

## ОПРЕДЕЛЕНИЕ НАБОРОВ ПРЕДИКТОРНЫХ ПЕРЕМЕННЫХ ПО ОРИЕНТИРОВАННОМУ ГРАФУ ЛОКАЛЬНЫХ ВЗАИМОСВЯЗЕЙ

На підставі орієнтованого графа локальних взаємодій запропоновані алгоритми побудови наборів предикторних змінних.

На основании ориентированного графа локальных взаимосвязей предложены алгоритмы построения наборов предикторных переменных.

On the basis of focused the column of local interrelations algorithms of construction of sets the predictor variables are offered.

В классических исследованиях по математическому моделированию физических, химических и других процессов, как правило, на основании опытных данных предполагались некоторые допущения о протекании таких процессов в «малом» (т. е., малость понималась то ли пространственная, то ли временная, то ли обе одновременно).

Такой подход позволял математическую модель представить в виде дифференциального уравнения или системы уравнений, решение которых давало возможность проследить протекание процесса как во времени, так и в пространстве.

Не умаляя достоинства данного подхода, необходимо отметить, что причинно-следственные связи не находили достаточного отражения. Особенно это касается управляемых процессов, так как выбор управляющих переменных в основном определяется постановщиком задачи.

Так как процесс электролитического покрытия представляет собой макроскопическую систему, не находящуюся в состоянии полного термодинамического равновесия, подвергающуюся внешним воздействиям и эволюционирующую во времени, то в общем случае макроскопическими величинами (температура, давление, плотность и т. д.) нельзя пользоваться для описания неравновесных систем.

Если макроскопическая система находится в состоянии равновесия, то к ней или к ее макроскопическим частям применимы термодинамические понятия. Тогда, разбив макроскопическую систему на малые, но макроскопические элементы и считая, что эти элементы находятся в состоянии локального равновесия, можно отвлечься от молекулярной структуры системы и

считать ее сплошной средой, т. е. в каждой точке раствора задаются такие термодинамические показатели, как плотность (концентрация) компонент их средние скорости и т. д.

Другими словами считаем, что между термодинамическими показателями существуют определенные причинно-следственные связи.

Пусть тот или иной процесс в любой момент времени характеризуется набором показателей  $\Omega = \{x_1, x_2, \dots, x_n\}$ .

**Определение 1.** Будем говорить, что  $x_i$  является причиной изменения  $x_j$  в малом, если изменение  $x_i$  на отрезке времени  $[t, t + \tau]$  приводит на этом же отрезке времени к изменению  $x_j$ .

В данном определении величина  $\tau$  – характеристическое время распространения возмущений (флуктуаций) и определяется макро- (микро-) уровнем рассматриваемого процесса [1].

**Определение 2.** Ориентированный граф  $G(V, E)$  [2] будем называть математической моделью локальных взаимосвязей, если имеет место:

1. Множество вершин  $V$  графа  $G(V, E)$  совпадает с множеством  $\Omega$ .

2. Множество вершин  $E$  – набор направленных дуг, отражающих причинно-следственные связи в малом.

Заметим, что в  $E$  могут входить дуга из  $x_i$  в  $x_j$  и ей противоположная из  $x_j$  в  $x_i$ .

В общем случае множество вершин  $V$  можно разложить на непересекающие множества:  $VI$  – множество истоков,  $VS$  – множество стоков и  $\tilde{V} = V \setminus (VI \cup VS)$ .

**Определение 3.** Множество  $M \subset \Omega$  будем

называть набором предикторных переменных если имеет место:

C1.  $M \subset \Omega$ .

C2.  $\forall x, y \in M$  нет дуг между ними.

C3.  $M$  – максимальный набор.

Под максимальнойностью множества  $M$  понимается ситуация, когда это множество не может быть пополнено элементами из  $\tilde{V}$  без нарушения свойства C2.

**Пример.** Пусть  $\Omega = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ , а граф  $G(V, E)$  имеет следующие ребра (дуги)  $E = \{e_{13}, e_{23}, e_{35}, e_{43}, e_{56}, e_{65}, e_{57}, e_{67}, e_{63}\}$ , где под дугой  $e_{ij}$  понимается дуга, начинающаяся в  $x_i$  и оканчивающаяся в  $x_j$ . На рис. 1 представлен данный граф.

