

И.В. СУРОВЦЕВ, д-р техн. наук, ст. науч. сотрудник,
отдел экологических цифровых систем,
Международный научно-учебный центр информационных технологий и систем НАН Украины
и МОН Украины, просп. Академика Глушкова, 40, Киев, 03187, Украина,
dep175@irtc.org.ua; igorsur52@gmail.com

О.В. БАБАК, канд. техн. наук, ст. науч. сотрудник,
отдел экологических цифровых систем,
Международный научно-учебный центр информационных технологий и систем НАН Украины
и МОН Украины, просп. Академика Глушкова, 40, Киев, 03187, Украина,
dep175@irtc.org.ua; babak@irtc.org.ua

КЛАССИФИКАЦИЯ ЗАГРЯЗНЕНИЯ УЧАСТКОВ ПОЧВЫ ТЯЖЕЛЫМИ МЕТАЛЛАМИ ПО РЕЗУЛЬТАТАМ КОМПЬЮТЕРНОГО ЭКСПЕРИМЕНТА

Разработан оригинальный алгоритм классификации участков загрязненной обширной области почвы. Рассматривая эти участки как класс, алгоритм позволяет по степени загрязнения разделить их на два подкласса. Эффективность алгоритма подтверждается результатами компьютерного эксперимента.

Ключевые слова: точное земледелие, тяжелые металлы, классификация, загрязненные участки почвы.

Введение

На современном этапе разработки прецизионных технологий в производстве растениеводческой продукции заметную роль играет точное земледелие [1]. Это прогрессивное направление в развитии земледелия связано с методами и средствами информационного обеспечения, в том числе с процедурами получения и интерпретации данных наземного, а также дистанционного зондирования Земли. При этом широко используются базы данных и знаний в процессе выработки и реализации оптимальных технологических решений. В настоящее время особую обеспокоенность вызывает увеличение загрязнения объектов окружающей среды тяжелыми металлами (ТМ) и в первую очередь эта проблема касается почвы.

Такое обстоятельство способствует попаданию ТМ в сельскохозяйственные культуры: в частности, зерно, овощи и фрукты, что наносит вред здоровью человека, а также животных. Поэтому первостепенной является задача наземного зондирования содержания ТМ в почве, а также последующее проведение агротехнических и агрохимических мероприятий по ее детоксикации [2].

В свою очередь такие действия предотвращают процесс деградации почвы и способствуют восстановлению ее плодородия, обеспечивая этим повышение урожайности сельскохозяйственных культур. Вполне естественно, что при организации процедуры исследования преследуется цель определенной оптимизации ее этапов. Она включает в себя минимизацию стоимости исследований и последующих дей-

ствий по детоксикации с учетом затрат материальных ресурсов.

В общем случае процедура исследования включает в себя три основных этапа:

- **этап 1** — разбиение объекта исследования, т.е. области загрязнения почвы, предназначенной для посева сельскохозяйственных культур, на отдельные участки;

- **этап 2** — измерение значений концентрации ТМ на участках;

- **этап 3** — анализ результатов исследования, включая их оценку.

При этом необходимо отметить следующее. Если этап 1 обычно не поддается формализации и зависит от специфических особенностей местности, то технические средства для эффективного выполнения этапа 2 есть.

Здесь имеется ввиду разработанный в Международном научно-учебном центре информационных технологий и систем НАН и МОН Украины прибор для измерения концентрации ТМ «Анализатор ИХП», обладающий в классе полярографических приборов более высокими в сравнении с известными метрологическими характеристиками. Такое преимущество достигнуто благодаря принципиально новым техническим решениям, предложенным в [3].

Менее освоенным с точки зрения формализации есть третий этап. Поэтому целью настоящей статьи является дальнейшее развитие в направлении формализации анализа результатов исследования.

Постановка задачи и ее решение

Используя компьютерный эксперимент, разработать принцип классификации результатов измерения концентрации ТМ на участках разбиения исследуемой области загрязнения почвы, а также подход к формальной оценке этих результатов.

