

УДК 519.95

## **МЕТОДИКА СОЗДАНИЯ КРИТЕРИЕВ РЕЛЕВАНТНОСТИ ИНФОРМАЦИИ НА ОСНОВЕ ФУНКЦИОНИРОВАНИЯ ПЕРЕХОДОВ СЕТЕЙ ПЕТРИ-МАРКОВА**

Г.Н. Никитина

*Предложен вариант решения задачи информационного поиска в массиве данных с использованием методики формирования базы критериев оценки релевантной информации.*

*Ключевые слова: сети Петри-Маркова, поисковая система, динамика поискового процесса, процесс выполнения шагов-переходов.*

Вследствие быстрого развития средств вычислительной техники и передачи данных, проектирование систем распределенной информации становится областью приложения усилий все большего числа не только научных, но и инженерных работников. Однако недостаточно полное отражение в литературе получили следующие важные направления проектирования систем: исследование переходных режимов функционирования; методы расчета не только моментов случайных величин (математического ожидания, дисперсии и т.д.), но и оценок их распределения, квантилей и других характеристик; инженерные подходы к определению характеристик времени достижения существенных различных состояний системы [1].

Сети Петри относятся к числу наиболее важных и распространенных математических моделей в области обработки информации и применяются для решения задач синтеза и оптимизации [2], целесообразно предложить их использование для определения оптимальных критериев оценки релевантности информационного поиска.

Анализ работы переходов сетей Петри-Маркова с целью их применения в разработке системы, критериев для оценки релевантности информации.

В процессе работы информационной системы создается определенный набор критериев оценки релевантности информационного поиска. Для наиболее точного поиска следует определить заданные характеристики объекта поиска, который обладает множеством свойств, и каждое из них характеризует объект с определенной долей вероятности. Данная вероятность зависит от набора параметров объекта поиска. Таким образом, следует подойти к задаче введения критериев наиболее точно описывающих объект поиска, что в свою очередь приведет к сокращению поискового процесса и получению наиболее точного и что не менее важно достоверного результата.

Каждая информационная система обладает собственными критериями, позволяющими осуществлять релевантный поиск. Для решения поставленной задачи можно использовать следующие критерии, которые позволяют осуществлять поиск информации в информационной системе.

Первоначально следует определить временной критерий циклических операций производимых в процессе поиска. Первоначальные условия, заданные поисковому процессу, могут отличаться, временной показатель выполнения задания также является различным. Отсюда следует, что данный временной критерий является величиной случайной.

Плотность распределения времени затрачиваемого на выполнение циклов поиска, запишется в виде

$$f(t) = \sum_{i=1}^n \delta(t - i\tau - T_{\min}), \quad (1)$$

где  $t$  - время;  $\tau$  - период следования циклов поиска;  $T_{\min}$  - минимальное время выполнения запроса.

Характеристики этой случайной величины определяются целым рядом вышеупомянутых факторов.

Время, затрачиваемое программой на весь процесс от нахождения информации до выдачи результата, может быть определено как взвешенная сумма вырожденных законов распределения:

$$f_{\bar{h}}(t) = \sum_{i=1}^I p_i \delta(t - u m_i), \quad (2)$$

где  $i$  - индексы для обозначения вариантов выполнения поиска;  $I$  - количество вариантов;  $p_i$  - вероятности появления соответствующего варианта;

$$0 < m_{\min} < m_i < m_{\max};$$

$m_i$  - количество циклов поиска при  $i$ -м варианте выполнения (целое число).

