

#### 4. ЛЕКЦИЯ. Скрытые Марковские модели

Скрытая марковская модель (Hidden Markov model – НММ) – это марковская цепь, в которой состояния не являются непосредственно наблюдаемыми. Если обратиться к примеру прогнозирования погоды, то погоду невозможно «измерить» напрямую, в действительности погода оценивается на основе последовательности сенсорных (воспринимаемых) показателей – температуры, давления, скорости ветра и т. п. Поэтому, как и во многих других явлениях, когда состояния не являются непосредственно наблюдаемыми, скрытые марковские модели предоставляют более подходящий и более мощный инструмент моделирования. Кроме того, можно объяснить скрытую марковскую модель как двойной стохастический (случайный) процесс:

1. скрытый стохастический процесс, который мы не можем наблюдать напрямую;
2. второй стохастический процесс, который создает последовательность наблюдений с учетом первого процесса.

Например, предположим, что имеются две «неправильные» или несимметричные монеты  $M_1$  и  $M_2$ . Для монеты  $M_1$  более высока вероятность выпадения орлов, а для монеты  $M_2$  более высока вероятность выпадения решек. Некто последовательно подбрасывает эти две монеты, но неизвестно, какая именно монета выбрана в каждом конкретном случае. Можно наблюдать только результаты бросков – орлы или решки:

O, R, R, O, R, O, O, O, R, O, R, O, R, R, O, R, O, O, ...

Предположим, что игрок, бросающий монеты, сам выбирает первую монету в последовательности (априорные вероятности), а следующую монету для броска выбирает с учетом предыдущего результата (вероятности переходов) с равной вероятностью. Кроме априорных вероятностей и вероятностей переходов для состояний (как и в марковской цепи), в скрытой марковской модели необходимо определить вероятности наблюдений, в рассматриваемом здесь случае это вероятности выпадения орлов и решек с учетом каждой монеты (конкретного состояния).

Предположим, что для монеты  $M_1$  вероятность выпадения орлов равна 80 %, а для монеты  $M_2$  вероятность выпадения решек равна 80 %. Далее для этого простого примера определены все требуемые параметры, объединенные в табл. 4.1, 4.2 и 4.3.

Таблица 4.1. Априорные вероятности ( $\Pi$ ) для примера с несимметричными монетами

	$M_1$	$M_2$
--	-------	-------

$\Pi =$	0,5	0,5
$\bar{\Pi} =$		

Таблица 4.2. Вероятности переходов (A) для примера с несимметричными монетами

		$M_1$	$M_2$
$A =$	$M_1$	0,5	0,5
	$M_2$	0,5	0,5

Таблица 4.3. Вероятности наблюдений (B) для примера с несимметричными монетами

		$M_1$	$M_2$
$B =$	$M_1$	0,8	0,5
	$M_2$	0,2	0,8

Диаграмма состояний для примера с несимметричными монетами изображена на рис. 4.1 с двумя переменными состояниями и двумя возможными наблюдениями, которые зависят от состояния.

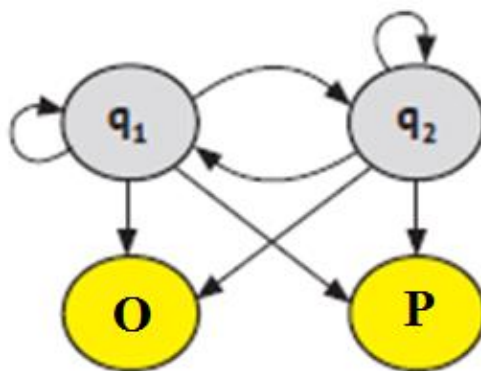


Рис. 4.1. Диаграмма состояния для скрытой марковской модели в примере с несимметричными монетами. Показаны два состояния  $q_1$  и  $q_2$  и два наблюдения  $O$  (орел) и  $P$  (решка) с дугами, представляющими переходы и вероятности наблюдений

Формально скрытая марковская модель определяется следующим образом:

- множество состояний:  $Q = \{q_1, q_2, \dots, q_n\}$ ;

Несмотря на то что состояния являются скрытыми, во многих случаях есть соответствие между состоянием модели и реальным состоянием процесса. В примере с подбрасыванием монеты каждое состояние соответствовало выбранной монете. В общем, переход в любое выбранное состояние возможен из любого состояния всей системы (в том числе и само в себя); с другой стороны, лишь определенные пути переходов представляют интерес в каждой конкретной модели.