Решение проблемы классификации результатов исследования при выполнении его этапов 1 и 2 на первый взгляд кажется простым и состоит в следующем.

Пусть исследуются значения x_{ij} концентраций четырех тяжелых металлов: x_{i1} — свинца

(Pb), x_{i2} — кадмия (Cd), x_{i3} — меди (Cu), x_{i4} — цинка (Zn), которые определены на множестве объектов $X_i \in X$. Таким образом, имеем выборку: $x_{ij} \in X_i$, $i = \overline{1, L}$, $j = \overline{1, N}$, где L — число участков (объектов), $N = 4$ — число измеряемых ТМ.

Стандартная задача классификации состоит в следующем: требуется разбить множество X_i на M подмножеств F_m , $m = \overline{1, M}$, чтобы были удовлетворены следующие два условия [4]:

1. Подмножества не пересекаются

$$F_m \cap F_k = 0, (m \neq k), m, k \in [1, M], \quad (1)$$

2. Любой элемент из X_i попадает в одно из подмножеств F_m , т.е.

$$X_i \in \bigcup_{m=1}^M F_m \quad (2)$$

и при этом каждое F_m состоит из *наиболее похожих* элементов.

Учитывая специфику подобной задачи, а именно то, что концентрация ТМ не должна превышать уровень ПДК (предельно допустимая концентрация металла — $C_{j\text{пдк}}$), нетрудно заметить следующее. Достаточно подготовленный оператор может легко разбить небольшое множество X_i на два класса, выделив класс, в котором имеет место $x_{ij} \geq C_{j\text{пдк}}$.

Однако, если множество X_i велико и нужно разбить его по уровню концентраций более чем на два класса, то такая задача становится уже затруднительной. При этом нельзя отбрасывать также и человеческий фактор.

В этой связи современный уровень развития системного анализа требует формализации процедуры классификации, применяя компьютерную обработку полученной информации. Обычно она заключается в том, что при сохранении условий классификации (1) и (2) выборку, представленную множеством X_i , разделяют на два подмножества. Запишем вышеуказанную выборку в виде:

$$x_{ij}, i = \overline{1, L}, j = \overline{1, N}. \quad (3)$$

Отметим, что стандартная задача классификации, при обучении распознаванию образов, за-

Код	Наименование	Свинец (Pb)	Кадмий (Cd)	Медь (Cu)	Цинк (Zn)
1	Объект_1	3.22	0.75	724.15	244.24
2	Объект_2	3.56	0.63	578.94	469.07
3	Объект_3	8.66	0.37	659.84	100.2
4	Объект_4	4.97	0.54	336	512.96
5	Объект_5	5.13	0.36	800.71	942.81
6	Объект_6	1.92	0.34	594.02	783.4
7	Объект_7	7.08	0.29	699.57	760.72
8	Объект_8	4.96	0.91	139.45	592.93
9	Объект_9	9.22	0.93	237.15	994.48
10	Объект_10	8.04	0.3	972.1	134.92
11	Объект_11	1	0.92	96.64	910.07
12	Объект_12	6.99	0.46	936.72	637.84
13	Объект_13	3.92	0.27	50.99	34.28
14	Объект_14	1.24	0.71	165.76	582.47
15	Объект_15	3.09	0.73	62.4	652.46
16	Объект_16	8.64	0.94	335.26	980.65
17	Объект_17	0.61	0.34	194.43	369.84
18	Объект_18	4.03	0.69	733.64	717.51
19	Объект_19	2.85	0.28	336.67	733.67
20	Объект_20	9.54	0.55	142.33	955.39
21	Объект_21	0.82	0.42	741.74	919.45
22	Объект_22	9.72	0.52	528.78	601.04
23	Объект_23	3.16	0.82	375.11	478.68
24	Объект_24	8.31	0.61	553	834.48
25	Объект_25	3.02	0.19	76.9	149.32
26	Объект_26	9.05	0.19	885.33	701.26
27	Объект_27	7.24	0.13	536.63	728.95
28	Объект_28	9.31	0.55	337.72	766.34
29	Объект_29	9.45	0.14	80.63	10.65
30	Объект_30	6.88	0.73	806.59	831.16

Рис. 1. Исходные значения концентраций, в мкг/дм³

ключается в отыскании гиперплоскости, разделяющей два множества X_1 и X_2 на два класса F_1 и F_2 (при условиях: $y = 1, x_{i1} \in F_1$ и $y_i = -1, x_{i2} \in F_2$).

