

Л.Г. Ахметшина, К.А. Ахметшин

ВОССТАНОВЛЕНИЕ ПРОСТРАНСТВЕННЫХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ И САМООРГАНИЗУЮЩЕЙСЯ КАРТЫ КОХОНЕНА

Аннотация. Рассмотрены возможности метода восстановления пропусков пространственных данных, основанного на использовании алгоритма нечеткой кластеризации и нейронной сети Кохонена. Представлены результаты проверки метода на модельных и реальных данных.

Ключевые слова: нечеткая кластеризация, самоорганизующаяся карта Кохонена, пространственные данные, интерполяция, неравномерная сетка.

Введение. При решении задач анализа по конечному набору имеющихся экспериментальных данных, определенных на неравномерной сетке (например, это данные экологического мониторинга, результаты георазведки, физические характеристики земной поверхности или атмосферы и т. д.), обычно необходимо восстановление неопределенной информации – получение прогнозных значений в дополнительных точках, в частности, на полном координатном пространстве заданного масштаба. В случаях, когда плотность пробелов высока и данных немного, что исключает возможность использования статистических характеристик, современный подход, основанный на нейро-фаззи технологиях, сочетающий возможности обучаемости нейронных сетей с прозрачностью и интерпретируемостью систем нечеткого вывода, обладает преимуществом перед детерминистическими моделями [1, 2]. В работе [3] рассмотрены информационные возможности метода двумерного проецирования нечетких кластеров, предназначенного для интерполяции пространственных данных, заданных на неравномерной сетке, который основан на использовании метода нечеткой кластеризации. Получаемая модель отражает обобщенное представление о реальной поверхности, поскольку решение обычно соответствует

локальному, а не глобальному максимуму и обеспечивает достаточно хорошее согласование с данными во внешних областях. Существует множество факторов, влияющих на точность получаемого результата, одним из которых является способ выполнения этапов фаззификации и дефаззификации, для которого на сегодня не существует теоретически обоснованного правила выбора.

Целью данной работы является повышение точности модели, получаемой по неравномерно распределенным результатам эксперимента методом двумерного проецирования нечетких кластеров, за счет использования для вычисления значений функций принадлежности в дополнительных точках поверхности с применением самоорганизующейся нейронной сети Кохонена (СКК).

Основная часть. Постановка задачи представляется в следующем виде: существует набор пространственных данных (сеть мониторинга): X, Y – координаты, Z_i ($i=1, \dots, k$) – зависящие от них переменные (целевые значения), интерполяцию которого необходимо произвести, $k+2$ – общее количество информативных признаков, N – ограниченное количество точек измерения (что не допускает использование статистического подхода). Требуется построить модель, которая обеспечивает получение прогнозных значений в дополнительных точках, в частности, для полного координатного пространства. В работе [4] для решения задачи интерполяции пространственных данных рассматривалась СКК, которая посредством процесса, именуемого самоорганизацией, образует отображение многомерной информации на двумерную плоскость и обладает прогнозирующими свойствами [5].

СКК состоит из одного слоя, имеет l нейронов, соединенных в некоторую структуру. В двумерном случае эта структура представляет собой триангуляцию Делоне (рис. 1 а) $T = \{T^1, \dots, T^M\}$, где $T^k = \{T_1^k, T_2^k, T_3^k\}$, $k = \overline{1, M}$, M – количество треугольников, T_p^k – номер нейрона, $p = \overline{1, 3}$. Частными случаями является прямоугольная или гексагональная решетки (рис. 1 б, в) размером $l_x \times l_y = l$. Каждый вход соединен со всеми нейронами сети с весовыми коэффициентами $W = (w_{ik})_{i=1, l}^{k=1, n}$. Вектор W_i , составленный из столбцов матрицы W , определяет положение i -ого нейрона в n -мерном евклидовом простран-

ве. Для случаев прямоугольной или гексагональной решетки координаты выходных нейронов определяются двумя переменными, и матрицей весовых коэффициентов $W = (w_{ijk})_{i=\overline{1,l_1}, j=\overline{1,l_2}}^{k=\overline{1,n}}$, причем вектор $W_{ij} = (w_{ij}^1, \dots, w_{ij}^n)$ определяет положение (i, j) -ого нейрона в n -мерном евклидовом пространстве, количество l_x, l_y задано изначально. Эти параметры непосредственно влияют на точность аппроксимации данных СКК и время её обучения.

