

50-я научная конференция МФТИ
Факультет проблем физики и энергетики
Секция методов решения обратных задач в физике

УДК 502.3

Свительман В.С.

Московский физико-технический институт (государственный университет)

Многоточечная геостатистика и её развитие
для непрерывных данных

В задачах анализа и обработки пространственно распределенной информации широко применяются методы геостатистики (пространственной статистики). Основой геостатистики является использование при анализе (интерполяции, стохастическом моделировании, вероятностном картировании) пространственной корреляционной структуры данных, то есть модели зависимости между данными в зависимости от их взаимного пространственного расположения.

В традиционной геостатистике для описания пространственной корреляции используется вариограмма:

$$\text{Var}\{V(x+h) - V(x)\} = E\{(V(x+h) - V(x))^2\} = 2\gamma(h), \quad \forall x.$$

Вариограмма определяется статистическим различием значений, разделенных заданным вектором, то есть, по сути, описывает двухточечные зависимости, что далеко не всегда достаточно.

В последние годы возникла идея многоточечной геостатистики, основной принцип которой состоит в построении модели пространственной корреляции на основе зависимостей между более чем двумя точками. Строится такая модель по возможным типам соседства, базируясь на тренировочном образе. Описывается модель многоточечной пространственной корреляции с помощью дерева вероятностей. Вероятности различных типов соседств оцениваются исходя из распределения характеристик в тренировочном образе.

Использование модели многоточечной пространственной корреляции ориентировано на стохастическое моделирование, то есть пространственную интерполяцию, основывающуюся на предположении, что анализируемые данные являются реализациями случайного процесса. Это позволяет, используя имеющиеся измерения как зафиксированные значения (условное моделирование), получать бесконечно много новых реализаций переменной в точке оценивания.

Алгоритмы многоточечной геостатистики были изначально разработаны для категориальных данных и показали на них хорошие результаты. В частности, многоточечная геостатистика применяется для моделирования каналов и трещин в задаче течения грунтовых вод, для моделирования нефтяных резервуаров и т. д.

Но в реальных задачах данные часто являются непрерывными (та же проводимость в задачах гидрогеологии или при моделировании нефтяных резервуаров), поэтому для применения к ним многоточечной геостатистики алгоритмы требуют модификации. Ранее предлагалось предварительное разбиение данных на предопределённые классы путём фильтрации. То есть каждое непрерывное значение заменялось вектором из 6 дискретных значений, а затем проводилась по этим векторам (алгоритм *filtersim*). Основная проблема этого метода в том, что такая классификация занимает очень много времени (из-за большого количества предопределённых классов).

В данной работе предлагается разбивать данные на классы непосредственно в процессе обучения, используя адаптивную технологию деревьев решений. Алгоритмы построения деревьев работают и для непрерывных данных, а технология деревьев решений достаточно разработана и широко используется в задачах многопеременной классификации и регрессии. Таким образом, разбиение производится на необходимое число классов, обусловленное структурой конкретного тренировочного образа.

СПИСОК ЛИТЕРАТУРЫ

1. *Caers J., Zhang T.* Multiple-point geostatistics: a quantitative vehicle for intergration geologic analogs into multiple reservoir models // Integration of outcrop and modern analog data in reservoir models. — 2004. — AAPG memoir 80. — P. 383–394.
2. *Frank E., Wang Y., Inglis S., Holmes G., Witten I.H.* Using model trees for classifications // Machine Learning. — 1998. — V. 32 (1). — P. 63–76.
3. *Yuhong Liu.* Using the Snesim program for multiple-point statistical simulation // Computers & Geosciences. — 2006. — V 32, N. 10.
4. *Zhang T., Switzer P., Journel A.* Filter-Based Classification of Training Image Patterns for Spatial Simulation // Mathematical Geology. — 2006. — V. 38, N. 1.
5. *Elomaa T., Kaariainen M.* An Analysis of Reduced Error Pruning // Journal of Articial Intelligence Research. — 2001. — V. 15.

Представленная выше версия доклада является ознакомительной.

Версию доклада, предназначенную для печати,
можно найти в факультетском сборнике трудов конференции.
Электронные материалы конференции публикуются по адресу
http://www.mipt.ru/nauka/conf50/plen_sections/