Математические структуры и моделирование
2015. № 4(36). С. 74–92

УДК 519.237.7

АЛГОРИТМЫ КОРА И ФОРДИАСИМПТ КАК МЕТОДЫ РАСПОЗНАВАНИЯ ДВУХ ОБРАЗОВ В ПРОСТРАНСТВЕ ДВОИЧНЫХ ПРИЗНАКОВ

В.В. Гольтяпин
к.ф.-м.н, доцент, e-mail: goltyapin@mail.ru
В.А. Шовин
научный сотрудник, e-mail: v.shovin@mail.ru

Омский филиал Федерального государственного бюджетного учреждения науки
Института математики им. С.Л. Соболева Сибирского отделения РАН

Аннотация. В данной статье в рамках теории латентного анализа сформулированы и доказаны утверждение, лемма и теорема, позволяющие находить апостериорные вероятности на базе альтернативных показателей с использованием ортогональной факторной структуры. На основе полученных теоретических выкладок построен вычислительный алгоритм ФОРДИАСИМПТ, позволяющий строить диагностические симптомокомплексы на базе вероятностного метода распознавания образов. Проведен сравнительный анализ алгоритмов КОРА и ФОРДИАСИМПТ как методов распознавания двух образов в пространстве двоичных признаков в случае независимых симптомокомплексов при адекватной статистической информации.

Ключевые слова: симптомокомплекс, факторная модель, латентная модель, корреляционный анализ, маргинальное распределение, маргинал, алгоритм КОРА, алгоритм ФОРДИАСИМПТ.

Введение

Основная цель работы заключается в обосновании практической целесообразности применения алгоритма формирования диагностических симптомокомплексов для случаев независимых симптомокомплексов. Ввиду длинного названия алгоритма предлагается использовать сокращение ФОРДИАСИМПТ. Для демонстрации достоинств и недостатков алгоритма ФОРДИАСИМПТ было проведено сравнительное исследование с алгоритмом КОРА.

Предлагаемый алгоритм решает следующие задачи. Во-первых, формирует набор симптомокомплексов, опираясь на ортогональную факторную структуру и на уровень значимости φ коэффициента по χ^2 критерию. Во-вторых, для каждого симптомокомплекса находит диагностическую шкалу на базе простейшей латентно-структурной модели.

В силу обоснованности использования факторного анализа для альтернативных показателей, считаем известным матрицу ортогонального факторного отображения. С полным изложением теоретических основ алгоритма ФОРДИАСИМПТ, относящихся к поиску факторной структуры, можно ознакомиться в работах [1-4].

Математический аппарат простейшей латентной модели и алгоритма ФОРДИАСИМПТ

Особое внимание в данной статье уделим непосредственно математическому аппарату, который используется в построении латентной модели на базе альтернативных данных. Обозначим количество объектов исследования n — объем выборки, а количество измеряемых параметров m — размерность выборки. Тогда исходные альтернативные данные представляются в виде матрицы $Y = \{y_{ij}\}$ размерности $m \times n$, столбцы которой — объекты исследования, а строки — значения измеряемых параметров у конкретного объекта.

Далее введём следующие обозначения:

p_i – отношение количества объектов к n, у которых i-ый показатель равен 1;

p_{ij} – отношение количества объектов к n, у которых i-ый и j-ый показатели равны 1;

$p_{i\bar{j}}$ – отношение количества объектов к n, у которых i-ый показатель равен 1, j-ый показатель равен 0;

$p_{\overline{ij}}$ – отношение количества объектов к n, у которых i-ый и j-ый показатели равны 0;

p_{ijk} – отношение количества объектов к n, у которых i-ый, j-ый и k-ый показатели равны 1;

$p_{i\bar{j}k}$ – отношение количества объектов к n, у которых i-ый и k-ый показатели равны 1, а j-ый показатель равен 0;

$p_{\overline{ij}k}$ – отношение количества объектов к n, у которых i-ый и j-ый показатели равны 0, а k-ый показатель равен 1;

$\tilde{\phi}(x_l)$ – частота, соответствующая относительному объёму l-го класса;

$\tilde{f}_i(x_l)$ – вероятность значения 1 по i-му показателю у объекта, находящегося в l-ом классе;

$\tilde{f}_{ik}(x_l)$ – вероятность значения 1 по i-му и k-му показателям у объекта, находящегося в l-ом классе;

$\tilde{f}_{ijk}(x_l)$ – вероятность значения 1 по i-му, j-му и k-му показателям у объекта, находящегося в l-ом классе.

Основываясь на теории латентного анализа, можно говорить об однозначном разделении объектов по трём альтернативным показателям на два латентных класса и сформировать разрешимую систему уравнений с дискретными переменными [5,6]:

$$\begin{cases} \tilde{\phi}(x_1) + \tilde{\phi}(x_2) = 1, \\ p_1 = \tilde{f}_1(x_1)\tilde{\phi}(x_1) + \tilde{f}_1(x_2)\tilde{\phi}(x_2), \\ p_2 = \tilde{f}_2(x_1)\tilde{\phi}(x_1) + \tilde{f}_2(x_2)\tilde{\phi}(x_2), \\ p_3 = \tilde{f}_3(x_1)\tilde{\phi}(x_1) + \tilde{f}_3(x_2)\tilde{\phi}(x_2), \\ p_{12} = \tilde{f}_{12}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{12}(x_2)\tilde{\phi}(x_2), \\ p_{13} = \tilde{f}_{13}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{13}(x_2)\tilde{\phi}(x_2), \\ p_{23} = \tilde{f}_{23}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{23}(x_2)\tilde{\phi}(x_2), \\ p_{123} = \tilde{f}_{123}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{123}(x_2)\tilde{\phi}(x_2). \end{cases} \tag{1}$$

Определение. *Параметры латентно-структурной модели отношения p_i, p_{ij}, p_{ijk} называются маргиналами.*

Определение. *Латентно-структурная модель называется простейшей, если для её построения используются три альтернативных показателя.*

Утверждение. *Нахождение частоты $\tilde{\phi}(x_i)$ в простейшей модели латентно-структурного анализа сводится к каноническому уравнению прямой с точкой $(p1, p2, p3)$ и направляющим вектором*
$\vec{n} = \left(\tilde{f}_1(x_i) - \tilde{f}_1(x_j), \tilde{f}_2(x_i) - \tilde{f}_2(x_j), \tilde{f}_3(x_i) - \tilde{f}_3(x_j)\right)$, где $i \neq j$.