Предлагается, в отличие от указанного подхода, рассматривать множество X_i и выборку (3) как один класс, в котором существует два подкласса [5]. При этом в качестве решающей функции для отыскания методом наименьших квадратов (МНК) гиперплоскости, разделяющей подклассы, используется

$$\sum_{j=1}^N a_j x_j = 1, \quad (4)$$

$$\text{т.е.} \quad \sum_{j=1}^N a_j x_{ij} = y_i, y_i = 1, i = \overline{1, L}, \quad (5)$$

где a_j — оценки коэффициентов параметрической идентификации.

Решающим правилом разделения на два подкласса, при обработке (5) МНК и упорядочения рассчитанных значений $\hat{y}(x)$, служит

$$1 - \hat{y}_i > 0, i_1 = \overline{1, l_1},$$

$$1 - \hat{y}_i < 0, i_2 = \overline{1, l_2} \text{ и } l_1 + l_2 = L \quad (6)$$

где l_1 и l_2 — длины выборок подклассов, полученные в процессе разделения.

Код	Наименование	Класс	Pb - Свинец	Cd - Кадмий	Cu - Медь	Zn - Цинк	Y	Y ₀	d = Y - Y ₀
25	Объект_25	A	0.302	0.190	0.077	0.149	1	0.353	+0.047
13	Объект_13	A	0.392	0.270	0.051	0.034	1	0.411	+0.589
17	Объект_17	A	0.061	0.340	0.194	0.370	1	0.444	+0.556
29	Объект_29	A	X 0.945	0.140	0.081	0.011	1	0.620	+0.380
19	Объект_19	A	0.285	0.280	0.337	0.734	1	0.680	+0.320
14	Объект_14	A	0.124	0.710	0.166	0.582	1	0.745	+0.255
15	Объект_15	A	0.309	0.730	0.062	0.652	1	0.808	+0.192
6	Объект_6	A	0.192	0.340	0.594	0.783	1	0.831	+0.169
11	Объект_11	A	0.100	X 0.920	0.097	X 0.910	1	0.901	+0.099
4	Объект_4	A	0.497	0.540	0.336	0.513	1	0.907	+0.093
27	Объект_27	A	0.724	0.130	0.537	0.729	1	0.921	+0.079
21	Объект_21	A	0.082	0.420	0.742	X 0.919	1	0.942	+0.058
23	Объект_23	B	0.316	X 0.820	0.375	0.479	1	1.011	-0.011
2	Объект_2	B	0.356	0.630	0.579	0.469	1	1.024	-0.024
8	Объект_8	B	0.496	X 0.910	0.139	0.593	1	1.051	-0.051
3	Объект_3	B	X 0.866	0.370	0.660	0.100	1	1.081	-0.081
7	Объект_7	B	0.708	0.290	0.700	0.761	1	1.117	-0.117
1	Объект_1	B	0.322	0.750	0.724	0.244	1	1.119	-0.119
20	Объект_20	B	X 0.954	0.550	0.142	X 0.955	1	1.132	-0.132
5	Объект_5	B	0.513	0.360	X 0.801	X 0.943	1	1.161	-0.161
28	Объект_28	B	X 0.931	0.550	0.338	0.766	1	1.191	-0.191
10	Объект_10	B	X 0.804	0.300	X 0.972	0.135	1	1.191	-0.191
18	Объект_18	B	0.403	0.690	0.734	0.718	1	1.230	-0.230
26	Объект_26	B	X 0.905	0.190	X 0.885	0.701	1	1.246	-0.246
22	Объект_22	B	X 0.972	0.520	0.529	0.601	1	1.266	-0.266
24	Объект_24	B	X 0.831	0.610	0.553	X 0.834	1	1.318	-0.318
12	Объект_12	B	0.699	0.460	X 0.937	0.638	1	1.331	-0.331
9	Объект_9	B	X 0.922	X 0.930	0.237	X 0.994	1	1.424	-0.424
16	Объект_16	B	X 0.864	X 0.940	0.335	X 0.981	1	1.455	-0.455
30	Объект_30	B	0.688	0.730	X 0.807	X 0.831	1	1.468	-0.468

