

ОБРАБОТКА ИНФОРМАЦИИ

УДК 519.283

Е.Н. Беккерман, С.С. Катаева

АЛГОРИТМ ОЦЕНИВАНИЯ ЧИСЛА СОСТОЯНИЙ И ЗНАЧЕНИЙ ИНТЕНСИВНОСТЕЙ МС-ПОТОКА СОБЫТИЙ

Приведен алгоритм оценивания числа состояний и интенсивностей состояний аппроксимирующего МС-потока исходя из выделенных в реализации потока участков постоянства интенсивности.

Ключевые слова: *МС-поток, интенсивность, интервал стационарности, оценка интенсивности, критерий согласия Пирсона, параметр состояния.*

МС-поток широко применяется в качестве модели входящего потока заявок в теории массового обслуживания. Это обусловлено тем, что модель МС-потока достаточно адекватно описывает входящий поток заявок в различных системах массового обслуживания.

Под МС-потоком будем понимать случайный поток событий, интенсивность которого представляет собой кусочно-постоянный марковский процесс $\lambda(t)$, принимающий значения из конечного множества констант $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ ($\lambda_i \neq \lambda_j$ при $i \neq j$), называемых состояниями процесса $\lambda(t)$. Длительности участков стационарности, т. е. участков, где $\lambda(t) = \text{const}$, распределены по экспоненциальному закону. Переход процесса из состояния i в состояние j определяется матрицей инфинитезимальных коэффициентов $\|\alpha_{ij}\|_{n \times n}$, где $\alpha_{ij} > 0$,

$i, j = \overline{1, n}$, $\alpha_{ii} = - \sum_{j=1, j \neq i}^n \alpha_{ij}$, $i = \overline{1, n}$. Величина α_{ij} , $i \neq j$, – интенсивность перехода

процесса $\lambda(t)$ из состояния i в состояние j . Введем в рассмотрение следующие величины: $\alpha_i = -\alpha_{ii}$, $i = \overline{1, n}$, $\alpha_i > 0$, – параметр экспоненциального распределения длительности пребывания процесса $\lambda(t)$ в состоянии i , $i = \overline{1, n}$;

$p_{ij} = \frac{\alpha_{ij}}{-\alpha_{ii}}$ $i \neq j$, $i, j = \overline{1, n}$, $0 \leq p_{ij} \leq 1$, $\sum_{j=1}^n p_{ij} = 0$ – вероятности перехода процесса

$\lambda(t)$ из состояния i в состояние j по окончании состояния i . Таким образом МС-поток событий с n состояниями полностью описывается следующими параметрами:

$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}, \quad \lambda_i > 0, \quad i = \overline{1, n}, \quad \lambda_i \neq \lambda_j, \quad i = \overline{1, n},$$

$$\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}, \quad \alpha_i > 0, \quad i = \overline{1, n},$$

$$P = \|p_{ij}\|_{n \times n}, \quad 0 \leq p_{ij} \leq 1, p_{ii} = 0, \quad \sum_{j=1}^n p_{ij} = 0, \quad i, j = \overline{1, n}.$$

На участке стационарности МС-поток ведет себя как простейший поток событий (пуассоновский поток с постоянной интенсивностью). Таким образом, реализация МС-потока состоит из отрезков простейших потоков; интенсивность потока на отрезке определяется состоянием МС-потока.

Для того чтобы использовать модель МС-потока при описании реальной системы массового обслуживания, необходимо определить характеристики МС-потока таким образом, чтобы модель хорошо аппроксимировала входящий поток событий этой системы. МС-поток, полученный в результате процедуры определения характеристик, будем называть аппроксимирующим МС-потоком. Характеристиками, определяющими МС-поток, являются количество состояний МС-потока, интенсивности состояний, а также интенсивности и вероятности переходов из состояния в состояние. Зачастую определить характеристики возможно лишь на основании информации о моментах наступления событий реального потока на некотором интервале времени.