- множество наблюдений:  $O = \{o_1, o_2, \dots, o_m\}$ ;

Есть количество возможных символов в наблюдаемой последовательности, размер алфавита наблюдаемой последовательности. В случае с подбрасыванием монеты — это 2 символа: орел и решка.

- вектор априорных вероятностей:  $\Pi = \{\pi_1, \pi_2, \dots, \pi_n\}$ , где  $\pi_i = P(S_0 = q_i)$ , то есть вероятность того,  $q_i$  — это начальное состояние модели.

- матрица вероятностей переходов:  $A = \{a_{ij}\}, i = [1..n], j = [1..n]$ , где  $a_{ij} = P(S_t = q_j | S_{t-1} = q_i)$ , то есть это вероятность того, что система, находящаяся в состоянии  $q_i$ , перейдет в состояние  $q_j$ . Если для любых двух состояний в модели возможен переход из одного состояния в другое, то  $a_{ij} > 0$  для любых  $i, j$ . В остальных скрытых моделях для некоторых  $i, j$  вероятность перехода  $a_{ij} = 0$ .

- матрица вероятностей наблюдений:  $B = \{b_{ik}\}, i = [1..n], j = [1..m]$ , где  $b_{ik} = P(O_t = o_k | S_t = q_i)$ .

$b_{ik}$  — вероятность того, что в момент времени  $t$ , система, находящаяся в  $i$ -ом состоянии, выдаст  $k$ -тый символ в наблюдаемую последовательность.

Здесь  $n$  — количество состояний,  $m$  — количество наблюдений,  $S_0$  — начальное состояние.

Заметим, что полное описание скрытой марковской модели состоит из двух параметров модели  $Q$  и  $O$ , описания символов наблюдаемой последовательности и трех массивов вероятностей —  $A, B$  и  $\Pi$ . Для удобства используется следующую запись в компактной форме:  $\lambda = \{A, B, \Pi\}$ .

Скрытая марковская модель (первого порядка) обладает следующими свойствами:

- марковское свойство:

$$P(S_t = q_j | S_{t-1} = q_i, S_{t-2} = q_k, \dots) = P(S_t = q_j | S_{t-1} = q_i);$$

- стационарный (устойчивый) процесс:

$$P(S_{t-1} = q_j | S_{t-2} = q_i) = P(S_t = q_j | S_{t-1} = q_i) \text{ и}$$

$$P(O_{t-1} = o_k | S_{t-1} = q_i) = P(O_t = o_k | S_t = q_i), \forall(t);$$

- независимость от наблюдений:

$$P(O_t = o_k | S_t = q_i, S_{t-1} = q_i, \dots) = P(O_t = o_k | S_t = q_i)$$

Марковское свойство подразумевает, что вероятность текущего состояния зависит только от предыдущего состояния и не зависит от остальной (предыдущей) хронологии состояний. Второе свойство означает, что вероятности переходов и наблюдений не изменяются со временем, т. е. процесс стационарный (устойчивый). Третье свойство определяет, что наблюдения зависят только от текущего состояния. Существуют расширения базовой

скрытой марковской модели, которые смягчают некоторые из этих положений.

С учетом описанных выше свойств графовая скрытая марковская модель показана на рис. 4.2. Эта модель содержит две последовательности случайных переменных: состояние в момент времени  $t$   $S_t$  и наблюдение в момент времени  $t$   $O_t$ .

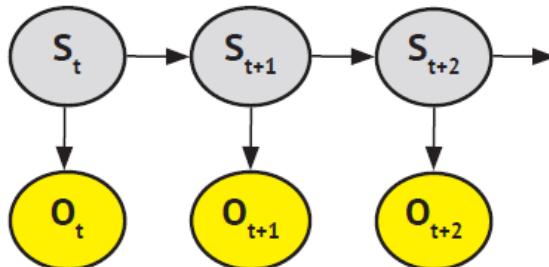


Рис. 4.2. Скрытая марковская модель, представленная в виде графа

С учетом представления конкретной предметной области в форме скрытой марковской модели возникают три основных вопроса (задачи), которые важны в большинстве приложений:

- 1) вычисление оценки: в рассматриваемой модели необходимо определить оценку вероятности последовательности наблюдений;
- 2) оптимальная последовательность: с учетом модели и конкретной последовательности наблюдений необходимо определить оценку наиболее вероятной последовательности состояний, которая создает эти наблюдения;
- 3) обучение параметров: с учетом количества последовательностей наблюдений необходимо отрегулировать (обучить) параметры модели.

