МАШИННЫЕ МЕТОДЫ ОБНАРУЖЕНИЯ ЗАКОНОМЕРНОСТЕЙ
(Вычислительные системы)
1981 год Выпуск 88

УДК 62-52

НОВЫЙ КЛАСС ЗАДАЧ ПОСТРОЕНИЯ ФУНКЦИЙ ПО ЭМПИРИЧЕСКИМ ДАННЫМ (В СВЯЗИ С ГЕОЛОГИЧЕСКОЙ ИНТЕРПРЕТАЦИЕЙ КОМПЛЕКСНЫХ ГЕОФИЗИЧЕСКИХ И ГЕОХИМИЧЕСКИХ ДАННЫХ)

С.К.Туренко

1. Первые представления о новом классе задач построения функций по эмпирическим данным, рассматриваемом в связи с геологической интерпретацией комплексных геофизических и геохимических данных были фиксированы в работах [1-4].

В силу ряда специфических черт, важных как с практической, так и с теоретической точек зрения, данный класс задач может представлять интерес и для других областей естествознания.

2. Цель данной работы - привлечь внимание научной общественности к обсуждению направления и результатов исследований [1-4] по вопросам построения функций по эмпирическим данным.

Детальное и строгое изложение методологического и теоретического обеспечения, разрабатываемого в рамках отмеченного класса задач, требует соответствующего языка. Необходимые элементы такого языка были разработаны в [3]. Так как данный язык не является общепринятым, то его описание и использование в этой работе весьма затруднительно. Однако попытаемся, опираясь на минимум необщепринятых понятий, кратко описать модель постановки рассматриваемого класса задач, зафиксировать наиболее важные соображения, положенные в основу выбора схемы решения этого класса задач, и отметить наиболее специфические его черты.

3. В общем случае под моделью построения функций по эмпирическим данным будем понимать упорядоченную пятерку $\langle A, Q, F, K, P \rangle$, где A - отображение множества эмпирических данных (Q) на множество целевых функций $F = \{f_i(x)\}, i = \overline{1, M1}$, в соответствии с некоторым критерием (K) и с учетом некоторых ограничений, априорной

информации и предположений (P). Область определения Q - D_1 и область определения функций из F - D_2.

Каждый элемент M фиксированной модели (A, Q, F, K, P) может в конкретной ситуации оказаться в одном из трех состояний "определенности": задан (M), задан с точностью до класса (M^{α}), не задан $(\overline{M})$.

В зависимости от конкретного содержания элементов принятой модели и их "определенности" можно выделять различные классы ситуаций построения функций по эмпирическим данным.

4. С учетом фиксированных представлений рассматриваемый класс задач характеризуется следующим образом.

Исходная модель рассматриваемого класса задач есть $\langle A^{\alpha}, Q^{\alpha}, \{f_i^{\alpha}(x)\}, K, P\rangle$, $i = i_0$, $i_0 \in I = \overline{1,Ml}$. Говоря, что $f_i(x)$ построена (задана), будем иметь в виду задание конечного множества пар $\{x_j, f_i(x_j)\}$, $x_j \in D_2$, $j = \overline{1,N}$, и процедуры вычисления $f_i(x_p)$ для каждого $x_p \in D_j$ (алгоритма интерполяции). Множество $\{x_j\}$, определенное в метрическом пространстве, будем называть сетью задания (измерения) функции $f_i(x)$, x_j - точкой задания (измерения), или узлом сети задания $f_i(x)$. Отношение числа узлов сети задания $f_i(x)$-N к "размеру" области определения $f_i(x)$ будем называть плотностью сети задания $f_i(x)$.

Через Q будем обозначать некоторое множество функций $\{e_p(x)\}$, полученных в результате эмпирических исследований. Каждая из этих функций задается таблицей размерности $\underline{m \times 3}$ вида $\{x_q;\ \theta_p(x_q),\ C_{\theta_p}(x_q),\ \Delta\theta_p(x_q)\}$, $x_q \in D_1^p \subset D_1$, $q=\overline{1,m}$, где x_q - координаты измерений; $\theta_p(x_q)$ - значение θ_p в точке x_q; $C_{\theta_p}(x_q)$ - стоимость измерения θ_p в точке x_q; $\Delta\theta_p$ - точность измерения θ_p в x_q; D_1^p - область определения θ_p.

Область определения Q-D_1 разбивается на две подобласти: $D_э$ ("область обучения") и $D_п = D_2$ ("область построения"); $D_1 = D_э \cup D_п$.

В $D_п$ по сети с необходимыми и достаточными плотностью и точностью измеряется совокупность функций $Q_п = (\{\varphi_1^{s_п l}(x^п)\}, f_i^{s_п}(x^п))$, $x^п \in D_п$, $l=\overline{1,L}$ ($\varphi_l(x)$ - неявно связанные с $f_i(x)$ функции, $s_п l$ и $s_п$ - схемы измерения сети и точности $\varphi_l(x)$ и $f_i(x)$ в $D_п$), необходимых и достаточных для построения $f_i(x)$ при фиксированных отображении A, критерии K и условиях P.