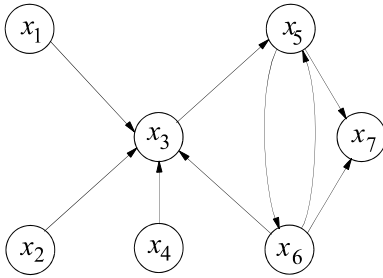


Рис. 1. Графическое представление  $G(V, E)$

В этом примере множество источников представляет собой  $VI = (x_1, x_2, x_4)$ , а множество стоков равно  $VS = \{x_7\}$ . Необходимо отметить, что множество истоков отражает влияние внешней среды на рассматриваемый процесс (явление), а множество стоков отражает влияние исследуемого процесса на внешнюю среду.

Для данного примера множества предикторных переменных представляют собой

$$\begin{aligned} M_1 &= \{x_1, x_2, x_4, x_5\} \\ M_2 &= \{x_1, x_2, x_4, x_6\} \end{aligned} \quad (1)$$

Тогда структуры математических моделей будут следующими:

$$\begin{aligned} M_1 &\rightarrow \left\{ \begin{array}{l} x_3 = f_3(x_1, x_2, x_4, x_6); \\ x_6 = f_6(x_5); \\ x_7 = f_7(x_5, x_6). \end{array} \right\}; \\ M_2 &\rightarrow \left\{ \begin{array}{l} x_3 = f_3(x_1, x_2, x_4, x_6); \\ x_5 = f_5(x_3, x_6); \\ x_7 = f_7(x_5, x_6). \end{array} \right\}. \end{aligned}$$

Заметим, что название переменных из  $M$

связано с тем, что знание  $x_i \in M$  позволяет определять остальные переменные по одной из представленных математических моделей.

В рассмотренном примере определение множеств  $M_1$  и  $M_2$  не вызывает затруднений. В более сложных ситуациях необходимо иметь алгоритм их построения.

Для разработки алгоритма построения набора предикторных переменных понадобится матрица смежности [2], которая отражает наличие ориентированных дуг.

Так, для графа, представленного на рис. 1, матрица смежности имеет вид

$$T := \begin{array}{c|cccccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ \hline x_1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline x_2 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline x_3 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ \hline x_4 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \hline x_5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ \hline x_6 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ \hline x_7 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

Элементы данной матрицы

$$T[i, j] = \begin{cases} 0, & \text{если из } x_i \text{ нет дуги в } x_j; \\ 1, & \text{если из } x_i \text{ есть дуга в } x_j. \end{cases}$$

Матрица смежности  $T$  позволяет определять, является ли данная вершина источником. Для этого необходимо вычислить сумму элементов по столбцу и если сумма равна 1, то данная вершина является источником. В рассмотренном примере источниками являются вершины  $x_1, x_2$  и  $x_4$ .

Если просуммировать элементы матрицы по строке и сумма будет равна 1, то вершина представляет собой сток. В примере из рис. 1 и ему соответствующей матрицы  $T$  таким свойством обладает строка  $x_7$ .

Так как набор источников входит в каждое множество предикторных переменных, то в это множество не войдут те вершины, в которых имеются дуги от источников (данное утверждение следует из определения 3). В рассмотренном примере таким множеством является  $W = \{x_3\}$ . И если из  $\tilde{V}$  удалить множество  $W$ , то получим множество

$$VK = \tilde{V} \setminus W = \{x_5, x_6\}.$$

Это множество будем называть кандидатами в предикторные переменные. В силу того,

что между вершинами  $x_5$  и  $x_6$  имеются дуги, то они одновременно не могут присутствовать в одном и том же наборе  $M$ , что в этом примере и приводит к порождению двух наборов  $M_1$  и  $M_2$ , которые были ранее приведены (1).

### Алгоритм построения наборов предикторных переменных

Для данного алгоритма исходная информация должна быть представлена в виде:  $n$  – количество вершин,  $V$  – перечень вершин графа локальных взаимосвязей и матрица смежности  $T$ .

