 показана общая схема системы автомат—среда, которая будет предметом нашего изучения. Среда в дальнейшем рассматривается как источник сигналов двух типов (0 и 1), подаваемых в моменты времени t на вход автомата. Сигнал ноль называется **нештрафом**, а сигнал единица—**штрафом**. Поступление сигналов нештраф и штраф на вход автомата носит вероятностный характер. Для каждого возможного выходного сигнала автомата Y_i, $i=1, 2, \ldots, r$, определена вероятность p^i того, что в очередной момент времени на вход автомата поступит сигнал штраф. Таким образом, среда задается вектором следующего вида:

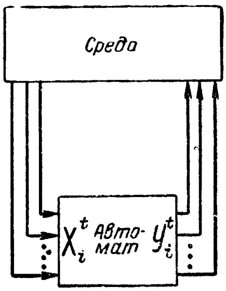


Рис. 22.

$$S=(p^1, p^2, \ldots, p^2), \qquad (37)$$

число компонент которого равно числу различных выходных сигналов автомата.

Если с течением времени значения p^i остаются неизменными и автомат не меняет своих характеристик, то система среда — автомат носит стационарный характер.

Мы рассмотрим в дальнейшем стационарную среду, в которой действуют автоматы следующих трех типов: обычный конечный автомат с памятью, стохастический автомат и автомат с переменной структурой. После рассмотрения поведения таких автоматов в стационарной среде мы рассмотрим поведение автоматов в нестационарной среде специального вида.

Сначала рассмотрим случай, когда в стационарной среде находится конечный автомат с памятью. Работа такого автомата определяется табл. 4. Мы рассмотрим частный случай такого автомата, у которого выходной сигнал в момент времени t однозначно определяется внутренним состоянием автомата в момент времени t. Другими словами, множество внутренних состояний автомата разбито на r непересекающихся подмножеств и элементам каждого подмножества однозначно сопоставлен некоторый выходной сигнал автомата. Будем задавать поведение такого автомата в случайной среде S с помощью двух матриц A_0 и A_1 размером $n\times n$, где n — число внутренних состояний автомата, и вектором — строкой B, число компонент которого также равно n. Элементы $a_{ij}(0)=0,1$ матрицы A_0 определяют переходы из одного состояния автомата в другое, если на вход автомата подан сигнал ноль (нештраф), а элементы матрицы $A_1 a_{ij}(1)$ определяют соответствующие переходы автомата из одних состояний в другие при подаче на вход автомата сигнала единица (штраф). При этом, конечно, должны выполняться условия

$$\sum_{i=1}^{n} a_{ij}(0) = 1 \quad \text{и}$$

$$\sum_{i=1}^{n} a_{ij}(1) = 1 \text{ для всех } i = 1, 2, \ldots, n.$$

Вектор B определяет выход автомата.

В качестве примера рассмотрим автомат, который задается с помощью следующих матриц A_0, A_1 и B:

$$A_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \; A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \; B = (0, 1).$$

На рис. 23 показана диаграмма смены состояний для этого автомата. Выходной сигнал автомата, как это следует из вида вектора B, совпадает с номером внутренне-

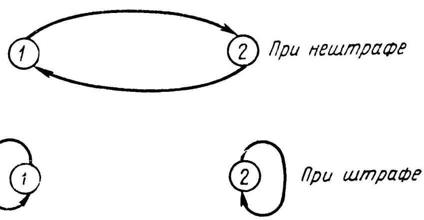


Рис. 23.

го состояния, в котором этот автомат находится в данный момент времени. Среда задается вектором с двумя компонентами по числу выходных сигналов автомата $S = (p^0, p^1)$.

Переходы автомата из одного внутреннего состояния в другое носят вероятностный характер, так как они определяются случайными воздействиями, поступающими на вход автомата. Однако вход автомата в некотором смысле определяется значением его выхода в предшествующий момент времени. Это позволяет описать систему автомат — среда как некоторую единую вероятностную систему.

Полная вероятность перехода автомата из состояния i в состояние j определяется следующим образом:

$$p_{ij}=\pi_0 a_{ij}(0)+\pi_1 a_{ij}(1); \qquad (38)$$

здесь π_0 соответствует вероятности появления на входе автомата сигнала нештраф, а π_1 — вероятности появления на входе автомата сигнала штраф. При этом $0\leqslant p_{ji}\leqslant 1$ и

$$\sum_i p_{ij}=\pi_0\sum_i a_{ij}(0)+\pi_1\sum_i a_{ij}(1)=\pi_0+\pi_1=1$$

для всех i.

Рассмотрим теперь матрицу

$$P=\begin{bmatrix} p_{11}, & p_{12}, & \ldots, & p_{1m} \\ p_{21}, & p_{22}, & \ldots, & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1}, & p_{m2}, & \ldots, & p_{mn} \end{bmatrix}.$$

Эта матрица заменяет собой матрицы A_i, B и вектор S и определяет некоторую дискретную марковскую цепь, характеризующую поведение системы автомат — среда. Будем предполагать, что эта дискретная марковская цепь имеет финальные вероятности. Через g_j обозначим финальную вероятность того, что автомат находится в состоянии с номером j. Через u_j обозначим сумму финальных вероятностей для всех тех состояний автомата, для которых выход автомата равен Y_j.

Определим теперь математическое ожидание штрафа для автомата, определяемого матрицей P в среде S:

$$M(P, S)=\sum_{j=1}^{r} p^j u_j. \qquad (39)$$

Если ввести обозначения

$$M_{\text{макс}}=\text{макс}\{p^1,\ p^2,\ \ldots,\ p^r\};$$

$$M_{\text{мин}}=\text{мин}\{p^1,\ p^2,\ \ldots,\ p^r\};$$

$$M_0=\frac{p^1+p^2+\ldots+p^r}{r},$$

то можно определить поведение автомата в среде S как целесообразное, если для этого автомата $M(P,\ S)<M_0$. В случае $M(P,\ S)=M_0$ автомат ведет себя как полностью необученный автомат, а при $M(P,\ S)>M_0$ автомат стремится к увеличению математического ожидания штрафа, который он получает от среды, т. е. ведет себя нецелесообразно. В случае равенства всех p^i $M_{\text{макс}}=M_0=M_{\text{мин}}$ и целесообразное поведение автомата совпадает с поведением необученного автомата. Во всех остальных случаях математическое ожидание штрафа для автомата с целесообразным поведением должно быть меньше математического ожидания штрафа для необученного автомата.

Для того чтобы стационарная среда и автомат реализовали при своем взаимодействии марковскую цепь, обладающую финальными вероятностями, достаточно выполнения следующих трех условий:

1. Все вероятности p^i удовлетворяют условию $0<p^i<1$. 2. Для любой пары состояний автомата можно указать такую последовательность двоичных сигналов, что если эту последовательность подать на вход автомата, то он перейдет из одного состояния в другое. 3. Множество состояний автомата нельзя разбить на такие непересекающиеся подмножества $Z_1,\ Z_2,\ \ldots,\ Z_q$, что переход из Z_i в Z_{i+1} совершается независимо от того, что подано на вход автомата: штраф или нештраф. Аналогичное требование верно и для перехода из Z_q в Z_1. Все среды и автоматы, рассматриваемые нами, удовлетворяют этим условиям и поэтому соответствующие им цепи Маркова всегда имеют финальные вероятности.

Рассмотрим с точки зрения целесообразности поведения автомат, который мы определили выше матрицами A_0, A_1, B, функционирующий в среде $S=(p^0,\ p^1)$ (диаграмма смены внутренних состояний этого автомата дана на рис. 23).

Строим матрицу P для такого автомата, используя соотношения вида (38):

$$p_{11}=(1-p^0)\cdot 0+p^0\cdot 1=p^0;$$

$$p_{12}=(1-p^0)\cdot 1+p^0\cdot 0=1-p^0;$$

$$p_{21}=(1-p^1)\cdot 1+p^1\cdot 0=1-p^1;$$

$$p_{22}=(1-p^1)\cdot 0+p^1\cdot 1=p_1$$

или

$$P=\begin{pmatrix} p^0 & 1-p^0 \\ 1-p^1 & p^1 \end{pmatrix}.$$

Для определения финальных вероятностей используем соотношение (32). Тогда получим:

$$\begin{cases} p_1=p_1 p^0+p_2(1-p^1); \\ p_2=p_1(1-p^0)+p_2 p^1, \end{cases}$$

$$p_1+p_2=1.$$

Отсюда

$$p_1=\frac{1-p^1}{(1-p^0)+(1-p^1)};$$

$$p_2=\frac{1-p^0}{(1-p^0)+(1-p^1)}.$$

На основании (39)

$$M(P, S)=p^0 p_1+p^1 p_2=\frac{p^0+p^1-2p^0p^1}{(1-p^0)+(1-p^1)}.$$

Если

$$\frac{p^0+p^1-2p^0p_1}{(1-p^0)+(1-p^1)}<\frac{p_0+p^1}{2},$$

то рассмотренный автомат обладает целесообразным поведением.

В качестве второго примера рассмотрим класс автоматов, которые были названы М. Л. Цетлиным автоматами с линейной тактикой. Автомат с линейной тактикой имеет $2n$ состояний и два выхода. При этом в случае состояний с номерами 1, 2, ..., n выход равен 0, а при состояниях $n+1$, $n+2$, ..., $2n$ он равен 1.

Матрицы переходов такого автомата имеют следующий вид:

$$A_0 = \begin{array}{c|cccccccccc|}
 & 1 & 2 & \cdots & n-1 & n & n+1 & n+2 & \cdots & 2n-1 & 2n \\
1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
n-1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
n & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
n+1 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
n+2 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
2n-1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
2n & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1
\end{array}$$

и

$$A_1 = \begin{array}{c|cccccccccc|}
 & 1 & 2 & \cdots & n-1 & n & n+1 & n+2 & \cdots & 2n-1 & 2n \\
1 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
n-1 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
n & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\
n+1 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
n+2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
2n-1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\
 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & 0
\end{array}.$$

Они иллюстрируются диаграммами переходов, показанными на рис. 24. Сплошные стрелки соответствуют пере-

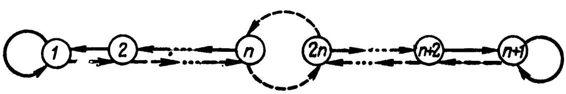


Рис. 24.