Рис. 1. Классификация объектов

Вышеуказанное определяет новизну алгоритма решения поставленной задачи классификации.

Далее рассмотрен пример разделения выборки (3), полученной в результате компьютерного эксперимента, на два подкласса (рис. 1, 2). На рис. 1 показан результат работы алгоритма, где в таблице приведены значения концентраций x_{ij} для металлов свинец, кадмий, медь, цинк, полученные в процессе компьютерного эксперимента для 30 участков (объектов). Эти значения, нормированные относительно значений ПДК, приведены в таблице на рис. 2.

По этим данным, используя случайную обучающую выборку (первые 20 объектов), восстановлена зависимость (5)

$$\hat{y}(x) = 0,5084x_{Pb} + 0,6463x_{Cd} + 0,5751x_{Cu} + 0,2195x_{Zn},$$

где $x_{Pb} = x_{i1}$, $x_{Cd} = x_{i2}$, $x_{Cu} = x_{i3}$, $x_{Zn} = x_{i4}$.

Согласно алгоритму, описанному соотношениями (4), (5) и (6), осуществлена процедура классификации 30 объектов, результаты которой представлены на рис. 2, где обозначено $\hat{y} = y_0$.

Данные измерения концентрации в экологических объектах целесообразно представлять в виде интервалов $[x_{ij}^-; x_{ij}^+]$ возможных значений измерительного устройства [6], в которых нижняя x_{ij}^- и верхняя границы x_{ij}^+ экспериментальных данных определяются значением относительной ошибки ε_j согласно методике измерения концентрации данного химического элемента в объекте окружающей среды

$$\begin{aligned} x_{ij}^- &= x_{ij} - x_{ij}\varepsilon_j, \\ x_{ij}^+ &= x_{ij} + x_{ij}\varepsilon_j, \end{aligned} \quad (7)$$

где x_{ij} — измеренное значение концентрации j элемента на i объекте (участке); ε_j — относительная ошибка измерения j элемента, которая считается постоянной для всех объектов. При определении в почве ТМ (*Pb*, *Cd*, *Cu*, *Zn*) относительная ошибка измерения концентрации этих элементов одинакова и составляет $\varepsilon_j = 20\%$ [3]. Для нормированных данных нижняя граница предельно допустимой концентрации равня-

ется $x_{ij}^- = 0,8$, а верхняя граница $x_{ij}^+ = 1,2$.

Из рис. 2 следует, что данные, приведенные на рис. 1, разделены на два подкласса объектов: **A = 12, B = 18, полагая, что объекты с нормированным значением $x_{ij} \geq 0,8$ являются загрязненными, а $x_{ij} < 0,8$ — чистыми.** Следовательно, в подклассе А, согласно этому условию, такими являются три объекта, а в подклассе Б их число составляет 14.

Если возникает необходимость выделения в подклассах групп *похожих* объектов поступают аналогичным образом. Например, пусть в подклассе А с длиной выборки l_1 возникает задача разбиения его на две группы, в этом случае используем подобную решающую функцию (4)

$$\sum_{j=1}^N a_j x_{ij} = y_i, \quad y_i = 1, \quad i = \overline{1, l_1}.$$

Тогда решающее правило имеет вид:

$$1 - \hat{y}_{i,1} > 0, \quad i_{1,1} = \overline{1, l_{1,1}},$$

$$1 - \hat{y}_{i,2} < 0, \quad i_{1,2} = \overline{1, l_{1,2}} \quad \text{и} \quad l_1 = l_{1,1} + l_{1,2}.$$

При увеличении объема информации может возникнуть необходимость разделения полученной группы на подгруппы. В этом случае указанный порядок разделения не меняется.