1. Общие принципы оценивания числа состояний и значений интенсивности

Если известны только моменты наступления событий в реализации потока, то первоначально необходимо разбить реализацию на отрезки, которые бы соответствовали участкам стационарности. В работе [1] представлен алгоритм выделения участков постоянства интенсивности в реализации случайного потока событий и оценки интенсивности потока на этих участках.

По выделенным из реализации случайного потока событий интервалам постоянства интенсивности оценим количество состояний аппроксимирующего МС-потока. Прежде чем оценивать количество состояний аппроксимирующего МС-потока, приведем несколько соображений.

Для расчета характеристик СМО важно, чтобы модель входящего потока, с одной стороны, достаточно точно описывала реальный поток, а с другой – имела не слишком большое количество параметров. В случае МС-потока, количество параметров модели определяется количеством состояний процесса $\lambda(t)$.

Вычислив оценки интенсивности на выделенных участках стационарности, можно определить количество состояний потока как количество различных по значению оценок. Однако понятно, что даже оценки интенсивности простейшего потока, полученные по разному количеству событий или на различных временных интервалах, не будут равны между собой. Таким образом, маловероятно получить на двух различных интервалах постоянства интенсивности равные значения оценок интенсивности потока.

Разумно будет исходить из того, что оценки интенсивности на разных интервалах стационарности, относящихся к одному и тому же состоянию МС-потока, различаются, так как являются случайными величинами, но, тем не менее, близки по значению. То есть участки стационарности реального потока с «близкими» интенсивностями относятся к одному и тому же состоянию. Определив критерий "близости" значений интенсивности на различных участках стационарности, можно будет разбить множество оценок на группы, в которые войдут оценки интенсивности интервалов стационарности, относящиеся к одному и тому же со-

стоянию аппроксимирующего МС-потока. Количество получившихся групп и будет определять количество состояний аппроксимирующего МС-потока.

Пусть с помощью алгоритма выделения интервалов стационарности МС-потока по реализации моментов наступления событий выделено некоторое число k интервалов постоянства интенсивности вида

$$\begin{aligned} [t_{i_1}, t_{j_1}], [t_{i_2}, t_{j_2}], [t_{i_3}, t_{j_3}], \dots, [t_{i_k}, t_{j_k}], \\ t_{i_l} < t_{j_l}, l = 1, k, \\ j_l < i_{l+1}, l = 1, k-1, \end{aligned}$$

где $t_{i_l}, t_{j_l}, l = \overline{1, k}$, есть определенные моменты наступления событий из реализации. Таким образом, в рассматриваемой реализации выделено k интервалов стационарности. Моменты времени $t_{i_l}, t_{j_l}, l = \overline{1, k}$, являются соответственно временами наступления первого и последнего событий на l -м интервале стационарности. Для каждого интервала стационарности найдем оценку интенсивности потока на данном интервале по формуле

$$\hat{\lambda}^{(l)} = \frac{j_l - i_l + 1}{t_{j_l} - t_{i_l}}, \quad l = \overline{1, k}.$$

Заметим, что такие оценки интенсивности есть оценки интенсивности простейшего потока, полученные по методу максимального правдоподобия, и они обладают всеми полезными свойствами таких оценок.

Далее, полученные оценки интенсивности на выделенных интервалах стационарности нужно разбить на классы, содержащие оценки интенсивности интервалов стационарности, относящихся к одному и тому же состоянию аппроксимирующего МС-потока. Очевидно, что для того, чтобы относиться к одному классу, оценки должны быть близкими по значению, то есть группироваться на числовой оси. Между группами же должно быть значительное расстояние. Если перенумеровать полученные оценки таким образом, чтобы они оказались упорядоченными по возрастанию, то очевидно, что процедура деления этого ряда на группы должна основываться на расстояниях между соседними значениями оценок ряда. В качестве критерия близости при группировке оценок интенсивности примем расстояние между соседними оценками в упорядоченной по возрастанию совокупности оценок интенсивностей выделенных интервалов стационарности.

