

ОСОБЕННОСТИ ПОИСКА ЗНАЧЕНИЙ ПАРАМЕТРОВ ПРИ АВТОМАТИЗИРОВАННОМ ПОСТРОЕНИИ ПРОГНОЗНЫХ КАРТ

Ковин Р.В., Копнов М.В.

Научный руководитель Н. Г. Марков д.т.н., профессор

Институт «Кибернетический центр» Томского политехнического университета

Email: kopnovmv@tpu.ru

В ряде прикладных задач по исходному набору точек наблюдения требуется оценить (спрогнозировать) значение исследуемого явления в областях, где наблюдения не проводились. Особенностью таких задач является, как правило, неравномерное распределение исходных точек наблюдения. Результаты оценивания традиционно представляются в виде прогнозных карт (карт рельефа местности, распределения температур, глубин залегания нефтеносных пластов и др.). В настоящее время для решения таких задач используются различные виды программных средств: математические пакеты, геоинформационные системы и специализированные системы.

В геоинформатике решение упомянутых задач сводится к решению задачи восстановления геополя [1]. Строго говоря, задача восстановления геополя является некорректной, поскольку не существует точного и единственного ее решения. По своей сути – это попытка по частному восстановить общую картину явления. Без дополнительных сведений о природе восстанавливаемого геополя такую задачу решить невозможно [2]. Поэтому существуют различные подходы и методы решения этой задачи. Наибольшее распространение получили детерминистические и геостатистические методы, использующие двумерную интерполяцию. Так или иначе, в каждом методе вводятся определенные представления о природе геополя. Для детерминистических методов на практике пользователь должен выразить их (представления) через конкретные значения параметров того или иного метода. В большинстве случаев это вызывает большие сложности у обычных пользователей – ведь они должны понимать математическую основу этих методов и влияние их параметров на результат. В геостатистических методах свои недостатки – требуется понимание основ геостатистики и сложность вариограммного анализа. В конечном итоге использование разных методов приводит к получению различных результатов и необходимости оценки достоверности этих результатов.

Зная специфику предметной области и располагая априорной информацией об исследуемом явлении, можно выбрать наиболее приемлемый метод восстановления гео-

поля. Стандартной методикой оценки результатов восстановления геополя является кросс-валидация (или кросс-проверка или скользящий контроль) – процедура эмпирического оценивания обобщающей способности алгоритмов, обучаемых по прецедентам [3].

Так как задача восстановления геополя является некорректной, кросс-валидация позволяет лишь косвенно оценить полученные результаты. Если результаты не удовлетворяют, то необходимо изменить параметры используемого метода и повторить восстановление. Подбор параметров зачастую является итеративным процессом, требующим полной перестройки модели геополя. Все это очень усложняет работу: необходимо проводить кросс-валидацию при каждом изменении параметра. В случае, когда параметров несколько, время поиска наилучшего решения существенно увеличивается.

На наш взгляд идеальной для пользователя является ситуация, когда он может явно задавать метод восстановления и значения параметров метода. При этом он также может задавать режим автоматического определения значения любого параметра, если затрудняется определить его значение самостоятельно. В этом случае система должна автоматически подобрать значения параметров, при которых достоверность результата будет наибольшей (а ошибка наименьшей). При автоматическом подборе значений параметров наиболее сложной с точки зрения формализации задачей является определение диапазона значений, которые будут использованы. Если поручить это пользователю, то сценарий его действий будет такой:

- 1) выбор метода восстановления геополя;
- 2) определение диапазонов изменения параметров метода;
- 3) установка требуемой точности восстановления;
- 4) запуск процесса построения прогнозной карты.

Данный сценарий можно представить в виде схемы, изображенной на рис. 1. Входными данными для процесса построения прогнозной карты являются диапазоны изменения параметров выбранного пользователем метода и требуемая точность. На каждой итерации выполняется кросс-валидация и

проводится оценка точности восстановления. Если требуемая точность не достигнута, то вычисляются новые значения параметров и итерация повторяется. При достижении заданной точности осуществляется восстановление геополя с использованием найденных значений параметров. По восстановленной таким образом модели строится прогнозная карта.