ходам при нештрафе, а пунктирные—переходам при штрафе. Пусть среда задается вектором $S=(p^0,\ p^1)$. Тогда

$$p_1=(1-p^0)\,p_1+(1-p^0)\,p_2;\quad p_{n+1}=(1-p^1)\,p_{n+1}+$$
$$+(1-p^1)\,p_{n+2};$$
$$p_2=p^0p_1+(1-p^0)\,p_3;\quad p_{n+2}=p^1p_{n+1}+(1-p^1)\,p_{n+3};$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$p_k=p^0p_{k-1}+(1-p^0)\,p_{k+1};\quad p_{n+k}=p^1p_{n+k-1}+$$
$$+(1-p^1)\,p_{n+k+1};$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$p_n=p^0p_{n-1}+p^1p_{2n};\quad p_{2n}=p^1-p_{2n-1}+p^0p_n;$$
$$p_1+p_2+\ldots+p_n=1,$$

отсюда

$$M(P,\ S)=$$
$$=\frac{\dfrac{1}{(p^0)^{n-1}}\,\dfrac{(p^0)^n-(1-p^0)^n}{p^0-(1-p^0)}+\dfrac{1}{(p^1)^{n-1}}\,\dfrac{(p^1)^n-(1-p^1)^n}{p^1-(1-p^1)}}{\dfrac{1}{(p^0)^n}\,\dfrac{(p^0)^n-(1-p^0)^n}{p^0-(1-p^0)}+\dfrac{1}{(p^1)^n}\,\dfrac{(p^1)^n-(1-p^1)^n}{p^1-(1-p^1)}}. \qquad (40)$$

Величина M убывает при увеличении n, т. е. с ростом памяти автомата, и $\lim\limits_{n\to\infty} M=M_{\text{мин}}$. Для справедливости последнего утверждения необходимо, чтобы было выполнено условие $\min\limits_{i}\{p^i\}<\frac{1}{2}$. Только для стационарных сред, удовлетворяющих этому условию, автомат с линейной тактикой будет обладать целесообразным поведением.

Рассмотрим теперь автомат диаграмма переходов для которого показана на рис. 25. Этот автомат имеет $2l+2m$ состояний и два выхода 0 и 1. Среда задается вектором $S=(p^0,\ p^1)$. Незакрашенные состояния на рис. 25 соответствуют тем состояниям автомата, которые определяют на выходе автомата сигнал 0, а закрашенные состояния — тем состояниям автомата, которые определяют на его выходе сигнал 1. Переходы, реализуемые после подачи на вход автомата нештрафа, показаны сплошными стрелками, а переходы, реализуемые после подачи сигнала штрафа, показаны пунктирными стрелками. Автомат такого типа был рассмотрен В. А. Пономаревым, который показал, что подобный автомат ведет себя оптимальным образом при $n \to \infty$ без учета ограничения на значения p^i среды, в которой этот

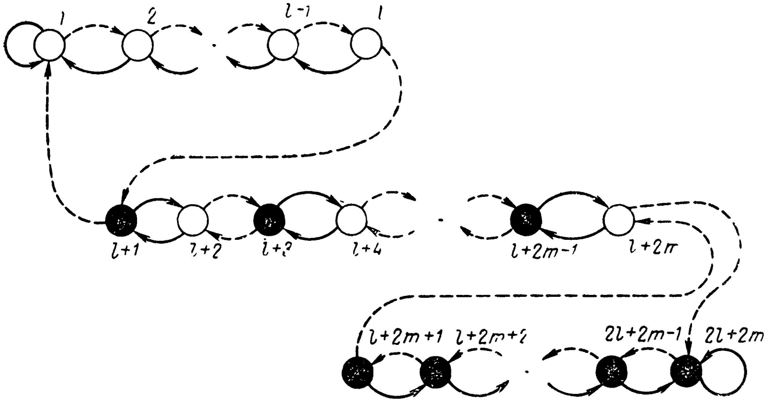


Рис. 25.

автомат находится. Финальные вероятности выходов 0 и 1 для него соответственно равны:

$$u_0 = \frac{\dfrac{v_1}{v_0}\,\dfrac{\lambda^m - 1}{\lambda - 1} + \dfrac{p^1}{p^0}\displaystyle\sum_{k=1}^{l}\frac{\lambda_1^k - 1}{\lambda_1 - 1}}{\left(1 + \dfrac{v_1}{v_0}\right)\dfrac{\lambda^m - 1}{\lambda - 1} + \dfrac{p^1}{p^0}\displaystyle\sum_{k=1}^{l}\frac{\lambda_1^k - 1}{\lambda_1 - 1} + \lambda^m\sum_{k=1}^{l}\frac{\lambda_2^k - 1}{\lambda_2 - 1}}$$

и

$$u_1 = \frac{\dfrac{\lambda^m - 1}{\lambda - 1} + \lambda^m\displaystyle\sum_{k=1}^{l}\frac{\lambda_2^k - 1}{\lambda_2 - 1}}{\left(1 + \dfrac{v_1}{v_0}\right)\dfrac{\lambda^m - 1}{\lambda - 1} - \dfrac{p^1}{p^0}\displaystyle\sum_{k=1}^{l}\frac{\lambda_1^k - 1}{\lambda_1 - 1} + \lambda^m\sum_{k=1}^{l}\frac{\lambda_2^k - 1}{\lambda_2 - 1}},$$

где $v_0 = 1 - p^0$; $v_1 = 1 - p^1$; $\lambda_1 = v_0/p^0$; $\lambda_2 = v_1/p_1$; $\lambda = \lambda_2/\lambda_1$.

Математическое ожидание штрафа

$$M(P,\ S)=p^0u_0+p^1u_1$$

и

$$\lim_{n\to\infty} M=M_{\text{мин}}.$$

Рассмотрим теперь задачу поведения в случайной среде для автоматов стохастического типа. Такие автоматы задаются аналогично обычным автоматам с конечной памятью в виде трех матриц A_0, A_1 и B. Отличие их от автоматов предыдущего типа состоит в том, что элементы матриц A_0 и A_1 $a_{ij}(0)$ и $a_{ij}(1)$ могут теперь принимать любые значения от нуля до единицы. При этом для $i=1,\ 2,\ldots,\ n$ выполнены равенства:

$$\sum_{j=1}^{n} a_{ij}(0)=1 \text{ и } \sum_{j=1}^{n} a_{ij}(1)=1.$$

Для стохастических автоматов соотношения (38) и (39) сохраняют свой смысл.

Рассмотрим два конкретных примера подобных автоматов. Первый автомат задается следующими матрицами:

$$A_0=\begin{pmatrix} 0 & 1 \\ 0{,}5 & 0{,}5 \end{pmatrix};\ A_1=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ B=(1,\ 0).$$

Среда имеет вид $S=(p^0,\ p^1)$. Диаграмма переходов автомата показана на рис. 26. Используя (38) и (39), получим:

$$p_{11}=(1-p^1)\cdot 0+p^1\cdot 1=p^1;$$

$$p_{12}=(1-p^1)\cdot 1+p^1\cdot 0=1-p^1;$$

$$p_{21}=(1-p^0)\cdot 0{,}5+p^0\cdot 0=\frac{1-p^0}{2};$$

$$p_{22}=(1-p^0)\cdot 0{,}5+p^0\cdot 1=\frac{1+p^0}{2}$$

и

$$P=\begin{pmatrix} p^1 & 1-p^1 \\ \frac{1-p^0}{2} & \frac{1+p^0}{2} \end{pmatrix},$$

отсюда финальные вероятности

$$\begin{cases} p_1 = p_1 p^1 + p_2 \dfrac{1-p_0}{2}; \\ p_2 = p_1(1-p^1) + p_2 \dfrac{1+p^0}{2}; \\ p_1 + p_2 = 1; \end{cases}$$

$$p_1 = \frac{1}{\dfrac{2(1-p^1)}{1-p^0}+1}; \quad p_2 = \frac{1}{\dfrac{1-p^0}{2(1-p^1)}+1}.$$

Математическое ожидание штрафа

$$M(P, S) = p^1 p_1 + p^0 p_2 = \frac{p^1}{\dfrac{2(1-p^1)}{1-p^0}+1} + \frac{p^0}{\dfrac{1-p^0}{2(1-p^1)}+1}.$$

Если, например, $p^0=0{,}25$, а $p^1=0{,}5$, то $M(P, S) = {}^5/_{14}$, а $M_0 = {}^3/_8$ и поведение рассматриваемого автомата в среде (0,25, 0,5) является целесообразным.

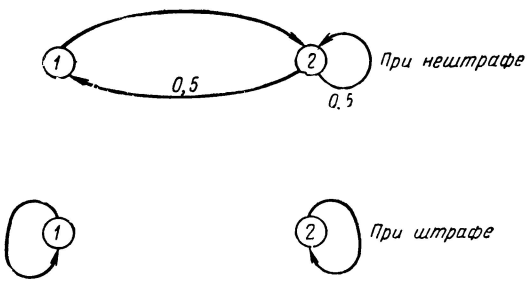


Рис. 26.

На рис. 27 показана диаграмма переходов для автомата, предложенного В. Ю. Крыловым. Этот автомат проще автомата Пономарева и имеет только $2n$ состояний. В состояниях 1, 2, ..., n он выдает на выходе сигнал y_1, а при состояниях $n+1$, $n+2$, ..., $2n$ его выходной сигнал есть y_2. Пунктирные стрелки означают смену состояний с вероятностью 0,5. Математическое ожидание штрафа для этого автомата

$$M(A_0, A_1, S) = \frac{\frac{1}{\lambda_1^{n-1}} \frac{\lambda_1^n - 1}{\lambda_1 - 1} + \frac{1}{\lambda_2^{n-1}} \frac{\lambda_2^n - 1}{\lambda_2 - 1}}{\frac{1}{p^1 \lambda_1^{n-1}} \frac{\lambda_1^n - 1}{\lambda_1 - 1} + \frac{1}{p^2 \lambda_2^{n-1}} \frac{\lambda_2^n - 1}{\lambda_2 - 1}},$$

где $\lambda_i = \frac{p^i}{2 - p^i}$, $i = 1, 2$.

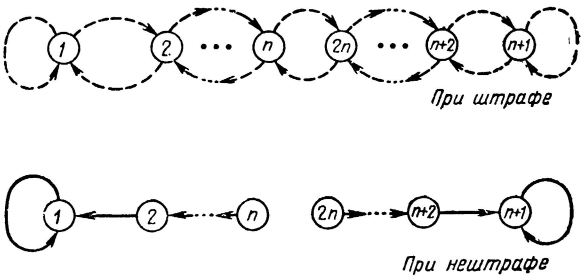


Рис. 27.

