

УДК 004.8

UDC 004.8

5.2.2. «Математические, статистические и инструментальные методы в экономике» (физико-математические науки, экономические науки)

5.2.2. Mathematical, statistical and instrumental methods in economics (physical and mathematical sciences, economic sciences)

ОСОБЕННОСТИ ПРИНЯТИЯ РЕШЕНИЙ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ СФЕРЕ

FEATURES OF DECISION-MAKING IN THE SOCIO-ECONOMIC SPHERE

Ганичева Антонина Валериановна
к.ф.-м.н., профессор кафедры “Физико-математических дисциплин и информационных технологий”,
SPIN-код: 9049-4545, AuthorID: 177856
tgan55@yandex.ru
Тверская государственная сельскохозяйственная академия, ул. Василевского, дом 7, поселок Сахарово, Тверь, 17131, Россия

Ganicheva Antonina Valerianovna
Cand.Phys-Math.Sci., Professor of the Department of Physical and Mathematical Sciences and Information Technologies
tgan55@yandex.ru
Tver state agricultural academy Vasilevskogo, 7, pos. Sakharovo, Tver, 171314, Russia

Ганичев Алексей Валерианович
старший преподаватель кафедры “Информатики и прикладной математики”,
SPIN-код: 4747-0880, AuthorID: 178091
alexej.ganichev@yandex.ru
Тверской государственный технический университет, 170026, Тверь, наб. Аф. Никитина, дом 22, Россия

Ganichev Alexey Valerianovich
Associate Professor of the Department of Computer Science and Applied Mathematics
RSCI SPIN-code: 4747-0880, AuthorID: 178091
alexej.ganichev@yandex.ru
Tver State Technical University, Nab. Nikitina, 22, Tver, 170026, Russia

В статье разработан метод использования апостериорных вероятностей в качестве априорных вероятностей гипотез. Получены оценки необходимого числа наблюдений для обеспечения требуемой точности метода. Для пояснения метода рассмотрен поясняющий пример

The article develops a method for using a posteriori probability as a priori probabilities of hypotheses. Estimates of the required number of observations to ensure the required accuracy of the method are obtained. An explanatory example is considered to explain the method

Ключевые слова: РАСПОЗНАВАНИЕ, ОПЫТ, ПОКАЗАТЕЛЬ, ВЕРОЯТНОСТЬ, ГИПОТЕЗА, ТОЧНОСТЬ, ОШИБКА

Keywords: RECOGNITION, EXPERIENCE, INDICATOR, PROBABILITY, HYPOTHESIS, ACCURACY, ERROR

<http://dx.doi.org/10.21515/1990-4665-210-013>

Введение

Методы теории распознавания образов находят все большее применение во всех сферах жизни общества: построении систем вооружения, технических, экономических, социальных систем, а также гуманитарной сфере, все шире распространяясь в ней [2].

В технических системах и системах вооружения часто имеется ограничение на число возможных наблюдений для принятия решений о классах объектов. В отличие от этого, в социально-экономической сфере

<http://ej.kubagro.ru/2025/06/pdf/13.pdf>

принятие решений возможно в течение более длительного времени, и поэтому можно получить больший объем выборки. Поэтому предложенный в работе метод в большей степени предназначен для социально-экономической области.

Методы и материалы

Рассмотрим следующую ситуацию. Относительно события A имеются гипотезы H_1 и H_2 с неизвестными вероятностями $c = P(H_1)$ и $d = P(H_2)$. Известны условные вероятности $a = P(A / H_1)$ и $b = P(A / H_2)$.

Примем, что $P_1(H_1) = P_1(H_2) = 0,5$. Рассчитаем:

$$P(H_1 / A) = \frac{ca}{ca + db}; \quad P(H_2 / A) = \frac{cb}{ca + db};$$

$$P_1(H_1 / A) = \frac{0,5a}{0,5a + 0,5b} = \frac{a}{a + b}; \quad P_1(H_2 / A) = \frac{0,5b}{0,5a + 0,5b} = \frac{b}{a + b}.$$

