

УДК 007:62-50

ЭКСПЕРИМЕНТЫ ПО ПЕРЕОТКРЫТИЮ
ПЕРИОДИЧЕСКОГО ЗАКОНА МЕНДЕЛЕЕВА
С ПОМОЩЬЮ ЭВМ

А.Н.Загоруйко, Н.Г.Загоруйко

У науки две главные цели:
предвидение и польза .

Д.И. Менделеев

Задачи обнаружения эмпирических закономерностей обычно связаны с отображением данных из исходного пространства описания X в некоторое другое пространство Y , более удобное для восприятия.

Пространство описания X может быть произвольным - любой размерности, с любым числом объектов наблюдения. Пространство же восприятия Y должно отвечать некоторым условиям, вытекающим из ограниченных возможностей человеческой системы восприятия информации. Одно из таких ограничений связано с размерностью пространства Y : она не должна быть большой, так как объектами в многомерном пространстве человек оперирует с большим трудом.

Например, если X - многомерное пространство симптомов, то Y - пространство небольшого числа групп признаков - синдромов или факторов, на языке которых проще и понятней формулируются те или иные медицинские закономерности. В другом случае X - исходное пространство большого числа характеристик объектов, а Y - одномерное пространство наименований распознаваемых образов. Распознать принадлежность объекта a к тому или иному образу - это значит отобразить его описание из пространства X в пространство Y .

Упрощение описания результатов наблюдения может делаться и путем уменьшения числа объектов в таблице "объект-свойство". Этот

прием используется, например, при замене множества объектов каждого образа одним "эталонным" объектом.

Иногда облегчить восприятие данных можно и не уменьшая количества объектов, а упорядочив их по одному или нескольким каким-нибудь свойствам. Если исходные свойства X при этом окажутся связанными простой зависимостью с этими новыми свойствами Y , то общие закономерности массива данных будут восприниматься легко. Для этого случая задача обнаружения закономерности может быть сформулирована так: построить пространство Y такое, чтобы связь между свойствами X объектов A , упорядоченных в пространстве Y , имела вид простой закономерности L .

Наиболее просто воспринимаются зависимости линейного типа. Если линейной зависимости между X и Y во всем диапазоне изменения их значений добиться не удается, можно вместо глобальной линейной зависимости попытаться использовать локальные линейные зависимости.

Сформулируем принцип "локальной линейной гладкости". В Δ -окрестности каждой точки u_0 по каждой j -й координате пространства восприятия Y свойства X объектов множества $A = \{a_1, a_2, \dots, a_n\}$ линейно (L) зависят от u_{0j} :

$$X(a'_0) = \frac{X(a'_1) + X(a'_2)}{2}, \quad (1)$$

где $X(a'_0)$ - свойства объекта a'_0 , расположенного в точке u_{0j} , $X(a'_1)$ - объекта a'_1 , расположенного в точке $u_{0j} - \Delta u_j$ ("левый сосед"), $X(a'_2)$ - объекта a'_2 , расположенного в точке $u_{0j} + \Delta u_j$ ("правый сосед").

Таблица I

A	x
a_1	5
a_2	1
a_3	3,2
a_4	7
a_5	7,9
a_6	2,2
a_7	6

Пользуясь критерием L , можно для некоторого множества A по свойствам X сконструировать свойства Y . Приведем простой числовой пример. Пусть множество A - набор 7 объектов, значения свойства x которых приведены в табл. I. Пространство восприятия Y будем строить в виде одномерного пространства с целочисленными значениями координаты y .

Поместим в точку $y = u_0$ в качестве объекта a'_0 , например, объект a_7 .

Подберем к нему ближайшего "левого соседа" (a'_1) по свойству x . Это, очевидно, будет объект a_1 . Из соотношения (I) можно найти прогнозные свойства ближайшего "правого соседа" (a'_2): $x(a'_2) = 2x(a'_0) - x(a'_1) = 12 - 5 = 7$. Этим объектом, очевидно, будет объект a_4 . Разместим эти три объекта (a_1, a_4, a_5) на оси u (рис. I). Теперь можно предсказать свойства x для "левого соседа" объекта a_1 и

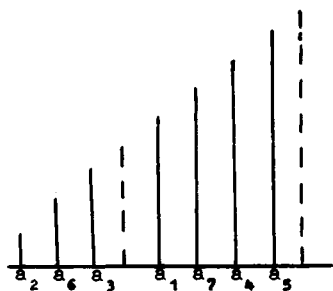


Рис. I

"правого соседа" объекта a_4 . Сосе - дом a_4 должен быть объект со свойст - вом $x(a'_2) = 2 \cdot 7 - 6 = 8$. Объекта с точ - но таким свойством нет. Ближайшим к этому значению является объект a_5 . Условимся, что разница $8 - 7,9 = 0,1$ меньше некоторого допустимого порога (δ) отклонения от линейности, и при - мем a_5 в качестве "соседа справа". Прогнозное свойство объекта $a_1 - x = 4$, ближайшее к этому значению $x = 3,2$ у объекта a_3 . Если $4 - 3,2 = 0,8$ больше