Выполнение каждого цикла поиска может быть охарактеризовано следующими параметрами: минимальным временем выполнения

$$T_{k \min} = u \min_i \{m_i\} = u m_{\min};$$

максимальным временем выполнения

$$T_{k \max} = u \max_i \{m_i\} = u m_{\max};$$

средним временем выполнения и дисперсией соответственно

$$T_k = u \sum_{i=1}^I p_i m_i; \quad (3)$$

$$D_k = u^2 \sum_{i=1}^I p_i m_i^2 - (T_k)^2. \quad (4)$$

Здесь и ниже под средним временем выполнения понимается математическое ожидание времени, вычисляемое для закона распределения  $f(t)$  по общеизвестной зависимости:

$$T = \int_0^{\infty} t f(t) dt. \quad (5)$$

Дисперсия времени для закона распределения  $f(t)$  также вычисляется по известной зависимости:

$$D = \int_0^{\infty} (t - T)^2 f(t) dt. \quad (6)$$

Данные вычисления показывают, что для вырожденного закона распределения  $\delta(t - u)$  математическое ожидание равно  $u$ . Дисперсия вырожденного закона равна нулю по определению.

Одним из существенных моментов любого вычислительного процесса является выбор одного из вариантов его продолжения в местах ветвления. Как правило, выбор вариантов осуществляется на основании обрабатываемых данных.

Данные, обрабатываемые поисковой системой, являются случайными величинами. Следовательно, переходы, которые осуществляются на основании их анализа, имеют случайный характер.

Вероятности выбора  $i$ -го продолжения вычислительного процесса определяются законом распределения данных, анализируемых при принятии решений и условиями перехода

$$p_i = \int_{X(i)} \varphi(x) dx, \quad (7)$$

где  $\varphi(x)$  - закон распределения обрабатываемых данных  $x$ ;  $X(i)$  - область действия  $i$ -х условий перехода,

$$1 \leq i \leq J(x);$$

$J(x)$  - количество условий перехода.

Таким образом, функционирование информационной системы можно представить в виде процесса выполнения последовательности циклических операций, производимых в поисковой программе. При этом характер переходов в местах передачи управления операторами принятия решения будет носить вероятностный характер.

Для определения эффективности запроса к системе данных, необходимо определить коэффициент эффективности запроса. Данный коэффициент характеризует то, насколько запрос пользователя будет эффективен по отношению к набору данных, над которыми должен производиться поиск.

Численно данный коэффициент определяется как сумма средних частот термов запроса, характеризующихся терминами коллекции. Под терминами следует понимать набор ключевых слов, выбранных из документов, выборка которых соответствует тематике документа. Для решения данной задачи необходимо определить средние частоты термов, характеризующие коллекцию. На основе тематических таблиц документов определяется среднее значение частоты термина в тематической коллекции:

$$\tilde{\beta}_{\zeta_D^T, t_j} = \frac{\sum_{D_i \in D} \beta_{j, D_i}}{\sum_{D_i \in D} |t_j \cap \zeta_{D_i}^T|}, \quad (8)$$

где  $\beta_{j, D_i}$  - частота вхождения термина  $t_j$  в документ  $D_i$ ;  $\zeta_{t_j}$  - терм, характеризующий запрос;  $\zeta_{D_i}^T$  - множество термов, характеризующих документы коллекции.

Таким образом, критерий оценки эффективности запроса определяется суммой средних частот термов коллекции.

$$\beta_{Q(T)} = \sum_{t_j \in X_{Q(T)}} \tilde{\beta}_{\zeta_D^T, t_j}. \quad (9)$$

В зависимости от полученной суммы, поиск начинается с той коллекции, которая характеризуется максимальным значением.

Поиск релевантной информации можно представить в виде ориентированного графа сеть Петри-Маркова. В соответствии с методологией анализа полумарковских процессов [3, 4], одним из способов оценки временных интервалов и стохастических характеристик в полумарковском процессе является выделение траекторий с заранее заданными свойствами и определение вероятности и плотности распределения времени движения по выделенным траекториям. В контексте рассматриваемой задачи и является вероятностными и временными характеристиками поискового процесса.

