

Н.М. Єрещенко

Дніпропетровський національний університет ім. О. Гончара

ОЦІНКА ІНФОРМАТИВНОСТІ І ЗВ'ЯЗКУ ПАРАМЕТРІВ У СИСТЕМІ ГІДРОХІМІЧНОГО МОНІТОРИНГУ

Запропоновано процедуру відбору інформативних ознак параметрів у системі гідрохімічного моніторингу.

Постановка проблеми. У системі гідрохімічного моніторингу актуальним є відбір інформативних ознак параметрів для подальшої обробки та прийняття рішень. Це дозволяє вирішувати задачі оперативного аналізу стану природного середовища, кластеризації, тобто формування класів за якими потрібно здійснювати прийняття рішень і прогнозування. У сукупності ці задачі дозволяють здійснити картографування місцевості та оцінити за інформативними ознаками вплив хімічних елементів на природне середовище.

Аналіз досліджень і публікацій. Існують різні методи відбору інформативних ознак.

В [1] запропонований три-етапний метод відбору інформативних ознак, заснований на розрахунку як локальних так і глобальних признаков.

В [2] розроблені та експериментально досліджені критерії суміщеності інформативних ознак та описані підходи до їхнього формування. Експериментально підтверджена ефективність застосування запропонованих критеріїв для відбору інформативних параметрів моделювання.

Постановка задачі. Задані результати спостережень у системі гідрохімічного моніторингу у вигляді масиву:

$$\{x_{ij}^{(k)}, t_k; i = \overline{1, n}; j = \overline{1, m}; k = \overline{1, N}\},$$

де n – кількість хімічних елементів; m – кількість свердловин;

$x_{ij}^{(k)}$, – концентрація i -го елемента в j -й момент часу для k -ї свердловини.

Моніторинг проводиться на території Північного гірничо-збагачувального комбінату.

Необхідно запропонувати процедуру відбору інформативних ознак параметрів у системі гідрохімічного моніторингу.

Вклад основного матеріалу. Наявність великої кількості ознак, що характеризують досліджувані гідрохімічні процеси, обумовлює необхідність скорочення розмірності простору ознак. Під формуванням простору інформативних ознак розуміють вибір або формування найменшої кількості ознак.

Тому, для скорочення розмірності локальної бази даних для оперативного аналізу пропонується процедура відбору інформативних ознак [3]:

Нехай A -таблиця образів X розміру $m \times n$, яка складена з нулів та одиниць, та в якій k -ий рядок $(x_{k1}, x_{k2}, \dots, x_{kn})$ відповідає об'єкту k , l - стовпчик $(x_{1l}, x_{2l}, \dots, x_{ml})$ відповідає ознаці l . При цьому, таблиця A повинна задовольняти наступним вимогам:

- $m > 2, n > 2$
- в таблиці немає однакових рядків і стовпців, тобто:
$$X_{ki} \neq x_{ki+1},$$
$$X_{jl} \neq x_{jl+1},$$
де $i = \overline{1, n}, j = \overline{1, m}$
- в таблиці A немає рядків і стовпців, створених суціль з нулів та одиниць.

1. Числова міра (інформаційна вага) для об'єктів і ознак задається у вигляді позитивних векторів. Як початкова оцінка ваги рядків і стовпців приймаються вектори:

$$\vec{y} = \{y_1, y_2, \dots, y_m\},$$

$$\vec{z} = \{z_1, z_2, \dots, z_n\},$$

$$y_i = \sum_{j=1}^n x_{ij},$$

$$z_j = \sum_{i=1}^m x_{ij}.$$

2. Далі визначають пару векторів:

$$\overline{V} = \{v_1, v_2, \dots, v_m\},$$

$$\overline{W} = \{w_1, w_2, \dots, w_n\},$$

де

$$v_i = \sum_{j=1}^n x_{ij} z_j, \quad i = \overline{1, m},$$

$$w_j = \sum_{i=1}^m x_{ij} y_i, \quad j = \overline{1, n}.$$

3. Тоді на другому кроці досліджень отримують вектори:

$$\overline{y_2} = M_1 \times \overline{V},$$

$$\overline{z_2} = M_2 \times \overline{W},$$

$$M_1 = \max |v_i|,$$

$$M_2 = \max |w_j|,$$

де M_1, M_2 – нормуючі множники, а норми векторів

$$\|\overline{y_2}\| = \|\overline{z_2}\| = 1$$

4. Таким чином отримують послідовність векторів

$$\overline{y_1}, \overline{y_2}, \dots, \overline{y_k}, \dots$$

$$\overline{z_1}, \overline{z_2}, \dots, \overline{z_k}, \dots$$

Кожний подальший вектор – вага рядка або стовпця визначається через попередній вектор. Знайдені послідовності сходяться до деяких y_0, z_0 .

Розташування ознак за убаванням інформативної ваги відповідає впорядкованості за їхньою важливістю для діагностики. Розташування об'єктів за убаванням інформаційної ваги виділяє найбільш інформативні поєднання ознак або «класичних» представників образу.

Реалізація процедури інформативних ознак здійснюється для наступних хімічних елементів: Cl^- , SO_4^{2-} , HCO_3^- , Na^+ , Ca^{2+} , Mg^{2+} .

Проводимо формування бінарної таблиці (1) з урахуванням гранично допустимої концентрації (2).

Таблиця 1

Бінарна таблиця

N / Bin	SO	S	SO4	Mg	Ca
1	1	1	0	1	1
2	1	1	0	1	1
3	1	1	0	1	1
4	1	1	0	1	0
5	1	1	0	0	1
6	1	1	0	0	1
7	1	1	0	0	1
8	1	1	0	0	1
9	1	1	0	1	1
10	1	1	0	1	1
11	1	1	0	1	1
12	1	1	0	0	1
13	1	1	0	0	1
14	1	1	0	0	1
15	1	1	0	0	1
16	1	1	0	0	1
17	1	1	0	0	1
...
...
52	1	1	0	0	1
53	1	1	0	1	1
54	1	1	0	1	1
55	1	1	0	1	1
56	1	1	0	1	1
57	1	1	0	0	1

Таблиця 2

Гранично припустимі концентрації показників

Хім. ел.	<i>Cl</i>	<i>SO₄</i>	<i>HCO₃</i>	<i>Na</i>	<i>Ca</i>	<i>Mg</i>
	x_1	x_2	x_3	x_4	x_5	x_6
ГПК:	350	500	1104,1	200	180	40

Розташовуємо ознаки за убунанням інформативної ваги. Показники важливості для кожної ознаки, отримані за методом «гойдалки» зведені в таблицю (табл. 3).

Таблиця 3

Таблиця відбору інформативних ознак

Показатель	Cl	Ca	S	SO	Mg	SO4
Значимость	1,000	0,563	0,550	0,540	0,297	0,023

Розташування параметрів за убунанням інформаційної ваги зображено на діаграмі.