δ , то будем считать, что "левого соседа" для a_1 среди объектов множества A нет. Сделаем предположение, что в природе объект со свойством $x = 4$ существует и случайно не оказался среди объек - тов множества A . Поставим условно его левее a_1 (пунктир на рис. I) и продолжим процедуру. Прогноз "соседа слева" для услов - ного объекта дает результат $x = 3$. Объект a_3 может быть принят в качестве "левого соседа". Правым к объекту a_3 должен быть объект со свойством $x = 2 \cdot 7,9 - 7 = 8,8$. Такого объекта в табл. I нет, изоб - разим его условно справа от a_5 пунктиром. Соседом слева от a_3 будет a_6 , а его соседом - a_2 . Все элементы множества A отобра - жены в пространство Y и теперь легко видно, что если объект a_2 поместить в точку $x = 1$, то между x и u будет иметь место простая связь $x \approx u$.

Если объекты множества A обладают числом свойств большим, чем одно, то прогноз и подбор соседей нужно делать по всем этим свой - ствам. В случае двумерного пространства восприятия Y соседями объекта a'_0 будут "левый" (a'_1), "правый" (a'_2), "верхний" (a'_3) и "нижний" (a'_4). Связь их свойств определяется соотношением

$$x(a'_0) = \frac{x(a'_1) + x(a'_2) + x(a'_3) + x(a'_4)}{4}. \quad (2)$$

Процедура двумерного линейного упорядочения аналогична той, что описана выше для одномерного случая: выбор начального элемента a_0' и ближайших к нему двух соседей - "слева" a_1' и "сверху" a_3' . Затем прогноз соседей "справа" a_2' и "снизу" a_4' , подбор объектов, свойства которых отличаются от прогнозных не более чем на δ , установка их на свои места и прогноз соседей для всех установленных объектов. На каждом этапе устанавливается один новый объект, максимально похожий на прогнозируемый и делается новая серия прогнозов и проверок.

На этом принципе построен алгоритм ПАМИР, предназначенный для решения задач многомерного упорядочения. Его версия для двумерного пространства восприятия Υ реализована в виде программы на языке PASCAL для микро-ЭВМ "Электроника-60".

С помощью программы ПАМИР был проведен ряд экспериментов по двумерному упорядочению объектов множества A различной природы. На рис. 2 и 3 показаны исходные множества чисел и их наилучшие отоб-

y_2	y_1					
	I	2	3	4	5	6
I	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Множество A:

2, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6,
6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9,
9, 10, 10, 10, 10, 11, 11, 12

Рис. 2

y_2	y_1					
	I	2	3	4	5	6
I	I	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16		24
5	5	10		20	25	30
6	6	12	18	24	30	36

Множество A:

I, 2, 2, 3, 3, 4, 4, 4, 5, 5, 6, 6, 6,
8, 8, 9, 10, 10, 12, 12, 12, 15, 16,
18, 18, 20, 24, 24, 25, 30, 30, 36

Рис. 3

ражения на двумерное пространство Υ . Наилучшим считается такое упорядочение, при котором достигается минимум суммы локальных отклонений τ значений истинных свойств $X(a_1)$ всех m объектов a_1 от значений $X(a_1')$, предсказываемых по интерполяционной формуле (2):

$$\tau = \sum_{i=1}^n |X(a_i) - X(a'_i)|.$$

Закономерность, полученная в результате упорядочения массива А на рис.2, выражается формулой $x_{1j} = y_{11} + y_{2j}$, а на рис.3 - формулой $x_{1j} = y_{11} \cdot y_{2j}$.

Возможности программы ПАМИР были испытаны в ходе экспериментов по переоткрытию периодического закона Д.И.Менделеева.

Множество А было представлено химическими элементами первых семи рядов - с I-го (водород) по 54-й (ксенон). Пространство описания X включало в себя три свойства элементов: x_1 - атомный вес, x_2 - валентность по кислороду, x_3 - валентность по водороду.