Пусть  $z_{j(z)}$ -я траектория, которая ведет из перехода  $z_j(z)$ , соответствующего первоначальному состоянию, в переход  $i$ , соответствующий результирующему состоянию поискового процесса состоящий из полушагов  $s_{1[j(w)]}, \dots, s_{j[j(w)]}, \dots, s_{J[j(w)]}$ . Количество таких траекторий равно  $J(w)$ , вероятность и плотность распределения времени выполнения соответствующего полушага определены как  $p_{j[j(w)]}$  и  $f_{j[j(w)]}(t)$ . Тогда вероятность и плотность распределения времени перемещения из перехода  $z_i(z)$  в переход  $z_j(z)$  по  $j(w)$ -й равны, соответственно

$$p_{j(w)} = \prod_{j[j(w)=1]}^{J[j(w)]} p_{j[j(w)]}; \quad (10).$$

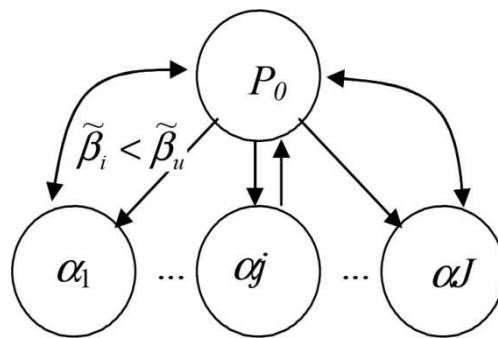
Вероятность и плотность достижения перехода  $z_j(z)$  из перехода  $z_i(z)$  равны, соответственно

$$p_{i(z),j(z)} = \prod_{j(w)=1}^{J(w)} p_{j(w)}; \quad (11)$$

$$f_{i(z),j(z)}(t) = \frac{\prod_{j(w)=1}^{J(w)} p_{j(w)} f_{j(w)}(t)}{\prod_{j(w)=1}^{J(w)} p_{j(w)}}. \quad (12)$$

Динамика поведения поискового процесса зависит от первоначально заданных условий  $P_0$ . Эти срабатывания соответствуют реализациям запросов, вследствие чего, поиск переходит в следующую стадию, завершения либо изменения параметров с последующим переходом в новую стадию поиска.

В зависимости от выбранных свойств объекта работа сети представляет собой совокупность локальных действий, приводимых к срабатыванию определенных переходов. На рис. 1 представлена работа поисковой программы, реализующая множество полушагов из  $a_{1(a)}, \dots, a_{j(a)}, \dots, a_{J(a)}$  переходов поиска.



**Рис. 1. Работа информационной системы. Начало поиска**

В том случае если полушаги обладают равнозначными коэффициентами значимости, то плотность распределения времени от запуска поискового процесса до полушага имеет вид:

$$f_{1,0(a)}(t) = \sum_{j(a)=1(a)}^{J(a)} f_{j(a),1}(t) \prod_{\substack{k(a)=1, \\ k(a) \neq j(a)}}^{J(a)} \left[ 1 - \int_0^t (\tau) d\tau \right]. \quad (13)$$

Слагаемые  $f_{1,0(a)}(t)$ ,  $\prod_{\substack{k(a)=1, \\ k(a) \neq j(a)}}^{J(a)} \left[ 1 - \int_{\tau=0}^t f_{k(a),1}(\tau) d\tau \right]$  представляют собой

взвешенные плотности распределения времени выполнения  $j(a)$ -го полушага из  $J(a)$ -возможных.

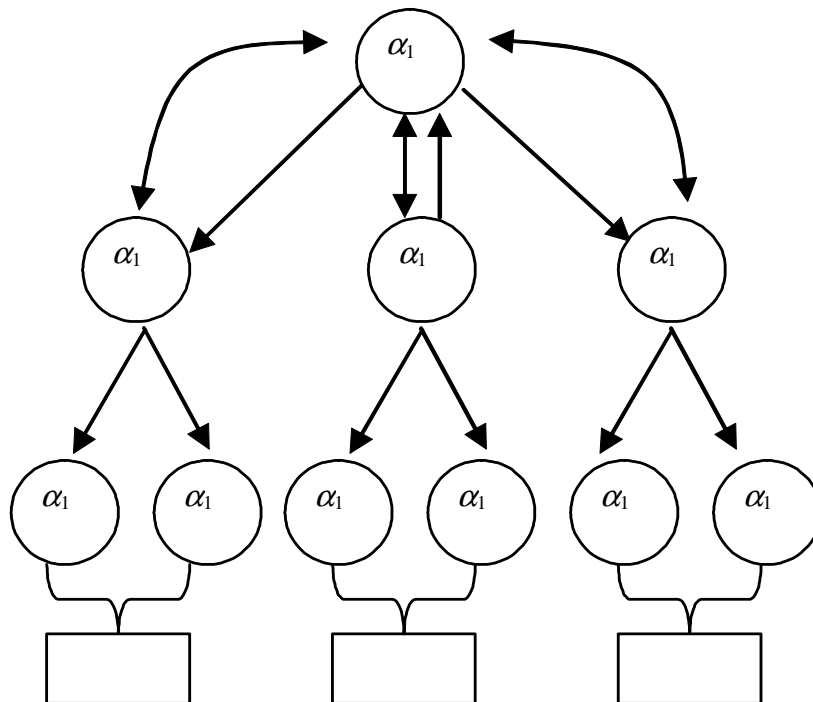
Вероятность того, что первым будет сделан  $J(a)$  полушаг