В. Ю. Крыловым показано, что при $n \to \infty$ получим $M(A_0, A_1, S) \to$ мин (p^1, p^2), т. е. такой автомат ведет себя асимптотически оптимально в любой стационарной среде.

Рассмотрим теперь, как ведет себя в стационарной случайной среде стохастический автомат с переменной структурой. Стохастический автомат с переменной структурой характеризуется тем, что в процессе его обучения в случайной среде происходит изменение значений переходных вероятностей $a_{ij}(q)$ в матрицах $A_q (q = 0,1)$. Это изменение происходит следующим образом. Если при переходе автомата из состояния с номером i в состояние с номером j под воздействием входа $x_q (x_q = 0,1)$ следует нештраф, то вероятность $a_{ij}(q)$ увеличивается. В противном случае она уменьшается. Все остальные вероятности в строке матрицы A_q пересчитываются так, чтобы сумма элементов строки оставалась бы равной нулю. Отметим, что автомат с переменной структурой описанного типа может не соответствовать цепи Маркова, обладающей финальными вероятностями состояний.

Однако для некоторых случаев автоматов с переменной структурой понятие финальной вероятности состояния имеет вполне определенный смысл. Проиллюстрируем это на следующем примере. Дан автомат

$$A_0 = \begin{pmatrix} p_{11}^{2} & 1 - p_{11}^{0} \\ 1 - p_{22}^{0} & p_{22}^{0} \end{pmatrix}; \quad A_1 = \begin{pmatrix} p_{11}^{1} & 1 - p_{11}^{1} \\ 1 - p_{22}^{1} & p_{22}^{1} \end{pmatrix}$$

и среда

$$S = (p^0,\ p^1).$$

Условимся, что если автомат перешел из состояния i в состояние $j\,(i,\ j = 0,\ 1)$ под влиянием входа $x_q\,(x_q = 0,\ 1)$, и был оштрафован, то строка матрицы $(p_{ij}^{q};\ 1 - p_{ij})$ заменяется на строку $(\alpha p_{ij}^{q};\ 1 - \alpha p_{ij}^{q})$. Если же при таком переходе штрафа не было, то эта же строка в матрице A_q заменяется на строку $[(1 - \alpha) + \alpha p_{ij}^{q};\ \alpha(1 - p_{ij}^{q})]$, где $0 < \alpha < 1$. Через $Q_i(t)$ обозначим вероятность состояния с номером i в момент времени t. Через $p_{ij}^{q}(t)$ обозначим значение соответствующих переходных вероятностей в момент времени t, а через $\overline{p}_{ij}^{q}(t+1)$ — математическое ожидание переходных вероятностей в момент времени $t+1$, тогда

$$\overline{p}_{11}^{0}(t+1) = p_{11}^{0}(t) - Q_1(t)(1 - p^0)(1 - \alpha) \times$$
$$\times [p_{11}^{0}(t)(p^0 + p^1) - p^1];$$

$$\overline{p}_{11}^{1}(t+1) = p_{11}^{1}(t) - Q_1(t)(1 - \alpha)p^1 \times$$
$$\times [p_{11}^{1}(t)(p^0 + p^1) - p^1];$$

$$\overline{p}_{22}^{0}(t+1) = p_{22}^{0}(t) - Q_2(t)(1 - p^1)(1 - \alpha) \times$$
$$\times [p_{22}^{0}(t)(p^0 + p^1) - p^0];$$

$$\overline{p}_{22}^{1}(t+1) = p_{22}^{1}(t) - Q_2(t)(1 - \alpha)p^1 \times$$
$$\times [p_{22}^{1}(t)(p^0 + p^1) - p^0].$$

Будем называть переходную вероятность p^q_{ij} стационарной, если p^q_{ij} удовлетворяет одному из следующих двух условий:

1) $\bar{p}_{ij}(t+1)=p^q_{ij}(t)=p^q_{ij}$;

если $p^q_{ij}(t)>p^q_{ij}$, то $\bar{p}^q_{ij}(t+1)-p^q_{ij}(t)<0$;

если $p^q_{ij}(t)<p^q_{ij}$, то $\bar{p}^q_{ij}(t+1)-p^q_{ij}(t)>0$.

2) $\lim\limits_{t\to\infty}[p^q_{ij}(t+1)-p^q_{ij}(t)]=0$ и знак разности, стоящей под знаком предела, не зависит от значения $p^q_{ij}(t)$.

Пусть для нашего примера выполняется первое требование стационарности. Тогда $\bar{p}^q_{ij}(t+1)=p^q_{ij}(t)$. Из этого требования вытекает, что

$$p^0_{11}=p^1_{11}=\frac{p^1}{p^0+p^1};\ p^0_{22}=p^1_{22}=\frac{p^0}{p^0+p^1}.$$

Но тогда, независимо от начальных значений переходных вероятностей, в матрицах A_q математическое ожидание штрафа имеет следующий вид:

$$M(A_0, A_1, S)=\frac{2p^0p^1}{p^0+p^1}.$$

Это выражение совпадает со значением математического ожидания штрафа для автомата с линейной тактикой, определяемого соотношением (40) при $n=2$. Другими словами, рассмотренный автомат с переменной структурой ведет себя аналогично автомату с линейной тактикой с двумя состояниями.

В качестве второго примера[1] можно рассмотреть автомат, определяемый теми же матрицами A_q и вектором S, но у которого изменение соответствующей пере-

[1] Оба примера автомата с переменной структурой взяты из работы В. И. Варшавского и И. П. Воронцовой (см. список литературы).

ходной вероятности (увеличение или уменьшение ее) происходит на величину, равную $\alpha p^q_{ij}(1 - p^q_{ij})$. Для такого автомата

$$
\begin{aligned}
\bar{p}^0_{11}(t+1) &= p^0_{11}(t) + Q_1(t)(1 - p^0)\alpha p^0_{11}(t)(1 - p^0_{11}(t)) \times \\
&\times [2(1 - p^0 - p^1)p^0_{11}(t) - (1 - 2p^1)]; \\
\bar{p}^1_{11}(t+1) &= p^1_{11}(t) + Q_1(t)p^0\alpha p^1_{11}(t)(1 - p^1_{11}(t)) \times \\
&\times [2(1 - p^0 - p^1)p^1_{11}(t) - (1 - 2p^1)]; \qquad (41) \\
\bar{p}^0_{22}(t+1) &= p^0_{22}(t) + Q_2(t)(1 - p^1)\alpha p^0_{22}(1 - p^0_{22}(t)) \times \\
&\times [2(1 - p^0 - p^1)p^0_{22}(t) - (1 - 2p^0)]; \\
\bar{p}^1_{22}(t+1) &= p^1_{22}(t) + Q_2(t)p^1_\alpha p^1_{22}(t)(1 - p^1_{22}(t)) \times \\
&\times [2(1 - p^0 - p^1)p^1_{22}(t) - (1 - 2p^0)].
\end{aligned}
$$

Если $p^0 < 0{,}5 < p^1$, то при $t \to \infty$ имеет место следующее:

$$p^0_{11} \to 1;\quad p^0_{22} \to 0;\quad p^1_{11} \to 1;\quad p^1_{22} \to 0;\quad Q_1(t) \to 1.$$

Отсюда следует, что $M(A_0, A_1, S) = p^0$ и рассматриваемый автомат имеет целесообразное поведение. Пусть теперь $p^0 + p^1 > 1$ и $p^0,\ p^1 > \frac{1}{2}$, тогда из условия стационарности переходных вероятностей вытекает, что должны удовлетворяться равенства:

$$p^0_{11} = p^1_{11} = \frac{1 - 2p^1}{2(1 - p^0 - p^1)};$$

$$p^0_{22} = p^1_{22} = \frac{1 - 2p^0}{2(1 - p^0 - p^1)}.$$

На рис. 28,*а* показан характер зависимости

$$p_{11}(t+1)-p_{11}(t)=f(p_{11});$$

$$p_{22}(t+1)-p_{22}(t)=f(p_{22}).$$

Как видно из этого рисунка, стационарность переходных вероятностей при выполнении условий, описанных выше, действительно имеет место, так как, если происходит отклонение $p(t)$ от стационарной точки и это отклонение приводит к увеличению математического ожидания штрафа, знак разности отрицателен и на следующем шаге происходит такое изменение переходных вероятностей, которое обеспечивает уменьшение математического ожидания штрафа для автомата. Если же отклонение от стационарной точки происходит в сторону уменьшения математического ожидания штрафа, то знак разности положителен и происходит дальнейшее изменение переходных вероятностей, обеспечивающее дальнейшее уменьшение математического ожидания штрафа. Для рассматриваемого случая

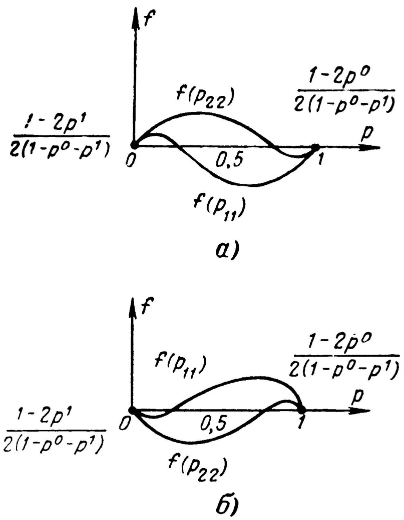


Рис. 28.

$$M(A_0, A_1, S)=\frac{p^0+p^1-4p^0p^1}{2(1-p^0-p^1)}.$$

Это значение совпадает со значением математического ожидания штрафа для автомата с линейной тактикой и среды, удовлетворяющей вышеприведенным условиям, но при условии бесконечной памяти автомата с линейной тактикой.

Рассмотрим, наконец, случай, когда p^0, $p^1<\frac{1}{2}$.

В этом случае

$$p_{11}^{0}=p_{11}^{1}=\frac{1-2p^{1}}{2(1-p^{0}-p^{1})};$$

$$p_{22}^{0}=p_{22}^{1}=\frac{1-2p^{0}}{2(1-p^{0}-p^{1})}.$$

Поведение разностей $f(p_{11})$ и $f(p_{22})$ показано на рис. 28,*б* в предположении, что $p^1>p^0$. Как видно из этого рисунка, точки $1-2p^1/2(1-p^0-p^1)$ и $1-2p^0/2(1-p^0-p^1)$ являются неустойчивыми. Целесообразность поведения автомата зависит от начальных значений переходных вероятностей в матрицах A_q.