Доказательство. Выразим из уравнения (1) одну из частот (например, $\tilde{\phi}(x_2)$) и подставим в остальные три уравнения:

$$p_1 = \tilde{f}_1(x_1)\tilde{\phi}(x_1) + \tilde{f}_1(x_2)\left(1 - \tilde{\phi}(x_1)\right) = \tilde{\phi}(x_1)\left(\tilde{f}_1(x_1) - \tilde{f}_1(x_2)\right) + \tilde{f}_1(x_2),$$

$$p_2 = \tilde{f}_2(x_1)\tilde{\phi}(x_1) + \tilde{f}_2(x_2)\left(1 - \tilde{\phi}(x_1)\right) = \tilde{\phi}(x_1)\left(\tilde{f}_2(x_1) - \tilde{f}_2(x_2)\right) + \tilde{f}_2(x_2),$$

$$p_3 = \tilde{f}_3(x_1)\tilde{\phi}(x_1) + \tilde{f}_3(x_2)\left(1 - \tilde{\phi}(x_1)\right) = \tilde{\phi}(x_1)\left(\tilde{f}_3(x_1) - \tilde{f}_3(x_2)\right) + \tilde{f}_3(x_2).$$

Осуществив последовательно элементарные преобразования, получим требуемое:

$$\tilde{\phi}(x_1) = \frac{p_1 - \tilde{f}_1(x_2)}{\tilde{f}_1(x_1) - \tilde{f}_1(x_2)} = \frac{p_2 - \tilde{f}_2(x_2)}{\tilde{f}_2(x_1) - \tilde{f}_2(x_2)} = \frac{p_3 - \tilde{f}_3(x_2)}{\tilde{f}_3(x_1) - \tilde{f}_3(x_2)}.$$

■

В целях дальнейшего изложения теоретического аппарата введём следующие обозначения:

$$A_{ij} = \begin{pmatrix} p_{ij} & p_{i\bar{j}} \\ p_{\bar{i}j} & p_{\overline{ij}} \end{pmatrix},$$

$$A_{ij|k} = \begin{pmatrix} p_{ijk} & p_{i\bar{j}k} \\ p_{\bar{i}jk} & p_{\overline{ij}k} \end{pmatrix},$$

$$A_{ij|\bar{k}} = \begin{pmatrix} p_{ij\bar{k}} & p_{i\overline{jk}} \\ p_{\bar{i}j\bar{k}} & p_{\overline{ijk}} \end{pmatrix}.$$

Тогда определители вышеуказанных матриц равны $\|A_{ij}\| = p_{ij}p_{\overline{ij}} - p_{\bar{i}j}p_{i\bar{j}}$, $\|A_{ij|k}\| = p_{ijk}p_{\overline{ij}k} - p_{\bar{i}jk}p_{i\bar{j}k}$, $\|A_{ij|\bar{k}}\| = p_{ij\bar{k}}p_{\overline{ijk}} - p_{\bar{i}j\bar{k}}p_{i\overline{jk}}$ и называются произведением i-го и j-го показателей при условии (или без такового), что k-ый показатель равен 0 или 1.

В основе поиска неизвестных вероятностей простейшей модели можно воспользоваться нижеследующими таблицами совместных распределений:

Таблица 1. Совместное распределение двух альтернативных показателей i и j.

i-ый показатель / j-ый показатель	1	0	
1	p_{ij}	$p_{i\bar{j}}$	p_i
0	$p_{\bar{i}j}$	$p_{\overline{ij}}$	$1 - p_i$
	p_j	$1 - p_j$	

Таблица 2. Совместное распределение двух альтернативных показателей i-го и j-го при условии, что k-ый показатель равен 1.

i-ый показатель / j-ый показатель	1	0	
1	p_{ijk}	$p_{i\bar{j}k}$	p_{ik}
0	$p_{\bar{i}jk}$	$p_{\overline{ij}k}$	$p_k - p_{ik}$
	p_{jk}	$p_k - p_{jk}$	

Таблица 3. Совместное распределение двух альтернативных показателей i-го и j-го при условии, что k-ый показатель равен 0.

i-ый показатель / j-ый показатель	1	0	
1	$p_{ij\bar{k}}$	$p_{i\overline{jk}}$	$p_{i\bar{k}}$
0	$p_{\bar{i}j\bar{k}}$	$p_{\overline{ijk}}$	$1 - p_k - p_{i\bar{k}}$
	$p_{j\bar{k}}$	$1 - p_k - p_{j\bar{k}}$	

Лемма. *Отношение определителей матриц $\|A_{ij}\|$ и $\|A_{ij|k}\|$ равно произведению вероятностей $\tilde{f}_i(x_l)$ и $\tilde{f}_j(x_l)$.*

Доказательство. Анализ таблицы 1 и таблицы 2 позволяет сформировать следующий список вспомогательных равенств:

$$p_{\bar{i}j} = p_j - p_{ij};$$

$$p_{i\bar{j}} = p_i - p_{ij};$$

$$p_{\overline{ij}} = 1 - p_i - p_j + p_{ij};$$

$$p_{\bar{i}jk} = p_{jk} - p_{ijk};$$

$$p_{i\bar{j}k} = p_{ik} - p_{ijk};$$

$$p_{\overline{ij}k} = p_k - p_{ik} - p_{jk} + p_{ijk}.$$

Используя эти равенства, можно преобразовать определители $\|A_{ij}\|$ и $\|A_{ij|k}\|$ к следующему виду:

$$\begin{aligned}
\|A_{ij}\| &= p_{ij}p_{\overline{ij}} - p_{\bar{i}j}p_{i\bar{j}} \\
&= p_{ij}\left(1 - p_i - p_j + p_{ij}\right) - \left(p_j - p_{ij}\right)\left(p_i - p_{ij}\right) \\
&= p_{ij} - p_{ij}p_i - p_{ij}p_j + p_{ij}^2 - \left(p_jp_i - p_jp_{ij} - p_{ij}p_i + p_{ij}^2\right) \\
&= p_{ij} - p_{ij}p_i - p_{ij}p_j + p_{ij}^2 - p_jp_i + p_jp_{ij} + p_{ij}p_i - p_{ij}^2 = p_{ij} - p_ip_j;
\end{aligned}$$

$$\begin{aligned}
\|A_{ij|k}\| &= p_{ijk}p\overline{ij}k - p_{\bar{i}jk}pi\bar{j}k \\
&= p_{ijk}\left(p_k - p_{jk} - p_{ik} + p_{ijk})\right) - \left(p_{ik} - p_{ijk}\right)\left(p_{jk} - p_{ijk}\right) \\
&= p_{ijk}\left(p_k - p_{jk} - p_{ik} + p_{ijk}\right) - \left(p_{ik}p_{jk} - p_{ik}p_{ijk} - p_{ijk}p_{jk} + p_{ijk}^2\right) \\
&= p_{ijk}p_k - p_{ijk}p_{jk} - p_{ijk}p_{ik} + p_{ijk}^2 - p_{ik}p_{jk} + p_{ik}p_{ijk} + p_{ijk}p_{jk} - p_{ijk}^2 \\
&= p_{ijk}p_k - p_{ik}p_{jk}.
\end{aligned}$$

