

5. Воронин А.Н. Многокритериальная оптимизация иерархических структур // Проблемы информатизации та управління: Зб.наук.пр. – К.:НАУ,2004. – Вип.11. – С. 101-104.

6. Воронин А.Н. Метод многокритериальной оценки оптимизации иерархических систем // Кибернетика и системный анализ. – 2007. – №3. – С. 84-92.

7. Воронин А.Н. Декомпозиция и композиция свойств альтернатив в многокритериальных задачах принятия решений // Кибернетика и системный анализ. – 2009. – №1. – С. 117-122.

8. Вишневикий В.М. Теоретические основы проектирования компьютерных сетей. – М.: Техносфера, 2003. – 512 с.

Поступила 04.02.2009

УДК 681.39

Козюра В.Д.

### МЕТОДЫ АППРОКСИМАЦИИ УСЛОВНЫХ РАСПРЕДЕЛЕНИЙ БИНАРНЫХ ПРИЗНАКОВ В СИСТЕМАХ РАСПОЗНАВАНИЯ

Структура статистической распознающей системы определяется, в основном, видом условных распределений вероятностей значений признаков распознаваемых классов  $f(\bar{x}/A_m)$ ,  $m = \overline{1, M}$ , где  $A_m$  – обозначение  $m$ -го класса,  $M$  – общее число классов;  $\bar{x} = (x_1, \dots, x_n, \dots, x_N)^T$  –  $N$ -мерный вектор измеряемых признаков [ 1 ].

Если каждый из признаков  $x_n$ ,  $n = \overline{1, N}$  является бинарным (т.е.  $x_n = 0$  или  $x_n = 1$ ), то общее число различных условных вероятностей  $f(\bar{x}/A_m) = \{P(b_j/A_m)\}_{j=1}^J$ ,  $m = \overline{1, M}$  для описания каждого класса составит  $J = 2^N$ . При числе признаков  $N \geq 20$  столько вероятностей  $P(b_j/A_m)$  оценить в процессе обучения распознающей системы практически нереально. В связи с этим возникает проблема в непараметрической оценке условных распределений условных распределений значений признаков.

Одним из эффективных путей непараметрической оценки искомых вероятностей является использование разложений дискретных распределений по системе многоортогональных функций [ 2 ]. Практически удобным является разложение Бахадура-Лазарсфельда, описывающее искомые распределения в виде:

$$f(\bar{x}) = \prod_{n=1}^N P_n^{x_n} \cdot [1 - P_n]^{1-x_n} \times \left\{ 1 + \sum_{n=1}^{N-1} \sum_{l=n+1}^N \rho_{nl} \cdot \frac{(x_n - P_n) \cdot (x_l - P_l)}{\sqrt{P_n \cdot (1 - P_n) \cdot P_l \cdot (1 - P_l)}} + \dots + \rho_{12\dots N} \cdot \frac{(x_1 - P_1) \cdot \dots \cdot (x_N - P_N)}{\sqrt{P_1 \cdot (1 - P_1) \cdot \dots \cdot P_N \cdot (1 - P_N)}} \right\},$$

где  $P_n = P(x_n = 1)$ ,  $n = \overline{1, N}$  – условные вероятности того, что  $n$ -й признак принимает значение 1;  $\rho_{nl}, \rho_{nlk}, \dots, \rho_{12\dots N}$  – коэффициенты корреляции двух, трех и т.д. бинарных переменных.

Аппроксимация распределения  $f(\bar{x})$  заключается в игнорировании статистических связей выше определенного порядка. Например, аппроксимацией первого порядка является выражение, описывающее независимые признаки

$$\hat{f}_1(\vec{x}) = \prod_{n=1}^N P_n^{x_n} \cdot [1 - P_n]^{1-x_n}.$$

Если учитываются коэффициенты корреляции между парами любых признаков, то получаем аппроксимацию второго порядка:

$$\hat{f}_2(\vec{x}) = \prod_{n=1}^N P_n^{x_n} \cdot [1 - P_n]^{1-x_n} \times \left\{ 1 + \sum_{n=1}^{N-1} \sum_{l=n+1}^N \rho_{nl} \cdot \frac{(x_n - P_n) \cdot (x_l - P_l)}{\sqrt{P_n \cdot (1 - P_n) \cdot P_l \cdot (1 - P_l)}} \right\}.$$