Для целесообразности поведения необходимо, чтобы эти вероятности были заключены между точками неустойчивого равновесия, ибо в этом случае при $t\to\infty$ $p_{11}^{0}\to 1$, $p_{11}^{1}\to 1$, $p_{22}^{0}\to 0$, $p_{22}^{1}\to 0$, $Q_1(t)\to 1$ и $M(A_0,\ A,\ S)=\underset{i}{\text{мин}}(p^i)$.

Интересно отметить, что для автоматов рассмотренного типа число состояний не играет роли и автомат с глубиной памяти, равной двум (т. е. имеющий два состояния), эквивалентен в смысле целесообразности поведения в стационарной случайной среде автомату с бесконечно большим числом состояний.

Более сложным является исследование поведения автоматов с постоянной и переменной структурой в случайных средах, которые не являются стационарными. В работах М. Л. Цетлина и В. И. Варшавского исследовалось поведение автомата в нестационарной среде следующего вида. Рассмотрим вектор $S=(S_1,\ S_2,\ \ldots,\ S_l)$, где S_i есть стационарная среда, и матрицу

$$\Delta=\begin{pmatrix}\delta_{11}\delta_{12}\ \ldots\ \delta_{1l}\\ \delta_{21}\delta_{22}\ \ldots\ \delta_{2l}\\ \ldots\ldots\ldots\\ \delta_{l_1}\delta_{l_2}\ \ldots\ \delta_{ll}\end{pmatrix},$$

элементы которой определяют вероятности перехода от одной стационарной среды к другой стационарной среде. Среды такого вида принято называть **составными** или переключающимися средами. Для простоты в дальнейшем мы ограничимся рассмотрением лишь сред вида $S=(S_1, S_2)$, для которых матрица Δ имеет вид:

$$\Delta=\begin{pmatrix}1-\delta & \delta \\ \delta & 1-\delta\end{pmatrix}. \tag{42}$$

Если рассматривать систему автомат — среда, показанную на рис. 22, как единую систему, то можно ввести понятие состояния этой системы. Состояние системы автомат — среда характеризуется вектором (i, j), где i — номер состояния, в котором в данный момент времени находится автомат, а j — номер стационарной среды, с которой в этот же момент времени совпадает переключающаяся среда. Если число различных внутренних состояний автомата равно m, а вектор S имеет l компонент, то число состояний системы автомат — среда равно ml. Переход из одного состояния системы в другое характеризуется матрицей следующего вида:

$$\Xi=\begin{pmatrix}\xi_{11} & \xi_{12} & \dots & \xi_{1\,(ml)} \\ \xi_{21} & \xi_{22} & \dots & \xi_{2\,(ml)} \\ \dots & \dots & \dots & \dots \\ \xi_{(ml)\,1} & \xi_{(ml)\,2} & \dots & \xi_{(ml)(ml)}\end{pmatrix}. \tag{43}$$

Каждый элемент этой матрицы сам представляется некоторой матрицей следующего вида:

$$\xi_{ij}=\begin{pmatrix}[a_{ij}(1)\,p^1_y+a_{ij}(0)(1-p^1_y)]\;[(1-\delta)\,[a_{ij}(1)\,p^1_y+ \\ +a_{ij}(0)\,(1-p^1_y)]\,\delta\,[a_{ij}(1)\,p^2_y+a_{ij}(0)\,(1- \\ -p^2_y)]\,\delta\,[a_{ij}(1)\,p^2_y+a_{ij}(0)\,(1-p^2_y)]\,(1-\delta);\end{pmatrix} \tag{44}$$

здесь $p^1{}_y$, $p^2{}_y$ означают вероятности штрафа для сред S_1, S_2 при условии, что автомат выдал на выходе воздействие y.

Обозначим через r^j_i финальные вероятности нахождения автомата в состоянии i в среде с номером j. Отметим, что система автомат — среда, поведение которой характеризуется матрицей (43), уже не может рассматриваться как стационарная дискретная цепь Маркова в обычном смысле этого понятия. Поэтому для такой системы неприменимы наши достаточные условия существования финальных вероятностей. Если предположить, что такие вероятности существуют, то эти финальные вероятности не зависят от начального состояния системы автомат — среда. Пусть автомат выдает на выходе два сигнала $y_1=0$, если он находится в состояниях с номерами 1, 2, ..., n, и $y_2=1$, если его внутренние состояния имеют номера $n+1$, $n+2$, ..., $2n$. Тогда математическое ожидание штрафа для такого автомата в переключающейся среде $S=(S_1, S_2)$ имеет следующий вид:

$$M(A_0, A_1, S)=p_0^1\sum_{i=1}^{n} r_i^1+p_1^1\sum_{i=n+1}^{2n} r_i^1+$$

$$+p_0^2\sum_{i=1}^{n} r_i^2+p_1^2\sum_{i=n+1}^{2n} r_i^2.$$

Вернемся теперь к автомату с линейной тактикой, диаграмма переходов которого показана на рис. 24. Пусть $p_0^1=p_0^2=p$ и $p_1^1=p_2^1=1-p$. Тогда из (45) получаем:

$$M(A_0, A_1\ S)=\frac{1}{2}-\frac{(\lambda-1)^2}{2}\times$$

$$\times\frac{\operatorname{ch} ny-1}{\frac{2n\delta}{1-2\delta}(\lambda+1)^2\operatorname{ch} ny+(\lambda-1)^2\operatorname{ch} ny+(\lambda-1)^2\operatorname{cth}\frac{y}{2}\operatorname{sh} ny},$$

где $\lambda=\frac{p}{p-1}$ и $\operatorname{ch} y=\frac{(1+\lambda)^2}{2\lambda}\,\frac{(1-\delta)}{1-2\delta}-1.$

На рис. 29,*а* показано изменение M при $p=0{,}33$ и различных значениях δ. На рис. 29,*б* при $\delta=0{,}01$ показано изменение M при различных значениях p. Приведенные графики показывают, что поведение автомата с линейной тактикой в переключающихся средах принципиально отлично от поведения таких же автоматов в стационарной среде. Это отличие состоит в том, что для случая стационарной среды при неограниченном увеличении памяти автомата его поведение становится асимптотически оптимальным. Для переключающихся сред это не так. При $n \to \infty$ автомат ведет себя нецелесообразно, что объясняется плохой гибкостью его приспособлений к новым условиям среды (большой «инерцией» перестройки). Наилучшее поведение автомата достигается при некоторой определенной глубине памяти n_0, являющейся функцией от δ и p. На рис. 29 этой глубине памяти соответствуют точки минимума на графиках.

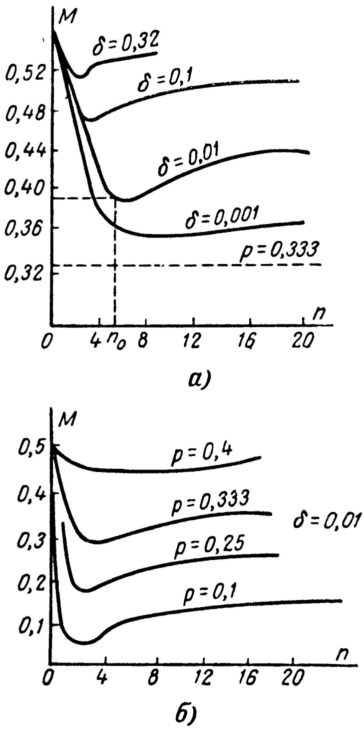


Рис. 29.

При $\delta \to 0$ $n_0 \to \infty$ и $M \to M_{\text{мин}}$. Введем параметр $d = 0{,}5 - M$, с помощью которого можно охарактеризовать различимость сред S_1 и S_2 для автомата с линейной тактикой. В работах М. Л. Цетлина приведены значения d и n_0 для некоторых значений δ и p. Эти значения показаны в нижеследующей таблице.

Эта таблица может использоваться при конструировании автоматов, обладающих целесообразным поведением в переключающихся средах.

Рассмотренные нами автоматы не учитывали свой опыт функционирования в переключающейся среде. Если при выборе выходного сигнала автомата использовать

p \ δ	0,001		0,01		0,032		0,1		0,32		0,45	
	n_0	d	n_0	d	n_0	d	n_0	d	n_0	d	n_0	d
0,1	3	0,396	2	0,372	2	0,336	1	0,256	1	0,115	1	0,032
0,2	5	0,294	3	0,266	2	0,223	2	0,157	1	0,065	1	0,018
0,25	6	0,224	4	0,212	3	0,172	2	0,116	1	0,045	1	0,012
0,33	8	0,158	5	0,125	3	0,091	2	0,055	1	0,020	1	0,006
0,4	11	0,089	6	0,056	4	0,037	2	0,020	1	0,007	1	0,002
0,45	16	0,036	7	0,017	4	0,010	4	0,005	1	0,002	1	0,001

некоторые накапливаемые в процессе работы автомата знания о функционировании среды, то поведение такого автомата станет более целесообразным, чем для автомата, не учитывающего свой опыт работы. А. В. Добровидовым и Р. Л. Стратоновичем была рассмотрена общая постановка проблемы нахождения оптимального автомата в переключающейся среде, который при своих поступках учитывает апостериорные (т. е. уточненные на основе опыта) вероятности штрафов в процессе работы автомата. Результаты их работы изложены в работе, которая указана в списке литературы в конце книги. Мы же рассмотрим проблему приспособления автомата к функционированию в переключающейся среде на примере автомата с переменной структурой.

Рассмотрим функционирование в переключающейся среде, определяемой матрицей (43), автомата с переменной структурой, для которого изменение переходных вероятностей в матрицах A_q $(q=0,1)$ происходит по следующему правилу. Если при переходе из состояния с номером i в состояние с номером j под воздействием входа x_q $(q=0,1)$ автомат был оштрафован, то переходная вероятность p_{ij}^q умножается на $0<\alpha<1$ и соответствующая строка в матрице A_q нормируется. Если штрафа не было, то все элементы i-й строки матрицы, кроме p_{ij}^q, умножаются на α и строка нормируется.

Для случая $p=0,25$ В. И. Варшавским и И. П. Воронцовой экспериментально путем моделирования на ЭВМ была получена зависимость математического ожидания штрафа в зависимости от δ для автомата с двумя устойчивыми состояниями. Эта зависимость показана на рис. 30. Пунктирной линией на этом рисунке показана аналогичная зависимость для автомата с линейной тактикой с двумя устойчивыми состояниями. Как следует из этого графика, при частоте переключения среды S, большей 0,5, автомат с линейной тактикой и двумя устойчивыми состояниями ведет себя нецелесообразно, а автомат с переменной структурой для всех δ, достаточно отличающихся от 0 и 1, ведет себя целесообразно.