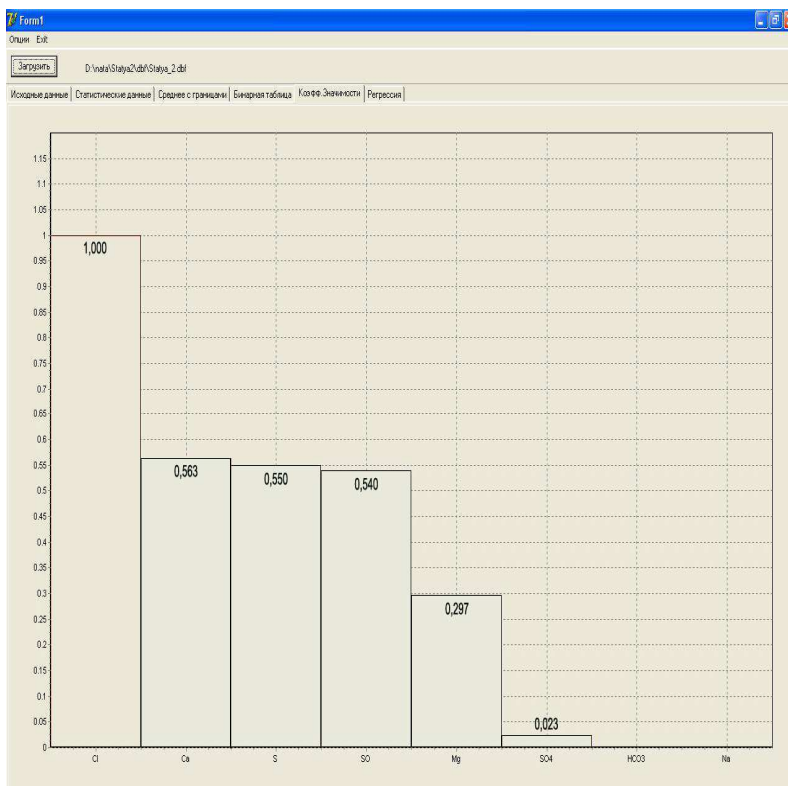


Рис. 1. Діаграма інформативної ваги

Були реалізовані дві моделі: лінійна та нелінійна модель.

Так для реалізації першої моделі на вхід поступають слідуєчі параметри:

Хлор (Cl^-) – X_1 ;

Кальцій (Ca^{2+}) – X_2 ;

Магній (Mg^{2+}) – X_3 .

Вклади змінних становлять:

X_1 : 734.14 : 2.09 %;

X_2 : 33263.14 : 94.87 %;

X_3 : 1066.16 : 3.04 %.

Рівняння лінії регресії для лінійної моделі

$$Y = A_0 + A_1 \cdot X_1 + A_2 \cdot X_2 + A_3 \cdot X_3$$

Таблиця 4

Оцінки параметрів (лінійна модель)

Оцінки параметрів лінії регресії		Довірчі інтервали та параметри	
		Нижній	Верхній
A_0	1297.20	635.37	1959.04
A_1	1.06	0	2.67
A_2	23.36	17.42	29.31
A_3	-3.51	-8.20	0

Середня похибка $E = 3.65$ % Модель адекватна у середньому.

Максимальна похибка $E = 20.17$ %

Критерій Дарбіна-Уотсона: 0.57 .

Автокореляція остатків присутня.

$S = 128.63$ $F = 1.49$ $Fa = 4.0$ Модель не значима.

де S – залишкова дисперсія

$$S^2 = \frac{1}{n-k-1} \sum_{i=1}^n \left[Y_i - \hat{Y}_i(X_{i1}, X_{i2}, X_{i3}, \bar{G}) \right]^2$$

Результати розрахунку приведені в таблиці 5.

Таблиця 5

Результати розрахунку для лінійної моделі

n / n	Остаток	\mathcal{E}_i	Y_i	Довірчі інтервали	
				Y_n	Y_e
1	183.24	2948.12	2764.87	2666.21	2863.53
2	270.72	2927.91	2687.18	2638.43	2735.93
3	63.06	2849.85	2786.78	2727.87	2845.69
4	-52.10	2915.89	2967.99	2894.20	3041.78
5	-108.97	2980.84	3089.81	3002.72	3176.90
6	10.10	2985.28	2975.17	2907.14	3043.20
7	11.02	2968.12	2957.09	2888.36	3025.82
8	99.03	2946.26	2847.22	2788.66	2905.77
9	74.04	2870.30	2796.25	2738.74	2853.75
10	136.70	804.69	2667.98	2616.88	2719.08
11	-3.88	2726.81	2730.69	2670.54	2790.84
12	-21.57	2732.14	2753.71	2695.13	2812.30
13	54.61	2881.17	2826.55	2756.96	2896.15
14	132.48	2947.52	2815.03	2743.34	2886.73
15	144.99	2908.98	2763.98	2700.44	2827.51
16	1333.06	2866.82	2733.75	2675.31	2792.20
17	-7.85	2860.41	2868.26	2815.33	2921.19
18	8.54	2880.93	2872.38	2819.36	2925.40
19	56.59	2921.16	2865.10	2814.86	2915.34
20	91.28	3012.85	2921.56	2840.85	300.26
21	-59.22	2977.57	3036.97	2959.22	314.72
...
...
55	75.33	2548.57	2473.23	2398.84	2547.62
56	38.04	2577.27	2539.22	2471.92	2606.51
57	-5.19	2661.67	2666.86	2609.38	2724.33

Для другої моделі (нелінійна) вклади змінних становлять:

X_1 : 0.00000329 : 0.15%,

X_2 : 0.0022 : 99.15%,

X_3 : 0.0000157 : 0,7%,

де X_1 – хлор (Cl⁻);

X_2 – кальцій (Ca²⁺);

X_3 – магній (Mg²⁺).