Подставив вместо маргиналов в данных определителях соответствующие выражения вероятностей $\tilde{f}_j(x_i)$ и частот $\tilde{\phi}(x_i)$ и осуществив элементарные преобразования, получим:

$$\|A_{ij}\| = p_{ij} - p_ip_j = \tilde{f}_{ij}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ij}(x_2)\tilde{\phi}(x_2) - \Big(\tilde{f}_i(x_1)\tilde{\phi}(x_1) + \tilde{f}_i(x_2)\tilde{\phi}(x_2)\Big) \times$$

$$\times \Big(\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{f}_j(x_2)\tilde{\phi}(x_2)\Big)$$

$$\begin{aligned}
&= \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) - \\
&- \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_j(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \\
&= \Big(\tilde{\phi}(x_1) + \tilde{\phi}(x_2)\Big)\Big(\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2)\Big) - \\
&- \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}^2(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1) \\
&- \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}^2(x_2) \\
&= \tilde{\phi}(x_1)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) + \tilde{\phi}(x_1)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) + \tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) \\
&+ \tilde{\phi}(x_2)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) - \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}^2(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \\
&- \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}^2(x_2) \\
&= \tilde{\phi}^2(x_1)\tilde{f}_i(x_1)\tilde{f}_j(x_1) + \tilde{\phi}(x_1)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) + \tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) \\
&+ \tilde{\phi}^2(x_2)\tilde{f}_i(x_2)\tilde{f}_j(x_2) - \tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}^2(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2) \\
&- \tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}^2(x_2)
\end{aligned}$$

$$= \tilde{\phi}(x_1)\tilde{f}_i(x_2)\tilde{f}_j(x_2)\tilde{\phi}(x_2) + \tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{f}_j(x_1)\tilde{\phi}(x_1) - \tilde{f}_i(x_1)\tilde{\phi}(x_1)\tilde{f}_j(x_2)\tilde{\phi}(x_2)$$
$$-\tilde{f}_i(x_2)\tilde{\phi}(x_2)\tilde{f}_i(x_1)\tilde{\phi}(x_1)$$
$$= \tilde{\phi}(x_1)\tilde{\phi}(x_2)\left(\tilde{f}_i(x_2)\tilde{f}_j(x_2) + \tilde{f}_i(x_1)\tilde{f}_j(x_1) - \tilde{f}_i(x_1)\tilde{f}_j(x_2) - \tilde{f}_i(x_2)\tilde{f}_i(x_1)\right)$$
$$= \tilde{\phi}(x_1)\tilde{\phi}(x_2)\left(\tilde{f}_i(x_2)\tilde{f}_j(x_2) + \tilde{f}_i(x_1)\tilde{f}_j(x_1) - \tilde{f}_i(x_1)\tilde{f}_j(x_2) - \tilde{f}_i(x_2)\tilde{f}_i(x_1)\right)$$
$$= \tilde{\phi}(x_1)\tilde{\phi}(x_2)\left(\tilde{f}_i(x_2)\left(\tilde{f}_j(x_2) - \tilde{f}_i(x_1)\right) - \tilde{f}_i(x_1)\left(\tilde{f}_j(x_2) - \tilde{f}_j(x_1)\right)\right)$$
$$= \tilde{\phi}(x_1)\tilde{\phi}(x_2)\left(\tilde{f}_i(x_2) - \tilde{f}_i(x_1)\right)\left(\tilde{f}_j(x_2) - \tilde{f}_j(x_1)\right).$$

$$\|A_{ij|k}\| = p_{ijk}p_k - p_{ik}p_{jk}$$
$$= \left(\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2)\right)\left(\tilde{f}_k(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\right) -$$
$$-\left(\tilde{f}_{ik}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\right)\left(\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{jk}(x_2)\tilde{\phi}(x_2)\right)$$
$$= \left(\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2)\right)\left(\tilde{f}_k(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\right) -$$
$$-\tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2)-$$
$$-\tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2)$$
$$= \tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2)+$$
$$+\tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2)$$
$$-\tilde{\phi}^2(x_1)\tilde{f}_{ik}(x_1)\tilde{f}_{jk}(x_1) - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2)-$$
$$-\tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{\phi}^2(x_2)\tilde{f}_{ik}(x_2)\tilde{f}_{jk}(x_2)$$
$$= \tilde{f}_k(x_1)\tilde{\phi}^2(x_1)\tilde{f}_{ijk}(x_1) + \tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2)+$$
$$+\tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1) + \tilde{f}_k(x_2)\tilde{\phi}^2(x_2)\tilde{f}_{ijk}(x_2)$$
$$-\tilde{\phi}^2(x_1)\tilde{f}_{ik}(x_1)\tilde{f}_{jk}(x_1) - \tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2)-$$
$$-\tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1) - \tilde{\phi}^2(x_2)\tilde{f}_{ik}(x_2)\tilde{f}_{jk}(x_2)$$
$$= \tilde{\phi}^2(x_1)\left(\tilde{f}_k(x_1)\tilde{f}_{ijk}(x_1)) - \tilde{f}_{ik}(x_1)\tilde{f}_{jk}(x_1)\right) +$$
$$+\tilde{\phi}^2(x_2)\left(\tilde{f}_k(x_2)\tilde{f}_{ijk}(x_2)) - \tilde{f}_{ik}(x_2)\tilde{f}_{jk}(x_2)\right) +$$
$$+\tilde{f}_k(x_1)\tilde{\phi}(x_1)\tilde{f}_{ijk}(x_2)\tilde{\phi}(x_2) + \tilde{f}_k(x_2)\tilde{\phi}(x_2)\tilde{f}_{ijk}(x_1)\tilde{\phi}(x_1)-$$
$$-\tilde{f}_{ik}(x_1)\tilde{\phi}(x_1)\tilde{f}_{jk}(x_2)\tilde{\phi}(x_2) - \tilde{f}_{ik}(x_2)\tilde{\phi}(x_2)\tilde{f}_{jk}(x_1)\tilde{\phi}(x_1)$$
$$= \tilde{f}_k(x_1)\tilde{f}_k(x_2)\tilde{\phi}(x_1)\tilde{\phi}(x_2)\left(\tilde{f}_{ij}(x_1) + \tilde{f}_{ij}(x_2) - \tilde{f}_i(x_1)\tilde{f}_j(x_2) - \tilde{f}_i(x_2)\tilde{f}_j(x_1)\right)$$
$$= \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|.$$