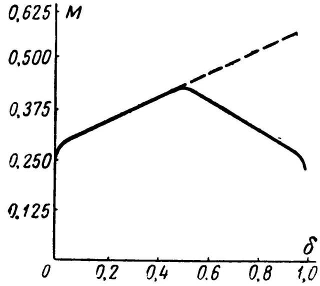


Рис. 30.

При моделировании на ЭВМ автомата с переменной структурой, имеющего четыре состояния и задаваемого в начале эксперимента матрицами A_q ($q=0,1$) вида

$$A_q = \begin{pmatrix} 0{,}25 & 0{,}25 & 0{,}25 & 0{,}25 \\ 0{,}25 & 0{,}25 & 0{,}25 & 0{,}25 \\ 0{,}25 & 0{,}25 & 0{,}25 & 0{,}25 \\ 0{,}25 & 0{,}25 & 0{,}25 & 0{,}25 \end{pmatrix},$$

уменьшение математического ожидания штрафа с течением времени имело вид, показанный на рис. 31,*а*.

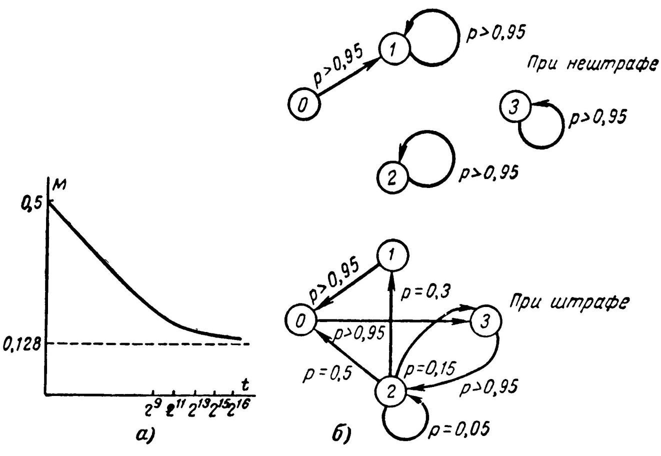


Рис. 31.

Время t показано в логарифмическом масштабе. На рис. 31,*б* показана диаграмма смены состояний автомата в конце эксперимента. На этой диаграмме не указаны переходы, реализация которых осуществляется с вероятностью, меньшей чем 0,01. Пунктирная прямая на рис. 31,*а* соответствует значению математического ожидания штрафа для автомата с линейной тактикой, имеющего бесконечное число внутренних состояний (т. е. обладающего асимптотически оптимальным поведением). Таким образом, автоматы с переменной струк-

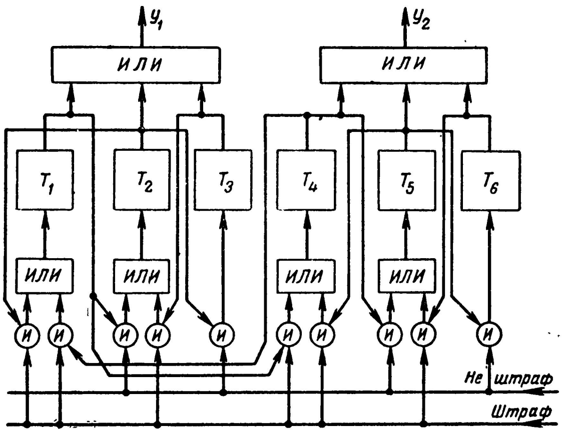


Рис. 32.

турой и конечным числом состояний в процессе «обучения» с ростом t стремятся к поведению, совпадающему с оптимальным.

Все описанные автоматы нетрудно реализовать схемно или в виде программ для вычислительных машин. На рис. 32 в качестве примера показана схемная реализация автомата с линейной тактикой, имеющего шесть состояний. В состояниях 1, 2, 3 автомат на выходе выдает сигнал y_1, а в остальных трех состояниях выходной сигнал автомата совпадает с y_2. Практически весь автомат реализован как реверсивный счетчик, выполненный на триггерах. Высокий потенциал на единичном выходе триггера соответствует тому, что автомат находится в состоянии, соответствующем этому триггеру. Для смены состояний в зависимости от сигналов штраф и нештраф, поступающим в автомат по входным шинам, предусмотрена несложная комбинационная схема, выполненная на клапанах (схема «И») и сборках (схема «ИЛИ»).

При большом числе состояний рассмотренная схема может оказаться неэкономичной. Однако усложнение схемы даже при весьма большом числе состояний происходит не слишком быстро, так как число триггеров в счетчике может быть выбрано как $[\log_2 m]+1$, где m — число состояний автомата с линейной тактикой, а объединение выходов триггеров для получения выходного сигнала автомата происходит так, как показано на рис. 30,*а* с помощью комбинационной схемы, сложность которой определяется числом выходов и состояний автомата[1]. Немногим сложнее реализация автомата, предложенного Крыловым, и автомата с переменной структурой. Правда, для этих автоматов необходимо еще иметь генератор случайных чисел, с помощью которого можно реализовать вероятностные переходы в матрицах A_q.

Построение схем автоматов стохастического типа вызывает большие трудности, чем построение детерминированных автоматов типа автомата с линейной тактикой. Одним из наиболее эффективных методов построения таких автоматов предложен В. Н. Ченцовым[2].

Рассмотрим вначале автономный стохастический автомат, т. е. такой стохастический автомат, у которого входное слово есть пустое слово (на вход автомата ничего не подается). Работа такого автомата полностью определяется заданием его начального состояния, матрицей смены состояний и строкой выходов, однозначно определяемых состоянием автомата в данный момент времени. Таким образом, автономный стохастический автомат задается с помощью матриц

$$C=\begin{bmatrix} q_{11} & q_{12} & \cdot & \cdot & \cdot & q_{1n} \\ q_{21} & q_{22} & \cdot & \cdot & \cdot & q_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ q_{n1} & q_{n2} & \cdot & \cdot & \cdot & q_{nn} \end{bmatrix}$$

и

$$B=(Y_{i_1}, Y_{i_2}, \ldots, Y_{i_n}).$$

---

[1] В квадратные скобки взята целая часть.

[2] Описанный метод не всегда дает минимальный по числу состояний автомат.

Как всегда, $0 \leqslant q_{ij} \leqslant 1$ и для всех $i = 1, 2, \ldots, n$ $\sum_{j=1}^{n} q_{ij} = 1$. Число компонент в векторе B равно числу внутренних состояний автомата.

Метод синтеза таких автоматов состоит в последовательном повторении следующих шагов:

1. В матрице C в каждой строке находится максимальный элемент. Этот элемент набран жирным шрифтом.

2. Среди всех отмеченных элементов находится наименьший.

3. Берется матрица, в которой единицы стоят на местах, соответствующих отмеченным элементам в исходной матрице, и нули на остальных местах, и умножается на минимальный из отмеченных элементов.

4. Из исходной матрицы вычитается матрица, полученная в п. 3.

5. Матрица-разность принимается за исходную матрицу. Если она ненулевая, то возвращение к п. 1, в противном случае — переход к п. 6.

6. Сумма матриц, полученных в п. 3, дает искомое разложение исходной матрицы в виде композиции простейших матриц.

Для иллюстрации этого алгоритма рассмотрим разложение для следующей матрицы:

$$C = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{8} & \frac{3}{8} & \frac{3}{8} \end{pmatrix}.$$

Применяя последовательно вышеописанный алгоритм, получим после первого шага

$$C = \begin{pmatrix} 0 & \mathbf{1} & 0 \\ \mathbf{\frac{1}{3}} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{8} & \frac{3}{8} & \mathbf{\frac{3}{8}} \end{pmatrix}.$$

В тех строках, в которых имеются одинаковые максимальные элементы, мы отмечаем жирным шрифтом любой из них.

Наименьший элемент среди отмеченных есть $\frac{1}{3}$. Вычитаем из исходной матрицу

$$\frac{1}{3}\cdot\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Получаем новую матрицу, в которой вновь отмечаем максимальные элементы:

$$\begin{pmatrix} 0 & \mathbf{\frac{2}{3}} & 0 \\ 0 & \mathbf{\frac{1}{3}} & \frac{1}{3} \\ \frac{2}{8} & \mathbf{\frac{3}{8}} & \frac{1}{24} \end{pmatrix}.$$

Вычитаемая матрица имеет вид:

$$\frac{1}{3}\cdot\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Новая матрица с отмеченными максимальными элементами:

$$\begin{pmatrix} 0 & \mathbf{\frac{1}{3}} & 0 \\ 0 & 0 & \mathbf{\frac{1}{3}} \\ \mathbf{\frac{2}{8}} & \frac{1}{12} & \mathbf{\frac{1}{24}} \end{pmatrix}.$$

Вычитаемая матрица

$$\frac{2}{8}\cdot\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

После очередного вычитания получается матрица

$$\begin{pmatrix} 0 & \mathbf{\frac{1}{12}} & 0 \\ 0 & 0 & \mathbf{\frac{1}{12}} \\ 0 & \mathbf{\frac{1}{24}} & \frac{1}{24} \end{pmatrix}.$$

Вычитаемая матрица

$$\frac{1}{24}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Наконец, после этого вычитания получаем матрицу

$$\begin{pmatrix} 0 & \frac{1}{24} & 0 \\ 0 & 0 & \frac{1}{24} \\ 0 & 0 & \frac{1}{24} \end{pmatrix},$$

из которой после вычитания матрицы

$$\frac{1}{24}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

получаем нулевую матрицу.

Для исходной матрицы нами найдено следующее разложение:

$$\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{8} & \frac{3}{8} & \frac{2}{8} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} +$$

$$+ \frac{1}{24}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} + \frac{1}{24}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Отождествим теперь синтезируемый автомат с конечным автоматом, имеющим три внутренних состояния. Определим входные слова $U_1, U_2, \ldots, U_s$, где число s совпадает с количеством членов в полученном разложении. Числовой коэффициент, стоящий перед матрицей из нулей и единиц в i-м члене разложения r_i, трактуется как вероятность появления на входе автомата входного слова U_i. Для рассмотренного примера мы имеем пять входных слов. Матрица переходов для этого автомата имеет вид:

$$\begin{array}{cc}
 & \begin{array}{ccc} z_1^{t+1} & z_2^{t+1} & z_3^{t+1} \end{array} \\
\begin{array}{c}
U_1^t z_1^t \\ U_1^t z_2^t \\ U_1^t z_3^t \\ U_2^t z_1^t \\ U_2^t z_2^t \\ U_2^t z_3^t \\ U_3^t z_1^t \\ U_3^t z_2^t \\ U_3^t z_3^t \\ U_4^t z_1^t \\ U_4^t z_2^t \\ U_4^t z_3^t \\ U_5^t z_1^t \\ U_5^t z_2^t \\ U_5^t z_3^t
\end{array} &
\begin{bmatrix}
0 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{bmatrix}
\end{array}$$

Схема автомата состоит из датчика квазиравномерно распределенных чисел, пересчетной схемы, число выходов которой равно числу слов U_i. Соответствующий выход U_i появляется с вероятностью r_i и детерминированного конечного автомата.