Пусть условия опыта не изменяются, а в качестве априорные вероятности примем определенные на первом шаге апостериорные вероятности. Тогда:

$$P(A) = \frac{ca^2}{ca + db} + \frac{db^2}{ca + db} = \frac{ca^2 + db^2}{ca + db};$$

$$P_1(A) = \frac{a^2}{a + b} + \frac{b^2}{a + b} = \frac{a^2 + b^2}{a + b}.$$

Имеем:

$$P(H_1 / A) \cdot \frac{ca^2}{ca + db} : \frac{ca^2 + db^2}{ca + db} = \frac{ca^2}{ca^2 + db^2};$$

$$P(H_2 / A) = \frac{db^2}{ca + db} : \frac{ca^2 + db^2}{ca + db} = \frac{db^2}{ca^2 + db^2};$$

$$P_1(H_1 / A) = \frac{a^2}{a + b} : \frac{a^2 + b^2}{a + b} = \frac{a^2}{a^2 + b^2};$$

$$P_1(H_2 / A) = \frac{b^2}{a + b} : \frac{a^2 + b^2}{a + b} = \frac{b^2}{a^2 + b^2}.$$

При повторении опыта в одних и тех же условиях n раз на n -ом шаге имеем:

$$P(H_1 / A) = \frac{ca^n}{ca^n + db^n}; \quad P(H_2 / A) = \frac{db^n}{ca^n + db^n}; \quad P(H_2 / A) = \frac{db^n}{ca^n + db^n}; \quad P_1(H_2 / A) = \frac{b^n}{a^n + b^n}. \quad (1)$$

Примем, что $b < a$. Разделив числитель и знаменатель первой дроби на ca^n , второй – на db^n , третьей – на a^n , четвертый – на b^n , получим:

$$\begin{aligned} P(H_1 / A) &= 1 / \left(1 + \frac{d}{c} \left(\frac{b}{a} \right)^n \right); & P(H_2 / A) &= 1 / \left(1 + \frac{c}{d} \left(\frac{a}{b} \right)^n \right); \\ P_1(H_1 / A) &= 1 / \left(1 + \left(\frac{b}{a} \right)^n \right); & P_1(H_2 / A) &= 1 / \left(1 + \left(\frac{a}{b} \right)^n \right). \end{aligned}$$

При большом количестве опытов $(b/a)^n \rightarrow 0$. В этом случае:

$$P(H_1 / A) \approx P_1(H_1 / A). \quad (2)$$

Сумма апостериорных вероятностей равна 1, поэтому:

$$P(H_2 / A) = 1 - P(H_1 / A), \quad P_1(H_2 / A) = 1 - P_1(H_1 / A). \quad (3)$$

Таким образом, при многократном повторении опыта (в одних и тех же условиях) незнание исходных априорных вероятностей гипотез и принятие за них рассчитанных апостериорных вероятностей гипотез почти не сказывается на конечный результат. Определим необходимое для этого число опытов. Обозначим через ε ($0 < \varepsilon < 1$) точность конечного результата при многократных повторениях опыта. Рассмотрим выполнение условия:

$$\frac{d}{c} \left(\frac{b}{a} \right)^n < \varepsilon, \quad (4)$$

или

$$\left(\frac{b}{a} \right)^n < \frac{c}{d} \varepsilon. \quad (5)$$

Возьмем логарифм от обеих частей неравенства (5):

$$\ln\left(\frac{b}{a}\right)^n < \ln\left(\frac{c}{d}\varepsilon\right).$$

После преобразования получим:

$$n \ln \frac{b}{a} < \left(\frac{c}{d}\varepsilon\right).$$

Тогда

$$n > \ln\left(\frac{c}{d}\varepsilon\right) / \ln \frac{b}{a} = \left(\ln \frac{c}{d} + \ln \varepsilon\right) / \ln \frac{b}{a},$$

то есть

$$n > \log_{\frac{b}{a}} \frac{c}{d} + \log_{\frac{b}{a}} \varepsilon. \quad (6)$$

Рассмотрим изменение n в зависимости от соотношения переменных в формуле (6):

- 1) при $c > d$ будет выполняться неравенство $\log_{\frac{b}{a}} \frac{c}{d} < 0$;
- 2) при $\varepsilon > \frac{b}{a}$ имеем $\log_{\frac{b}{a}} \varepsilon < 0$.

Если выполняются оба записанные выше условия, то n может быть любым натуральным числом, при котором выполняются (2), (3), (6).