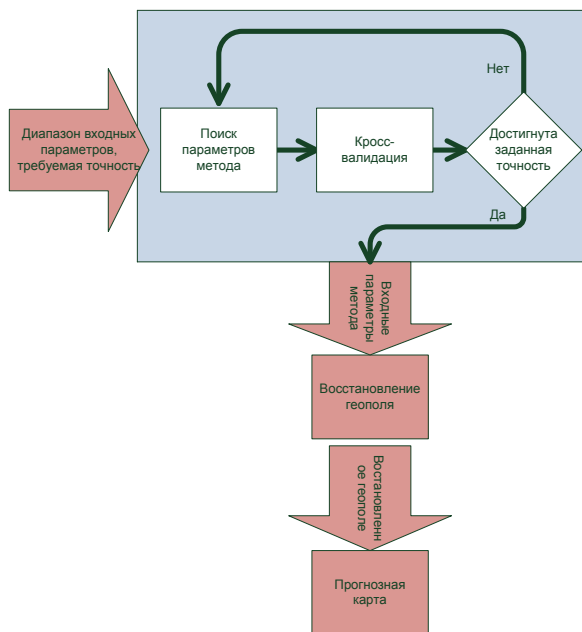


Рис. 1 Схема автоматизированного построения прогнозных карт

Поиск значений параметров является задачей многомерной оптимизации. Для ее решения можно использовать один из классических методов многомерной оптимизации (метод градиентного спуска, метод покоординатного спуска, метод сопряженных градиентов или др. [4]).

В качестве примера рассмотрим процесс поиска входных значений параметров, когда число параметров равно двум (рис. 2). По заданным пользователем диапазонам ($P_{1 \min} \dots P_{1 \max}$) и ($P_{2 \min} \dots P_{2 \max}$) система находит искомые значения параметров $P_{1 \text{ res}}$ и $P_{2 \text{ res}}$, при которых величина ошибки будет минимальной. Такое значение будет находиться в одном из локальных минимумов (точки 1, 2, 3).

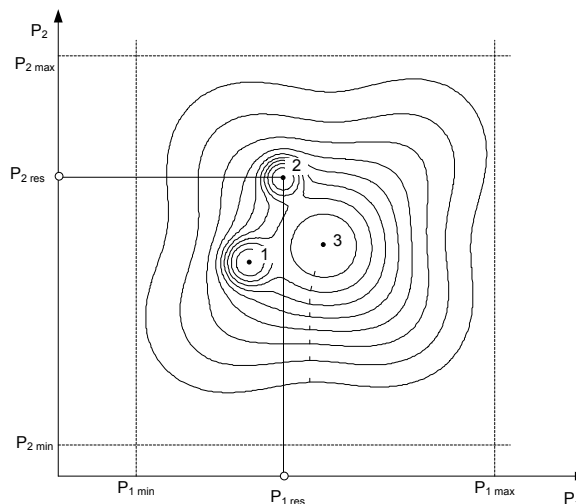


Рис. 2 Схема нахождения значений входных параметров

Из примера видно, что найденное решение может быть хорошим, но не лучшим. Поэтому при использовании методов многомерной оптимизации следует учитывать, что если локальных минимумов несколько, то перечисленные выше методы в зависимости от начальных значений параметров могут дать в качестве решения любой из существующих локальных минимумов.

В случае если требуется лучшее решение необходимо последовательно с заданным шагом перебрать все возможные варианты значений параметров из диапазона. Если число всех возможных значений параметров невелико, то решение задачи поиска значений входных параметров может быть найдено простым перебором за приемлемое время. В противном случае такой перебор потребует больших временных затрат.

Список литературы

1. Ковин Р.В., Марков Н.Г., Геоинформационные технологии для анализа двумерных геополей. – Томск: Изд-во Том. ун-та, 2006. – 166 с.
2. Тихонов А.Н., Арсенин В.Я., Методы решения некорректных задач – М.: Наука, 1979.
3. Воронцов К. В., Обзор современных исследований по проблеме качества обучения алгоритмов. – ВЦ РАН, <http://www.ccas.ru/frc/papers/voron04twim.pdf>.
4. Гилл Ф., Мюррей У., Райт М., Практическая оптимизация. Пер. с англ. — М.: Мир, 1985.