Заключение

Приведен подход к решению задачи классификации участков области почвы загрязненной тяжелыми металлами, которая возникает в условиях точного земледелия. Инструмент классификации — разработанный оригинальный алгоритм разделения класса участков почвы по степени загрязнения их тяжелыми металлами на два подкласса. Подытоживая сказанное можно констатировать: эффективность разработанного алгоритма подтверждается результатами компьютерного эксперимента; применение предложенного алгоритма на реальных объектах позволит формализовать процедуру классификации участков загрязнения обширной области почвы, что важно при точном земледелии для оптимизации затрат по детоксикации.

ЛИТЕРАТУРА

1. Якушев В.В. Точное земледелие: теория и практика. СПб.: ФГБНУ АФИ, 2016. 364 с.
2. Мотузова Г.В., Безуглова О.С. Экологический мониторинг почвы: учебник. М.: Академический проект, Гаудеамус. 2007. 237 с.
3. Суровцев И.В., Галимов С.К., Татаринов А.Э. Информационная технология определения токсичных элементов в объектах окружающей среды. Кибернетика и вычислительная техника. 2018. № 9. С. 5–31.
4. Алгоритмы и программы восстановления зависимостей. Под ред. В.И. Вапника. М.: Наука, 1984. 816 с.
5. Бабак О.В., Татаринов А.Э. Об одном подходе к решению задач классификации в условиях неполноты информации Кибернетика и системный анализ. 2005. №6. С. 116–123.
6. Дивак М.П. Ідентифікація дискретних моделей систем з розподіленими параметрами на основі аналізу інтервальних даних: монографія Дивак, Н.П. Порплиця, Т.М. Дивак. Тернопіль: ВПЦ «Економічна думка ТНЕУ», 2018. 220 с. <http://dSPACE.tneu.edu.ua/handle/316497/30569>

Поступила 23.05.2019

REFERENCES

1. Yakushev, V.V., 2016. *Tochnoe zemledelie: teoriya i praktika*. Precision farming: theory and practice. Saint-Petersburg: FGBIU AFI, 364 p. (In Russian).
2. Motuzova, G.V., Bezuglova, O.S., 2007. *Ecological monitoring of soils*. M.: Academic Project; Gaudeamus, 2007. 237 p. (In Russian).
3. Surovtsev, I.V., Galimov, S.K., Tatarinov, A.E., 2018. "Information technology for determining the concentration of toxic elements in environmental objects". *Cybernetics and Information Engineering*, 191, pp. 5–31 (In Ukrainian).
4. *Algorithms and Programs for Reconstructing Dependencies*, 1984. By ed. V.I. Vapnik. M.: Nauka, 816 p. (In Russian).
5. Babak, O.V., Tatarinov, A.E., 2005. "An approach to classification problem solution under incompleteness of information". *Kibernetika i sistemny analiz*, 6, pp. 116–123. (In Ukrainian).
6. Dyvak, M.P., 2018. *Identification of discrete models of systems with distributed parameters on the basis of analysis of interval data: monograph* M.P. Dyvak, N.P. Porplytsya, T.M. Dyvak. Ternopil: VPTS «Ekonomichna dumka TNEU», 220 p. <http://dSPACE.tneu.edu.ua/handle/316497/30569>. (In Ukrainian).

