

Л.Г. Ахметшина, Т.С. Ямнич

## **ПОВЫШЕНИЕ ТОЧНОСТИ ИНТЕРПОЛЯЦИИ ПРОСТРАНСТВЕННЫХ ДАННЫХ МЕТОДОМ ДВУМЕРНОГО ПРОЕЦИРОВАНИЯ НЕЧЕТКИХ КЛАСТЕРОВ**

*Аннотация* Рассмотрены информационные возможности метода нечеткой интерполяции пространственных данных, заданных на неравномерной сетке.

*Ключевые слова:* пространственные данные, интерполяция, неравномерная сетка, нечеткие модели, кластеризация, функция принадлежности.

Постановка проблемы. Точность моделирования реальных систем на основе экспериментальных данных существенно зависит от их пространственного распределения. Наилучшие результаты можно получить, если они определены на равномерной сетке. Однако, во многих практических случаях, измерения возможны лишь в некоторых точках, которые распределены неравномерно и покрывают область значений не полностью. Вопросам, связанным с обработкой неполной информации в настоящее время уделяется все большее внимание, поскольку полностью определенные задачи уже получили решение с приемлемой на практике точностью. Для создания моделей используются различные алгоритмы интерполяции, суть которых состоит в использовании имеющихся данных для получения прогнозных значений в дополнительных точках [1].

Важной практической областью, требующей использования методов обработки неполной неравномерно распределенной информации, характер которой обусловлен исключительно объективными причинами, являются задачи моделирования и анализа пространственных данных (интерполяция поверхностей, оценка объемов рудных тел и оседания земной поверхности на основе результатов бурения или измерений различных физических характеристик и т. д.). Особенность работы с пространственными данными осложняются необходимостью учета их особенностей, обусловленными свойствами пространственной топологии и корреляции, учет которых требует

применения специфических методов. Наиболее сложной задачей является моделирование подземных поверхностей, для которых отсутствует возможность получения дополнительной экспертной или картографической информации.

В работе [1] демонстрируется возможность использования интерполяции пространственных данных заданных на неравномерной сетке, методом двумерного проецирования нечетких кластеров. Возможность использования нечеткой логики для решения подобных задач была отмечена в [2]. Основное преимущество нечетких моделей, по сравнению с традиционными математическими моделями, связано с возможностью использования для их разработки значительно меньших объемов информации о системе, причем она может носить приближенный характер.

Алгоритм метода включает последовательность следующих действий.

1. Определение центров кластеров  $m_1, m_2, \dots, m_c$ , неравномерных экспериментальных измерений с применением метода FCM в пространстве признаков  $X_1, X_2, \dots, X_n, Y$  ( $Y$  – целевой, интерполяция которого производится), вычисляемых итеративно по формуле

$$m_{xl}(t+1) = \frac{\sum_{j=1}^N \mu_{ij}^q(t) \cdot x_{lj}}{\sum_{j=1}^N \mu_{ij}^q(t)}, \quad (1)$$

где  $(\forall i \in \{2, \dots, c\})$ ,  $c$  – количество классов;  $N$  – число объектов кластеризации,  $l$  – количество информативных признаков, причем  $x_{n+1} = y$ ;  $q > 1$  – параметр фаззификации.

2. Определение функций принадлежности отдельно для входных и целевого параметров, с использованием проекций  $m_i^{X_j}$  центров кластеров на координатное пространство, где  $j = 1, \dots, n$ ,  $i = 1, \dots, c$ , которые и рассматриваются в качестве «существенных» точек, являющимися центрами функций принадлежности  $\mu_{ij}^X$  каждого из классов:

$$\mu_{ij}^X(x) = \frac{1}{\sum_{k=1}^c \left[ \frac{d_{ij}^X(x)}{d_{kj}^X} \right]^{\frac{2}{q-1}}} = \mu_i^X(x_j), \quad (2)$$

где расстояние между вектором  $x_i$  и центром кластера  $m_i^X$  определяется только во входном пространстве по формуле:

$$d_{ij}^X(x) = |x - m_i^X| = \sqrt{(x_j - m_i^X)^T A (x_j - m_i^X)}, \quad (3)$$

размерность положительно определенной симметричной матрицы  $A$  равна  $n \times n$ .

3. Вычисление выхода нечеткой модели  $y$  на основе вычисленных функций принадлежности к заданным кластерам каждой точки полного координатного пространства интерполируемой поверхности и вычисляется по формуле.

$$y(x_j) = \frac{\sum_{i=1}^c m_i^Y \cdot \mu_i^X(x_j)}{\sum_{i=1}^c \mu_i^X(x_j)}. \quad (4)$$

4. Итеративное уточнение модели – корректировка центров кластеров на основе использования значения ошибки модели.

Целью работы является повышение точности интерполяционной поверхности, формируемой методом двумерного проецирования нечетких кластеров, за счет использования нескольких решений с различным уровнем обобщения.

Центральным элементом при построении нечеткой модели является база правил типа «Если–то», которая содержит основную информацию о поверхности решения. Способ ее формирования определяет адекватность и точность модели, при этом наиболее значимыми являются следующие факторы.

1. Условие каждого нечеткого правила соответствует определенной точке пространства входных значений (определяется п. 1 алгоритма в методе двумерного проецирования нечетких кластеров) и влияет на условие всех правил. Изменение каждого условия приводит к смещению всех опорных точек, таким образом, определяя глобальные характеристики поверхности решения.

2. Изменение заключения правил, определяемое точками пространства выходных значений, приводит к локальному изменению участков поверхности модели, прилегающей к данной точке измерения, и может не оказывать влияния на удаленные участки поверхности решения.

Количество используемых правил с одной стороны, повышает возможность получения высокой точности нечеткой модели, с другой стороны, увеличивает стоимость вычислений и приводит к преобладанию «локальных» тенденций, определяемых каждой точкой измерения. Использование различных модификаций метода нечеткой кластеризации, изменение задаваемых априори количества кластеров, способ предобработки данных (нормализация, масштабирование, формирование нового пространства входных векторов на основе данных измерений и дополнительной, например картографической или экспертной оценки) позволяет влиять на точность модели. В предельном случае, если правила задаются для каждой точки измерения, ошибка модели в них стремится к нулю. В то же время, теоретически, для построения оптимальной поверхности решения, должно выполняться условие разбиения единицы для каждого элемента  $x_j$ :

$$\sum_{i=1}^c \mu_i^x(x_j) \equiv 1. \quad (5)$$

Алгоритм FCM обеспечивает выполнение условия (5) только для набора входных значений. Для остальных точек интерполируемой поверхности данное условие нарушается. Это связано с тем, что функции принадлежности имеют бесконечный носитель с большой областью влияния отдельных заключений. Если указанная сумма  $>1$  или  $<1$ , осуществляется формирование более рельефной, или плоской поверхности, соответственно, что и отражается в общей ошибке интерполяции.

Важным фактором, влияющим на поверхность отклика, является параметр фаззификации  $q$ , увеличение которого приводит к усилению нечеткости кластеров (увеличение площади под графиком функции принадлежности, и, соответственно, увеличению площади поверхности влияния точки). Кроме этого, при построении модели, следует учитывать вид поверхности, в частности, степень крутизны и характер присутствующих нелинейностей, как в целом по области определения, так и по отдельным областям, ее составляющих.

В данной работе, для повышения точности модели, получаемой с применением метода нечеткого проецирования кластеров, предлагается новый способ формирования результирующей поверхности на основе нескольких решений с различным уровнем обобщения. Использование различного количества кластеров  $c$  и различных значе-

ний параметров фаззификации  $q$ , для каждого из них, позволяет учесть и сбалансировать глобальные и локальные тенденции присутствующие в данных: выделить общий тренд по направлениям и незначительные пространственно коррелированные поверхностные структуры.

Экспериментальные результаты были проведены для модельных и реальных данных. В качестве модели использовалось гравитационное поле поверхности Земли  $Z = f(x, y)$ , трехмерное изображение которого приведено на рис. 1 а.  $Z$  – нормированное значение, определенное для 57 экспериментальных точек с координатами  $(x, y)$ , расположение которых представлено на рис. 1 б. Координаты точек соответствуют скважинам бурения, предназначенным для определения глубины залегания и толщины угольных пластов на шахтном поле шахты «Павлоградская». В данном приложении модель необходима для оценки последствий закрытия шахт – выявления зон затопления, расчета объема грунта для засыпки, оценки ущерба, наносимого горнодобывающими предприятиями.

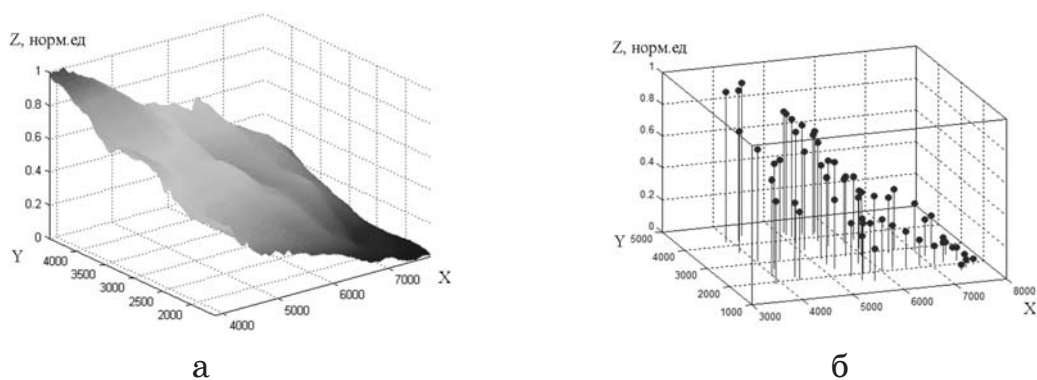


Рисунок 1 - Модельное поле (а) и координаты скважин бурения

На рис. 2 а представлен результат моделирования с применением метода нечеткого проецирования кластеров при значении  $c = 37$  и  $q = 2$ . Поверхность, изображенная на рис. 2 б получена на основе использования трех решений, полученных со значениями параметров кластеризации  $c = 8, q = 1.78$ ;  $c = 30, q = 1.85$  и  $c = 51, q = 1.95$ . Визуальный анализ и оценка ошибки интерполяции свидетельствует о повышении точности получаемого результата.

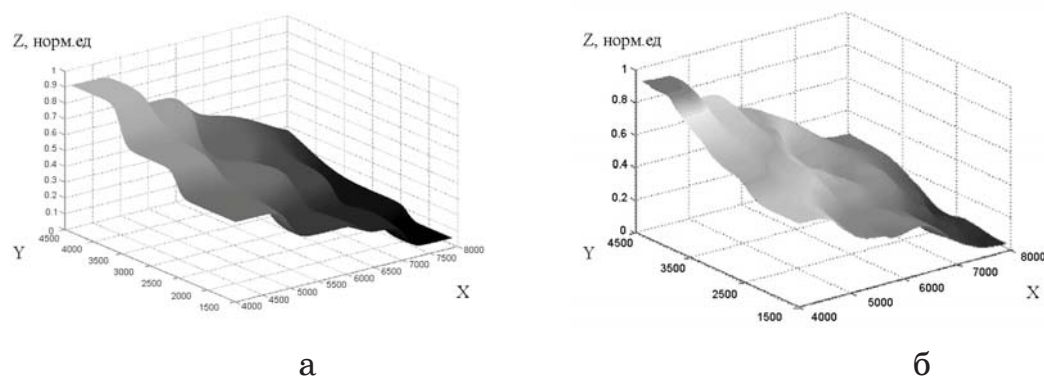


Рисунок 2 - Нечеткая модель: а –  $c=37$  и  $q=2$ ; б – комбинированное решение с параметрами  $c=8$ ,  $q=1.78$ ;  $c=30$ ,  $q=1.85$  и  $c=51$ ,  $q=1.95$

На рис. 3 а представлена интерполяция поверхности трех угольных пластов, полученная с использованием описанного метода, которая позволяет визуально оценить их форму и взаимное расположение. Точность формируемой модели достаточна для решаемой задачи и находится в пределах точности измерения экспериментальных данных (следует отметить, что достаточно хорошим показателем для данной предметной области является точность модели в пределах 0,1 м по вертикальной составляющей).

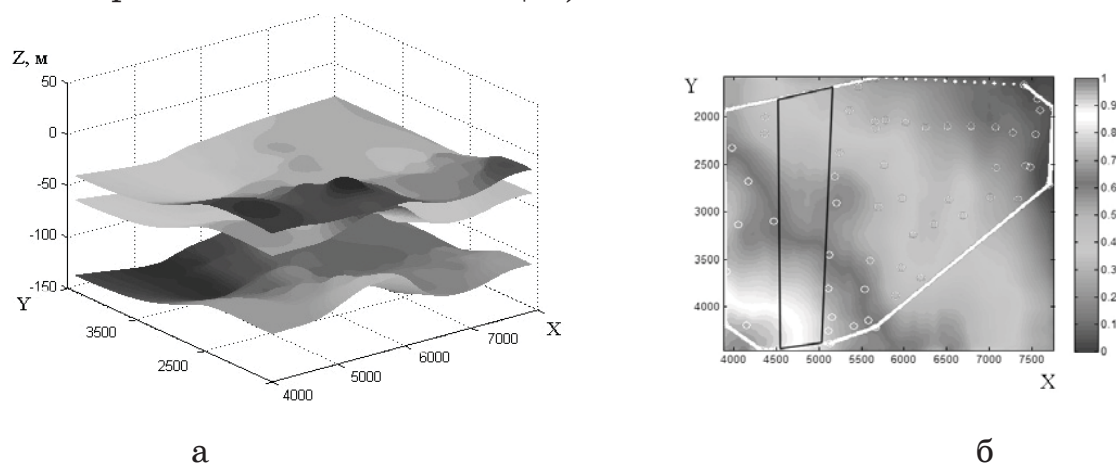


Рисунок 3 - Модель угольных пластов (а) и расположение скважин (б)

Наибольшая погрешность наблюдается в областях, не покрываемых элементами выборки, выделенными на рис. 3 б, отображающего двумерную проекцию модельного поля и расположение экспериментальных точек, белыми линиями – области внешней экстраполяции.

Выводы. В результате анализа экспериментов, проведенных на модельных и реальных данных, сделаны следующие выводы.

1. Комбинирование нескольких поверхностей решения, полученных с использованием метода двумерного проецирования нечетких кластеров с различным уровнем обобщения, повышает точность интерполяции пространственных данных заданных на неравномерной сетке.

2. Наибольшие значения ошибки сосредоточены не только в области пространства, для которых измерения отсутствуют (т.е. достоверность модели не подтверждена и должна решаться задача внешней экстраполяции), но и в области, с «большим» расстоянием между точками измерения (выделена черной линией на рис. 3 б), которую можно определить как область «внутренней экстраполяции».

3. Дальнейшее повышение точности модели требует решения задачи внешней и внутренней экстраполяции.

#### ЛИТЕРАТУРА

1. Л.Г. Ахметшина., Т.С. Ямнич. Интерполяция пространственных данных методом двумерного проецирования нечетких кластеров // Искусственный интеллект, –2010. № 3. – С. – 433 –438.
2. А. Пегат. Нечеткое моделирование и управление. – М.: «Бином». 2009, 798 с.

Получено 16.01.2011г.