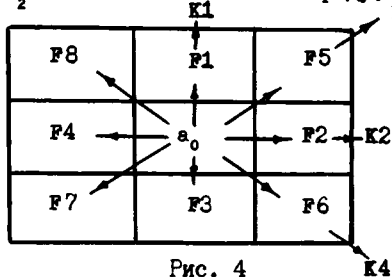


Рис. 4

В качестве пространства восприятия Y была взята плоскость с дискретными значениями переменных $y_1 = 1-25$, $y_2 = 1-25$. Относительно каждой клетки y_{11} , y_{2j} поля Y рассматривались "соседи" по восьми направлениям F1-F8 (рис.4). Программа может анализировать локальную гладкость свойства X вдоль

любого из четырех направлений K1-K4, по двум перпендикулярным направлениям: (K1, K2) или (K3, K4), или сразу по всем четырем направлениям K1-K4. В последнем случае проверяется равенство

$$X(a_0) = \frac{\sum_{i=1}^8 X(a_i)}{8} . \quad (3)$$

На рис.5 представлен результат работы программы ПАМИР на множестве А. В качестве начального элемента a_0 был взят элемент с номером 15 (фосфор). Рамкой показана граница таблицы в стандартном современном представлении [1]. Отклонения от стандартного вида имеют место для элементов длинного VI ряда III группы (элементы 44, 45, 46).

На рис.6 показан результат упорядочения множества элементов таблицы с I по 54 за исключением элементов с номерами 12, 22, 27, 28, 35, 39, 45, 46. Программа, начав с элемента № 17, расставила все элементы на те места, которые они занимают в таблице Менделеева, ос-

I	1								2
II	3	4	5	6	7	8	9	10	
III	11	12	13	14	15	16	17	18	
IV	19	20	21	22	23	24	25	26	27 28
V	29	30	31	32	33	34	35	36	45 51 44
VI	37	38	39	40	41	42	43	46	
VII	47	48	49	50	51	52	53	54	

Рис. 5

I	1								2
II	3	4	5	6	7	8	9	10	
III	11		13	14	15	16	17	18	
IV	19	20	21		23	24	25	26	
V	29	30	31	32	33	34		36	
VI	37	38		40	41	42	43	44	
VII	47	48	49	50	51	52	53	54	

Рис. 6

тавив пустые клетки в местах, соответствующих пропущенным элементам. Свойства пропущенных элементов просто определяются по формуле (2) или (3).

Как видно из этих примеров, если столбцам таблицы на рис. 5 и 6 сопоставить номера групп, а строкам - номера рядов, то результат программы ПАМИР почти совпадает с результатом, полученным Д.И. Менделеевым. Однако следует отметить различия в пространствах X, использованных Д.И. Менделеевым и программой ПАМИР.

Д.И. Менделеев, кроме атомного веса и валентностей, использовал большое число других свойств химических элементов: виды соединений с хлором, растворимость сернистых соединений, температуру плавления, цвет, запах летучих соединений и многое другое [2,3]. Все эти глубокие знания облегчали ему поиск "соседей" при упорядочении элементов. Следует отметить, что задача группировки многих элементов решалась еще и предшественниками Д.И. Менделеева - Ленсеном и Дюма. Так, было известно, что в одну группу следует относить элементы с современными номерами: 7-15-33-51; 9-16-34-52; 9-17-35-53; 12-20-38-56; и т.д. Подсказка такого типа существенно облегчила бы работу программы.

С другой стороны, машина использовала современные данные об атомных весах. До Менделеева атомные веса были известны не точно, Менделееву приходилось исправлять их путем предсказания на основании своей таблицы. Ошибки в значениях атомных весов могут существенно исказить упорядочение, если использовать такое бедное описание X, которое было в данных машинных экспериментах.

Результаты данной работы можно сформулировать так: принцип локальной линейной гладкости является эффективным средством отображения пространства описания X в пространство восприятия Y.

Свойством локальной линейной гладкости обладают пространства восприятия, построенные для многих известных эмпирических закономерностей количественного характера – законов Ома, Ньютона, Менделеева и др.

Если пространство исходного описания X наблюдаемых объектов множества A достаточно информативно, чтобы можно было построить пространство восприятия Y со свойством локальной линейной гладкости, то сам процесс построения пространства Y с помощью ЭВМ принципиального затруднения не вызывает. Это тем более справедливо для таких удачно найденных пространств X , для которых пространство Y удовлетворяет свойству глобальной линейной гладкости.

Л и т е р а т у р а

1. НЕКРАСОВ Б.В. Основы общей химии. – М.: Химия, 1973.
2. МЕНДЕЛЕЕВ Д.И. Периодический закон. Основные статьи. – М.: Изд-во АН СССР, 1958.
3. КЕДРОВ Б.М. Открытие галлия – первое химическое открытие нового типа. – В кн.: Прогнозирование в учении о периодичности. М. 1976.

Поступила в ред.-изд.отд.
23 марта 1984 года