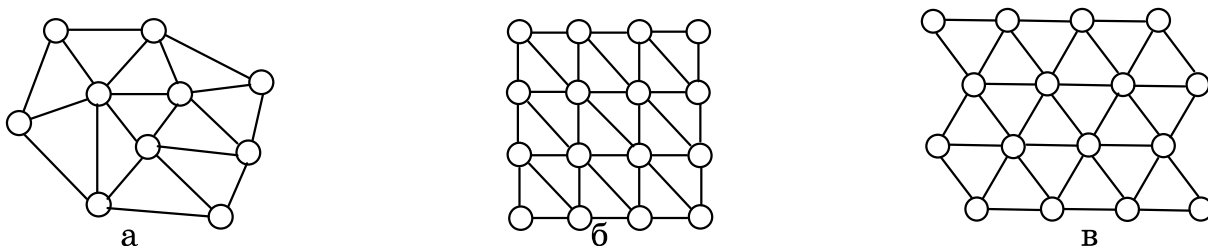


Рисунок 1 – Структуры двумерной СКК:

а – триангуляция Делоне, б – прямоугольная, в – гексагональная

В начале обучения значения матрицы \tilde{W} формируются случайным образом, обычно равномерно на области определения. Для прямоугольной и гексагональной СКК матрица $\tilde{W} = (w_{ijk})_{i=\overline{1,l_1}, j=\overline{1,l_2}}^{k=\overline{1,2}}$ определяется соответственно по формулам

$$\tilde{w}_{ij1} = i, \quad \tilde{w}_{ij2} = j, \quad i = \overline{1,l_1}, \quad j = \overline{1,l_2}, \quad (1)$$

$$\tilde{w}_{ij1} = \begin{cases} i, & \text{при } j:2 \\ i + 0,5 & \text{при } j \not: 2 \end{cases}, \quad \tilde{w}_{ij2} = j, \quad i = \overline{1,l_1}, \quad j = \overline{1,l_2}. \quad (2)$$

Для прямоугольной СКК триангуляция Делоне задается выражениями

$$\begin{cases} T^{2((j-1)l_1+i)} = \{(i, j), (i, j+1), (i+1, j)\}, \\ T^{2((j-1)l_1+i)+1} = \{(i+1, j+1), (i, j+1), (i+1, j)\}, \end{cases} \quad i = \overline{1, l_1 - 1}, \quad j = \overline{1, l_2 - 1}. \quad (3)$$

Для гексагональной СКК триангуляция Делоне следующим образом

$$\begin{cases} \begin{cases} T^{2((j-1)l_2+i)} = \{(i, j), (i, j+1), (i+1, j+1)\}, \\ T^{2((j-1)l_2+i)+1} = \{(i, j), (i+1, j), (i+1, j+1)\}, \end{cases} & \text{при } j:2 \\ \begin{cases} T^{2((j-1)l_2+i)} = \{(i, j), (i, j+1), (i+1, j)\}, \\ T^{2((j-1)l_2+i)+1} = \{(i, j+1), (i+1, j), (i+1, j+1)\}, \end{cases} & \text{при } j \not: 2 \end{cases} \quad (4)$$

$$i = \overline{1, l_1 - 1}, j = \overline{1, l_2 - 1}.$$

Для того, чтобы СКК не имела топографических дефектов и аппроксимировала многомерные данные за минимальное число итераций, нейроны равномерно располагаются на плоскости главных компонент внутри квадрата с вершинами

$$\{ \{ \alpha_{\min}, \beta_{\min} \}, \{ \alpha_{\max}, \beta_{\min} \}, \{ \alpha_{\max}, \beta_{\max} \}, \{ \alpha_{\min}, \beta_{\max} \} \}, \quad (5)$$

где $\alpha_{\min} = \min_{c=1, N} (A^1, X^c - A^0)$, $\alpha_{\max} = \max_{c=1, N} (A^1, X^c - A^0)$,

$$\beta_{\min} = \min_{c=1, N} (A^2, X^c - A^1 (A^1, X^c - A^0)), \quad \beta_{\max} = \max_{c=1, N} (A^2, X^c - A^1 (A^1, X^c - A^0)).$$

Таким образом, для двумерной СКК

$$W_i = A^0 + (\alpha_{\min} + \tilde{w}_{i1} (\alpha_{\max} - \alpha_{\min})) A^1 + (\beta_{\min} + \tilde{w}_{i2} (\beta_{\max} - \beta_{\min})) A^2, \quad i = \overline{1, l} \quad (6)$$

Для прямоугольной и гексагональной СКК

$$W_{ij} = A^0 + \left(\alpha_{\min} + \frac{\tilde{w}_{ij1} - 1}{l_1 - 1} (\alpha_{\max} - \alpha_{\min}) \right) A^1 + \left(\beta_{\min} + \frac{\tilde{w}_{ij2} - 1}{l_2 - 1} (\beta_{\max} - \beta_{\min}) \right) A^2, \quad (7)$$

$$i = \overline{1, l_1}, \quad j = \overline{1, l_2}.$$

В процессе обучения форма сети претерпевает существенное искажение (рис. 2), что затрудняет восстановление данных для равномерной координатной сетки.

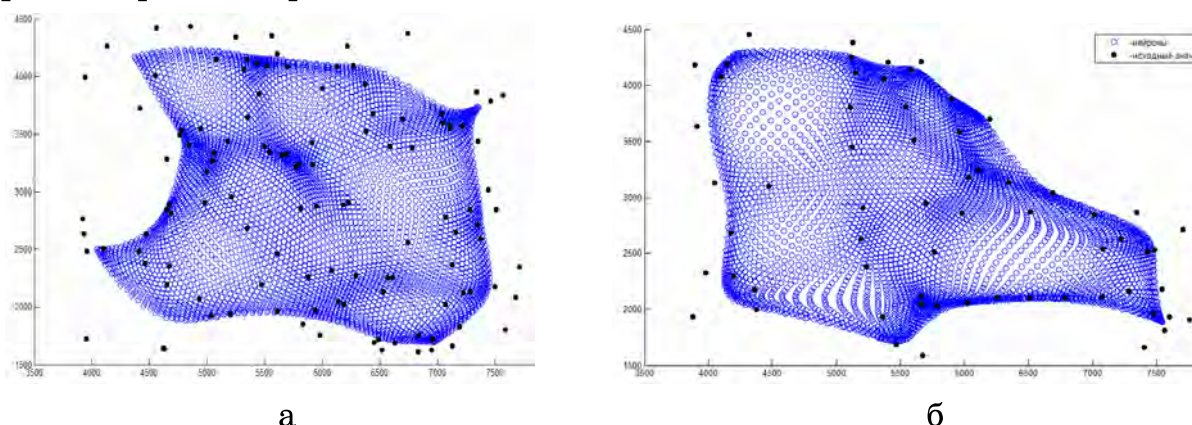


Рисунок 2 – Структура двумерной СКК после обучения и координаты исходных данных: а – генерированы случайным образом; б – положение точек измерений

Для достижения цели работы по восстановлению неизвестных целевых значений для всего координатного пространства на основе многомерных неравномерно распределенных экспериментальных значений предлагается следующий алгоритм.

1. Выполнение нечеткой кластеризации исходных данных на c классов. Результатом являются значения центров кластеров m_1, m_2, \dots, m_c и функции принадлежности μ_{ij}^c каждого образца с координатами x_i, y_j к ним.

2. Определение функций принадлежности отдельно для входных и целевых параметров (в практических задачах требуется определить целевое значение z_{ij} , при заданном значении координат x_i, y_j). Данная цель достигается построением проекций $m_i^{X_j}$ центров кластеров на координатную плоскость, где $j=1, \dots, n$, $i=1, \dots, c$, которые рассматриваются в качестве «существенных» точек, являющимися центрами функций принадлежности μ_{ij}^X каждого из классов.

3. Использование СКК, для интерполяции значений функций принадлежности в дополнительных точках с координатами, соответствующих положению нейронов. Входными значениями сети являются результаты, полученные в п.1, причем значения функций принадлежности для проекции точек центров кластеров m_1, m_2, \dots, m_c на координатную плоскость принимается равной 1.

4. Вычисление выхода нечеткой модели по значениям функций принадлежности для каждой точки координатного пространства поверхности

$$z(x_j) = \frac{\sum_{i=1}^c m_i^Y \cdot \mu_i^X(x_j)}{\sum_{i=1}^c \mu_i^X(x_j)} \quad (8)$$

Экспериментальные результаты. На рис. 3 а представлено модельное изображение и точки, которые генерировались случайным образом, на основе которых выполнялось восстановление неизвестных значений (в данном случае яркости) для всех координат. Количество входных данных в ходе эксперимента варьировалось в пределах 1-5 % от общего числа пикселей исходного изображения.

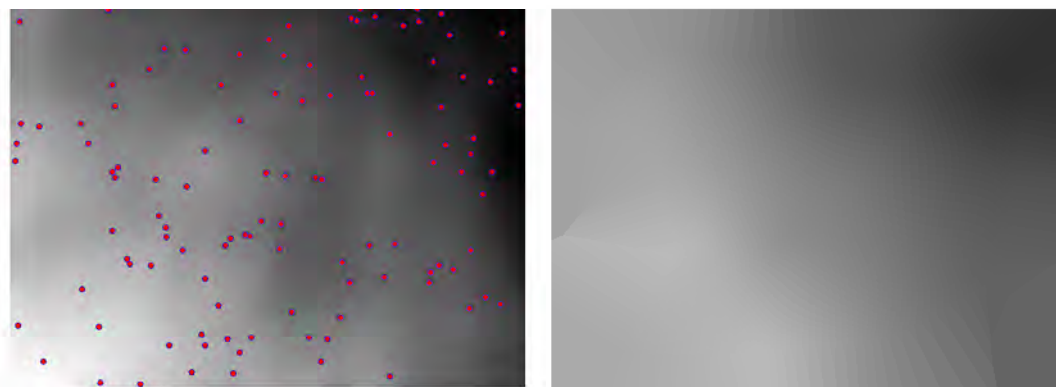
Число нейронов СКК определяется шагом дискретизации реальных исходных данных и диапазоном их изменения: $dx = \text{round}((\max(x) - \min(x))/h)$; $dy = \text{round}((\max(y) - \min(y))/h)$, где x, y – пространственные координаты, h – величина шага, зависящая от

масштаба измерения. На точность конечного результата существенное влияние оказывает количество исходных данных N , а также такие параметры алгоритма, как вид и значение функции соседства, количество кластеров при нечеткой кластеризации и нейронов СКК.

На рис. 3 б приведен результат восстановления яркости, который был получен при использовании следующих значений параметров алгоритма: СКК с прямоугольной структурой, гауссовой функцией соседства, радиус соседства изменялся в ходе обучения от 3-х до 1, количество нейронов сети $78*57$, количество кластеров при нечеткой кластеризации выбиралось в зависимости от количества исходных данных и при малом их значении $c = N$.

Полученный результат свидетельствует о достаточно хорошем качественном согласовании модели, что позволяет применять данный подход при оценочном анализе и прогнозе пространственных данных.

На рис. 4 а показан вид проекций двух произвольных функций принадлежности построенных с использованием СКК, на основе которых производилось восстановление неопределенных значений целевого параметра в соответствии с выражением (8). На рис. 4 б представлен результат моделирования поверхности залегания угольного слоя по сведениям из 57 разведочных скважин, месторасположение которых обозначены точками, минимальный шаг по координатам для них составляет значение 5.



а

б

Рисунок 3 - Модельные данные: а – расположение точек, используемых при интерполяции; б – восстановленное изображение

Выводы. Особенностью предложенного алгоритма является создание модели достаточной точности для проведения оценочного анализа и прогноза пространственных данных на основе неравномерно распределенных результатов измерений и возможность восстановления целевых значений для координатной плоскости различного масштаба.

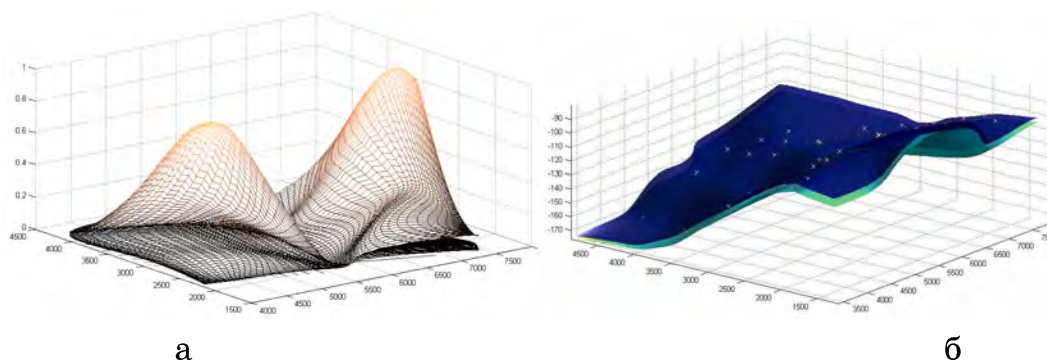


Рисунок 4 - Восстановление реальных данных:

а – проекции функций принадлежности на координатную ось;
б – поверхность угольного пласта

ЛИТЕРАТУРА

1. Рутковский Л. Методы и технологии искусственного интеллекта / Рутковский Л. – М., 2010. – 600 с.]
2. А. Егоров, Л. Ахметшина. Оптимизация яркости изображений на основе нейро-фаззи технологий / Монография. Изд. Lambert. – 2015. –139 с.
3. Ахметшина Л.Г. Интерполяция пространственных данных методом двумерного проецирования нечетких кластеров / Ахметшина Л.Г., Т.С. Ямнич // Искусственный интеллект, –2010. № 3. – С. – 433 –438.
4. Метод интерполяции неравномерных пространственных данных на основе сети Кохонена // Л.Г.Ахметшина, А.М.Ахметшин, И.М.Удовик Системні технології, –2013, –Дніпропетровськ.– Вип.1–(84).
5. Кохонен Т. Самоорганизующиеся карты / Кохонен Т. [пер. 3-го англ. изд. В.Н. Агеева под ред. Ю.В. Тюменцева]. – М.: Бином. Лаборатория знаний, 2008. – 665 с.