Основные элементы алгоритма представляют собой:

Э1. Определение множества источников  $VI$ .

```
VI := {}: for x in V do s := 0:
for i from 1 to n do s := s + T[j, x]
end do: if s = 1 then VI := VI union {x} end do:
```

Э2. Определение множества стоков.

В данном элементе можно использовать процесс как в Э1, только  $VI$  необходимо заменить на  $VS$ , а сумму  $S$  вычислять по формуле

$$s := s + T[x, j].$$

Э3. Формирование множества  $\tilde{V}$  по формуле

$$\tilde{V} := V \text{ minus } (VI \text{ union } VS)$$

Э4. Определение множества  $W$  – набор вершин, в которые имеются дуги из источников.

Реализация данного элемента в среде Maple [3] будет следующей

```
W := {}: for z in VI do for i
from 1 to n do if T[z, i] = 1 then
W := W union {i} end if: end do:
end do: W := W minus VI:
```

Э5. Формирование множества  $VK$  -кандидатов в предикторные переменные по формуле

$$VK := V \text{ minus } (VI \text{ union } VS \text{ union } W)$$

Э6. Формирование множества  $Wmm$  из элементов множества  $VK$ . Элементы из  $Wmm$  представляют собой списки со свойствами С2 и С3.

Э7. Формирование наборов предикторных переменных по схеме

```
for w in Wmm do M := VI union
{op(w)} end do:
```

Из перечисленных элементов основным является Э6, который при своей реализации

использует две несложные процедуры: определение последнего элемента в списке и вторая процедура – исключение из множества, элементами которого являются списки, списков по признаку вложенности в другие списки этого множества.

Суть работы элемента Э6 покажем на примере, когда граф локальных связей имеет такой вид, как на рис. 2.

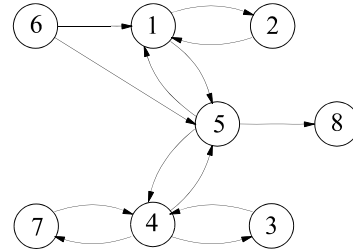


Рис. 2. Граф локальных взаимосвязей

В этом примере  $VI = \{6\}$ ;  $VS = \{8\}$ ;  $W = \{1, 5\}$ , тогда множество кандидатов в предикторные переменные представляет собой

$$VK = \{2, 3, 4, 7\}.$$

Данное множество является исходной информацией для Э6. Полагаем  $x = 2$  первому элементу из  $VK$  и среди множества  $VK \setminus \{x\}$  определяем тех, которые не имеют дуг с вершиной 2. Таковыми будут  $\{3, 4, 7\}$  и формируем множество списков

$$WR = \{[2, 3], [2, 4], [2, 7]\}.$$

Далее для каждого списка из этого множества смотрим, каким элементом его можно пополнить. Так, список  $[2, 3]$  можно пополнить элементом 7, а  $[2, 7]$  – элементом 3. Тогда для  $x = 2$  имеем

$$WR = \{[2, 3, 7], [2, 4], [2, 7, 3]\}.$$

Но третий элемент как множество  $\{2, 7, 3\}$  содержится во множестве  $\{2, 3, 7\}$ , получаемого из элементов первого списка, поэтому список  $[2, 7, 3]$  должен быть исключен из  $WR$ , т. е. полагаем

$$WR := WR \setminus \{[2, 7, 3]\}$$

и запоминаем его во множестве

$$Wm = \{[2, 3, 7], [2, 4]\}.$$

Далее берем следующий элемент из  $VK$  и для него строим очищенное по принадлежности множество  $WR = \{[3, 2, 7]\}$ , которым попол-

няем  $Wm$  и тут же его очищаем по принадлежности, что не приводит к изменению  $Wm$ . Аналогично поступаем и с остальными элементами из  $VK$ , что приводит к окончательному варианту наборов предикторных переменных в виде списков из  $Wm = \{[2,3,7],[2,4]\}$ . Пополняя каждый список множеством  $VI$ , получаем следующие варианты наборов предикторных переменных для графа на рис. 2

$$\begin{aligned} M_1 &= \{2,3,6,7\}; \\ M_2 &= \{2,4,6\}. \end{aligned}$$

Реализацию Э6 можно выполнить, если рассматривать граф на рис. 2 после удаления источников  $VI$  и стоков  $VS$  (рис. 3).