■

Теорема. *Наличие всех маргиналов в простейшей латентно-стуктурной модели позволяет свести поиск всех неизвестных вероятностей к решению трёх квадратных уравнений.*

Доказательство. В первую очередь обратим внимание на, то что из вышедоказанной леммы следует выполнение следующих двух равенств:

$$\|A_{ij|k}\| = \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|,\ \|A_{ij|\bar{k}}\| = \tilde{f}_{\bar{k}}(x_1)\tilde{f}_{\bar{k}}(x_2)\|A_{ij}\|.$$

Или в более наглядной форме

$$\|A_{ij|k}\| = \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|,$$
$$\|A_{ij|\bar{k}}\| = \Big(1 - \tilde{f}_k(x_1)\Big)\Big(1 - \tilde{f}_k(x_2)\Big)\|A_{ij}\|.$$

Раскрыв скобки и осуществив элементарные преобразования, получим следующий вид системы уравнений с двумя неизвестными:

$$\|A_{ij|k}\| = \tilde{f}_k(x_1)\tilde{f}_k(x_2)\|A_{ij}\|,$$
$$\|A_{ij|\bar{k}}\| = \Big(1 - \tilde{f}_k(x_1) - \tilde{f}_k(x_2) + \tilde{f}_k(x_1)\tilde{f}_k(x_2)\Big)\|A_{ij}\|.$$

Поделив обе стороны уравнений данной системы с двумя неизвестными на определитель $\|A_{ij}\|$, получим

$$\frac{\|A_{ij|k}\|}{\|A_{ij}\|} = \tilde{f}_k(x_1)\tilde{f}_k(x_2),$$
$$\frac{\|A_{ij|\bar{k}}\|}{\|A_{ij}\|} = \Big(1 - \tilde{f}_k(x_1) - \tilde{f}_k(x_2) + \tilde{f}_k(x_1)\tilde{f}_k(x_2)\Big).$$

Если осуществить подстановку первого уравнения во второе и разместить неизвестные в левой части уравнений, а известные – в правой, то получим классическую теорему Виета:

$$\tilde{f}_k(x_1)\tilde{f}_k(x_2) = \frac{\|A_{ij|k}\|}{\|A_{ij}\|},$$
$$\tilde{f}_k(x_1) + \tilde{f}_k(x_2) = \left(1 - \frac{\|A_{ij|\bar{k}}\|}{\|A_{ij}\|} + \frac{\|A_{ij|k}\|}{\|A_{ij}\|}\right),$$

где $\tilde{f}_k(x_1)$ и $\tilde{f}_k(x_2)$ – корни квадратного уравнения

$$x^2 - \left(1 - \frac{\|A_{ij|\bar{k}}\|}{\|A_{ij}\|} + \frac{\|A_{ij|k}\|}{\|A_{ij}\|}\right)x + \frac{\|A_{ij|k}\|}{\|A_{ij}\|} = 0.$$

Если же рассматривать каждый показатель как условный, то соответственно получим ещё два квадратных уравнения:

$$y^2 - \left(1 - \frac{\|A_{jk|\bar{i}}\|}{\|A_{jk}\|} + \frac{\|A_{jk|i}\|}{\|A_{jk}\|}\right)y + \frac{\|A_{jk|i}\|}{\|A_{jk}\|} = 0,$$

$$z^2 - \left(1 - \frac{\|A_{ik|\bar{j}}\|}{\|A_{ik}\|} + \frac{\|A_{ik|j}\|}{\|A_{ik}\|}\right) z + \frac{\|A_{ik|j}\|}{\|A_{ik}\|} = 0,$$

где $\tilde{f}_i(x_1)$ и $\tilde{f}_i(x_2)$ – корни первого уравнения, а $\tilde{f}_j(x_1)$ и $\tilde{f}_j(x_2)$ – второго.

Далее предполагается совместное использование латентной модели и ортогональной факторной структуры для построения алгоритма метода ФОРДИАСИМПТ вероятностного метода распознавания на базе альтернативных показателей.

Во-первых, необходимо сформулировать определение диагностической шкалы, симптомокомплекса и независимости симптомокомплексов в методе ФОРДИАСИМПТ.

Определение. *Диагностической шкалой называется набор апостериорных вероятностей, полученных с помощью простейшей латентно-структрной модели и формулы Баейса, позволяющей отнести объект исследования к одному из двух сформированных классов.*

Определение. *Симптомокомплекс – тройка альтернативных показателей, используемых для построения диагностической шкалы в методе ФОРДИАСИМПТ.*

Определение. *Два симптомокомплекса считаются зависимыми, если они содержат один и более общих параметров, в противном случае они независимы.*

Во-вторых, в целях упрощения дальнейшего изложения, введем функцию

$$\gamma_l(y_{i_k j}) = \begin{cases} \tilde{f}_i^{(k)}(x_l) \text{ если } y_{i_k j} = 1, \\ 1 - \tilde{f}_i^{(k)}(x_l) \text{ если } y_{i_k j} = 0, \end{cases}$$

где l – номер класса и может принимать значение 1 или 2, k – номер номер симптомокомплекса, $\tilde{f}_i^{(k)}(x_l)$ – вероятность значения 1 по i_k-му показателю у объекта из l-ого класса. Выбранный исследователем i_k-ый показатель входит в состав k-го симптомокомплекса, полученного на основании анализа факторной структуры.

Тогда вероятность принадлежности к первому классу можно определить посредством формулы Баейса с использованием введённой функции

$$P(1|y_{a_k j}, y_{b_k j}, y_{c_k j}) = \frac{\gamma_1(y_{a_k j})\gamma_1(y_{b_k j})\gamma_1(y_{c_k j})\tilde{\phi}_k(x_1)}{\sum_{i=1}^{2} \gamma_i(y_{a_k j})\gamma_i(y_{b_k j})\gamma_i(y_{c_k j})\tilde{\phi}_k(x_i)}, \tag{2}$$

где a_k, b_k, c_k – номера трёх параметров k-го симптокомплекса.

