

Л.Г. Ахметшина, А.М. Ахметшин, И.М. Удовик
**МЕТОД ИНТЕРПОЛЯЦИИ НЕРАВНОМЕРНЫХ
ПРОСТРАНСТВЕННЫХ ДАННЫХ
НА ОСНОВЕ СЕТИ КОХОНЕНА**

Аннотация. Рассмотрены информативные возможности метода интерполяции пространственных данных заданных на неравномерной сетке, основанной на результатах кластеризации с использованием самоорганизующейся сети Кохонена. Представлены результаты проверки метода на модельных и реальных данных.

Ключевые слова: самоорганизующаяся карта Кохонена, параметры обучения, пространственные данные, интерполяция, неравномерная сетка.

Постановка проблемы. Построение моделей пространственного распределения различных свойств природных явлений требуется при решении прикладных задач, например, в геосистемах при оценке границ возможных зон подтопления и загрязнения, определения объемов рудного тела, прогнозом картировании.

Наиболее сложной является задача построения поверхности по малому конечному набору исходных значений с неравномерной плотностью сетки мониторинга, которые обычно и являются результатом реальных измерений: (результаты бурения, измерение уровня вод в контрольных скважинах и др). Использование традиционных математических подходов проблематично, а единственным способом оценки достоверности прогнозирования является эмпирическое тестирование на модельных данных.

Современный подход решения таких задач возможен на базе нечеткого и нейросетевого подходов [1, 2]. Преимущество нечетких моделей связано с возможностью использования малых объемов исходной информации, причем она может носить приближенный характер. В работе [3] демонстрируется возможность интерполяции экспериментальных пространственных данных заданных на неравномерной сетке методом двумерного проецирования нечетких кластеров.

Недостатком метода является значительная ошибка на границе области достоверности и сложность настройки вида функций принадлежности

Использование нейронных сетей обычно требует значительного объема исходных данных, которые необходимы для их обучения и тестирования. Однако, самоорганизующаяся нейронная сеть Кохонена (СКК) позволяет работать с небольшими объемами данных исходных данных и имеет ряд преимуществ перед другими архитектурами – отражает их топологические особенности, что особенно важно при обработке пространственных данных, и обладает прогнозирующими свойствами.

Целью данной работы является описание информативных возможностей метода интерполяции пространственных данных с использованием самоорганизующейся сети Кохонена, который позволяет в условиях малого объема нерегулярно расположенных исходных данных интерполировать поверхность на регулярную сетку произвольного масштаба.

Основная часть. На рис. 1 показана архитектура нейронной сети СКК – самоорганизующейся карты, которая является относительно новым методом вычислительного интеллекта [4].

(x_1, x_2, \dots, x_n) – входной вектор

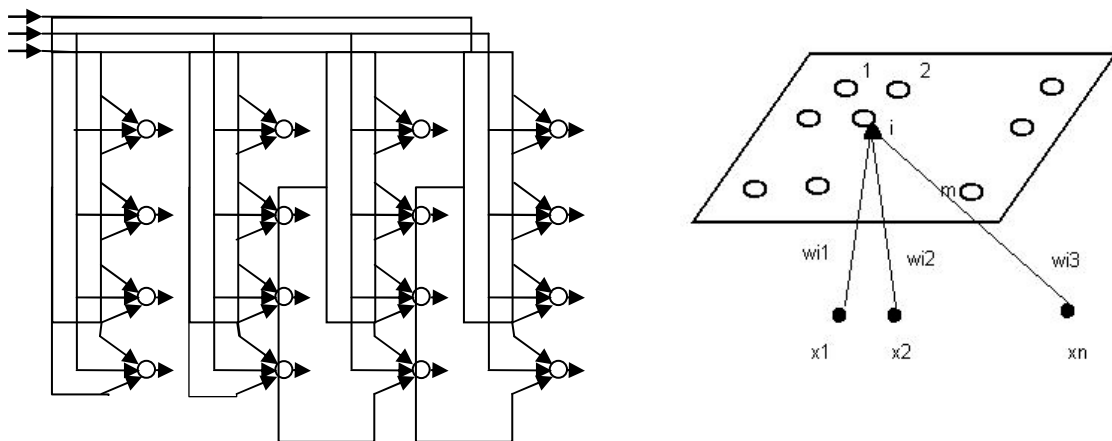


Рисунок 1 - Структурная схема карты Кохонена: $(n_x \times n_y -$ общее число узлов в карте; (x_1, x_2, \dots, x_n) – входы;