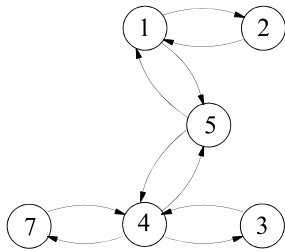


Рис. 3. Граф без источников и стоков

Удалив из этого графа вершины  $W$ , которые связаны с источниками получим следующий граф (рис. 4).

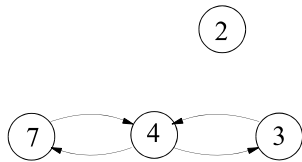


Рис. 4. Граф с вершинами из множества  $VK$

Теперь к этому графу построим дополнительный граф (рис. 5).

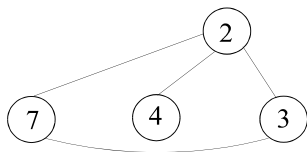


Рис. 5. Дополнительный граф к графу на рис. 4

Набор  $T$  называется классом толерантности, если между любыми двумя вершинами из  $T$  имеется дуга и этот набор не пополняем без нарушения предыдущего требования [3].

Тогда классы толерантности дополнительного графа будут следующими

$$T_1 = \{2,4\}; \quad T_2 = \{2,3,7\},$$

которые пополним вершинами из множества  $VI$  получим необходимый набор предикторных переменных для графа, представленного на рис. 2.

Заметим, что графу, представленному на рис. 5, можно поставить в соответствие матрицу толерантности  $\tau$ , элементы которой представляет собой

$$\tau_{ij} = \begin{cases} 1, & \text{если между вершинами} \\ & x_i \text{ и } x_j \text{ имеется дуга;} \\ 0, & \text{в противном случае.} \end{cases}$$

Для рассматриваемого примера матрица толерантности будет следующей

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
|       | $x_2$ | $x_3$ | $x_4$ | $x_7$ |
| $x_2$ | 1     | 1     | 1     | 1     |
| $x_3$ | 1     | 1     | 0     | 1     |
| $x_4$ | 1     | 0     | 1     | 0     |
| $x_7$ | 1     | 1     | 0     | 1     |

Рассматривая элементы матрицы как булевы переменные и используя правила упрощения булевых выражений:

$$x + x = x; \quad x \times x = x; \quad x + xy = x$$

изложим алгоритм М. Е. Пиша [4] построения классов толерантности. Для каждой строки матрицы  $\tau$  и по элементам выше главной диагонали строим  $x_i$  плюс произведение тех переменных, которые равны нулю. Если в строке нет нулей, то полагаем, что сумма равна 1. В рассматриваемом примере имеем

$$x_2 \rightarrow x_2 = 1; \quad x_3 \rightarrow x_3 + x_4;$$

$$x_4 \rightarrow x_4 + x_7; \quad x_7 \rightarrow x_7 = 1.$$

Далее строим произведение полученных сумм

$$I = 1 \cdot (x_3 + x_4) \cdot (x_4 + x_7) \cdot 1.$$

Перемножив с учетом правил упрощения булевых переменных, получим

$$S = x_4 + x_3x_7.$$

Теперь строим сумму  $\bar{S}$ , слагаемые которой являются дополнениями к слагаемым суммы  $S$ , тогда

$$\bar{S} = x_2x_3x_7 + x_2x_4,$$

а классы толерантности будут

$$\left. \begin{aligned} T_1 &= \{x_2, x_4\}; \\ T_2 &= \{x_2, x_3, x_7\}. \end{aligned} \right\}$$

Изложенное можно представить в виде следующего алгоритма:

III. Введение верхней треугольной части матрицы  $\tau$ .

П2. Формирование сумм  $S_{x_i}$  по формуле

$$S_i = x_i + \prod_{k=1}^n x_{ik},$$

где  $i_k$  – номер переменной, для которой  $\tau_{x_i x_{i_k}} = 0$

П3. Построение суммы  $S = \prod_{i=1}^n S_i$  и приведение ее к виду

$$S = P_1 + P_2 + \dots + P_k,$$

где  $P_k$  – произведение некоторого набора переменных.