### **Вычисление оценки.**

Вычисление оценки заключается в определении вероятности последовательности наблюдений  $O = \{o_1, o_2, o_3, \dots\}$  с учетом модели  $\lambda = \{A, B, P\}$ , т. е. в оценке вероятности  $P(O|\lambda)$ . Будут рассмотрены два метода. Сначала будет представлен прямой метод, простейший алгоритм, который стал причиной разработки более эффективного метода.

#### *Прямой метод.*

Последовательность наблюдений  $O = \{o_1, o_2, o_3, \dots, o_T\}$  можно сгенерировать по различным последовательностям состояний  $Q_i$ , так как состояния неизвестны для скрытых марковских моделей. Таким образом, для вычисления вероятности последовательности наблюдений можно вычислить ее оценку для конкретной последовательности состояний, а затем прибавить вероятности для всех возможных последовательностей состояний:

$$P(O|\lambda) = \sum_i P(O, O_i|\lambda) \quad (4.1)$$

Чтобы получить  $P(O, O_i|\lambda)$ , нужно просто умножить вероятность

начального состояния  $q_1$  на вероятности переходов для последовательности состояний  $q_1, q_2, \dots$  и вероятности наблюдений для последовательности наблюдений  $o_1, o_2, \dots$ :

$$P(O, O_i | \lambda) = \pi_1 b_1(o_1) a_{12} b_2(o_2) \dots a_{(T-1)T} b_T(o_T) \quad (4.2)$$

Таким образом, вероятность  $O$  с учетом суммирования всех возможных последовательностей состояний  $Q$  равна:

$$P(O | \lambda) = \sum_Q \pi_1 b_1(o_1) a_{12} b_2(o_2) \dots a_{(T-1)T} b_T(o_T) \quad (4.3)$$

Для модели с  $N$  состояниями и длиной (последовательности) наблюдений  $T$  существует  $N^T$  возможных последовательностей состояний. Каждый член в выражении суммирования требует выполнения  $2T$  операций. В результате для вычисления оценки требуется количество операций порядка  $2T \times N^T$ .

Например, если рассматривается модель с пятью состояниями  $N = 5$  и с длиной последовательности наблюдений  $T = 100$  – это обычные параметры для приложений скрытых марковских моделей, – то количество требуемых операций будет иметь порядок 1072. Очевидно, что необходим более эффективный метод.

#### *Итеративный метод*

Основная идея итеративного метода, также известного как алгоритм Forward, заключается в определении оценки вероятностей состояний/наблюдений для каждого интервала времени. То есть вычисляется вероятность частичной последовательности наблюдений до времени  $t$  (начиная с момента времени  $t = 1$ ), и на основе этого неполного (промежуточного) результата вычисляется вероятность последовательности наблюдений для времени  $t + 1$  и т. д.

Сначала необходимо определить вспомогательную переменную, обозначенную как *forward*:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, S_t = g_i | \lambda) \quad (4.4)$$

Это вероятность частичной последовательности наблюдений до момента времени  $t$ , находящаяся в состоянии  $q_i$  в момент времени  $t$ .

Итеративный алгоритм состоит из трех основных частей: инициализация, индукция и завершение. На этапе инициализации определяются переменные  $\alpha$  для всех состояний в начальный момент времени. На этапе индукции вычисляются значения  $\alpha_{t+1}(i)$  по значениям  $\alpha_t(i)$ . Эти вычисления повторяются от  $t = 2$  до  $t = T$ . Наконец, на этапе завершения вычисляется значение  $P(O | \lambda)$  путем суммирования всех значений  $\alpha_T$ . Рассмотрим полную процедуру алгоритма 4.1.

**Требования:** Скрытая марковская модель  $\lambda$ ; Последовательность наблюдений  $O$ ; Количество состояний  $N$ ; Количество наблюдений  $T$

**for**  $i = 1$  **to**  $N$  **do**

$\alpha_1(i) = P(o_1, S_1 = q_i) = \pi_i b_i(O_1)$  (Инициализация)

**end for**

**for**  $t = 2$  **to**  $T$  **do**

**for**  $j = 1$  **to**  $N$  **do**

$\alpha_t(j) = [\sum_i \alpha_{t-1}(i) a_{ij}] b_j(O_t)$  (Индукция)

**end for**

**end for**

$P(O) = \sum_i \alpha_T(i)$  (Завершение)

**return**  $P(O)$

Теперь проанализируем сложность по времени этого итеративного метода. На каждой итерации требуется (приблизительно)  $N$  операций умножения и  $N$  операций сложения, так что для  $T$  итераций количество операций будет иметь порядок  $N^2 \times T$ . Таким образом, сложность по времени от экспоненциальной по  $T$  для прямого метода снизилась до линейной по  $T$  и квадратичной по  $N$  для итеративного метода. Это существенное уменьшение сложности. Следует отметить, что в большинстве приложений  $T \gg N$ .