$W_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ – вектор весов для узла i

Нейроны сети образуют двумерную решетку и связаны со всеми узлами входного слоя: n_x, n_y – число узлов в карте по соответствующим координатным осям; (x_1, x_2, \dots, x_n) – исходные данные размерности n ; $W_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ – весовые коэффициенты i -го узла. При использовании метода «интеллектуальной» аппроксимации на основе данной архитектуры вид функциональной зависимости автоматически определяется данными и некоторыми заранее заданными ограничениями.

Обеспечение процедуры самоорганизации происходит за счет использования трех процессов:

конкуренция – рассчитывается значение дискриминантной функции, на основе которой определяется нейрон-победитель – имеющий наименьшее различие с рассматриваемым образцом;

кооперация – победивший нейрон определяет пространственное положение топологической окрестности нейронов-соседей, параметры которых подлежат корректировке;

синаптическая адаптация – усиление значений нейронов-соседей посредством корректировки их синаптических весов (самоусиление) (рис. 2).

Кратко структуру алгоритма можно представить в следующем виде [1].

1. Задание значений n_x, n_y и инициализация, например, случайным образом, начальных значений исходных векторов синаптических весов w_j для каждого нейрона m_j .

2. Выбор вектора X из входного пространства.

3. Поиск нейрона-победителя (ближайшего узла) m_c для вектора X с использованием критерия подобия. Для евклидова расстояния, он определяется по формуле $\arg(\min \|x - w_j\|), j = 1, \dots, k$, где $k = n_x \times n_y$ – количество нейронов сети.

4. Коррекция синаптических весов всех нейронов-соседей (рис. 2): $w_j(t+1) = w_j(t) + \eta(t)h_{j,i(x)}(t)(x - w_j(t))$, а для остальных узлов $w_j(t+1) = w_j(t)$, где t – шаг итерации; $0 < \eta(t) < 1$ – параметр скорости обучения, убывающий с каждым шагом; $h_{i,j(x)}$ – функция окрестности с центром победившего нейрона, определяющая силу взаимного влияния любой пары узлов сети, т.е. расстояние, на которое переда-

ется коррекция. Эти параметры динамически изменяются во время обучения с целью получения лучшего результата.

5. Если разность целевых функций текущего и предыдущего меньше ε , где ε – пороговое значение, то переход к пункту 2.

После построения СКК на нее отображаются входные данные. Результат – исходные многомерные данные аппроксимируются к $n_x \times n_y$ узлам сети, что позволяет получить проекцию всего пространства наблюдений на двумерную плоскость и отобразить важные статистические характеристики входного пространства, которые можно использовать при решении задачи прогноза неопределенных значений.

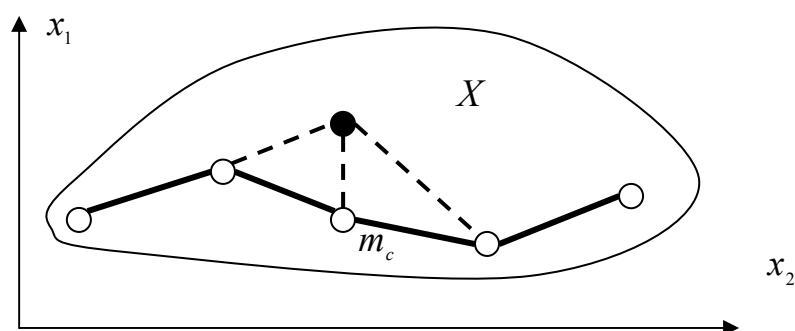


Рисунок 2 – Шаг коррекции синаптических весов для двумерного входного вектора $X = (x_1, x_2)$

После построения СКК на нее отображаются входные данные. Результат – исходные многомерные данные аппроксимируются к $n_x \times n_y$ узлам сети, что дискриминантных функций позволяет получить проекцию всего пространства наблюдений на двумерную плоскость и отобразить важные статистические характеристики входного пространства, которые можно использовать при решении задачи прогноза неопределенных значений.