На рис. 33,*а* показана общая схема такого автомата. Вход на пересчетную схему необходим в том случае, если синтезируемый автомат не является автономным. В этом случае под влиянием активного входа (например, среды, посылающей на вход автомата сигналы 0 и 1) пересчетная схема меняет свою структуру и выдает на автомат слова U_i с вероятностями, соответствующими данному активному значению входа. На рис. 33,*б* показана диаграмма переходов для автомата, соответствующего разложению, полученному выше. Рядом с входными словами, обеспечивающими переход данного типа в скобках, указаны вероятности появления этих слов на выходе пересчетной схемы.

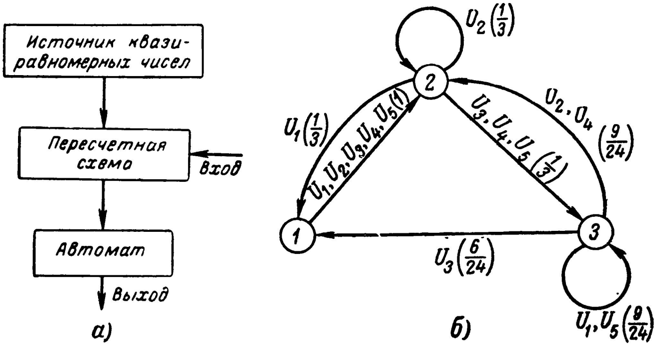


Рис. 33.

8. ИГРЫ АВТОМАТОВ

Развивая идеи предыдущего параграфа, мы перейдем теперь к рассмотрению поведения нескольких автоматов, одновременно действующих в среде S. При этом действия каждого автомата на среду влияют на входной сигнал (штраф или нештраф), поступающий на входы всех автоматов, находящихся в данной среде (рис. 34). Такое «сожительство» автоматов можно рассматривать как игру автоматов между собой. При этом выходные сигналы i-го автомата $y_j^i(t)$ будут трактоваться нами, как выбор этим автоматом в данной партии игры t своей

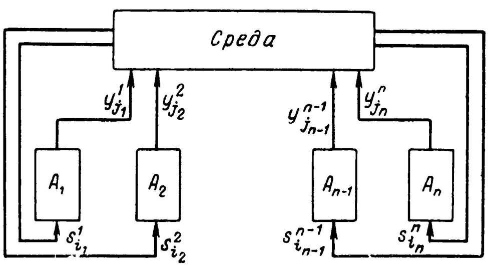


Рис. 34.

j-й стратегии. Следовательно, общее число различных выходов автомата совпадает с числом различных стратегий, имеющихся у него в игре. На вход каждого автомата поступает в момент времени $t+1$ сигнал $s^i=0,1$. Поступление сигнала 0 трактуется как сообщение о выигрыше, а сигнала 1 — как сообщение о проигрыше автомата. Сама величина выигрыша — проигрыша будет считаться в дальнешем фиксированной. Поэтому нас в дальнейшем будут интересовать не количественные значения выигрышей и проигрышей автомата в данной партии игры, а лишь сам факт выигрыша или проигрыша автомата в данной партии.

Вектор $S=(s^1, s^2, \ldots, s^n)$ представляет собой вектор, с помощью которого автоматы получают информацию о результатах своих действий в игре. Роль среды сводится к формированию вероятностного (или в частном случае детерминированного) ответа на выбор автоматами их полных стратегий. Другими словами, заданы значения $p(Y(t), S(t+1))$ вероятностей появления вектора $S(t+1)$ при условии, что на выходах автоматов в момент времени t появился вектор $Y(t)=(y^1_{j1}, y^2_{j2}, \ldots, y^n_{jn})$.

При наличии двух автоматов эта вероятность может быть задана, например, следующим образом (автоматы имеют по два выхода):

Стратегия автомата A_1		Стратегии автомата A_2								
		y_1^2					y_2^2			
y_1^1	S	(0,0)	(0,1)	(1,0)	(1,1)	S	(0,0)	(0,1)	(1,0)	(1,1)
	p	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	p	0	0	0	1
y_2^1	S	(0,0)	(0,1)	(1,0)	(1,1)	S	(0,0)	(0,1)	(1,0)	(1,1)
	p	$\frac{1}{2}$	0	0	$\frac{1}{2}$	p	$\frac{3}{4}$	$\frac{1}{4}$	0	0

Если в момент времени t автомат A_1 выбрал свою вторую стратегию, т. е. на его выходе появился сигнал y_2^1, а автомат A_2 выдал в этот же момент времени сигнал y_1^2, то в момент времени $t+1$ оба автомата с равной вероятностью получат сигнал о выигрыше или проигрыше. Аналогично рассматриваются и остальные случаи выбора автоматами своих стратегий. Таким образом, партия игры автоматов состоит из выбора ими в момент времени t одной из возможных для них стратегий и получения в момент $t+1$ перед очередным выбором сообщения о результатах своего выбора (и выбора всех остальных автоматов) в момент времени t.

Информацию о выигрышах и проигрышах автоматов при $n=2$ можно задавать и в более удобной для описания форме. Если k есть число полных стратегий автомата A_1, а l — число полных стратегий автомата A_2, то матрица игры есть совокупность двух матриц порядка $k \times l$. Элементами первой из них являются вероятности выигрышей (или проигрышей) для первого автомата при данной комбинации полных стратегий, выбираемых автоматами в игре, а элементами второй матрицы являются вероятности выигрыша (проигрыша) второго автомата при различных выборах полных стратегий автоматами A_1 и A_2.

Для рассмотренной нами таблицы эти матрицы выглядят как

$$\begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{pmatrix} \text{ и } \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{3}{4} \end{pmatrix}.$$

Рассматриваемая нами игра автоматов не обязательно является игрой с нулевой суммой, так как проигрыш одного автомата в данной партии игры не всегда равен выигрышу другого автомата в данной партии. Если, например, для описанной выше игры оба автомата в данной партии игры выбирают свои первые стратегии, то каждый из них с вероятностью 0,5 получает выигрыш или проигрывает с такой же вероятностью. Это значит, что в результате такого выбора стратегий в конце партии оба автомата могут одновременно выиграть или одновременно проиграть, что противоречит условиям игр с нулевой суммой.

Теория игр с ненулевой суммой пока еще развита недостаточно. Для таких игр неверна основная теорема об оптимальных смешанных стратегиях, которая широко использовалась нами при решении игр с нулевой суммой. Это утверждение мы проиллюстрируем на следующем примере игры двух лиц. Пусть в игре участвуют игроки $И_1$ и $И_2$, каждый из которых имеет в своем распоряжении три полные стратегии. Матрицы игры задаются следующим образом:

$$M_1 = \begin{pmatrix} 1 & 6 & 2 \\ 3 & 0 & 3 \\ 4 & 2 & 0 \end{pmatrix}; \quad M_2 = \begin{pmatrix} 6 & 0 & 3 \\ 4 & 4 & 3 \\ 0 & 1 & 4 \end{pmatrix}.$$

Элементы первой матрицы указывают выигрыш первого игрока при соответствующем выборе стратегий игроками $И_1$ и $И_2$, а элементы второй матрицы определяют выигрыши (в данном случае проигрыш второго игрока при том же выборе стратегий). Можно условно считать, что в такой игре участвует третий игрок — природа, которая выплачивает игрокам $И_1$ и $И_2$ необходимые выигрыши и поглощает в случае необходимости «избыточные» проигрыши игроков.

В левой матрице стратегии природы соответствуют выбору соответствующих столбцов матрицы M_1, а в правой матрице выбор стратегии природы совпадает с выбором строки матрицы M_2. Таким образом, если, например, игрок $И_1$ выбирает свою вторую стратегию, а игрок $И_2$ свою третью стратегию, то природа в отношении игрока $И_1$ должна выбрать третью стратегию, а в отношении игрока $И_2$ свою вторую стратегию. При этом выигрыш первого игрока будет равен трем единицам, а проигрыш $И_2$ — также трем единицам.

Если игрок $И_1$, руководствуясь матрицей M_1 и считая, что имеет место игра с нулевой суммой, найдет на основании методов, изложенных нами в § 1, оптимальную смешанную стратегию, то она будет равна $\left(\frac{4}{11}, \frac{6}{11}, \frac{1}{11}\right)$. Для игрока $И_2$ оптимальная смешанная стратегия будет равна соответственно $\left(0 \ \frac{1}{4} \ \frac{3}{4}\right)$. При этом выигрыш $И_1$

будет равен в среднем $2\frac{4}{11}$, а проигрыш $И_2$ будет равен в среднем $3\frac{1}{4}$.

Однако если бы оба игрока вместо найденных таким образом оптимальных смешанных стратегий пользовались чистыми стратегиями J_1^1 и J_2^2, то выигрыш $И_1$ стал бы равен 6 единицам, а проигрыш $И_2$ сократился бы до нуля.

Этот пример показывает, что для игр с ненулевой суммой игрокам может оказаться выгодным образовывать коалиции с договорным распределением суммарного выигрыша между участниками коалиции. Пример такой игры автоматов будет рассмотрен нами в конце этого параграфа.