В $D_э$ по сети с достаточными точностью и плотностью измеряются как достаточный набор $\{\varphi_h^{sэh}(x_э)\}$, $h = \overline{1,hl}$, $(\{\varphi_1^{s\,l}(x_п)\} \in \{\varphi_h^{s\,h}(x_э)\}$, $hl \geq L)$, так и $f_i^{sэ}(x_э)$, $x_э \in D_э$. Здесь $s\,эh$

и эп – схемы измерения $\varphi_h(x)$ и $f_i(x)$ в $D_э$. Данные $Q_э$ служат для выбора $Q_п$ и A (из A_α). Говоря о достаточности $\{\varphi_h^{*э}(x_э)\}$, понимаем, что существует $A \in A^\alpha$, которое с использованием некоторого подмножества $\{\varphi_h^*(x_э)\}$ позволит построить $f_i(x)$ при фиксированных К и Р.

Для выбора схемы решения исходят из предположения: $C_{\varphi_h} \ll C_{f_i}$, $D_э \approx D_п$ ($\approx$ – аналогично), так как построить $f_i(x)$ с необходимой точностью на основе измерения только $\{\varphi_h(x)\}$ затруднительно. Кроме того, при формировании целевой установки на построение, а следовательно, и выборе схемы ее реализации, могут использоваться предположения о том, что строить "точно" $f_i(x)$ нужно только в подобластях $D_п^{j+}$, "полезных" по некоторому критерию; вне этих подобластей ее можно строить приближенно; таким образом, эффективность построения может существенно зависеть от порядка построения $f_i(x)$ на подобластях $D_п^{j+}$.

Для выбора решения используется критерий эффективности типа $K(W\,V) - \max\limits_{П.АП}$, $C < C_0$; $П \geq П_0$, где П.АП – параметры алгоритма построения (параметры отображения А, совокупность функций $\{\varphi_i(x)\}$ параметры схем измерения функций в Q); С – затраты на построение C_0 – ассигнования; П – приобретения от построения; $П_0$ – плановые приобретения; V – приведенные затраты на построение; W – показатель качества построения типа $W = \sum\limits_{j=1}^{M} \Psi(x_j)\Phi(f_i(x_j)-\hat{f}_i(x_j))$, где $\Psi(x_j)$ – некоторая весовая функция, определяющая важность точки измерения; Φ – некоторый функционал отклонения построенной функции ($\hat{f}_i$) от "истинной" (f_i).

Отображение А позволяет в рамках "формальной схемы" эффективно на основе ЭВМ организовать взаимодействие формальных и неформальных элементов схемы решения.

На целевую функцию $f_i(x)$ никаких специальных ограничений не накладывается, кроме требования "физической" измеримости.

5. Суть решения заключается в том, что функции $\{\varphi_i(x)\}$ используются не для непосредственного построения $f_i(x)$, а для исследования ее структуры (построения весовой функции $\psi(x_i)$ в соответствии с целевой установкой). На основе результатов этих исследований осуществляется измерение $f_i(x)$ по оптимальной схеме измерений для фиксированного критерия и алгоритма "интерполяции".

Общая схема построения включает этапы: выделения "полезных" подобластей с избытком $D_п^j$, разделения их на "полезные" $D_п^{j+}$ и "неполезные" D_{Π}^{j-} , упорядочения D_{Π}^{j+} по "полезности", оценивания сложности строения $f_i(x)$ на D^{j+} ,измерения и восстановления $f_i(x)$ по оптимальной схеме. Этапы выделения, разделения, упорядочения и оценивания реализуются на основе косвенных данных как подзадачи рассматриваемой задачи.

6. Итак рассматриваемый класс задач характеризуется:

- взаимосвязанностью процессов получения и преобразования эмпирических данных;

- единством методологических и теоретических основ построения функций, измеряемых в различных шкалах;

- теоретическим обеспечением, позволяющим работать с различными по количеству и качеству данными;

- использованием косвенных данных ($\varphi_i(x)$) не для непосредственного построения целевой функции $f_i(x)$, а для планирования "прямого эксперимента" (выделения "полезных" подобластей, выбора порядка их обследования и сети измерения целевой функции на них).

В настоящее время для рассматриваемого класса задач разработано методологическое, теоретическое и программное обеспечение первого уровня.

Литература

1. ВОРОНИН Ю.А., ТУРЕНКО С.К., ФЕЙГЕНБЕРГ С.Д. О постановке и решении основной задачи геологической интерпретации комплексных геофизических данных. -В сб.: Математические методы при поиске и разведке полезных ископаемых. Новосибирск, 1978, с.71-91.

2. ТУРЕНКО С.К. Общая схема решения на ЭВМ основной задачи геологической интерпретации комплексных геофизических данных.-Новосибирск, 1980. - 24 с. (Препринт/ВЦ СО АН СССР; 224).

3. ТУРЕНКО С.К. О системном подходе к построению функций по эмпирическим данным на основе применения формальных методов и ЭВМ.-В сб.:Применение математических методов и ЭВМ при поисках полезных ископаемых. Новосибирск, 1980, с. 74-87.

4. ВОРОНИН Ю.А., ТУРЕНКО С.К. О новом классе задач построения функций по эмпирическим данным. -Там же, с. 95-107.

Поступила в ред.-изд.отд.
2 февраля 1981 года