Если вернуться к примеру с  $N = 5$  и  $T = 100$ , то теперь количество операций приблизительно равно 2500.

### Оценка состояния

Поиск наиболее вероятной последовательности состояний для некоторой последовательности наблюдений  $O = \{o_1, o_2, o_3, \dots\}$  можно интерпретировать двумя способами:

- получение наиболее вероятного состояния  $S_t$  в каждом интервале времени  $t$ ;
- получение наиболее вероятной последовательности состояний  $S_0, S_1, \dots, S_T$ .

Следует отметить, что Сначала решается задача поиска наиболее вероятного или оптимального состояния для конкретного момента времени  $t$ , затем задача поиска оптимальной последовательности состояний.

В первую очередь необходимо определить несколько дополнительных вспомогательных переменных. Переменная *backward* аналогична переменной *forward*, но в данном случае мы начинаем с конца последовательности, т. е.

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T, S_t = g_i | \lambda) \quad (4.5)$$

Это вероятность частичной последовательности наблюдений от  $t + 1$  до  $T$ , при нахождении в состоянии  $q_i$  в момент времени  $t$ . Тем же способом, что и для  $\alpha$ , можно выполнить итеративные вычисления, но теперь в обратном направлении:

$$\beta_t(i) = \sum_j \beta_{t+1}(j) a_{ij} b_j(o_t) \quad (4.6)$$

Переменные  $\beta$  для момента времени  $T$  определяются как  $\beta_T(j) = 1$ .

Таким образом можно решить и задачу вычисления оценки в определении вероятности последовательности наблюдений, используя  $\beta$  вместо  $\alpha$ , начиная с конца последовательности наблюдений и выполняя итерации обратно по времени. Или можно объединить обе переменные и выполнять итерации в прямом и обратном направлениях, встретившись в некоторый промежуточный момент времени  $t$ , т. е.:

$$P(O, s_t = g_i | \lambda) = \alpha_t(i) \beta_t(i) \quad (4.7)$$

Тогда:

$$P(O | \lambda) = \sum_i \alpha_t(i) \beta_t(i) \quad (4.8)$$

Теперь определим дополнительную переменную  $\gamma$  – это условная вероятность нахождения в определенном состоянии  $q_i$  с учетом последовательности наблюдений:

$$\gamma_t(i) = P(s_t = g_i | O, \lambda) = P(s_t = g_i, O | \lambda) / P(O) \quad (4.9)$$

Эту формулу можно записать с использованием переменных  $\alpha$  и  $\beta$  в следующем виде:

$$\gamma_t(i) = \alpha_t(i) \beta_t(i) / \sum_i \alpha_t(i) \beta_t(i) \quad (4.10)$$

Эта переменная  $\gamma$  дает решение первой подзадачи о нахождении наиболее вероятного состояния (MPS – most probable state) в момент времени  $t$ : необходимо просто определить, для какого состояния эта переменная принимает максимальное значение, т. е.

$$MPS(t) = ArgMax_i \gamma_t(i) \quad (4.11)$$

Теперь решим вторую подзадачу – нахождение наиболее вероятной последовательности состояний  $Q$  с учетом последовательности наблюдений  $O$ , такой, что требуется максимизация  $P(Q | O, \lambda)$ . По правилу Байеса:  $P(Q | O, \lambda) = P(Q, O | \lambda) / P(O)$ . Учитывая, что  $P(O)$  не зависит от  $Q$ , это равнозначно максимизации  $P(Q, O | \lambda)$ .

Метод получения оптимальной последовательности состояний известен как алгоритм Витерби (Viterbi), который, как и алгоритм Forward, решает задачу итеративно. Прежде чем подробно рассматривать сам алгоритм, необходимо определить дополнительную переменную  $\delta$ . Эта переменная дает максимальное

значение вероятности частичной последовательности состояний и наблюдений до момента времени  $t$  при нахождении в состоянии  $q_i$  в момент времени  $t$ , т. е.

$$\delta_t(i) = \text{MAX}[P(s_1, s_2, \dots, s_t = q_i, o_1, o_2, \dots, o_t | \lambda)] \quad (4.12)$$

Это значение также можно получить итеративным способом:

$$\delta_{t+1}(i) = [\text{MAX}_j \delta_t(j) a_{ij}] b_i(o_{t+1}) \quad (4.13)$$

Алгоритм Витерби требует выполнения четырех этапов: инициализация, рекурсия, завершение и обратный проход (поиск с возвратом). Для алгоритма Витерби требуется дополнительная переменная  $\psi_t(i)$ , в которой для каждого состояния  $i$  в каждый момент времени  $t$  хранится предыдущее состояние, для которого была определена максимальная вероятность. Эта переменная используется для восстановления последовательности при обратном проходе после этапа завершения. С помощью алгоритма Витерби можно получить наиболее вероятную последовательность состояний, даже если они невидимы для скрытых марковских моделей.

