

МЕТОДЫ, ИНСТРУМЕНТЫ, ПРИМЕРЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ЦИФРОВЫХ ПОЛЕЙ

Таранчук В. Б.

Кафедра компьютерных технологий и систем, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: taranchuk@bsu.by

Обсуждаются вопросы обработки, аппроксимации, интеллектуального анализа, функционального описания цифровых полей. Рассмотрены различные варианты приближений, включая метод Кригинга. Введено понятие составного цифрового поля, предложены и рассмотрены различные способы автоматического разбиения исходных данных на подгруппы для построения составного цифрового поля и его функционального описания с реализацией алгоритма адаптивной аппроксимации в системе Wolfram Mathematica.

ВВЕДЕНИЕ

Развитие цифровых технологий, усовершенствование аппаратных средств проведения измерений и диагностики предоставили доступ всевозможным сферам к большому объему информации, накопленных наблюдений и статистик. Как и любой вид информации, ценность она предоставляет лишь при корректном дальнейшем использовании. Именно поэтому сейчас актуальны вопросы исследования и анализа данных, правильного построения оценок и прогнозов. Значимую роль в современном мире играет исследование пространственных переменных, в частности, измерение геологических и геофизических показаний, проведение геоэкологических исследований для оценки состояния местности, сбор гидрометрических показателей для изучения океанов. Характерно для перечисленных и в отличие от других сфер – ограниченность возможности проведения наблюдений; геофизические и геоэкологические исследования, отправка экспедиций предполагают существенные затраты. Правильное прогнозирование, как распределений в пространстве на основе полученных наблюдений, так и составление оптимального плана организации наблюдений, позволяют реальному сектору сократить значимую часть расходов на исследования, добиться лучших результатов. В ракурсе отмеченного широкое распространение получила геостатистика; одновременное развитие цифровых технологий позволило ей проводить все более сложные вычисления, обеспечивающие качественные прогнозы.

I. ЦИФРОВЫЕ ПОЛЯ.

Основные позиции предложенной и обсуждаемой методики заимствованы из основ исследований и интерпретаций пространственной информации с помощью статистических моделей. Цифровым полем далее обозначается трехмерная цифровая модель подлежащей аппроксимации изучаемой пространственной переменной. Обычно использование цифровой модели не предполагает наличие аналитического описания пространственной переменной – исходные данные, как правило, получают на основе собранной в циклах

отдельных наблюдений информации, часто на рассеянном (нерегулярном) множестве точек.

Современные технологии позволяют строить модели, наглядно визуализировать исходные, получаемые результаты и показатели различными способами. Для решения задач анализа пространственных данных разработаны, используются геоинформационные системы и системы компьютерной алгебры.

В данной работе при исследовании цифровых полей выбор сделан в пользу систем компьютерной алгебры, конкретно – одной из ведущих – системы компьютерной алгебры Wolfram Mathematica, содержащей множество статистических и математических функций, инструментов когнитивной компьютерной графики.

II. ТИПЫ ЦИФРОВЫХ ПОЛЕЙ, ИХ АНАЛИЗ.

Основой цифрового поля являются: структурированное или неструктурированное множество точек наблюдений; область восстановления пространственной переменной; построенная по точкам наблюдения на области определения аппроксимирующая (по принятому методу, алгоритму) функция. Под структурированным множеством точек наблюдений подразумевается множество точек, расположенных на сетке или ограниченном числе прямых, либо в вершинах геометрических примитивов. Неструктурированные множества могут задаваться хаотическим набором пунктов наблюдений, неравномерно распределенных по области восстановления (рассеянное множество). Большинство прикладных задач связано с построением цифрового поля по хаотически расположенным пунктам наблюдений. Обычно, для случаев структурированных наборов проведение вычислений и построение цифровой модели значительно проще. Ввиду возможного выбора различных способов построения аппроксимирующей функции, цифровые поля для одного множества точек наблюдений, могут значительно отличаться.