Received 23.05.2019

I.V. Surovtsev, Dr (Eng.), Senior Researcher, Department of ecological digital systems, International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine, Acad. Glushkov ave., 40, Kiev, 03187, Ukraine, dep175@irtc.org.ua; igorsur52@gmail.com

O.V. Babak, PhD (Eng.), Senior Researcher, International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine, Acad. Glushkov ave., 40, Kiev, 03187, Ukraine, dep175@irtc.org.ua; babak@irtc.org.ua

CLASSIFICATION OF SOIL PLOTS POLLUTION WITH HEAVY METALS ACCORDING TO THE RESULTS OF A COMPUTER EXPERIMENT

Introduction. At the present stage of development of precision technologies in crop production begins to play a significant role precision farming. This progressive direction in the development of agriculture is associated with the methods and means of information support, including the procedures for obtaining and interpreting from the ground-based sensing, as well as remote sensing the Earth. Currently, particular concern is the increase in environmental pollution by heavy metals. Specified circumstance necessitates ground-based sensing of soil plots polluted with heavy metals, which requires the solution of the problem of their classification.

Purpose. The purpose of this article is to develop a principle for classifying the results of measuring the concentration of heavy metals on the sections of the fragmentation of the studied soil pollution area, as well as an approach to the formal assessment of these results, using a computer experiment.

Methods. An algorithm for solving the classification problem on the basis of presenting a sample of data as one class divided into two subclasses is developed.

Results. The effectiveness of the approach and the algorithm for solving the problem of classification of soil pollution with heavy metals is confirmed by a computer experiment.

Conclusion. The proposed approach and the algorithm for solving the problem of classifying plots soil polluted with heavy metals can be used for ground-based sensing of the soil area.

Keywords: *precise farming, heavy metals, classification, polluted plots of soil.*

И.В. Суровцев, доктор техн. наук, ст. науковий співробітник, відділ екологічних цифрових систем, Міжнародний науко-навчальний центр інформаційних технологій та систем НАН України та МОН України, просп. Академіка Глушкова, 40, Київ, 03187, Україна, dep175@irtc.org.ua; igorsur52@gmail.com

О.В. Бабак, канд. техн. наук, ст. науковий співробітник, відділ екологічних цифрових систем, Міжнародний науко-навчальний центр інформаційних технологій та систем НАН України та МОН України, просп. Академіка Глушкова, 40, Київ, 03187, Україна, dep175@irtc.org.ua; babak@irtc.org.ua

КЛАСИФІКАЦІЯ ЗАБРУДНЕННЯ ДІЛЯНОК ҐРУНТУ ВАЖКИМИ МЕТАЛАМИ ЗА РЕЗУЛЬТАТАМИ КОМП'ЮТЕРНОГО ЕКСПЕРИМЕНТУ

Вступ. На сучасному етапі розробки прецизійних технологій у виробництві рослинницької продукції починає помітну роль грати точне землеробство. Цей прогресивний напрямок у розвитку землеробства пов'язаний з методами і засобами інформаційного забезпечення, в тому числі з процедурами отримання та інтерпретації даних наземного, а також дистанційного зондування Землі. В даний час особливу стурбованість викликає зростання забруднення об'єктів навколишнього середовища важкими металами (ВМ). Зазначена обставина викликає необхідність наземного зондування забруднених ВМ ділянок ґрунту, при якому необхідно рішення задачі їх класифікації.

Мета статті – використовуючи комп'ютерний експеримент, розробити принцип класифікації результатів вимірювання концентрації ВМ на ділянках розбивки досліджуваної області забруднення ґрунту, а також підхід до формальної оцінки цих результатів.

Методи. Розроблено алгоритм вирішення задачі класифікації на основі подання вибірки даних в якості одного класу, поділеного на два підкласи.

Результати. Ефективність підходу і алгоритм вирішення задачі класифікації забруднення ділянок ґрунту ВМ підтверджується комп'ютерним експериментом.

Висновки. Запропонований підхід і алгоритм вирішення задачі класифікації забруднення ВМ ділянок ґрунту може бути використаний при наземному зондуванні області ґрунту.

Ключові слова: *точне землеробство, важкі метали, класифікація, забруднені ділянки ґрунту.*