- 3) если $\varepsilon = \frac{b}{a}$, то $n > \log_{\frac{b}{a}} \frac{c}{d} + 1$. Может быть два случая данной ситуации:

3а) если $\log_{\frac{b}{a}} \frac{c}{d} + 1 \leq 0$, то $\log_{\frac{b}{a}} \frac{c}{d} + \log_{\frac{b}{a}} \frac{b}{a} \leq 0$ (так как $1 = \log_{\frac{b}{a}} \frac{b}{a}$); тогда

$$\log_{\frac{b}{a}} \frac{c}{d} \leq \log_{\frac{b}{a}} \frac{a}{b} \text{ и } \frac{c}{d} \geq \frac{a}{b}.$$

Если $\frac{c}{d} \geq \frac{a}{b}$, то $\log_{\frac{b}{a}} \frac{c}{d} + 1 \leq 0$, то n может быть любым (в том числе равным 1 или 2);

3б) при $\frac{c}{d} < \frac{a}{b}$ и для случая 4) $\varepsilon < \frac{b}{a}$ составляются специальные таблицы.

Специальная таблица (табл. 1) для определения n (при $\log_b \frac{c}{d} = k$),

Расчеты для $b/a = 0/5$ (табл. 2).

Таблица 1

ε	$(b/a)^2$	$(b/a)^3$	$(b/a)^4$	$(b/a)^5$	$(b/a)^6$
n	$k + 2$	$k + 3$	$k + 4$	$k + 5$	$k + 6$

Таблица 2

Значение показателя степени	2	3	4	5	6	7	8	9
Значение ε	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$	$\frac{1}{256}$	$\frac{1}{512}$

Точность $\varepsilon = \frac{1}{16}$ соответствует 6 % ошибки при замене вероятности

$P(H_1 / A)$ на вероятность $P_1(H_1 / A)$; $\varepsilon = \frac{1}{32}$ - 3 % ошибки; $\varepsilon = \frac{1}{64}$ - 2 %

ошибки, т.е. обеспечивается высокая точность вычисления.

При $b/a = s$ получаем табл. 3:

Таблица 3

Значение показателя степени	2	3	4	5	6	7	8	9
Значение ε	S^2	S^3	S^4	S^5	S^6	S^7	S^8	S^9

Второе слагаемое формулы (6) получается из уравнения по известным

$S = \frac{b}{a}$ и ε (x – неизвестный показатель степени):

$$S^x = \varepsilon.$$

Отсюда $x = \log_s \varepsilon$.

Для ε близкого к 0 $S < 1$, то максимальное значение x будет достигаться при значении S сколь угодно близким к 1 (например, при $S = 0,99$). В этом случае для $\varepsilon = 0,02$ получим $x = 9$. Значение $x = 9$ будет удовлетворять и большей точности, например, $\varepsilon = 0,006$.

Результаты и их обсуждение

Теперь получим оценку для максимального числа шагов n при различной точности $\varepsilon = (0,1)^m$, где $m=2,3$ и т.д., причем $b < a$.

Преобразуем формулу (6) к виду:

$$n > \frac{\ln(c/d)}{\ln(b/a)} + \frac{\ln \varepsilon}{\ln(b/a)} = \frac{\ln(c/d) + \ln \varepsilon}{\ln(b/a)}. \quad (7)$$

Так как $c = 1 - d$ и $b = 1 - a$, то выражение (7) будет:

$$n > \frac{\ln(c/(1-c)) + \ln \varepsilon}{\ln(b/(1-b))} \quad (8)$$

или

$$n > \frac{\ln c - \ln(1-c) + \ln \varepsilon}{\ln b - \ln(1-b)}. \quad (9)$$

В правой части выражения (9) стоит функция $f(c, b)$, зависящая от c и b .

Найдем ее на экстремум. Частные производные будут:

$$f'_c = \frac{1}{\ln b - \ln(1-b)} \times \left(\frac{1}{c} + \frac{1}{1-c} \right);$$

$$f'_b = -(\ln c - \ln(1-c) + \ln \varepsilon) \frac{1}{(\ln b - \ln(1-b))^2} \times \left(\frac{1}{b} + \frac{1}{1-b} \right).$$

Производная f'_c не существует в следующих случаях: 1) $c=0$, 2) $c=1$, 3) $b=0$, 4) $b=1$, 4) $\ln b - \ln(1-b) = 0$ (т.е. $\frac{1}{1-b} = 1$ и $b=0,5$). Эта производная в 0 не обращается.

Производная f'_b не существует в следующих случаях: 1) $b=0$, 2) $b=1$, 3) $b=0,5$, 4) $c=0$, 4) $c=1$. Данная производная равна нулю, если $\ln c - \ln(1-c) + \ln \varepsilon = 0$ (т. е. $\frac{1}{1-c} = -\varepsilon$), но этого не может быть т.к. $0 < c < 1$.