В аппроксимации третьего порядка учитываются коэффициенты корреляции между тройками любых признаков:

$$\hat{f}_3(\vec{x}) = \prod_{n=1}^N P_n^{x_n} \cdot [1 - P_n]^{1-x_n} \times \left\{ 1 + \sum_{n=1}^{N-1} \sum_{l=n+1}^N \rho_{nl} \cdot \frac{(x_n - P_n) \cdot (x_l - P_l)}{\sqrt{P_n \cdot (1 - P_n) \cdot P_l \cdot (1 - P_l)}} + \sum_{n=1}^{N-2} \sum_{l=n+1}^{N-1} \sum_{k=l+1}^N \rho_{nlk} \cdot \frac{(x_n - P_n) \cdot (x_l - P_l) \cdot (x_k - P_k)}{\sqrt{P_n \cdot (1 - P_n) \cdot P_l \cdot (1 - P_l) \cdot P_k \cdot (1 - P_k)}} \right\}.$$

Если для случая статистически независимых признаков при обучении необходимо определение  $N$  вероятностей  $P_n = P(x_n = 1)$ ,  $n = \overline{1, N}$  для каждого распознаваемого класса, то при учете попарных статистических связей признаков ( $\rho_{nl}$ ,  $n, l = \overline{1, N}$ ,  $n \neq l$ ) число определяемых параметров увеличивается на  $N \cdot (N - 1) / 2$ .

В общем случае при аппроксимации  $s$ -го порядка количество параметров распределений, определяемых при обучении для каждого распознаваемого класса, составляет

$$J_s = \sum_{k=1}^s \frac{N!}{k!(N-k)!}, \quad s = \overline{1, N}.$$

В таблице 1 приведены данные по количеству параметров, определяемых в процессе обучения при различных порядках аппроксимации и различном числе признаков.

Таблица 1

Порядок аппроксимации (s) \ Число признаков (N)	2	5	10	20	50	100
1	2	5	10	20	50	100
2	3	15	55	210	1275	5050
3	-	25	175	1350	20875	166750
Полное описание	3	32	1024	$\sim 10^6$	$\sim 10^{15}$	$\sim 10^{30}$

Эти результаты показывают, что при числе признаков  $N \geq 50$  использование аппроксимации выше второго порядка приводит к увеличению количества определяемых параметров условных распределений более чем в 20 раз по сравнению с аппроксимацией первого порядка.

Для случая двух бинарных признаков ( $N = 2$ ) выражение аппроксимации Бахадура-Лазарсфельда совпадает с точной формулой двумерного распределения:

$$f(x_1, x_2) = \prod_{n=1}^2 P_n^{x_n} \cdot [1 - P_n]^{1-x_n} \times \left\{ 1 + \rho_{12} \cdot \frac{(x_1 - P_1) \cdot (x_2 - P_2)}{\sqrt{P_1 \cdot (1 - P_1) \cdot P_2 \cdot (1 - P_2)}} \right\}. \quad (1)$$

Таким образом, аппроксимация условных распределений сводится к тому, что перестают учитывать корреляционные связи выше определенного порядка. Однако такой подход обладает одним, весьма серьезным недостатком. Дело в том, что при некоторых значениях векторов измеряемых признаков  $\vec{x} = (x_1, \dots, x_n, \dots, x_N)^T$  в расчетах по формулам аппроксимации могут быть получены отрицательные значения вероятностей. Причина этого кроется в игнорировании объективно существующих коэффициентов корреляции высоких порядков.

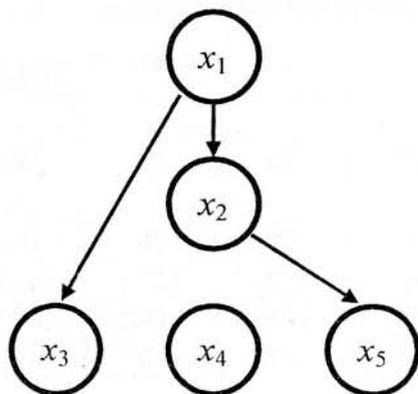
Чтобы устранить эту проблему, предлагается использовать комбинацию метода Бахадура-Лазарсфельда и метода Чоу.