При постановке задачи интерполяции пространственных данных исходная информация рассматривается в виде равнозначных результатов измерений, в которых выделяются две части X, Y – пространственные координаты и $Z_i, i=1, \dots, n$ – любая другая характеристика, восстановление значения которой необходимо произвести в точках равномерной координатной сетки. Обработка таких данных предполагает необходимость сохранения отношения соседства, обеспечения однозначности и непрерывности при преобразовании на рав-

номерную сетку системы координат и возможности работы с различными масштабами (в зависимости от задачи – оценочная, уточняющая).

Применение для решения такой задачи СКК показало следующее:

– карта не покрывает всей области значений измерений, что приводит к существенным ошибкам для координат, находящихся за ее пределами;

– при существенном различии диапазона изменения компонент исходных векторов карта имеет существенно скошенный вид, усложняющий формирование на выходе равномерной сетки (рис. 1 а);

– за счет отсутствия всех соседей у граничных нейронов за ее пределами наблюдается значительное увеличение ошибки интерполяции (рис. 1 б).

Для устранения этих недостатков предлагается выполнение дополнительных операций.

- расширение области определения модели за границы области определения системы;
- инициализация начальных значений нейронов для компонент X , Y на равномерной координатной сетке;
- инициализация начальных значений исходных векторов синаптических весов w_j для каждого нейрона на основе анализа значений зависимой переменной;
- определение лучшего аппроксимирующего вектора только по значениям координат (пространственная близость).

Экспериментальные результаты были получены с использованием СКК, архитектура которой выбиралась с учетом того, что на результаты кластеризации существенное влияние оказывает как общее число нейронов в сети, так и соотношение количества нейронов n_x , n_y по осям X , Y .

В качестве исходных для интерполяции модельных данных выбирались значения известного гравитационного поля участка Земной поверхности $Z = f(x, y)$ (рис. 2 а) в координатах (x_i, y_i) $i = 1, \dots, 57$, соответствующих координатам реальных разведочных скважин, предназначенных для оценки залегания угольных пластов (рис. 2 б).

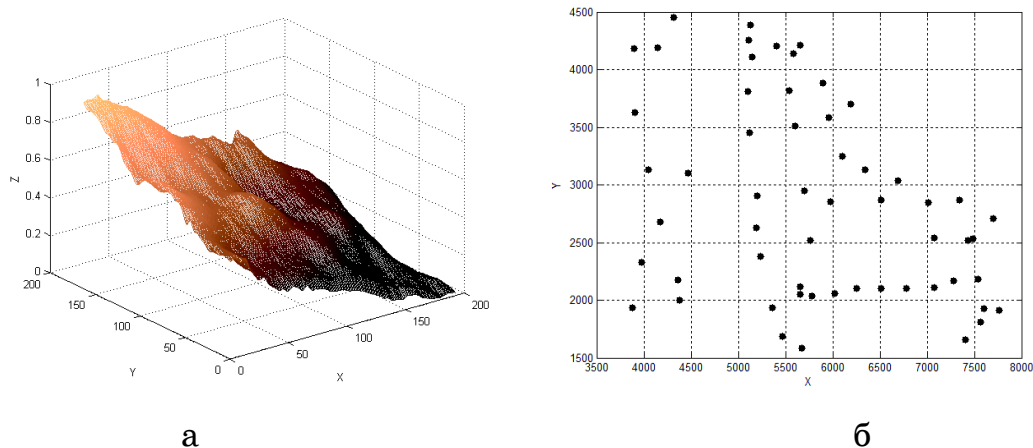


Рисунок 3 - а – геофизическое поле, используемое для формирования модельных данных ;б – неравномерно расположенные экспериментальные данные, используемые при интерполяции

После обучения сети, целевое значение, которое подлежит восстановлению, ассоциируется с ближайшим узлом сетки, что означает замену точки на центр локального сгущения точек данных в пространстве, которое хранится в кодовом блокноте (codebook). Для оценочного моделирования вполне достаточна непосредственная визуализация последнего.

Существенное значение на точность и гладкость результирующей модельной поверхности оказывает величина используемого радиуса соседства. При увеличении его значения повышается гладкость поверхности решения, при этом сужается область интерполяции, увеличивается ошибка в граничной зоне. Уменьшение радиуса соседства с одной стороны, повышает точность аппроксимации каждого экспериментального значения (в пределе при значении равном 1, ошибка стремится к нулю), однако поверхность приобретает ступенчатый вид.

Точность восстановления пропущенных данных можно также повысить, за счет увеличения размерности карты. Такого подход имеет существенные вычислительные ограничения и, кроме того, фактически начинает моделировать случайный шум.

На рис. 3 приведен результат интерполяции модели модельной поверхности с расширенной областью определения и реальных данных – интерполяция угольного пласта по результатам бурения. Целевыми значениями являются глубина его залегания и толщина слоя.

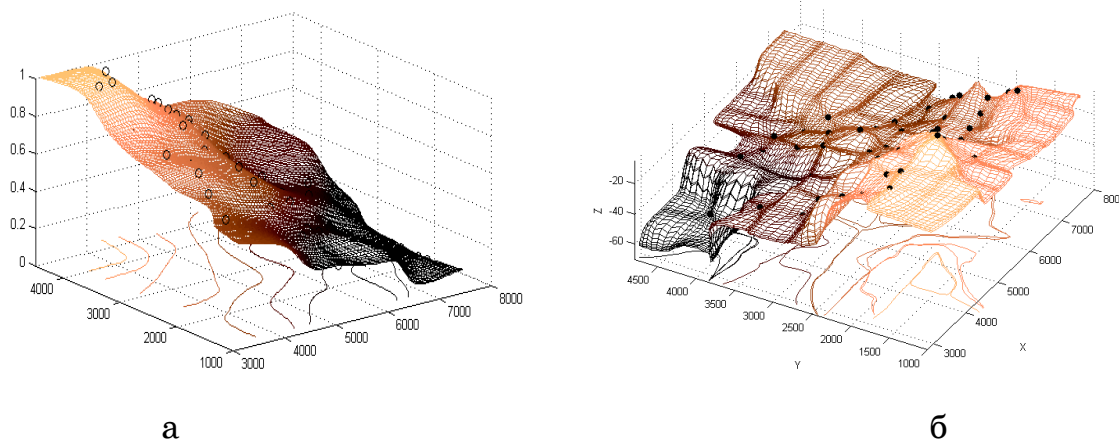


Рисунок 4 – Результаты интерполяции:
а – геофизическое поле ;б – угольный пласт

Выводы.

1. Важнейшей свойством самоорганизующихся карт Кохонена, которое обуславливает их эффективность при обработке именно пространственных данных, является то, что они обеспечивают адаптивное топологическое отображение входного пространства признаков произвольного размера в двумерное представление, в качестве которого может рассматриваться система координат.

2. Точность моделирования реальных систем на основе экспериментальных данных существенно зависит от их пространственного распределения, при этом единственным способом оценки правдоподобности процедуры восстановления является эмпирическое тестирование модели.

3. Для оценочного моделирования поверхности достаточна непосредственная визуализация кодового блокнота.

4. При малом значении неравномерных экспериментальных данных и значительном различии расстояний между ними, важно исследование на результат влияния метода их предобработки, способа инициализации начальных значений, параметров обучения сети.

ЛИТЕРАТУРА

1. Кохонен Т. Самоорганизующиеся карты / Кохонен Т. [пер. 3-го англ. изд. В.Н. Агеева под ред. Ю.В. Тюменцева]. – М.: Бинوم. Лаборатория знаний, 2008. – 665 с.
2. Рутковский Л. Методы и технологии искусственного интеллекта / Рутковский Л. – М., 2010. – 600 с.
3. Л.Г. Ахметшина., Т.С. Ямнич. Интерполяция пространственных данных методом двумерного проецирования нечетких кластеров // Искусственный интеллект, –2010. № 3. – С. – 433 –438.