Внутри же группы, которая включает оценки интервалов, относящихся к одному и тому же состоянию аппроксимирующего МС-потока, оценки должны быть в некотором смысле однородны. Вряд ли возможна единственно верная формулировка критерия однородности оценок в группе. Однако представляется естественным в качестве критерия однородности взять соотношение между размахом значений в группе (разницей между минимальным и максимальным значением в группе) и средним значением оценок в группе, то есть как сильно крайние значения отличаются от среднего.

Пусть $L_{ij} = \{\hat{\lambda}^{(i)}, \hat{\lambda}^{(i+1)}, \dots, \hat{\lambda}^{(j)}\}$ – группа оценок интенсивностей, следующих по порядку. Назовем величину

$$c(L_{ij}) = (\hat{\lambda}^{(j)} - \hat{\lambda}^{(i)}) / \left(\frac{1}{j-i+1} \sum_{l=i}^j \hat{\lambda}^{(l)} \right)$$

показателем однородности группы L_{ij} . Здесь числителем является размах значений в группе, знаменателем – среднее значений в группе. Будем говорить, что группа оценок однородна, если $c(L_{ij}) \leq c_0$, и неоднородна в противном случае, где c_0 – пороговое значение показателя однородности группы, некоторая наперед заданная положительная величина, выбираемая исходя из того, что разброс значений оценок интенсивности в однородной группе сопоставим с величиной среднего элементов группы. Заметим при этом, что группа, состоящая из одного значения, однородна.

Если группа оценок интенсивностей однородна, то группа соответствует состоянию интенсивности предполагаемого МС-потока событий, при этом оценка интенсивности, соответствующая данному состоянию, рассчитывается по формуле

$$\hat{\lambda} = \hat{\lambda}(L_{ij}) = \sum_{l=i}^j \frac{n_l}{\sum_{s=i}^j n_s} \hat{\lambda}^{(l)},$$

где $\hat{\lambda}^{(l)}$ – l -я оценка интенсивности в группе, n_s – количество событий на интервале стационарности, соответствующем s -й оценке интенсивности в группе. То есть оценка интенсивности состояния есть средневзвешенная по числу событий оценок интенсивности на участках постоянства интенсивности, отнесенных к этому состоянию.

Группы оценок из всей совокупности предлагается выделять следующим образом. Для начала нужно проверить, не является ли вся совокупность оценок L_{1k} однородной. В случае, если совокупность L_{1k} не является однородной, разбить ее на две части: L_{1j} и L_{j+1k} , причем так, чтобы граница между двумя группами пролегла в том месте набора значений оценок интенсивности, где разница между двумя соседними значениями максимальна. То есть граница между частями определяется расстоянием между соседними значениями упорядоченного ряда интенсивностей выделенных интервалов стационарности. Каждую из групп вновь проверяем на однородность и, если они не однородны, снова разбиваем на части. Процесс заканчивается, когда все выделенные группы будут однородны, то есть для всех выделенных групп показатель однородности будет меньше единицы. Однородные группы и будут содержать оценки интенсивности на участках стационарности, которые можно отнести к одному и тому же состоянию аппроксимирующего МС-потока.

Число групп оценок, полученное в результате классификации, будет являться оценкой числа состояний интенсивности аппроксимирующего МС-потока событий. В качестве оценки интенсивности состояния МС-потока естественно будет взять величину, равную взвешенному среднему оценок внутри класса, соответствующего данному состоянию.

2. Алгоритм оценивания числа состояний и значений интенсивности МС-потока событий

С учетом приведенных выше рассуждений, сформулируем алгоритм группировки оценок интенсивностей.