Суть метода Чоу [ 3 ] заключается в том, что распределения  $f(\vec{x}/A_m)$ ,  $m = \overline{1, M}$  представляются в виде произведения условных распределений признаков

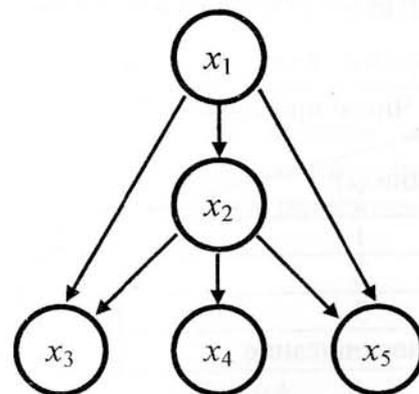
$$f(\vec{x}/A_m) = f(x_1, x_2, \dots, x_N/A_m) = f(x_1/A_m) \cdot f(x_2/x_1, A_m) \cdot \dots \cdot f(x_N/x_{N-1}, \dots, x_1, A_m) = \prod_{n=1}^N f(x_n/x_{n-1}, \dots, x_1, A_m), \quad m = \overline{1, M}.$$

Такое представление определяет все статистические связи, существующие между признаками. Если предположить, что существуют сильные связи, которые нельзя игнорировать, и слабые связи, не оказывающие заметного влияния на ошибки аппроксимации, то можно статистические зависимости признаков  $\vec{x} = (x_1, \dots, x_n, \dots, x_N)^T$  представить в виде некоторого графа зависимостей. Примеры таких графов приведены на рис. 1.

Особенностью одномерного графа (рис. 1, а) является то, что любой признак  $x_n, n = \overline{1, N}$  может зависеть не более чем от одного предшествующего признака  $x_{k(n)}, 0 \leq k(n) < n$ ; для двумерных графов (рис. 1, б) такая зависимость распространяется не



а) одномерный граф  
 $f(\vec{x}) = f(x_1) \cdot f(x_2/x_1) \cdot f(x_3/x_1) \cdot f(x_4) \cdot f(x_5/x_2)$



б) двумерный граф  
 $f(\vec{x}) = f(x_1) \cdot f(x_2/x_1) \cdot f(x_3/x_2, x_1) \cdot f(x_4/x_2) \cdot f(x_5/x_2, x_1)$

Рис. 1. Графы зависимости признаков

более чем на два предшествующих признака  $x_{k(n)}, x_{l(n)}$   $0 \leq k(n), l(n) < n$ .

Использование одномерного графа позволяет аппроксимировать многомерное распределение вероятностей бинарных признаков выражением вида:

$$\hat{f}_1(\vec{x}) = P_1^{x_1} \cdot [1 - P_1]^{1-x_1} \cdot \prod_{n=2}^N \left[ \gamma_{1n}^{x_n} \cdot (1 - \gamma_{1n})^{1-x_n} \right]^{x_{k(n)}} \cdot \left[ \gamma_{0n}^{x_n} \cdot (1 - \gamma_{0n})^{1-x_n} \right]^{1-x_{k(n)}},$$

где  $\gamma_{0n} = P(x_n = 1/x_{k(n)} = 0)$ ,  $\gamma_{1n} = P(x_n = 1/x_{k(n)} = 1)$ ,  $n = \overline{2, N}$  - условные вероятности зависимых признаков.

Определение параметров  $\gamma_{0n}, \gamma_{1n}$ ,  $n = \overline{2, N}$  свяжем с использованием разложения (1).

Так как  $P(x_n/x_{k(n)}) = \frac{P(x_n, x_{k(n)})}{P(x_{k(n)})}$ ,  $n = \overline{2, N}$ , то легко показать, что искомые

коэффициенты рассчитываются по формулам:

$$\gamma_{0n} = P(x_n = 1/x_{k(n)} = 0) = P_n - \rho_{k(n),n} \cdot \sqrt{\frac{1 - P_n}{1 - P_{k(n)}} \cdot P_n \cdot P_{k(n)}};$$

$$\gamma_{1n} = P(x_n = 1/x_{k(n)} = 1) = P_n + \rho_{k(n),n} \cdot \sqrt{\frac{1 - P_n}{1 - P_{k(n)}} \cdot P_n \cdot P_{k(n)}}, \quad n = \overline{2, N}.$$

Аналогичные выражения могут быть получены для двумерных графов зависимости бинарных признаков.

Процесс обучения для описанного комбинированного метода аппроксимации условных распределений вероятностей значений признаков сводится к определению максимум  $2 \times N - 1$  параметров для каждого распознаваемого класса ( $N$  значений одномерных вероятностей  $P_n$  и  $N - 1$  наиболее существенных коэффициентов корреляции  $\rho_{k(n),n}$ ).