Во всех указанных точках функция $f(c, b)$ не существует. Рассмотрим значения c и b , близкие к указанным значениям, т.е. $c=0,0...01$ или $c=0,99...9$; $b=0,0...1$ либо $b=0,99...9$, либо $b=0,49...9$. Значение $c=0,99...9$ не подходит (этот случай рассматривался в п.1); значение $b=0,99...9$ тоже не подходит, поскольку по условию $b < a$; при $\varepsilon > (0,1)^m$, $m \geq 2$ значение $b=0,0...1$ не подходит, т. к. в этом случае

$$\frac{b}{a} = \frac{(0,1)^{m+1}}{\underbrace{(9...9)}_{m=1}} = (0,1)^{m+1}, m = 2, 3, \dots, \text{ а согласно п. 4) } \varepsilon < \frac{b}{a} \text{ (точность } \frac{b}{a}$$

берется на один разряд большая, чем у ε).

Итак, возможны 2 случая:

$$\text{а) } b = \underbrace{0,49...9}_m, c = 0,001 \times (0,1)^m, (m \geq 2);$$

$$\text{б) при } \varepsilon < (0,1)^m, (m \geq 2) b = (0,1)^{m+1}, c = (0,1)^{m+1}.$$

В случае а) формула (10) примет вид:

$$\begin{aligned} n &> \frac{\ln \frac{(0,1)^{m+1}}{0,9...9} + \ln(0,1)^m}{\ln \frac{0,49...99}{0,5...01}} = \frac{\ln(0,1)^{m+1} + \ln(0,1)^m}{\ln 0,9...96} = \frac{(2m+1)\ln 0,1}{\ln 0,9...96} = \\ &= \frac{-2,31(2m+1)}{\ln 0,9...96} = 2,31 \times (2m+1) \times 8 \times 8^m, \text{ т.е. в этом случае} \\ n &> 18,48 \times (2m+1) \times 8^m \end{aligned} \quad (10)$$

Рассмотрим случаи, когда $m=2$ и $m=3$.

При $m=2$ значение $18,48 \times 5 \times 8^2 \approx 5914$. При $m=3$ будет $18,48 \times 7 \times 8^3 \approx 66233$.

В случае б) формула (8) запишется в виде:

$$n > \frac{(2m+1) \ln 0,1}{\ln \frac{(0,1)^{m+1}}{0,99...9}} = \frac{(2m+1) \ln 0,1}{\ln(0,1)^{m+1}} = \frac{(2m+1) \ln 0,1}{(m+1) \ln 0,1} = 2 - \frac{1}{m+1},$$

т.е. n достаточно положить равным 2.

Пусть теперь $b/a = v$, где $0 < v < 1$. Тогда согласно (8) и с учетом (2) имеем:

$$n > \frac{-2,31(2m+1)}{\ln v} = \frac{2,37}{\ln(1/v)} \cdot (2m+1). \quad (11)$$

Значения n , при которых с заданной точностью выполняется условие (6), а тем самым и соотношения (2) и (3), для различных m и v приведены в табл. 4. В этой таблице на пересечении строки « m » и столбца « v » стоит соответствующее значение n .

Таблица 4

$m \backslash v$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
2	5	8	10	13	17	23	33	53	116
3	8	11	14	18	24	32	45	74	162
4	10	13	18	21	31	41	58	95	208

Заключение

Полученное значение числа опытов невелико для социально-экономических систем, при анализе надежности которых используются выборки объема 100 и большего. Для многих таких систем возможно моделирование опытов на ЭВМ. Дальнейшим развитием данных исследований является учет неопределенности в принятии решений [1, 3].

Литература

1. Ганичева А.В., Ганичев А.В. Адаптивный байесовский метод принятия решений // В мире научных открытий. 2014. № 2. С. 618.
2. Горелик А.А., Скрипкин В.А. Методы распознавания. М.: Высшая школа, 2004. 260 с.
3. Ganicheva A., Ganichev A.B. Bayesian approach to object recognition under the conditions of fuzzy information // Applied methods of statistical analysis. nonparametric approach. Proceedings of the international workshop. 2015. PP. 456-460.

References

1. Ganicheva A.V., Ganichev A.V. Adaptivnyy bayyesovskiy metod prinyatiya resheniy // V mire nauchnykh otkrytiy. 2014. № 2. S. 618.
2. Gorelik A.A., Skripkin V.A. Metody raspoznavaniya. M.: Vysshaya shkola, 2004. 260 s.
3. Ganicheva A., Ganichev A.B. Bayesian approach to object recognition under the conditions of fuzzy information // Applied methods of statistical analysis. nonparametric approach. Proceedings of the international workshop. 2015. PP. 456-460.