Шаг 1. Задаются величины, $\hat{n} = 0$, $i = 1$, $j = k$, где $\hat{n}$ – оценка количества состояний МС-потока, k – количество интервалов постоянства интенсивности, выделенных в реализации исследуемого потока событий $L = \{\hat{\lambda}^{(1)}, \dots, \hat{\lambda}^{(k)}\}$.

Шаг 2. Вычислить показатель однородности группы оценок L_{ij} по формуле

$$c(L_{ij}) = \frac{\hat{\lambda}^{(j)} - \hat{\lambda}^{(i)}}{\frac{1}{j-i+1} \sum_{l=i}^j \hat{\lambda}^{(l)}}.$$

Шаг 3. Если $L_{ij} < 1$, то

- 3.1. Оценку количества состояний аппроксимирующего МС-потока $\hat{n}$ увеличить на 1;
- 3.2. Оценку интенсивности потока в состоянии, определенном группой оценок L_{ij} , вычислить по формуле

$$\hat{\lambda}(L_{ij}) = \sum_{l=i}^j \frac{n_l}{\sum_{s=i}^j n_s} \hat{\lambda}^{(l)};$$

в противном случае:

- 3.3. Найти такое l^* , что $l^* = \arg \max_{l=i, j-1} (\hat{\lambda}^{(l+1)} - \hat{\lambda}^{(l)})$;
- 3.4. Положить $i = i, j = l^*$ и для группы L_{ij} проделать шаги 2 и 3;
- 3.5. Положить $i = l^* + 1, j = j$ и для группы L_{ij} проделать шаги 2 и 3.

Шаги 2 и 3 алгоритма представляют собой рекурсивную процедуру (пункты 3.3, 3.4, 3.5 шага 3 производят повторение проверки однородности группы с новыми данными, пункты 3.1, 3.2 определяют условие останова рекурсии). Результатом работы алгоритма являются:

- величина $\hat{n}$ – оценка числа состояний;
- значения $\hat{\lambda}_1, \dots, \hat{\lambda}_{\hat{n}}$ – значения оценок интенсивности состояний аппроксимирующего МС-потока событий.

3. Алгоритм оценивания значений интенсивности МС-потока событий при ограничении числа состояний

В некоторых случаях количество состояний входящего потока заявок заранее ограничено условиями задачи. Зачастую количество состояний должно быть не более некоторого наперед заданного числа. В таких случаях также можно руководствоваться приведенными выше соображениями с той лишь разницей, что процедура деления ряда оценок на группы заканчивается по достижении заранее известного количества групп (состояний потока) или когда все группы признаны однородными.

Шаг 1. Задать величины, $\hat{n} = 1, i = 1, j = k, n^*$, где $\hat{n}$ – оценка количества состояний МС-потока, k – количество интервалов постоянства интенсивности, выделенных в реализации исследуемого потока событий $L = \{\hat{\lambda}^{(1)}, \dots, \hat{\lambda}^{(k)}\}$, n^* – максимально возможное по условию задачи количество состояний аппроксимирующего МС-потока. Сформировать список групп оценок $\tilde{L} = \{L_{1k}\}$.

Шаг 2. Если $\hat{n} = n^*$, то перейти к шагу 7, в противном случае – к шагу 3.

Шаг 3. Найти такое $L_{i'j'}$ из $\tilde{L}$, что $c(L_{i'j'}) = \max_{\tilde{L}}(c(L_{ij}))$.

Шаг 4. Если $c(L_{i'j'}) \leq 1$, то перейти к шагу 7, в противном случае – к шагу 5.

Шаг 5. Найти такое l' , что $l' = \arg \max_{l=i', j'-1}(\hat{\lambda}^{(l+1)} - \hat{\lambda}^{(l)})$.

Шаг 6. Удалить из списка $\tilde{L}$ группу $L_{i'j'}$ и добавить группы $L_{i'l'}$ и $L_{l'+1j'}$, оценку количества состояний $\hat{n}$ увеличить на единицу и перейти к шагу 2.

Шаг 7. Принять в качестве оценки количества состояний аппроксимирующего МС-потока значение $\hat{n}$. По группам из списка $\tilde{L} = \{L_{i_1j_1}, \dots, L_{i_nj_n}\}$ вычислить значения интенсивности для каждого состояний по формуле

$$\hat{\lambda}_k = \hat{\lambda}(L_{i_kj_k}) = \sum_{l=i_k}^{j_k} \frac{n_l}{\sum_{s=i_k}^{j_k} n_s} \hat{\lambda}^{(l)},$$

где $k = \overline{1, \hat{n}}$, $\hat{\lambda}^{(l)}$ – l -я оценка интенсивности в группе, n_s – количество событий на интервале стационарности, соответствующем s -й оценке интенсивности в группе.

4. Пример работы алгоритма оценивания числа состояний и значений интенсивности МС-потока событий на имитационной модели МС-потока

Для иллюстрации работы алгоритма приведем пример.

В качестве исходных данных для имитационной модели МС-потока событий были заданы следующие значения параметров:

$$n = 3, \quad \lambda = \{20, 100, 200\}, \quad \alpha = \{0, 5, 0, 5, 0, 5\},$$

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

В результате моделирования была получена реализации МС-потока с числом событий 1000. Первое и последнее события и интенсивность потока на каждом из реализовавшихся интервалов стационарности представлены в следующей таблице:

k	i_k	j_k	λ_k
1	1	54	20
2	55	107	100
3	108	109	20
4	110	676	200
5	677	716	20
6	717	871	100
7	872	979	200
8	980	1000	20

В результате применения алгоритма выделения участков постоянства интенсивности получено следующее разбиение реализации на интервалы стационарности:

k	i_k	j_k	$\hat{\lambda}_k$
1	1	54	20,37306
2	55	106	104,29035
3	107	674	181,50631
4	675	711	19,25258
5	712	858	87,82144
6	859	978	192,47803
7	979	999	22,66016

Упорядочим полученные оценки по возрастанию: $\hat{\lambda}^{(1)} = 19,25258$, $\hat{\lambda}^{(2)} = 20,37306$, $\hat{\lambda}^{(3)} = 22,66016$, $\hat{\lambda}^{(4)} = 87,821$, $\hat{\lambda}^{(5)} = 104,29035$, $\hat{\lambda}^{(6)} = 181,50631$, $\hat{\lambda}^{(7)} = 192,47803$.

На первом шаге алгоритма зададим начальные значения величин: $\hat{n} = 0$, $i = 1$, $j = 7$, $c_0 = 1$, $L_{1,7} = \{\hat{\lambda}^{(1)} = 19,25258$, $\hat{\lambda}^{(2)} = 20,37306$, $\hat{\lambda}^{(3)} = 22,66016$, $\hat{\lambda}^{(4)} = 87,821$, $\hat{\lambda}^{(5)} = 104,29035$, $\hat{\lambda}^{(6)} = 181,50631$, $\hat{\lambda}^{(7)} = 192,47803\}$.

Вычислим показатель однородности группы оценок $L_{1,7}$ по формуле

$$c(L_{ij}) = \frac{\hat{\lambda}^{(j)} - \hat{\lambda}^{(i)}}{\frac{1}{j-i+1} \sum_{l=i}^j \hat{\lambda}^{(l)}}.$$

$c(L_{1,7}) = 1,929683$. Это больше заданного $c_0 = 1$, поэтому переходим к шагу 3.3.

Найдем $l^* = \arg \max_{l=1,6} (\hat{\lambda}^{(l+1)} - \hat{\lambda}^{(l)})$. $\hat{\lambda}^{(2)} - \hat{\lambda}^{(1)} = 1,12048$, $\hat{\lambda}^{(3)} - \hat{\lambda}^{(2)} = 2,2871$, $\hat{\lambda}^{(4)} - \hat{\lambda}^{(3)} = 65,16128$, $\hat{\lambda}^{(5)} - \hat{\lambda}^{(4)} = 16,46891$, $\hat{\lambda}^{(6)} - \hat{\lambda}^{(5)} = 77,21596$, $\hat{\lambda}^{(7)} - \hat{\lambda}^{(6)} = 10,97172$. Таким образом $l^* = 5$ и группа $L_{1,7}$ распадается на две группы: $L_{1,5}$ и $L_{6,7}$. Проверим получившиеся группы на однородность: $c(L_{1,5}) = 1,671356$, $c(L_{6,7}) = 0,058674$.

Группа $L_{6,7}$, согласно выбранному c_0 , однородна. Согласно шагам 3.1 и 3.2, необходимо увеличить счетчик количества состояний $\hat{n} = 1$ и вычислить оценку интенсивности состояния

$$\hat{\lambda}(L_{6,7}) = \sum_{l=6}^7 \frac{n_l}{\sum_{s=6}^7 n_s} \hat{\lambda}^{(l)}, \quad \hat{\lambda}(L_{6,7}) = 183,42.$$

Для группы $L_{1,5}$ снова проделаем шаг 3, для чего необходимо найти $l^* = \arg \max_{l=1,4} (\hat{\lambda}^{(l+1)} - \hat{\lambda}^{(l)})$. Величина $l^* = 3$, группа $L_{1,5}$ распадается на две группы $L_{1,3}$ и $L_{4,5}$, и обе они однородны: $c(L_{1,3}) = 0,16412$, $c(L_{4,5}) = 0,17145$. Прделав шаг 3.1 для обеих однородных групп, получим $\hat{n} = 3$, $\hat{\lambda}(L_{1,3}) = 20,431$, $\hat{\lambda}(L_{4,5}) = 92,124$.

Таким образом, по окончании работы алгоритма выделено три состояния МС-потока, интенсивности состояний: $\hat{\lambda}_1=20,431$, $\hat{\lambda}_2=92,124$, $\hat{\lambda}_3=183,42$.

Сопоставляя результаты работы алгоритмов выделения интервалов стационарности и оценивания числа состояний и значений интенсивности потока событий с исходными данными и результатами имитационного моделирования, отметим, что

- оценка числа состояний $\hat{n}$ совпадает с истинным значением $n=3$ числа состояний МС-потока;

- среднее относительное отклонение оценок $\hat{\lambda}_i$ значений интенсивности от ближайших к ним истинных значений λ_i интенсивностей, рассчитанное по формуле

$\Delta = \frac{1}{n} \sum_{i=1}^n \frac{|\lambda_i - \hat{\lambda}_i|}{\lambda_i}$, составляет 0,061 (то есть оценка отличается от истинного значения величины интенсивности состояния в среднем на 6,1% от величины истинного значения).

Таким образом, результаты работы алгоритма можно признать удовлетворительными.

Заключение

В работе предложен алгоритм оценивания числа состояний и значений интенсивностей МС-потока событий по выделенным в реализации потока интервалам постоянства интенсивности. Приведен пример работы алгоритма с использованием имитационной модели МС-потока, что позволило сравнить результаты работы алгоритма с исходными данными и результатами работы имитационной модели. Кроме того, предложена процедура оценивания интенсивностей состояний МС-потока в тех случаях, когда количество состояний должно быть не более некоторого наперед заданного числа.

ЛИТЕРАТУРА

1. Беккерман Е.Н., Катаев С.Г., Катаева С.С. Об одном алгоритме распознавания МС-потока событий // Вестник ТГУ. 2001. № 271. С. 41 – 45.

Беккерман Екатерина Николаевна

Катаева София Семеновна

E-mail: bekkermankn@tspu.edu.ru; kataeva@fpmk.tsu.ru

Томский государственный университет

Поступила в редакцию 12 марта 2009 г.