Преимущества рассмотренного комбинированного метода реализуются только в том случае, когда удастся определить функцию зависимостей признаков  $k(n)$ ,  $n = \overline{2, N}$ . В общем случае для  $N$ -мерного вектора бинарных признаков  $\vec{x} = (x_1, \dots, x_n, \dots, x_N)^T$  существует  $N^{N-2}$  возможных вариантов построения функции  $k(n)$ ,  $n = \overline{2, N}$ . При больших значениях  $N$  определить оптимальную функцию  $k(n)$  практически невозможно путем прямого перебора различных вариантов. Поэтому попытаемся определить квазиоптимальную функцию  $\hat{k}(n)$ , используя понятие дивергенции: для каждого признака  $x_n$ ,  $n = \overline{2, N}$  выбирается такой предшествующий ему признак  $x_{k(n)}$ ,  $0 \leq k(n) < n$ , дивергенция совместного распределения которых максимальна, т.е.

$$\hat{k}(n) = \arg \max_{s, l=0, n-1} \{ \text{Div}[f(x_s, x_n), f(x_l, x_n)] \}, \quad n = \overline{2, N},$$

где значение дивергенции рассчитывается по формуле

$$\text{Div}[f(x_s, x_n), f(x_l, x_n)] = \sum_{x_s, x_l} [f(x_s, x_n) - f(x_l, x_n)] \log \frac{f(x_s, x_n)}{f(x_l, x_n)}.$$

Таким образом, использование комбинированного метода аппроксимации условных распределений бинарных признаков (методы Бахадура-Лазарсфельда и Чоу) позволяет, во-первых, сократить на несколько порядков количество определяемых при обучении распознающей системы параметров распределений, во-вторых, учесть наиболее сильные корреляции между признаками, что ведет к повышению вероятности правильного

распознавания объектов, и, в-третьих, избавиться от проблемы получения отрицательных значений вероятностей, что характерно для метода аппроксимации Бахадура-Лазарсфельда.

**Список литературы**

1. Козюра В.Д. Статистическая модель распознавания речевых сигналов. Вісник ДУІКТ, Том 6, № 3, 2008. - с. 229-234.
2. Дуда Р., Харт П. Распознавание образов и анализ сцен. Пер с англ. – М.: Мир, 1976. Chow C.K., Liu C.N. Approximating Discrete Probability Distributions with Dependence Trees. IEEE Transactions of Information Theory, Vol. IT-14, No. 3, May 1968. - p. 462-467.

Поступила 28.01.2009

УДК 004.681.3:519.67

Дахно Н.Б., Тискина Е.О., Хорошко В.А.

**АЛГОРИТМЫ ОБРАЩЕНИЯ МАТРИЦ ДЛЯ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ СИСТЕМ ЗАЩИТЫ ИНФОРМАЦИИ**

**Вступление**

Повышение эффективности математического моделирования сложных технических систем, к которым относятся и системы защиты информации, можно обеспечить за счет использования вычислительных структур (ВС) с параллельной обработкой данных. Эта необходимость стимулирует разработку методов и алгоритмов решения систем уравнений большой размерности и алгоритмов обращения матриц, допускающих распараллеливания вычислений. Однако вопросы параллельной реализации обращения матриц в литературе уделено недостаточное внимание.

**Основная часть**

Исходя из этого и для решения этой задачи рассматриваются алгоритмы обращения невыраженной квадратной матрицы  $A=[a_{ij}](i,j=1,2,\dots,n)$  большой размерности. Эти алгоритмы базируются на использовании метода «цифра за цифрой» [1,2] с целью организации параллельных вычислений.

Алгоритм 1. Обращение матрицы с помощью алгоритма Гаусса по методу «цифра за цифрой».

Для нахождения обратной матрицы  $A^{-1}=[x_{ij}]$  используется основное соотношение  $AA^{-1}=I$ , где  $I$ -единичная матрица размером  $n \times n$ . Так, для матрицы размером  $4 \times 4$  основное соотношение в матричной форме имеет вид:

$$A \times A^{-1} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \times \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix} = \begin{bmatrix} 1 & - & - & - \\ - & 1 & - & - \\ - & - & 1 & - \\ - & - & - & 1 \end{bmatrix},$$

из которого получается четыре системы уравнений с 16 неизвестными элементами  $x_{ij}$  обратной матрицы  $A^{-1}=[x_{ij}] (i,j=1,2,3,4)$ . Матричную форму основного соотношения представим следующим образом: