

ПРИМЕНЕНИЕ ТЕОРИИ РАСПОЗНАВАНИЯ ОБРАЗОВ В ЗАДАЧАХ
ОЦЕНКИ РИСКОВ НАРУШЕНИЯ ФИНАНСОВОЙ И ИНФОРМАЦИОННОЙ
БЕЗОПАСНОСТИ

В современных условиях решение задачи оценки субъектов финансовой и информационной безопасности с точки зрения своевременного выявления рисков правонарушений является одним из приоритетных направлений исследований. Ниже приведен пример статистического оценивания благонадежности организаций на основе теории распознавания образов.

Согласно методологии распознавания образов, выбираются соответственно m_1 лучших и m_2 худших организаций. Лучшими в рамках настоящего исследования будем считать организации - члены Российского союза промышленников и предпринимателей, худшими - организации, ликвидированные по решению арбитражного суда. ликвидированные по решению арбитражного суда.

№	Признак	Организации - члены Российского союза промышленников и предпринимателей S_1					Организации, ликвидированные по решению суда S_2				
		$X_1^{(1)}$	$X_2^{(1)}$	$X_3^{(1)}$	$X_4^{(1)}$	$X_5^{(1)}$	$X_1^{(2)}$	$X_2^{(2)}$	$X_3^{(2)}$	$X_4^{(2)}$	$X_5^{(2)}$
1	Количество филиалов и дочерних компаний	2	6	0	11	17	0	0	0	0	0
2	Количество организаций, имеющих номер телефона, указанный в документах исследуемой компании	54	1	1	1	1	1	1	1	1	1
3	Количество организаций, в которых руководитель совмещает должность руководителя и учредителя	1	0	0	2	0	4	1	19	8	8
4	Количество компаний, которыми руководит директор исследуемой организации	4	3	2	4	2	6	2	20	8	12

Рассмотрим вначале признаковое пространство, состоящее из первых трех признаков.

Для групп предприятий S_1 и S_2 составим векторы средних (соответственно a_1 и a_2), а также их разность ($a_1 - a_2$) и сумму ($a_1 + a_2$):

$$a_1 = \frac{1}{m_1} \sum_{i=1}^{m_1} x_i^{(1)} = \begin{bmatrix} 7,2 \\ 11,6 \\ 0,6 \end{bmatrix}, a_2 = \frac{1}{m_2} \sum_{i=1}^{m_2} x_i^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 8 \end{bmatrix}.$$

$$a_1 - a_2 = \begin{bmatrix} 7,2 \\ 10,6 \\ -7,6 \end{bmatrix}, a_1 + a_2 = \begin{bmatrix} 7,2 \\ 12,6 \\ 8,6 \end{bmatrix}.$$



Вычислим ковариационные матрицы M_1 и M_2 , где m_1 и m_2 количество организаций - членов Российского союза промышленников и предпринимателей и организаций, ликвидированных по решению суда соответственно (здесь $m_1 = m_2 = 5$).

Ковариационная матрица M_1

$$M_1 = \frac{1}{m_1 - 1} \sum_{i=1}^{m_1} (X_i^{(1)} - a_1)(X_i^{(1)} - a_1)^T. \quad (1)$$

$$M_1 = \begin{pmatrix} 47,7 & -68,9 & 0,6 \\ -68,9 & 561,8 & 5,3 \\ 0,6 & 5,3 & 0,8 \end{pmatrix}.$$

Ковариационная матрица M_2

$$M_2 = \frac{1}{m_2 - 1} \sum_{i=1}^{m_2} (X_i^{(2)} - a_2)(X_i^{(2)} - a_2)^T. \quad (2)$$

$$M_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 46,5 \end{pmatrix}.$$

Найдем общую ковариационную матрицу M

$$M = \frac{1}{m_1 + m_2 - 2} * (m_1 M_1 + m_2 M_2) \quad (3)$$

$$M = \begin{pmatrix} 29,813 & -43,063 & 0,375 \\ -43,063 & 351,125 & 3,313 \\ 0,375 & 3,313 & 29,563 \end{pmatrix}$$

и обратную ковариационную матрицу M^{-1}

$$M^{-1} = \begin{pmatrix} 0,041 & 5,014 * 10^{-3} & -1,079 * 10^{-3} \\ 5,014 * 10^{-3} & 3,467 * 10^{-3} & -4,521 * 10^{-4} \\ -1,079 * 10^{-3} & -4,521 * 10^{-4} & 0,034 \end{pmatrix}.$$

Для определения достоверности D_3 вычислим расстояние Махалонбиса

$$d_3^2 = \frac{1}{2} * (a_1 - a_2)^T * M^{-1} * (a_1 - a_2)$$

$$d_3^2 = 2,6569, d_3 = 1,63.$$

Найдем σ_1 и σ_2 , где $m_1 = m_2 = 5$, а значение n для упрощения вычислений выберем минимальным $n=1$

$$\sigma_1^2 = \frac{1}{m_1} + \frac{1}{m_2}; \sigma_1^2 = \frac{1}{m_1} + \frac{1}{m_2} + \frac{4}{n}; \quad (4)$$

$$\sigma_1 = 0,632; \sigma_2 = 2,121.$$

Найдем вероятности ошибок распознавания $\alpha_3 = \beta_3$ по формуле

$$\alpha_3 = \beta_3 = F\left(-\frac{d_3}{\sigma_2}\right) + \left(\frac{\sigma_2}{d_3 \sqrt{2\pi}}\right) * \exp\left(-\frac{d_3^2}{2\sigma_2^2}\right) * \left[\left(\frac{\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)^{\frac{p-1}{2}} - 1\right].$$

Достоверность распознавания равна

$$D_3 = 1 - \alpha_3 = 1 - \beta_3 = 0,73.$$

Поскольку достоверность распознавания D_3 при трех признаках оказывается недостаточно высокой, добавим еще один показатель (N_4) - количество компаний, которыми руководит директор исследуемой организации.