Алгоритм ФОРДИАСИМПТ:

1. Из матрицы Y путём элементарного преобразования получаем стандартизованную матрицу Z размерности $m \times n$ [1-3].

2. Вычисляем корреляционную матрицу R размерности $m \times m$ [1-3].

3. С целью исключения незначимых показателей вычисляем вероятностные значения уровней зависимости по формуле $\chi^2 = n \cdot \varphi$ при единичной степени свободы.

4. Определяем наименьшее количество выделяемых факторов (критерий Гуттмана, критерий «каменной осыпи» или другой адекватный критерий) [1-3].

5. Находим общности любым из известных методов (лучше взять метод минимальных остатков) [1-3].

6. Вычисляем первичную ортогональную матрицу весовых нагрузок факторов A размерности $m \times r$ (метод главных факторов, метод минимальных остатков или любой другой адекватный метод) [1-3].

7. Полученную на предыдущем шаге матрицу весовых нагрузок подвергаем ортогональномоу вращению в соответствии с варимакс критерием [2-3,7,8].

8. Осуществляем анализ ортогональной факторной структуры, полученной после вращения и формируем зависимые и независимые симптомокомплексы [3].

9. Для каждого симптомокомплекса формируем диагностическую шкалу, вычисляя маргиналы и решая систему уравнений (1), используя результаты теоремы.

10. По формуле (2) вычисляем частные апостериорные вероятности для всех объектов исследования.

Утверждение. *Алгоритм ФОРДИАСИМПТ используется для распознавания двух образов в пространстве двоичных признаков при совпадении количества выделенных измеряемых факторов и полученных независимых симптомокомплексов.*

Доказательство. В силу ортогональности выделяемых факторов получаем, что группы параметров, наполняющие тот или иной фактор, очень слабо коррелируют между собой. Тогда можно использовать условия независимости частных апостериорных вероятностей для получения формулы общей апостериорной вероятности:

$$P(1|y_{a_1j}, y_{b_{1j}}, y_{c_{1j}}, \dots, y_{a_rj}, y_{b_rj}, y_{c_rj}) = \prod_{i=1}^{r} P(1|y_{a_ij}, y_{b_ij}, y_{c_ij}). \tag{3}$$

■

Замечание. *Объект распознавания в случае независимых симптомокомплексов относится к первому классу, если общая апостериорная вероятность из формулы (3) меньше или равна 0,5, и ко второму классу – в противоположном случае.*

Вычислительный алгоритм КОРА

Алгоритм Кора (комбинаторного распознавания) — алгоритм классификации (взвешенного голосования правил), предложенный М. Вайнцвайгом и М. Бонгардом в 1973 г. [9-10], который применяется как метод распознавания двух образов в пространстве двоичных признаков путём построения и анализа набора конъюнктивных закономерностей.

Полагаем, что даны две таблицы «объект-свойство». Обозначим количество объектов исследования в каждой таблице за n и k соответственно, а количество измеряемых параметров за m — размерность выборки. Тогда исходные альтернативные данные представляются в виде таблиц, столбцы которой — объекты исследования, а строки — значения измеряемых параметров у конкретного объекта. На первом шаге алгоритма осуществляется объединение исходных таблиц в одну матрицу Y размерности $m \times (n+k)$ с сохранением информации о классовой принадлежности объектов. Далее просматриваются всевозможные матрицы Y_{ruv} размерности $3 \times (n+k)$, получаемые извлечением из матрицы Y, что составляет C_m^3 извлечения.

$$Y_{ruv} = \begin{pmatrix} y_{r1} & y_{r2} & \cdots & y_{rn} & | & y_{r(n+1)} & \cdots & y_{r(n+k)} \\ y_{u1} & y_{u2} & \cdots & y_{un} & | & y_{u(n+1)} & \cdots & y_{u(n+k)} \\ y_{v1} & y_{v2} & \cdots & y_{vn} & | & y_{v(n+1)} & \cdots & y_{v(n+k)} \end{pmatrix},$$

На следующем шаге среди первых n столбцов матрицы Y_{ruv} выделяются и фиксируются все тройки, не совпадающие ни с одной из троек в столбцах с $(n+1)$ по $(n+k)$. Аналогичная операция осуществляется со столбцами с $(n+1)$ по $(n+k)$. Из полученных двух совокупностей троек формируются два множества, которые обозначим за $A_{ruv} = \{(y_{ri}, y_{ui}, y_{vi})\}$ и $B_{ruv} = \{(y_{rj}, y_{uj}, y_{vj})\}$. Множества A_{ruv} и B_{ruv} называются характеристиками классов A и B. Эти характеристики формируются по всевозможным матрицам Y_{ruv}.

Основной этап алгоритма КОРА заключается в распознавании некоторого объекта $x = (x_1 \dots x_r \dots x_u \dots x_v \dots x_m)$ относительно классов A и B. Число совпадений $(x_r, x_u, x_v) = (y_{ri}, y_{ui}, y_{vi})$ обозначим $V(x, A)$ – число голосов, поданных для x за класс A, аналогично $V(x, B)$ – число голосов $(x_r, x_u, x_v) = (y_{rj}, y_{uj}, y_{vj})$. Если $V(x, A) > V(x, B)$, то объект относится к первому классу, если $V(x, A) < V(x, B)$, то к второму, при равенстве алгоритм отказывается от классификации.

Некоторые авторы предлагают вводить пороговое значение L. В этом случае если $V(x, A) - L > V(x, B)$, то объект относится к первому классу, если $V(x, A) < V(x, B) - L$, то к второму, при равенстве или нулевом значении голосов алгоритм отказывается от классификации.

Сравнительный анализ алгоритмов КОРА и ФОРДИА-СИМПТ на независимых симптомокомплексах

Отметим, что алгоритм ФОРДИАСИМПТ статистического типа, а алгоритм КОРА относится к категории эмпирических алгоритмов. В этой связи не имеет особого смысла сравнивать их на объёмах малых и средних выборок. На очень малых выборках алгоритм ФОРДИАСИМПТ нецелесообразно использовать из-за низкого уровня значимости коэффициента φ по χ^2 критерию. Вычислительные эксперименты на независимых симпотомокомплексах с средними выборками показали, что результаты обоих алгоритмов приблизительно одинаковы. А в том случае, когда обрабатывается выборка большого объёма, выявляются существенные расхождения. В качестве примера таких исходных данных для вычислительного эксперимента предлагаются альтернативные показатели с объёмом $n = 150$. Основные результаты работы алгоритма ФОРДИАСИМПТ представлены в нижеследующих таблицах (для краткости будем обозначать симптомокомплекс как С-комплекс).