Рівняння ліній регресії для нелінійної моделі

$$Y = A_0 \cdot X_1^{A_1} \cdot X_2^{A_2} \cdot X_3^{A_3}.$$

Таблиця 6

Оцінки параметрів (нелінійна модель)

Оцінки параметрів ліній регресії		Довірчі інтервали та параметри	
		Нижній	Верхній
A_0	5.74	0	13972.32
A_1	0.1	0	34.08
A_2	0.63	0	125.98
A_3	-0.21	-99.33	0

Середня похибка $E = 98.15\%$ Модель не адекватна у середньому.

Максимальна похибка $E = 98.3\%$.

Критерій Дарбіна-Уотсона: $8.31 \cdot 10^{-4}$.

Автокореляція остатків присутня.

$S = 2714.6$ $F = 0.99$ $Fa = 4.0$ Модель не значима.

Результати розрахунку приведені в таблиці 7:

Таблиця 7

Результати розрахунку для нелінійної моделі

n/n	Залишок	\hat{Y}_i	Y_i	Довірчі інтервали	
				Y_n	Y_g
1	2896.82	2948.12	51.29	0	2133.39
2	2878.38	2927.91	49.52	0	1078.23
3	2798.56	2849.85	51.28	0	1294.48
4	2861.41	2915.89	54.47	0	1611.67
5	2924.31	2980.84	56.52	0	1894.42
6	2930.75	2985.28	54.52	0	1490.16
7	2913.81	2968.12	54.3	0	1504.72
8	2893.89	2946.26	52.36	0	1287.97
9	2818.76	2870.3	51.53	0	1265.01

Закінчення табл. 7

10	2755.52	2804.69	49.16	0	1127.54
11	2676.54	2726.81	50.26	0	1319.62
12	2681.44	2732.14	50.69	0	1286.94
13	2829.21	2881.17	51.95	0	1520.59
14	2895.77	2947.52	51.74	0	1564.67
15	2858.11	2908.98	50.87	0	1391.70
16	2816.47	2866.82	50.34	0	1283.76
17	2807.65	2860.41	52.75	0	1169.81
18	2828.11	2880.93	52.82	0	1171.69
...
...
55	2503.23	2548.12	45.33	0	1615.11
56	2530.62	2577.27	46.64	0	1466.84
57	2612.55	2661.67	49.11	0	1262.02

З аналізу випливає, що найбільш адекватна лінійна модель, що її показали розрахунки.

Висновки. 1. Запропоновано процедуру відбору інформативних ознак параметрів гідрохімічного моніторингу.

2. Найбільш інформативними показниками виявилися хлор, кальцій, мінералізація, сухий залишок, магній. Вага кожного з показників складає: хлор (Cl) – 1; мінералізація (S) – 0,55; сухий залишок (SO) – 0,54; магній (Mg) – 0,297.

Бібліографічні посилання

1. **Полякова М.П.** Об использовании метода «голосования» признаков в алгоритмах распознавания /М.П. Полякова, М. Н. Вайнцвайг// — Моделирование обучения и поведения. – М., 1975, С. 25–28.
2. **Голуб С.В.** Формування критерію відбору інформативних параметрів об'єктів моделювання в інформаційних системах багаторівневого моніторингу /С.В. Голуб // Математичні машини і системи. – 2007. – № 3, 4. – С. 218 – 226.
3. **Дмитриев А.М.** О математических принципах классификации предметов и явлений /А.М. Дмитриев, Ю.И. Журавлев, Ф.В. Кренделев // Дискретный анализ. – Новосибирск, 1966. – Вып.7. – С. 14 – 27.

Надійшла до редколегії 25.08.08