Полная процедура показана в алгоритме 4.2.

*Алгоритм 4.2. Алгоритм Витерби*

**Требования:** Скрытая марковская модель  $\lambda$ ; Последовательность наблюдений  $O$ ; Количество состояний  $N$ ; Количество наблюдений  $T$

```

for  $i = 1$  to  $N$  do
    (Инициализация)
     $\delta_1(i) = \pi_i b_i(O_1)$ 
     $\psi_1(i) = 0$ 
end for
for  $t = 2$  to  $T$  do
    for  $j = 1$  to  $N$  do
        (Рекурсия)
         $\delta_t(j) = \text{MAX}_i [\delta_{t-1}(i) a_{ij}] b_j(O_t)$ 
         $\psi_t(j) = \text{ARGMAX}_i [\delta_{t-1}(i) a_{ij}]$ 
    end for
end for
(Завершение)
 $P^* = \text{MAX}_i [\delta_T(i)]$ 
 $q_T^* = \text{ARGMAX}_i [\delta_T(i)]$ 
for  $t = T$  to  $2$  do
    (Обратный проход)
     $q_{t-1}^* = \psi_t(q_t^*)$ 
end for

```



## Обучение

Рассмотрим, как можно обучить скрытую марковскую модель по имеющимся данным с использованием алгоритма Баума–Велша (Baum-Welch). Необходимое предварительное замечание: этот метод предполагает, что структура модели известна: предварительно определено количество состояний и наблюдений, следовательно, метод оценивает только параметры. Обычно наблюдения определяются предметной областью приложения, но количество состояний, которые скрыты, не так-то просто определить. Иногда количество скрытых состояний можно определить на основе знаний предметной области, иногда это делается экспериментально, методом проб и ошибок: тестируется производительность модели с различными количествами состояний (2, 3, ...), и выбирается то количество, при котором получены наилучшие результаты. Следует отметить, что при таком выборе неизбежен определенный компромисс, так как увеличение количества состояний дает более точные результаты, но увеличивает сложность вычислений.

Алгоритм Баума–Велша определяет параметры скрытой марковской модели  $\lambda = A, B, \Pi$  с учетом количества последовательностей наблюдений  $O = O_1, O_2, \dots, O_K$ . Для этого алгоритм максимизирует вероятность модели с учетом наблюдений:  $P(O|\lambda)$ . Для скрытой марковской модели с  $N$  состояниями и  $M$  наблюдениями необходимо оценить  $N + N^2 + N \times M$  параметров для  $\Pi, A$  и  $B$  соответственно.

Нужно определить еще одну вспомогательную переменную  $\xi$  – вероятность перехода из состояния  $i$  в момент времени  $t$  в состояние  $j$  в момент времени  $t + 1$  с учетом последовательности наблюдений  $O$ :

$$\xi_t(i, j) = P(s_t = q_i, s_{t+1} = q_j | O, \lambda) = P(s_t = q_i, s_{t+1} = q_j | \lambda) / P(O) \quad (4.14)$$

Эту формулу можно записать с использованием переменных  $\alpha$  и  $\beta$ :

$$\xi_t(i, j) = \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) / P(O)$$

Выражение  $P(O)$  также можно записать с использованием переменных  $\alpha$  и  $\beta$ :

$$\xi_t(i, j) = \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) / \sum_i \sum_j \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \quad (4.15)$$

Формулу для переменной  $\gamma$  еще можно записать с применением переменной  $\xi$ :  $\gamma_t(i) = \sum_j \xi_t(i, j)$ .

Добавляя  $\gamma_t(i)$  для всех моментов (интервалов) времени, т. е.  $\sum_t \gamma_t(i)$ , получаем оценку количества случаев, когда цепь находилась в состоянии  $i$ , а при накопительном суммировании  $\xi_t(i, j)$  по времени  $t$ , т. е.  $\sum_t \xi_t(i, j)$ , оценивается количество переходов из состояния  $i$  в состояние  $j$ .

Таким образом, процедура Баума-Велша для оценки параметров скрытых марковских моделей обобщена в алгоритме 4.3.