Таблица 4. Матрица коэффициентов корреляции.

1	2	3	4	5	6
1	0,293	0,362	-0,0611	-0,0215	-0,0159
0,293	1	0,418	-0,173	0,0189	0,0586
0,362	0,418	1	-0,0161	0,0509	0
-0,0611	-0,173	-0,0161	1	0,23	0,207
-0,0215	0,0189	0,0509	0,23	1	0,39
-0,0159	0,0586	0	0,207	0,39	1

Таблица 5. Матрица значимости φ коэффициентов по χ^2 распределению.

1	2	3	4	5	6
1,000	0,999	0,999	0,545	0,207	0,154
0,999	1,000	0,999	0,167	0,183	0,527
0,999	0,999	1,000	0,156	0,466	0,000
0,545	0,167	0,156	1,000	0,995	0,988
0,207	0,183	0,466	0,995	1,000	0,999
0,154	0,527	0,000	0,988	0,999	1,000

При использовании алгоритма КОРА получены следующие характеристики классов:

Таблица 6. Матрица ортогонального факторного отображения после варимакс вращения.

	Фактор №1	Фактор №2
1	0,699	0,0441
2	0,769	0,012
3	0,785	-0,061
4	-0,198	-0,597
5	0,0517	-0,784
6	0,0505	-0,765

Таблица 7. Основные показатели симптомокомплекса №1.

Маргиналы	Значения частот и априорных вероятностей	Варианты ответов			Апостериорная вероятность
$p_1 = 0,266$	$\tilde{\phi}_1 = 0,770$	1	1	1	0,007
$p_2 = 0,166$	$\tilde{\phi}_2 = 0,230$	0	1	1	0,086
$p_3 = 0,3$	$\tilde{f}_1(x_1) = 0,048$	0	0	1	0,705
$p_{12} = 0,113$	$\tilde{f}_2(x_1) = 0,174$	0	0	0	0,992
$p_{13} = 0,153$	$\tilde{f}_3(x_1) = 0,096$	1	0	0	0,914
$p_{23} = 0,1$	$\tilde{f}_1(x_2) = 0,849$	1	1	0	0,294
$p_{123} = 0,08$	$\tilde{f}_2(x_2) = 0,564$	1	0	1	0,162
	$\tilde{f}_3(x_2) = 0,723$	0	1	0	0,837

Таблица 8. Основные показатели симптомокомплекса №2.

Маргиналы	Значения частот и априорных вероятностей	Варианты ответов			Апостериорная вероятность
$p_1 = 0,66$	$\tilde{\phi}_1 = 0,746$	1	1	1	0,995
$p_2 = 0,7$	$\tilde{\phi}_2 = 0,254$	0	1	1	0,972
$p_3 = 0,51$	$\tilde{f}_1(x_1) = 0,842$	0	0	1	0,486
$p_{12} = 0,546$	$\tilde{f}_2(x_1) = 0,858$	0	0	0	0,046
$p_{13} = 0,393$	$\tilde{f}_3(x_1) = 0,615$	1	0	0	0,220
$p_{23} = 0,406$	$\tilde{f}_1(x_2) = 0,125$	1	1	0	0,913
$p_{123} = 0,333$	$\tilde{f}_2(x_2) = 0,235$	1	0	1	0,857
	$\tilde{f}_3(x_2) = 0,214$	0	1	0	0,643

Таблица 9. Таблица исходных объектов исследования с частотой встречаемости в каждом классе.

Класс А						Частота	Класс В						Частота
0	1	1	1	1	1	0.015625	0	0	0	1	1	1	0.372093
0	0	0	1	0	0	0.078125	1	0	0	0	1	1	0.034884
1	1	1	1	1	1	0.046875	1	0	0	1	1	1	0.058140
1	0	1	1	0	1	0.046875	0	1	0	0	1	0	0.011628
0	0	0	0	0	0	0.171875	0	0	0	0	1	0	0.046512
1	0	0	0	0	1	0.015625	0	0	0	0	1	1	0.197674
0	0	0	0	0	1	0.093750	0	0	1	1	0	1	0.011628
1	0	0	1	0	0	0.015625	1	0	0	1	1	0	0.023256
0	1	0	0	0	1	0.031250	0	0	1	1	1	0	0.034884
1	1	1	0	1	1	0.093750	0	0	0	1	0	1	0.058140
1	0	1	0	0	1	0.015625	0	0	0	1	1	0	0.034884
1	0	1	0	0	0	0.031250	1	0	0	1	0	1	0.011628
1	1	0	0	1	1	0.015625	0	0	1	1	1	1	0.034884
1	0	0	0	0	0	0.078125	0	1	0	1	1	1	0.011628
1	0	1	1	1	1	0.062500	0	0	1	0	1	1	0.034884
1	1	1	0	0	0	0.015625	0	1	0	1	0	1	0.011628
1	1	0	0	0	1	0.015625	1	0	0	0	1	0	0.011628
1	1	0	1	1	1	0.015625							
1	1	1	0	1	0	0.015625							
0	1	1	0	1	1	0.015625							
0	0	1	0	1	0	0.015625							
0	1	1	0	0	1	0.015625							
0	1	1	0	0	0	0.031250							
0	0	1	1	0	0	0.015625							
1	1	1	1	1	0	0.015625							
1	0	1	0	1	1	0.015625							

$$A_{123} = \{(011), (111), (101), (110)\}, A_{124} = \{(111), (110)\}, A_{125} = \{(111), (110)\},$$
$$A_{126} = \{(111), (110)\}, A_{134} = \{(111), (110)\}, A_{135} = \{(111), (110)\},$$
$$A_{136} = \{(111), (110)\}, A_{145} = \{(000), (100)\}, A_{156} = \{(000), (100)\},$$
$$A_{234} = \{(111), (110)\}, A_{235} = \{(111), (110)\}, A_{236} = \{(111), (110)\},$$
$$A_{245} = \{(000), (100)\}, A_{246} = \{(101), (110)\}, A_{256} = \{(000), (100)\},$$
$$A_{345} = \{(000), (100)\}, A_{346} = \{(100)\}, A_{356} = \{(000), (100)\},$$
$$A_{456} = \{(100), (000), (001)\},$$
$$B_{135} = \{(001)\}, B_{235} = \{(001)\}, B_{236} = \{(100)\}, B_{245} = \{(110)\}, B_{356} = \{(010)\}.$$

В таблицах №11 и №12 вместо количества голосов в алгоритме КОРА для нераспознанных объектов стоят вопросы.