В основе любого поля лежит аппроксимирующая функция, которая в прикладных задачах обеспечивает предсказания или оценки пространственной переменной в областях, не покрытых

наблюдениями. Как правило, наиболее часто для аппроксимации используется интерполяция, часто – в варианте, при котором значения аппроксимирующей функции в точках наблюдения совпадают с действительными. Сглаживающие методы аппроксимации, в отличие от интерполяционных, могут не сохранять значения в точках наблюдения. Методы интерполяции делятся на детерминированные и статистические. Первые – алгоритмы численного анализа. Вторые основаны на статистических методах и являются главным образом инструментами прогнозирования значений в неисследованных точках; рассматриваются в первую очередь статистические свойства – несмещенность оценки, дисперсия ошибки. Третьей, отдельной группой, является аппроксимация при помощи линии тренда. Она имеет сглаживающий характер и отражает общую тенденцию изменения данных.

Вариограммы. На первый взгляд множество наблюдений не содержит полезной информации о пространственной переменной помимо позиции точек наблюдения и значений в них. Методы геостатистики позволяют извлечь дополнительную информацию. Обеспечивают это вариограммы – Wolfram Mathematica содержит широкий набор инструментов для работы с вариограммами, функция VariogramModel позволяет задать предпочтительную из более 20 предопределенных моделей; функция BinnedVariogramList позволяет вычислить эмпирическую вариограмму на основе конкретного множества наблюдений.

Интерполяция Кригинга. Mathematica содержит набор инструментов для проведения интерполяции Кригинга. Ключевой является функция SpatialEstimate – принимает значения наблюдений и их позиций, возвращает функцию аппроксимации вместе с дополнительной информацией о ключевых показателях результата и процесса поиска (вариограмма, порядок тренда, порог шума), реализует "Универсальный метод Кригинга". В качестве параметров алгоритма можно уточнить функцию вариограммы и функцию тренда. При отсутствии явных значений, вариограмма вычисляется на наборе наблюдений при помощи функции EstimatedVariogramModel, автоматически подбирается функция тренда. При помощи параметра SpatialNoiseLevel можно определить порог перехода между локальным и глобальным трендом. Отдельно в докладе будут отмечены *Факторы качества восстановления*, влияние на точность цифрового поля плотности наблюдений, параметра *Удельная полезность*, критерия *Точность вариограммы*.

Составные цифровые поля. Изучение результатов аппроксимации рядов нестационарных про-

странственных переменных, позволило сформулировать идею и главное преимущество составных цифровых полей. Основное – возможности составления отдельных вариограмм, модели которых строятся по разным подмножествам наблюдений, для отдельных подобластей, минуя неустойчивый и ненадежный этап прогнозирования глобального тренда.

Адаптивный алгоритм составления модели поля. План работы алгоритма. На вход алгоритму подается начальное множество точек наблюдений их соответствующих значений. Отдельно, указывается количество этапов улучшения результата и соответствующие количества наблюдений, которые можно провести на каждом из этапов. Методом универсальной интерполяции Кригинга алгоритм выстраивает первоначальную функцию аппроксимации по начальному множеству наблюдений. В последующем, перед каждым из этапов, алгоритм рассчитывает оценку вариации точек наблюдения, а затем оценивает возможности разбиения множества наблюдений на отдельные (подобласти).

Если разбиение невозможно либо определяется недостоверным, дополнительные измерения проводятся в областях наименьшей плотности покрытия. Покрытие проводится отдельно в каждой области в пропорции площади и оценки улучшения ошибки (этап равномерного покрытия).

При возможности разбиения общей области исследования на отдельные подобласти, выбираются части с наилучшей оценкой улучшения ошибки. Эти подобласти добавляются в составное цифровое поле, включая все точки наблюдения, расположенные в них. Дополнительные расчеты проводятся в новых областях в пропорции оценки улучшения ошибки (этап образования подобласти). Наблюдения в существующих областях при вычислении оценки вариации и поиске новых областей не учитываются.

Особенности, трудности настройки алгоритма. Детали в докладе, только ключевые позиции – определение порогового значения для оценки точек возмущения, способ (метод, метрика) определения кластеров.

ЗАКЛЮЧЕНИЕ

Эффективность алгоритма. В финальной части проведены вычислительные эксперименты для оценок предложенного алгоритма на серии представительных (типовых) примеров с случайным набором наблюдений и данными наблюдений на сетке. Результаты будут представлены иллюстрациями и в таблице.