В силу вышесказанного естественно в начале ограничиться играми автоматов с нулевой суммой. Для случая двух автоматов условие того, что данная игра является игрой с нулевой суммой, выглядит следующим образом:

$$\begin{cases} p(Y(t), (0, 0)) = p(Y(t), (1, 1)) = 0; \\ p(Y(t), (0, 1)) + p(Y(t), (1, 0)) = 1. \end{cases} \qquad (45)$$

Рассмотрим вначале случай, когда автомат A_2 не учитывает результатов предыдущих партий для автомата A_1 и выбирает свои чистые стратегии, сообразуясь только с заданными вероятностями выбора стратегий r_1, r_2, ..., r_l. Тогда вероятности выигрыша и проигрыша автомата A_1 в партии, происходящей в момент времени t, выражаются следующими соотношениями:

$$p_i = \sum_{j=1}^{l} p_{ij} r_j$$

и

$$q_i = \sum_{j=1}^{l} q_{ij} r_j; \qquad (46)$$

здесь p_{ij} — вероятность выигрыша автомата A_1 при выборе им полной стратегии с номером i и выборе автоматом A_2 полной стратегии с номером j, а q_{ij} — вероятность проигрыша автоматом A_1 при том же выборе полных стратегий автоматами. Соотношения (46) показывают, что поставленная задача есть просто задача об оптимальном поведении автомата в случайной среде $S=(q_1, q_2, \ldots, q_k)$ при условии, что матрица P имеет вид:

$$P=\begin{pmatrix} p_{11}-q_{11} & p_{12}-q_{12} & \ldots & p_{1l}-q_{1l} \\ p_{21}-q_{21} & p_{22}-q_{22} & \ldots & p_{2l}-q_{2l} \\ \cdot & \cdot & \cdot & \cdot \\ p_{k_1}-q_{k_1} & p_{k_2}-q_{k_2} & \ldots & p_{kl}-q_{kl} \end{pmatrix}, \tag{47}$$

где p_{ij}—q_{ij} есть математическое ожидание выигрыша автомата A_1 при соответствующем выборе автоматами полных стратегий. Для нахождения оптимальной стратегии автомата A_1 можно использовать методы, рассмотренные нами в § 7. Рассмотрим в качестве автомата A_1 автомат с линейной тактикой, число состояний которого равно nk. Обозначим эти состояния как z_j^i, где $i=1, 2, \ldots, k$, а $j=1, 2, \ldots, n$. В состоянии z_j^i автомат выдает на выходе сигнал с номером i (выбирает полную стратегию с номером i).

Если в данной партии получен выигрыш (нештраф), то смена состояний автомата происходит по правилу $z_j^i \to z_{j-1}^i$, где $j=2, 3, \ldots, n$ и $z_1^i \to z_1^i$. При наличии проигрыша в данной партии (при штрафе) смена состояний происходит по правилу $z_j^i \to z_{j+1}^i$, где $j=1, 2, \ldots, n-1$, $z_n^i \to z_n^{i+1}$ и $z_n^k \to z_n^1$. Из результатов, сформулированных в § 7, следует, что математическое ожидание выигрыша для автомата A_1 в игре рассматриваемого типа имеет вид:

$$M(P, S)=\sum_{i=1}^{k}(1-\lambda_i)^n\left(\sum_{i=1}^{k}\frac{1-\lambda_i^n}{v_i}\right)^{-1},$$

где

$$v_i=\sum_{j=1}^{k}(p_{ij}-q_{ij})\,r_j \text{ и } \lambda_i=\frac{p_i}{q_i},$$

и

$$\lim_{n\to\infty} M(P, S)=\text{макс}\,(v_1, v_2, \ldots, v_k). \tag{48}$$

Однако это условие выполняется только тогда, когда макс$(v_1, v_2, \ldots, v_k) \geqslant 0$. Если же это не так, то при $n \to \infty$ математическое ожидание выигрыша для автомата A_1 имеет вид среднего гармонического $(v_1^{-1} + v_2^{-1} + \ldots + v_k^{-1})^{-1}$. Таким образом, автомат A_1 играет иногда хуже своего партнера, но если партнер (автомат A_2) использует оптимальную смешанную стратегию, то математическое ожидание выигрыша для автомата A_1 будет равно на основании (48) (при условии, что выполнено необходимое для этого условие на элементы матрицы P) цене игры.

Если предположить, что оба автомата A_1 и A_2 есть автоматы с линейной тактикой и числом состояний kn и ln соответственно, то при правилах перехода для $s=0$:

$$z^i_j \to z^i_{j-1} \; (j=2, 3, \ldots, n), \quad z^i_1 \to z^i_1,$$

а для $s=1 \; (j=1, 2, \ldots, n-1)$:

$$z^i_j \longrightarrow z^i_{j+1}; \quad z^i_n \longrightarrow \left\{ \begin{matrix} z^1_n \\ z^2_n \\ \vdots \\ z^q_n \end{matrix} \right. \quad \text{с вероятностями } \frac{1}{q},$$

где $q=k; \; l$.

можно показать следующие свойства математического ожидания выигрыша для любого из этих автоматов: 1. Если в матрице P вида (47) есть хотя бы одна строка из неотрицательных элементов, то математическое ожидание выигрыша для автомата A_1 стремится с ростом n к среднему гармоническому из элементов этой строки. 2. Если такой строки в матрице P нет, то $\underset{n \to \infty}{M(P, S)} \to 0$, что соответствует ничьей. Автомат A_1 аналогичен игроку с осторожной тактикой и стремится не к максимальному выигрышу, а к его гарантированному значению.

Исследование игр большого числа автоматов аналитическими методами довольно затруднительно. Поэтому в настоящее время основные результаты в теории таких игр для автоматов получены из экспериментального моделирования довольно простых игр на вычислительных машинах дискретного действия. Ниже мы приведем результаты такого моделирования конкретных задач, но прежде нам необходимо рассмотреть ряд положений теории игр со многими участниками, которые могут в процессе игры обмениваться некоторой информацией и кооперироваться между собой.

Пусть в игре участвует n игроков. Игрок $И_j$ имеет в своем распоряжении k_j полных стратегий. Функция выигрыша этого игрока зависит не только от той полной стратегии, которую он выбрал, но и от выбора своих полных стратегий другими игроками. Интересен частный случай, когда все функции выигрыша для игроков $И_j$ одинаковы с точностью до переименования аргументов в этих функциях.

Партией игры называется вектор $F=(f^1_{i_1}, f^2_{i_2}, \ldots, f^i_{i_n})$, где $f^j_{i_j}$ есть полная стратегия, выбранная игроком $И_j$. Каждый вектор F определяет вектор $S=(s^1_{i_1}, s^2_{i_2}, \ldots, s^n_{i_n})$ платежей игрока.

Таким образом, для каждого игрока определяется платежная функция $s^j=\varphi(F)$, определяемая на множестве партий игры. Отметим, что при дальнейшем изложении не предполагается, что в каждой отдельной партии мы имеем игру с нулевой суммой, т. е. мы не предполагаем, что сумма выигрышей одной группы игроков совпадает с суммой проигрышей другой группы игроков. Таким образом, в игре существует некоторый банк, за счет которого происходит выплата недостающих выигрышей или исчезновение из игры «лишних» проигрышей.

Цель каждого игрока, так выбрать свою стратегию, чтобы обеспечить себе максимально возможный выигрыш.

Среди всевозможных партий игры выделим такие, которые характеризуются тем свойством, что никому из игроков, участвующих в игре, невыгодно изменять свою стратегию при условии, что никто из остальных игроков не меняет свою стратегию. Если в игре существуют такие партии, то это соответствует некоторому «равновесному состоянию» в игре. Партии такого типа называются партиями Нэша в честь Дж. Нэша, первого исследователя игр с такими партиями.

Пусть, например, матрица игры с нулевой суммой двух игроков имеет вид:

$$\begin{pmatrix} 0{,}25 & 0{,}9 \\ -0{,}1 & 0{,}25 \end{pmatrix}.$$

В этой игре партия, в которой оба игрока выбирают свои первые стратегии, является партией Нэша. В самом деле, если игрок $И_1$ захочет изменить свою стратегию, то его проигрыш станет равным 0,1, а если игрок $И_2$ изменит свою стратегию, то его проигрыш достигнет значения 0,9. Другим примером игры, имеющей партии Нэша, может служить симметричная игра двух игроков, в распоряжении каждого из которых имеются две полные стратегии и платежные функции обоих игроков совпадают (поэтому игра и называется симметричной). Пусть эта платежная функция имеет следующий вид:

$$\varphi(1{,}1)=0{,}4;\ \varphi(1{,}2)=\varphi(2{,}1)=0{,}1;\ \varphi(2{,}2)=0{,}8.$$

В этой игре есть две партии Нэша, которые соответствуют выбору игроками одинаковых стратегий.

Однако партия Нэша не всегда соответствует устремлениям игроков. Если игрок стремится максимизировать свой выигрыш, то партия Нэша может оказаться для него не самой оптимальной. Во втором из вышеприведенных примеров партия Нэша, в которой оба игрока выбрали первую стратегию, не оптимальна для игроков, ибо максимальный выигрыш игроков достигается при одновременном выборе ими вторых стратегий. Партии Нэша, для которых платежные функции всех игроков достигают одновременно максимального значения выигрыша, соответствуют партиям, которые обладают «абсолютной устойчивостью» для игроков. Если, следуя терминологии отечественных работ, называть такие партии партиями Мора, то можно сказать, что если в данной игре есть партии Мора, то цель игроков, участвующих в игре, состоит в нахождении и разыгрывании партии Мора. Во втором из рассмотренных нами примеров партия Нэша, при которой оба игрока выбирают вторые статегии, есть партия Мора.

В качестве примера такой игры, в которой участвует группа автоматов, мы рассмотрим игру в размещения. Игра в размещение описывается следующим образом. Пусть в игре принимают участие n автоматов A_1,

$A_2, \ldots, A_n$. Каждый из автоматов имеет $k>n$ стратегий, которыми он может воспользоваться. Эти стратегии мы обозначим как $s_1, s_2, \ldots, s_k$. Имеется k положительных оценок — констант $1 \geqslant a_1 \geqslant a_2 \geqslant \ldots \geqslant a_k \geqslant 0$.

Оценка a_i называется мощностью стратегии s_i. Если в некоторой партии m_j игроков — автоматов выбрали одинаковую стратегию s_j, то выигрыш любого из них определяется по формуле

$$M = \frac{a_j}{m_j}.$$

Пусть m_i автоматов выбрали стратегию a_i и $\sum\limits_i m_i = n$.

Если m_i удовлетворяют неравенствам

$$\frac{a_i}{m_i} > \frac{a_i}{m_j + 1} \quad (i, j) \text{ — любые,}$$

то ни одному из автоматов невыгодно менять свою стратегию и игра в размещение имеет $n!$ партий Нэша. Если бы автоматы могли договариваться между собой так, чтобы получить наибольший суммарный выигрыш, то они должны были бы выбрать первые n стратегий и распределиться по этим стратегиям по одному. Это случай общей кассы игроков, когда все выигрыши игроков суммируются и делятся между игроками поровну. Однако если договоренности нет, то автомат, который выбрал стратегию s_i $(i>1)$, вообще говоря, хочет сменить ее на стратегию с меньшим номером, рассчитывая получить больший выигрыш. Поэтому при отсутствии общей кассы партия Нэша $(\overbrace{1, 1, \ldots, 1}^{n}, \overbrace{0, \ldots, 0}^{k-n})$ не является партией Мора.