П4. Построение классов толерантности по формуле

$$T_j = \Omega \setminus \{P_j\}, \quad j=1, \nu,$$

где  $\Omega$  – перечень переменных, по которым строилась матрица толерантности  $\tau$ .

В данном алгоритме наиболее трудоемкой операцией является операция приведения суммы  $S = \prod_{i=1}^n S_i$  к нормальному виду.

Если  $A$  и  $B$  – некоторые множества, соответствующие слагаемым  $S$  после выполнения произведений, то имеем

$$A + B = \begin{cases} A, & \text{если } A \subset B; \\ B, & \text{если } B \subset A; \\ A + B, & \text{если } (A \not\subset B) \wedge (B \not\subset A). \end{cases} \quad (*)$$

При раскрытии  $S$  в общем виде получаем

$$S = A_1 + A_2 + \dots + A_m,$$

где  $A_i$  – некоторые произведения  $x_k$  и  $m \leq 2^{n-1}$ .

Операцию (\*) необходимо выполнить не более, чем

$$N \leq \frac{m(m-1)}{2},$$

тогда порядок алгоритма М. Е. Пиша будет

$$Q(n) \approx 2^{2n},$$

что при достаточно больших  $n$  – числа вершин графа может быть затруднительно реализуемым на ЭВМ.

Напомним, что все изложенное относится к случаю, когда ориентированный граф локальных связей  $G(V, E)$  является связным [2]. В противном случае, когда он распадается на графы  $G_1(V_1, E_1), \dots, G_m(V_m, E_m)$ , для

которых имеет место

$$\left. \begin{aligned} V &= \bigcup_{i=1}^m V_i; & V_i \cap V_j &= \emptyset, \\ E &= \bigcup_{i=1}^m E_i; & E_i \cap E_j &= \emptyset, \quad i \neq j, \end{aligned} \right\}$$

моделирующий процесс распадается на  $m$  независимых процессов, которые можно рассматривать в отдельности от других.

Следовательно, в первую очередь исходный граф локальных взаимосвязей (связей) должен быть проверен на связность и определены его связные подграфы.

Так как связность ориентированного графа определяется через соответствующий неориентированный граф, который соответствует симметризованной матрице смежности исходного графа, то затем задача проверки на связность сводится к проверке соответствующего неориентированного графа. Так, например, граф, представленный на рис. 1, имеет соответствующий неориентированный граф с матрицей смежности

|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
|       | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ |
| $x_1$ | 1     | 0     | 1     | 0     | 0     | 0     | 0     |
| $x_2$ | 0     | 1     | 1     | 0     | 0     | 0     | 0     |
| $x_3$ | 1     | 1     | 1     | 1     | 1     | 1     | 0     |
| $x_4$ | 0     | 0     | 1     | 1     | 0     | 0     | 0     |
| $x_5$ | 0     | 0     | 1     | 0     | 1     | 1     | 1     |
| $x_6$ | 0     | 0     | 1     | 0     | 1     | 1     | 0     |
| $x_7$ | 0     | 0     | 0     | 0     | 1     | 1     | 1     |

и проверка на связность сводится к определению простого пути между любыми двумя вершинами по матрице  $TS$ .

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Левич В. Г. Курс теоретической физики. Т. 2. / В. Г. Левич, Ю. А. Вдовин, В. А. Мямлин. – М.: Наука, 1971. – 936 с.
2. Андерсон Д. А. Дискретная математика и комбинаторика. – М.: Изд. дом «Вильямс», 2004. – 960 с.
3. Босов А. А. Підвищення ефективності роботи транспортної системи на основі структурного аналізу. / А. А. Босов, Н. А. Мухіна, Б. П. Піх. – Д.: ДІТ, 2006. – 200 с.
4. Емелечев В. А. Лекции по теории графов. / В. А. Емелечев, О. И. Мельников, В. И. Сарванов, Р. И. Тышкевич. – М.: Наука, 1990. – 383 с.

Надійшла до редколегії 20.01.2007.