*Алгоритм 4.3. Алгоритм Баума–Велша*

1. Оценка априорных вероятностей – количество случаев нахождения в состоянии  $i$  в момент времени  $t$ .

$$\pi_i = \gamma_1(i)$$

2. Оценка вероятностей переходов – количество переходов из состояния  $i$  в состояние  $j$  по отношению к количеству случаев нахождения в состоянии  $i$ .

$$a_{ij} = \sum_t \xi_t(i, j) / \sum_t \gamma_t(j)$$

3. Оценка вероятностей наблюдений – количество случаев нахождения в состоянии  $j$  при количестве наблюдений  $k$  по отношению к количеству случаев нахождения в состоянии  $j$ .

$$b_{jk} = \sum_{t, O=k} \gamma_t(j) / \sum_t \gamma_t(j)$$

Следует отметить, что вычисление переменных  $\gamma$  и  $\xi$  выполняется с использованием переменных  $\alpha$  и  $\beta$ , для которых требуются параметры скрытой марковской модели  $\Pi, A, B$ . Здесь возникает проблема «курицы и яйца» – необходимы параметры модели для алгоритма Баума–Велша, который оценивает параметры этой модели. Решение данной проблемы основано на принципе ЕМ (expectation-maximization – ЕМ-алгоритм). Принцип (алгоритм) ЕМ состоит в том, чтобы начать с некоторых исходных параметров для модели (Е-этап), т. е.  $\lambda = \{A, B, \Pi\}$ , которые можно инициализировать случайным образом или на основе каких-либо знаний предметной области. Затем по алгоритму Баума–Велша эти параметры переоцениваются (М-этап). Цикл повторяется, пока не будет обеспечена сходимость, т. е. до тех пор, пока различия между параметрами модели на последней и предшествующей итерациях не станут меньше определенного порогового значения.

Напомню, что ЕМ-алгоритм используется в математической статистике для нахождения оценок максимального правдоподобия параметров вероятностных моделей, в случае, когда модель зависит от некоторых скрытых переменных. Каждая итерация алгоритма состоит из двух шагов. На Е-шаге (expectation) вычисляется ожидаемое значение функции правдоподобия, при этом скрытые переменные рассматриваются как наблюдаемые. На М-шаге (maximization) вычисляется оценка максимального правдоподобия, таким

образом увеличивается ожидаемое правдоподобие, вычисляемое на Е-шаге. Затем это значение используется для Е-шага на следующей итерации. Алгоритм выполняется до сходимости.

### Расширения

Несколько расширений стандартных скрытых марковских моделей были предложены для устранения конкретных проблем в отдельных приложениях. Кратко рассмотрим некоторые из этих расширений (рис. 4.3.).

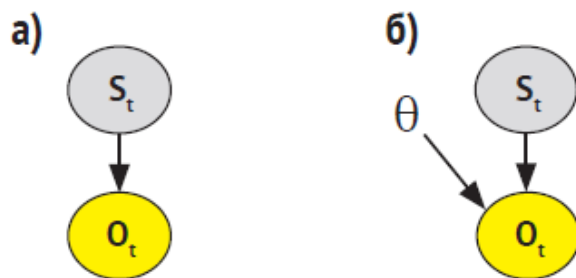


Рис. 4.3.1 Представление графовых моделей: а) стандартная (базовая) модель, б) параметрические скрытые марковские модели.

Параметрические скрытые марковские модели представляют предметные области, в которых предполагается использование нескольких вариаций скрытых марковских моделей. В параметрических скрытых марковских моделях переменные наблюдений обусловлены переменной состояния и одним или несколькими параметрами, значимыми для таких переменных (см. рис. 4.3.1.б). Значения параметров известны и постоянны на этапе тренировочного обучения. На этапе тестирования значения, которые максимизируют правдоподобие параметрических скрытых марковских моделей, восстанавливаются (корректируются) с помощью адаптированного ЕМ-алгоритма.