Заканчивая нашу книгу, отметим, что теория игр автоматов находится сейчас на стадии интенсивного развития и каждый день приносит новые результаты в этом многообещающем направлении теоретической кибернетики.

ЛИТЕРАТУРА

К § 1

1. Дрешер М., Стратегические игры, изд-во «Советское радио», 1964.
2. Льюс Р., Райфа Х., Игры и решения, Изд-во иностранной литературы, 1961.
3. Мак Кинси Дж., Введение в теорию игр, Физматгиз, 1960.

К § 3

1. Амбарян С. Л., Брутян Х. К., Тер-Микаэлян Т. М., Об автоматической выработке оценки ситуации при игре в крестики и нулики, «Проблемы кибернетики», вып. 13, 1965.
2. Брудно А. Л., Ландау И. Я., Одномастка, «Проблемы кибернетики», вып. 13, 1965.
3. Ганзен В. А., Электронная модель игры, «Нервная система», издание Ленинградского государственного университета, вып. 3, 1962.
4. Срапян Ш. О., Тер-Микаэлян Т. М., Об одном методе оценки ситуации при игре в крестики и нулики, «Проблемы кибернетики», вып. 9, 1963.
5. Шеннон К., Играющие машины, «Кибернетический сборник», вып. 1, 1960.

К § 4

1. Безбородов Ю. М., Орлов В. Б., Машина играет в шахматы, «Математическое просвещение», вып. 6, 1961.
2. Брудно А. Л., Грани и оценки для сокращения перебора вариантов, «Проблемы кибернетики», вып. 10, 1963.
3. Гаазе-Рапопорт М. Г., Автоматы и живые организмы, Физматгиз, 1961.
4. Кадушин В. П., Об алгоритмизации игры в нарды, Вероятностные методы и кибернетика, вып. 2, издание КГУ, 1963.
5. Цермело Э., О применении теории множеств к теории шахматной игры, Сб. «Матричные игры», Физматгиз, 1961.
6. Baylor G. W., A heuristic approach to chess programs, Carneque Technique, v. 25, № 1, 1960.
7. Braffort P., Lussan A., Rencontre de problémes numériques et non-numériques lors de l'elaboration d'un programme destiné à la simulation du jeu de Go-Bang. «Deuxieme Congress Association franc. calcul et traitement information AFCALTI, Paris 1961», Paris 1962.

8. Constantinéscu P., Lulea C., Niculescu S., Algoritm pentru determinarea nucleului grafului asociat jocului NIM, Studii si cercetari mat. Acad., RPR, t. 15, № 1, 1964.
9. Findler N. V., Some remarks on the game «dama» which can be played on a digital computer, Computer, J., v. 3, № 1, 1960.
10. Kozdrowicki E. W., Johnson D. L., Automated chess playing: moving towards simulated intelligence, Trend Engng. Univ., Washington, v. 15, № 4, 1963.
11. Marill Th., A note on pattern-recognition techniques and game-playing programs, Inform. and Control, v. 6, № 3, 1963.
12. Michie D., Experiments on the mechanization of game-learning, Computer J., v. 3, № 3, 1963.
13. Pelikan P., Machine performing the game NIM, Stroje na zpracování informačí, Sb. 8, 1962.
14. Remus H., Simulation of a learning machine for playing GO, Information Process, 1962, Amsterdam, 1963.
15. Samuel A. L., Programming computers to play games, Advances in computers, v. 1, 1960.
16. Samuel A. L., Some studies in machine learning using the game of checkers, Lernende Automaten, München, 1961.
17. Schlang A., A digital NIM computer, Radio Electronics, v. 25, № 6, 1954.
18. Schliebs G., Über die Grundzüge eines Programms für eine Schachspielende, Funk und Ton, v. 7, № 5, 1953.
19. Stein P., Ulam S., Experiments in chess on electronic computing machines, Computers and Automatisierung, v. 6, № 9.
20. Veenker G., Eine Programm zur Lösung von Schachaufgaben, Elektronische Rechenanlagen, H. 1, 1965.
21. Weizenbaum J., How to make a computer appear intelligent, Datamation, v. 8, № 2, 1962.

К § 5

1. Генин М. А., Решение одной дискретной игры двух лиц с блефом, Вестник ЛГУ, № 1, 1964.
2. Первин Ю. А., Об алгоритмизации и программировании игры в домино, «Проблемы кибернетики», вып. 3, 1960.
3. Bastian A. L., Foley J. P., Petrick S. R., On the implementation of a language for contract bridge bidding, Symbolic Languages Data Process, 1962.
4. Berlekamp E. R., Program for double-dummy bridge problems — a new strategy for mechanical game playing, J. of ACM, v. 10, № 3, 1963.
5. Findler N. V., Computer model of gambling and bluffing, IRE Transaction on Electronic Computing, v. 10, № 1, 1961.
6. Gardner M., Mathematical games, How to play dominoes in two and three dimensions, Scient. Amer., v. 204, № 3, 1961.
7. Thiele T. N., Lemke R. R., Fu K. S., A digital computer card-playing program, Behav. Sci., v. 8, № 4, 1963.
8. Wenzel G., Spiele bridg—it gegen die IBM 1401, Bürotechn. Automat., v. 4, № 3, 1963.

К § 6

1. Хагельбергер Д. У., СИИР — автомат, экстраполирующий последовательности, «Кибернетический сборник», вып. 1, 1960.
2. Уайт Дж., Машины, играющие в сравнение монет, «Кибернетический сборник», вып. 3, 1961.
3. Griesmer J. H., Shubik M., Toward a study of bidding processes: some constant sum games, Naval Res. Logist. Quart, v. 10, № 1, 1963.
4. Thorp E., A favorable strategy for twenty-one, Proc. Nat. Acad. Sci. USA, v. 47, № 1, 1961.
5. Wall E., Brown R. M., A pennymatching programm, Communs of ACM, v. 6, № 6, 1963.

К § 7

1. Бардзинь Я. М., Емкость среды и поведение автоматов, Доклады АН СССР, т. 160, 1965, № 2.
2. Бланк А. М., Целесообразность автоматов, «Автоматика и телемеханика», 1964, № 10.
3. Брызгалов В. И., Пятецкий-Шапиро И. И., Шик М. Л., О двухуровневой модели взаимодействия автоматов, Доклады АН СССР, т. 160, 1965, № 5.
4. Варшавский В. И., Воронцова И. П., О поведении стохастических автоматов с переменной структурой, «Автоматика и телемеханика», № 3, 1963.
5. Варшавский В. И., Воронцова И. П., Цетлин М. Л., Обучение стохастических автоматов, Сб. «Биологические аспекты кибернетики», 1962.
6. Добровидов А. В., Стратонович Р. Л., О синтезе оптимальных автоматов, функционирующих в случайных средах, «Автоматика и телемеханика», 1964, № 10.
7. Крылов В. Ю., Об одном автомате, асимптотически оптимальном в случайной среде, «Автоматика и телемеханика», 1963, № 9.
8. Милютин А. А., Об автоматах с оптимальным целесообразным поведением в стационарной среде, «Автоматика и телемеханика», 1965, № 1.
9. Пономарев В. А., Об одной конструкции автомата, асимптотически оптимального в стационарной случайной среде, «Биофизика», 1964, вып. 1.
10. Стефанюк В. Л., Пример задачи на коллективное поведение двух автоматов, «Автоматика и телемеханика», 1963, № 6.
11. Цетлин М. Л., Конечные автоматы и моделирование простейших форм поведения, «Успехи математических наук», 1963, вып. 4.
12. Цетлин М. Л., Некоторые задачи о поведении конечных автоматов, Доклады АН СССР, т. 139, 1961, № 4.
13. Цетлин М. Л., О поведении конечных автоматов в случайных средах, «Автоматика и телемеханика», 1961, № 10.
14. Яровицкий Н. В., Предельное поведение замкнутой системы автоматов со случайным входом, «Кибернетика», 1965, № 1.

К § 8

1. Гельфанд И. М., Пятецкий-Шапиро И. И., Цетлин М. Л., О некоторых классах игр и игр автоматов, Доклады АН СССР, т. 152, 1963, № 4.
2. Кринский В. И., Пономарев В. А., Об играх вслепую, «Биофизика», 1964, вып. 3.
3. Кринский В. И., Об одной конструкции последовательности автоматов и ее поведении в играх, Доклады АН СССР, т. 156, 1964, № 6.
4. Крылов В. Ю., Цетлин М. Л., Об играх автоматов, «Автоматика и телемеханика», 1963, № 7.
5. Цетлин М. Л., Замечание об игре конечного автомата с партнером, использующим смешанную стратегию, Доклады АН СССР, т. 149, 1963, № 1.
6. Цетлин М. Л., Крылов В. Ю., Примеры игр автоматов, Доклады АН СССР, т. 149, 1963, № 2.

СОДЕРЖАНИЕ

БИБЛИОТЕКА ПО АВТОМАТИКЕ

ГОТОВЯТСЯ К ПЕЧАТИ

Богорад Г. З., Киблицкий В. А., Цифровые регуляторы и измерители скорости

Борцов Ю. А., Суворов Г. В. Методы исследования динамики сложных систем электропривода

Бунаков В. Л., Гаспаров Р. Г., Полупроводниковые регуляторы частоты и напряжения электрических машин

Гиршберг В. В. и др., Единая серия полупроводниковых логических элементов промышленной автоматики

Веников Г. В., Сверхбыстродействующие вычислительные устройства

Дружинин Г. В., Реле времени

Зимин Е. Н., Преображенский В. И., Соколов Н. Г., Элементы и схемы бесконтактного управления металлорежущими станками

Карандеев К. Б., Гриневич Ф. Б., Новик А. И., Емкостные самокомпенсированные уровнемеры

Коротков А. М., Мочалов В. Д., Счетчики импульсов на переключающих диодах

Лабунцов В. А., Ривкин Г. А., Шевченко Г. И., Автономные инверторы и преобразователи частоты

Лисичкин Д. А., Транзисторные усилители с обратными связями для следящих систем

Макаров А. К., Свердлин В. М., Автоматические сигнализаторы уровня

Тун А. Я., Тахогенераторы для управления электроприводами

Туркельтауб Р. М., Метод исследования точности и надежности схем аппаратуры

Тутевич В. Н. и др., Временные системы телеуправления на магнитных элементах

Форейт И., Емкостные датчики неэлектрических величин

Фролов Л. Б., Измерение крутящего момента

Эйгенброт В. М., Многоканальные регуляторы технологических процессов

Цена 35 коп.