Т. о. при работе алгоритма КОРА имеются объекты, которые остались не распознанными, а алгоритм ФОРДИАСИМПТ позволяет распознать все объекты исследования. За исключением нераспознанных объектов, алгоритмы сработали одинаково. Для того чтобы понять, почему алгоритм КОРА не смог распознать некоторые объекты, достаточно сравнить нижеследующие распознанные объекты из класса A с первым из нераспознанных объектов из класса B.

Анализ данной таблицы показывает, что объект класса B не распознан в силу полного перекрытия тройками объектов класса A. Аналогичная ситуация наблюдается и у остальных нераспознанных объектов. Возможно, подобная ситуация редко встречается в геологии или сейсмологии [11-12], однако, в медицине и социологии этот случай встречается довольно часто [3,5].

Резюмируя изложенное, отметим, что преимуществом алгоритма ФОРДИАСИМПТ является возможность подтвердить или опровергнуть саму возможность разбиения объектов на два класса, основываясь на полученной статистической информации. Требование минимального пересечения классов является, например, ключевым для медицинских исследований [13,14].

Выводы

В рамках теории латенто-структурной модели сформулированы и доказаны:

1. Утверждение, позволяющее находить относительный объем соответствующего класса через каноническое уравнение прямой и утверждение, позволяющее использовать алгоритм ФОРДИАСИМПТ для распознавания двух образов в пространстве двоичных признаков при независимых симптомокомплексах.

2. Лемма о связи отношения определителей матриц $\|A_{ij}\|$ и $\|A_{ij|k}\|$ с произведением вероятностей $\tilde{f}_i(x_l)$ и $\tilde{f}_j(x_l)$

3. Теорема о сведении решения системы уравнений латентно-структурной модели к решению трёх квадратных уравнений.

На базе полученных теоретических выкладок построен и апробирован вычислительный алгоритм ФОРДИАСИМПТ, позволяющий строить диагностические симптомокомплексы на базе альтернативных данных, ортогональной

Таблица 10. Таблица исходных объектов класса А с результатом распознавания алгоритмами ФОРДИАСИМПТ и КОРА.

Класс А						Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Баейсовская вероятность	Число голосов в КОРА
0	1	1	1	1	1	0.0861	0.9952	0.0856	4
0	0	0	1	0	0	0.9924	0.2199	0.2182	4
1	1	1	1	1	1	0.0076	0.9952	0.0075	10
1	0	1	1	0	1	0.1622	0.8471	0.1374	4
0	0	0	0	0	0	0.9924	0.0458	0.0455	7
1	0	0	0	0	1	0.9138	0.4856	0.4437	4
0	0	0	0	0	1	0.9924	0.4856	0.4819	4
1	0	0	1	0	0	0.9138	0.2199	0.2009	4
0	1	0	0	0	1	0.8375	0.4856	0.4067	5
1	1	1	0	1	1	0.0076	0.9725	0.0073	11
1	0	1	0	0	1	0.1622	0.4856	0.0788	8
1	0	1	0	0	0	0.1622	0.0458	0.0074	12
1	1	0	0	1	1	0.2942	0.9725	0.2861	5
1	0	0	0	0	0	0.9138	0.0458	0.0419	7
1	0	1	1	1	1	0.1622	0.9952	0.1614	4
1	1	1	0	0	0	0.0076	0.0458	0.0003	18
1	1	0	0	0	1	0.2942	0.4856	0.1429	9
1	1	0	1	1	1	0.2942	0.9952	0.2928	4
1	1	1	0	1	0	0.0076	0.6427	0.0049	11
0	1	1	0	1	1	0.0861	0.9725	0.0837	5
0	0	1	0	1	0	0.7053	0.6427	0.4533	1
0	1	1	0	0	1	0.0861	0.4856	0.0418	9
0	1	1	0	0	0	0.0861	0.0458	0.0039	12
0	0	1	1	0	0	0.7053	0.2199	0.1551	4
1	1	1	1	1	0	0.0076	0.9134	0.0069	11
1	0	1	0	1	1	0.1622	0.9725	0.1577	4

Таблица 11. Таблица исходных объектов класса B с результатом распознавания алгоритмами ФОРДИАСИМПТ и КОРА.

Класс B						Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Баейсовская вероятность	Число голосов в КОРА
0	1	1	1	1	1	0.0861	0.9952	0.0856	4
0	0	0	1	0	0	0.9924	0.2199	0.2182	4
1	1	1	1	1	1	0.0076	0.9952	0.0075	10
1	0	1	1	0	1	0.1622	0.8471	0.1374	4
0	0	0	0	0	0	0.9924	0.0458	0.0455	7
1	0	0	0	0	1	0.9138	0.4856	0.4437	4
0	0	0	0	0	1	0.9924	0.4856	0.4819	4
1	0	0	1	0	0	0.9138	0.2199	0.2009	4
0	1	0	0	0	1	0.8375	0.4856	0.4067	5
1	1	1	0	1	1	0.0076	0.9725	0.0073	11
1	0	1	0	0	1	0.1622	0.4856	0.0788	8
1	0	1	0	0	0	0.1622	0.0458	0.0074	12
1	1	0	0	1	1	0.2942	0.9725	0.2861	5
1	0	0	0	0	0	0.9138	0.0458	0.0419	7
1	0	1	1	1	1	0.1622	0.9952	0.1614	4
1	1	1	0	0	0	0.0076	0.0458	0.0003	18
1	1	0	0	0	1	0.2942	0.4856	0.1429	9
1	1	0	1	1	1	0.2942	0.9952	0.2928	4
1	1	1	0	1	0	0.0076	0.6427	0.0049	11
0	1	1	0	1	1	0.0861	0.9725	0.0837	5
0	0	1	0	1	0	0.7053	0.6427	0.4533	1
0	1	1	0	0	1	0.0861	0.4856	0.0418	9
0	1	1	0	0	0	0.0861	0.0458	0.0039	12
0	0	1	1	0	0	0.7053	0.2199	0.1551	4
1	1	1	1	1	0	0.0076	0.9134	0.0069	11
1	0	1	0	1	1	0.1622	0.9725	0.1577	4

Таблица 12. Таблица объектов независимого контроля с результатом распознавания алгоритмами ФОРДИАСИМПТ и КОРА.