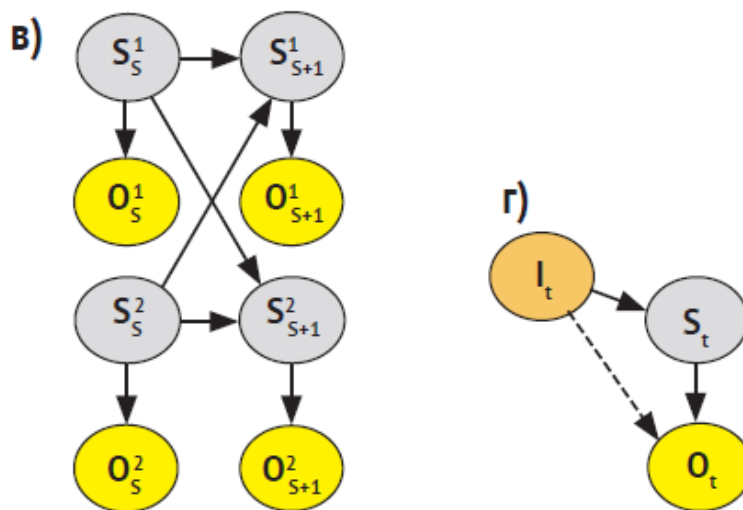


Рис. 4.3.2. Представление графовых моделей: в) связанные, г) ввода-вывода

Связные скрытые марковские модели объединяют скрытые марковские модели, вводя условные зависимости между переменными состояния (см. рис.

4.3.2.в). Эти модели вполне подходят для представления взаимовлияний между подпроцессами, которые протекают параллельно.

Скрытые марковские модели ввода-вывода рассматривают специальный добавочный входной параметр, который воздействует на состояния марковской цепи и дополнительно (но не обязательно) на переменные наблюдения. Эти типы моделей показаны на рис. 4.3.2.г. Входная переменная соответствует наблюдениям. Выходным сигналом скрытых марковских моделей ввода-вывода является класс моделей, которые должны выполняться. Отдельная модель может описывать полное множество классов.

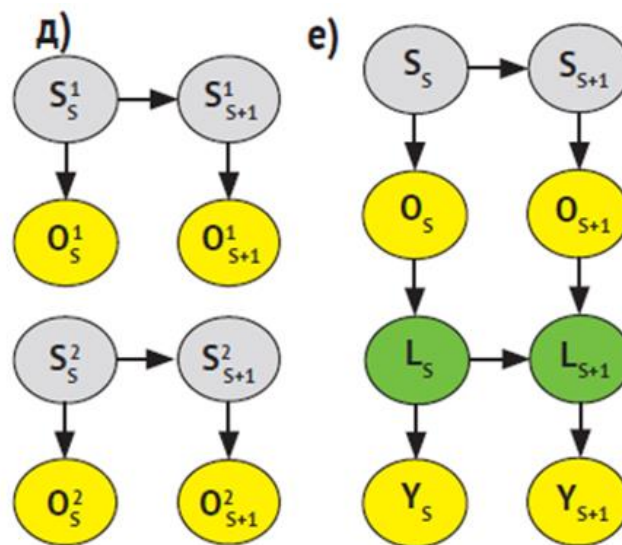


Рис. 4.3.3. Представление графовых моделей: д) параллельные, е) иерархические .

Параллельные скрытые марковские модели требуют меньше скрытых марковских моделей, чем связные для сложных составных процессов с предположением о взаимной независимости между скрытыми моделями (см. рис. 4.3.3.д). Основной принцип заключается в создании независимых скрытых марковских моделей для двух (и более) независимых параллельных процессов и объединении их с помощью умножения их отдельных собственных правдоподобий. Параллельные скрытые марковские модели с наиболее вероятным совместным правдоподобием определяют требуемый класс.

Иерархические скрытые марковские модели создают многослойную структуру скрытых марковских моделей на различных уровнях абстракции (см. рис. 4.3.3.е). В двухуровневых иерархических скрытых марковских моделях нижний уровень – это множество скрытых марковских моделей, которое представляет последовательность подмоделей. Верхний уровень – это марковская цепь, которая управляет динамикой этих подмоделей. Разделение на уровни позволяет повторно использовать стандартные скрытые марковские

модели, просто заменяя верхние уровни.

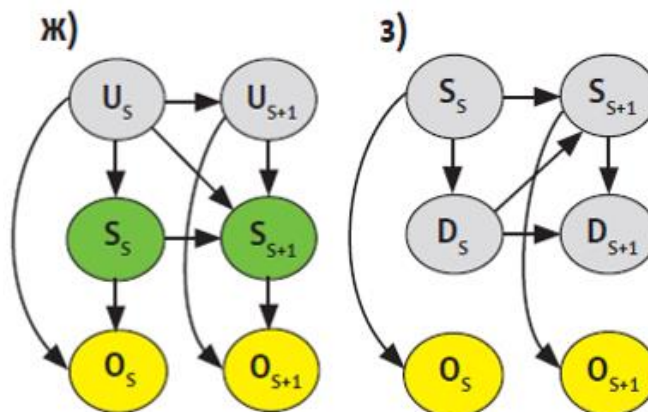


Рис. 4.3.4. Представление графовых моделей: ж) динамические байесовские сети со смешанным состоянием, з) скрытые полумарковские модели