$$p_{j(a),1} = \int_0^{\infty} f_{j(a),1}(t) \prod_{\substack{k(a)=1, \\ k(a) \neq j(a)}}^{J(a)} \left[ 1 - \int_{\tau=0}^{\infty} f_{k(a),1}(\tau) d\tau \right] dt. \quad (14)$$

Плотность распределения времени выполнения  $j(a)$ -го полушага первым

$$\varphi_{j(a),1} = \frac{f_{j(a),1}(t)}{p_{j(a),1}} \prod_{\substack{k(a)=1, \\ k(a) \neq j(a)}}^{J(a)} \left[ 1 - \int_{\tau=0}^{\infty} f_{k(a),1}(\tau) d\tau \right]. \quad (15)$$

Следовательно, процесс выполнения условий шагов-переходов в сети Петри будет иметь вид:



**Рис. 2. Работа информационной системы.  
Срабатывание переходов**

Таким образом, при создании базы критериев оценки релевантности информации, целесообразно воспользоваться следующей методикой.

1. Построить сеть Петри, соответствующую структуре информационной системы.
2. Определить временной интервал поискового пути.
3. Определить коэффициент эффективности запроса.
4. Определить временные и вероятностные характеристики поискового процесса.
5. Определяются логические условия выполнения всех шагов-переходов в процессе работы системы.

### Список литературы

1. Котов В.Е. Сети Петри. М.: Наука. Главная редакция физико-математической литературы, 1984. 158 с.
2. Кочегаров В.А., Фролов Г.А. Проектирование систем распределения информации. Марковские и немарковские модели. М.: Радио и связь, 1991. 216 с.
3. Никитина Г.Н. Анализ сетей Петри-Маркова в концепции работы информационной системы. // Известия Тульского государственного университета. Серия: Вычислительная техника. Информационные технологии. Системы управления. Технические науки. Вып. 5. Ч. 1. Изд-во ТулГУ, 2011. С. 217 - 221.
4. Никитина Г.Н. Анализ временных характеристик алгоритмов функционирования ЭВМ на основе полумарковских процессов // Известия Тульского государственного университета. Вычислительная техника. Информационные технологии. Системы управления. Технические науки. Вып. 2. Ч. 2. Тула: Изд-во ТулГУ, 2010. С. 162 - 166.

*Никитина Галина Николаевна, инженер [galinanikki@mail.ru](mailto:galinanikki@mail.ru), Россия, Тула, Тульский государственный университет*

### *METHODS OF ESTABLISHING CRITERIA OF RELEVANCE OF INFORMATION ON THE BASIS OF THE FUNCTIONING OF THE TRANSITIONS PETRI NETS AND MARKOV*

*G.N. Nikitina*

*The proposed solution to the problem of information search in the data array, using the method of forming the base evaluation criteria relevant information.*

*Key words: Petri nets and Markov, the search system, the dynamics of the search process, the process execution steps-transitions.*

*Nikitina Galina Nikolaevna, engineer, [galinanikki@mail.ru](mailto:galinanikki@mail.ru), Russia, Tula, Tula state University*