Класс В						Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Баейсовская вероятность	Число голосов в КОРА
1	1	0	0	0	0	0.2942	0.0458	0.0135	11(A)>1(B)
0	0	1	0	1	1	0.7053	0.9725	0.6859	?
0	0	1	0	0	1	0.7053	0.4856	0.3425	4(A)> 0(B)
1	1	1	0	1	1	0.0076	0.9725	0.0073	11(A)> 0(B)
0	0	0	1	1	1	0.9924	0.9952	0.9877	0(A)< 2(B)
0	0	0	0	1	1	0.9924	0.9725	0.9651	0(A)< 2(B)
0	0	0	0	0	0	0.9924	0.0458	0.0455	7(A)> 0(B)
1	0	0	1	1	0	0.9138	0.9134	0.8347	0(A)< 2(B)
0	0	0	0	1	0	0.9924	0.6427	0.6378	0(A)< 3(B)
1	1	1	1	1	1	0.0076	0.9952	0.0075	10(A)> 0(B)
1	0	1	1	1	0	0.1622	0.9134	0.1481	4(A)> 0(B)
1	0	0	0	1	0	0.9138	0.6427	0.5872	0(A)< 2(B)
0	1	1	1	0	0	0.0861	0.2199	0.0189	9(A)> 1(B)
0	0	1	1	0	0	0.7053	0.2199	0.1551	4(A)> 0(B)
0	1	0	1	1	1	0.8375	0.9952	0.8335	0(A)< 1(B)
0	1	1	1	0	1	0.0861	0.8471	0.0729	4(A)> 1(B)
1	0	1	0	0	0	0.1622	0.0458	0.0074	12(A)> 0(B)
1	0	0	0	0	0	0.9138	0.0458	0.0419	7(A) > 0(B)

Таблица 13. Первый нераспознанный объект класса B и объекты класса A, имеющие пересечение по тройкам.

Часть объектов класса А и нераспознанные объекты класса В						Вероятность в первом С-комплексе	Вероятность во втором С-комплексе	Баейсовская вероятность	Число голосов в КОРА
1	**0**	**1**	**1**	**0**	**1**	0.1622	0.8471	0.1374	4
0	**0**	**1**	0	1	0	0.7053	0.6427	0.4533	1
0	**0**	**1**	**1**	**0**	0	0.7053	0.2199	0.1551	4
0	1	**1**	**1**	1	**1**	0.0861	0.9952	0.0856	4
0	**0**	0	0	**0**	**1**	0.9924	0.4856	0.4819	4
0	1	0	0	**0**	**1**	0.8375	0.4856	0.4067	5
0	1	**1**	0	1	**1**	0.0861	0.9725	0.0837	5
0	**0**	**1**	**1**	**0**	**1**	0.7053	0.8471	0.5974	?

факторной структуры, простейшей латентно-структурной модели и формулы Баейса.

Проведён сравнительный анализ алгоритмов КОРА и ФОРДИАСИМПТ на выборке объёмом 150 объектов. Показана целесообразность применения алгоритма ФОРДИАСИМПТ для независимых симптомокомплексов при адекватной статистической информации.

Литература

1. Иберла К. Факторный анализ. М. : Статистика, 1989.
2. Харман Г. Современный факторный анализ. М. : Статистика, 1972.
3. Гольтяпин В.В. Реализация вычислительного алгоритма метода ФОРДИАСИМПТ на примере альтернативных показателей артериальной гипертензии // Современные наукоёмкие технологии 2014. № 11. С. 50–55.
4. Гольтяпин В.В. Вероятностный метод формирования симтомокомплексов. // Математические структуры и моделирование. 2014. № 4(32). С. 53–59.
5. Осипов Г.В. Методы измерения в социологии. М. : Наука, 2003.
6. Lazarsfeld P.F. The logical and mathematical foundation of latent structure analysis. 1950 In: Measurement and Prediction. N.Y.
7. Kaiser H.F. The varimax criterion for analytic rotation in factor analysis // Psyhometrika. 1958. № 23. С. 187–200.
8. Saunders D. The rationale for an “oblimax” method of transformation in factor analysis // Psyhometrika. 1961. № 26. С. 317–324.
9. Вапник В.Н. Алгоритм обучения распознаванию образов. М. : Советское радио, 1973.
10. Журавлев Ю.И. Математические основы теории прогнозирования (курс лекций). 2008, МГУ.

11. Платоненко И.М. Исследование и реализация алгоритмов распознавания по представительным наборам на базе решения специальных систем булевых уравнений: диссертация на соискание учёной степени кандидата физико-математических наук: 01.01.09. 1973. Москва.
12. Завьялов А.Д. Среднесрочный прогноз землетрясений по комплексу признаков: диссертация на соискание ученой степени доктора физико-математических наук: 25.00.10. 2003. Москва.
13. Гольтяпин В.В., Друк И.В., Нечаева Г.И. Возможности практической реализации принципа динамического наблюдения пациентов с недифференцированной дисплазией соединительной ткани и риском развития неблагоприятных сердечно-сосудистых проявлений // Справочник врача общей практики. 2014. № 19. С. 27–37.
14. Гольтяпин В.В., Шовин В.А. Косоугольная факторная модель артериальной гипертензии первой стадии // Вестник Омского университета. 2010. № 4. С. 120–128.

KORA AND FORDIASIMPT ALGORITHMS AS METHODS OF RECOGNITION OF THE TWO IMAGES IN THE SPACE OF BINARY FEATURES

V.V. Goltyapin
Ph.D.(Phys.-Math.), Associate Professor, Senior Reseacher, e-mail: goltyapin@mail.ru
V.A. Shovin
Researcher, e-mail: v.shovin@mail.ru

Omsk Branch of the Federal State budget institution Science Institute of Mathematics S.L. Soboleva of Siberian Branch of RAS

Abstract. In this article we proved the assertions, lemma and theorem allowing us to find posterior probabilities based on binary data of orthogonal factor structure. On the basis of theoretical calculations there was build FORDIASIMPT computational algorithm allowing us to build diagnostic symptom complexes based on probabilistic method of pattern recognition. A comparative analysis of FORDIASIMPT and KORA algorithms as a methods of recognition of the two images in the space of binary characters in the case of independent symptom complexes with adequate statistical information is conducted.

Keywords: symptom, factor model, the latent model, correlation analysis, KORA algorithm, the algorithm FORDIASIMPT.