Динамические байесовские сети со смешанным состоянием объединяют дискретные и непрерывные пространства состояний в двухуровневую структуру. Динамические байесовские сети со смешанным состоянием состоят из скрытой марковской модели на верхнем уровне и линейной динамической системы на нижнем уровне. Линейная динамическая система используется для моделирования переходов между действительно-значными (или вещественно-значными) состояниями. Выходные значения скрытой марковской модели управляют этой линейной системой. Графовое представление динамических байесовских сетей со смешанным состоянием показано на рис. 4.3.4ж. В динамических байесовских сетях со смешанным состоянием скрытые марковские модели могут описывать дискретные концепции высокого уровня, такие как грамматика, тогда как линейная динамическая система описывает входные сигналы в пространстве непрерывных состояний.

Скрытые полумарковские модели используют временные знания, относящиеся к текущему процессу, определяя явную продолжительность каждого состояния (см. рис. 4.3.4з). Скрытые полумарковские модели вполне пригодны для того, чтобы избежать экспоненциального убывания вероятностей наблюдений при моделировании больших последовательностей наблюдений.

### Алгоритм PageRank

Всемирную сеть WWW (World Wide Web) можно мысленно представить как весьма большую марковскую цепь, такую, что каждая веб-страница является состоянием, а гиперссылки между веб-страницами соответствуют переходам между этими состояниями. Предположим, что существует  $N$  веб-страниц. Любая конкретная веб-страница  $w_i$  имеет  $m$  исходящих гиперссылок. Если некоторый пользователь находится на веб-странице  $w_i$ , то может выбрать здесь любую

гиперссылку для перехода на другую веб-страницу. Обоснованное предположение: каждая исходящая гиперссылка может быть выбрана с равной вероятностью.

Таким образом, вероятность перехода с веб-страницы  $w_i$  на любую веб-страницу, на которую указывает одна из существующих гиперссылок,  $w_j$  равна  $A_{ij} = 1/m$ . Для других веб-страниц, на которые нет гиперссылок с текущей веб-страницы, вероятность перехода равна нулю. При таком подходе в соответствии со структурой WWW можно сформировать матрицу вероятностей переходов  $A$  для соответствующей марковской цепи. Диаграмма состояний для небольшого примера с тремя веб-страницами показана на рис. 4.4.

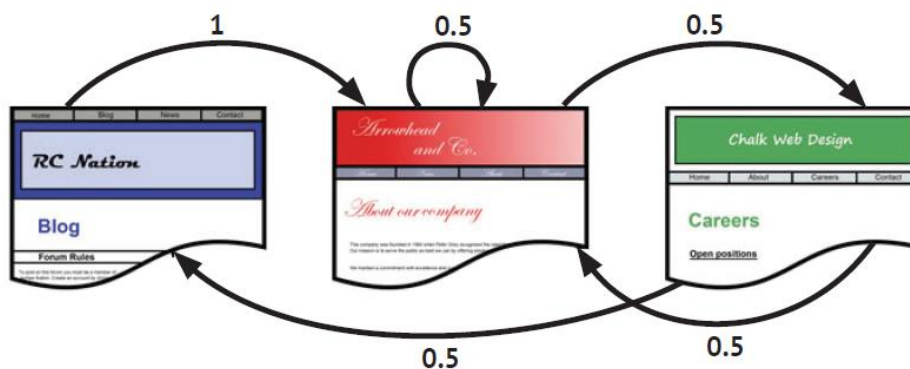


Рис. 4.4. Небольшой пример для WWW с тремя страницами

Сходимость вероятности для конкретной веб-страницы можно считать равнозначной вероятности посещения этой веб-страницы некоторым пользователем, произвольно перемещающимся по интернету. Теоретически можно предположить, что веб-страницы с бóльшим количеством входящих гиперссылок с веб-страниц, также содержащих большее количество входящих гиперссылок, будут иметь более высокую вероятность их посещения.

На основе изложенных выше принципов Л. Пэйдж (L. Page) и др. разработали алгоритм PageRank, представляющий собой основу для упорядочения (ранжирования) веб-страниц при выполнении поиска с помощью сервиса Google. Веб-страницы, извлекаемые с помощью алгоритма поиска, предъявляются пользователю в соответствии со сходимостью их вероятностей. Здесь основная идея состоит в следующем: более важные веб-страницы, как правило, обладают более высокой сходимостью их вероятности.