Для групп предприятий S_1 и S_2 составим векторы средних (соответственно a_1 и a_2), а также их разность $(a_1 - a_2)$ и сумму $(a_1 + a_2)$

$$a_1 = \frac{1}{m_1} \sum_{i=1}^{m_1} x_i^{(1)} = \begin{bmatrix} 7,2 \\ 11,6 \\ 0,6 \\ 3 \end{bmatrix}, a_2 = \frac{1}{m_2} \sum_{i=1}^{m_2} x_i^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 8 \\ 9,6 \end{bmatrix}.$$

и вычислим ковариационные матрицы: M_1 и M_2 по формулам (1) и (2):

$$M_1 = \begin{pmatrix} 47,7 & -68,9 & 0,6 & -1 \\ -68,9 & 561,8 & 5,3 & 13,25 \\ 0,6 & 5,3 & 0,8 & 0,75 \\ -1 & 13,25 & 0,75 & 1 \end{pmatrix},$$

$$M_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 46,5 & 45,5 \\ 0 & 0 & 45,5 & 46,8 \end{pmatrix}.$$

Найдем общую ковариационную матрицу M по формуле (3):

$$M = \begin{pmatrix} 29,813 & -43,063 & 0,375 & -0,625 \\ -43,063 & 351,125 & 3,313 & 8,281 \\ 0,375 & 3,313 & 29,563 & 28,906 \\ -0,625 & 8,281 & 28,906 & 29,875 \end{pmatrix}$$

и обратную ковариационную матрицу M^{-1}

$$M^{-1} = \begin{pmatrix} 0,041 & 0,00489 & -0,011 & 0,009905 \\ 0,00489 & 0,003569 & 0,007521 & -0,008164 \\ -0,011 & 0,007521 & 0,656 & -0,637 \\ 0,009905 & -0,008164 & -0,637 & 0,653 \end{pmatrix}$$

Для определения достоверности D_4 вычислим расстояние Махаланобиса:

$$d_4^2 = \frac{1}{2} * (a_1 - a_2)^T * M^{-1} * (a_1 - a_2),$$

$$d_4 = 1,666.$$

Значения σ_1 и σ_2 совпадают с (4), т.к. $m_1 = m_2 = 5$ и $n = 1$ не изменяются.

Вероятность ошибок распознавания $\alpha_4 = \beta_4$ вычисляются по формуле

$$\alpha_4 = \beta_4 = F\left(-\frac{d_4}{\sigma_2}\right) + \left(\frac{\sigma_2}{d_4 \sqrt{2\pi}}\right) * \exp\left(-\frac{d_4^2}{2\sigma_2^2}\right) * \left[\left(\frac{\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)^{\frac{p-1}{2}} - 1\right],$$

$$D_4 = 1 - \alpha_4 = 1 - \beta_4 = 0,73.$$

Для установления оптимальной размерности ρ признакового пространства вычислим отношение расстояний Махаланобиса d_4^2 (при $\rho = 4$) и d_3^2 (при $\rho = 3$).

$$\frac{d_4^2}{d_3^2} = 1,045 < \sqrt{p} = \sqrt{4} = 2.$$

Т.к. $\frac{d_4^2}{d_3^2} = 1,045 < \sqrt{p} = \sqrt{4} = 2$, то оптимальная размерность признакового пространства $\rho = 4$.

Вычислим теперь обобщенный показатель благонадежности организации $\ln L$ при $n = 1$.

$$\ln L = \frac{1}{2} * (a_1 - a_2)^T * M^{-1} [2 X_i - (a_1 + a_2)],$$

$$\frac{1}{2} * (a_1 - a_2)^T * M^{-1} = (0,18 \ 0,036 \ -0,324 \ 0,197).$$

Организация №1.

№	Признак	Значение
1	Количество филиалов и дочерних компаний	120
2	Количество организаций, имеющих номер телефона, указанный в документах исследуемой компании	1

3	Количество организаций, в которых руководитель совмещает должность руководителя и учредителя	0
4	Количество компаний, которыми руководит директор исследуемой организации	1

X_i	X_i
$2 X_i - (a_1 + a_2)$	232,8
	-6,6
	-12,6
	1,706
$\ln L$	46,187

Организация №2.

№	Признак	Значение
1	Количество филиалов и дочерних компаний	0
2	Количество организаций, имеющих номер телефона, указанный в документах исследуемой компании	1
3	Количество организаций, в которых руководитель совмещает должность руководителя и учредителя	45
4	Количество компаний, которыми руководит директор исследуемой организации	35

X_i	X_i
$2 X_i - (a_1 + a_2)$	-7,2
	-6,6
	77,4
	69,706
$\ln L$	-12,884

Таким образом, с большой вероятностью можно утверждать, что Организация №1 является благонадежной, а Организация №2 — не благонадежной.

СПИСОК ЛИТЕРАТУРЫ:

1. Фомин Я. А., «Распознавание образов. Теория и применение»; изд. «Фазис», 2010 г.

