

АППРОКСИМАЦИЯ ДАННЫХ ПОВЕРХНОСТИ РАСТРОВЫХ КАРТ В ГЕОИНФОРМАЦИОННОЙ СИСТЕМЕ

Рассматриваются растровое представление данных в геоинформационных системах и задача аппроксимации имеющихся данных на растровой карте. Обосновывается выбор метода сплайн-аппроксимации для географической карты.

Ключевые слова: аппроксимация, растр, поверхность, геоинформационная система.

Географическая информационная система (ГИС) – это совокупность аппаратных средств, программного обеспечения, географических данных и персонала (операторов, программистов, системных аналитиков), предназначенная для эффективного ввода, хранения, обновления, обработки, анализа и визуализации всех видов географически привязанной информации.

Аппаратные средства включают в себя компьютеры (платформы), на которых работает ГИС. Такие ГИС, как ARC/INFO, функционируют на достаточно большом числе платформ – мощных серверов, обслуживающих клиентские машины в локальных сетях и сети Интернет, на рабочих станциях и отдельных персональных компьютерах. Кроме того, географические информационные системы используют разнообразное периферийное оборудование: дигитайзеры (для оцифровки карт), лазерные принтеры, плоттеры (для печати карт) и т. п.

Программное обеспечение позволяет вводить, сохранять, анализировать и отображать географическую информацию. Ключевыми компонентами программного обеспечения являются:

- средства для ввода и манипулирования географическими данными;
- система управления базой данных;
- программные средства, обеспечивающие поддержку запросов, географический анализ и визуализацию информации;
- графический интерфейс пользователя, облегчающий использование программных средств [1].

Геоинформационные системы предоставляют мощные средства для анализа экологической информации, однако сами по себе они не порождают новых знаний о состоянии окружающей среды, а являются только инструментом для исследователя.

В последние годы быстро развиваются информационные технологии, ориентированные на формирование знаний о состоянии окружающей среды. Программные модули, основанные на таких технологиях, используются в наиболее мощных географических информационных системах, но значительно чаще они интегрируются в экспертные системы, ориентированные на прогнозирование состояния окружающей среды, оценку риска хозяйственной деятельности и поддержку принятия решений, обеспечивающих устойчивое развитие экономики [2].

Географические данные являются главным компонентом ГИС. Эти системы работают с данными двух типов:

– описательными (атрибутивными, табличными) данными о географических объектах, состоящими из наборов чисел, текстов и т. п.;

– пространственными (картографическими, векторными) данными, описывающими положение и форму географических объектов и их пространственные связи с другими объектами.

Описательная информация организуется в реляционную базу данных. Кроме того, в ГИС описательная информация связывается с пространственными данными. Отличие ГИС от стандартных систем управления базами данных (СУБД) состоит именно в том, что ГИС позволяют работать с пространственными данными.

В настоящее время в организациях, занимающихся сбором и обработкой информации о территориях, накоплено и используется огромное количество картографических материалов на бумажной и жесткой основе. В большинстве современных геоинформационных систем имеется возможность использования растровых подложек, которые представляют собой отсканированные изображения картографического материала в различных масштабах.

Пространственные данные представлены в ГИС двумя основными моделями: векторной и растровой. Векторная модель основывается на представлении карты в виде точек, линий и плоских замкнутых фигур. В растровой модели территория отображается в виде совокупности регулярно организованных площадных объектов типа квадратного пиксела.

Пространственные данные могут быть организованы различными способами, определяемыми в первую очередь целью их использования, а также способом сбора и хранения. Специальные атрибуты могут хранить дополнительную информацию относительно местоположения, топологии и геометрии пространственных объектов.

В современных ГИС топологические атрибуты формируются автоматически при создании графической базы данных. Информация о пространственном положении объектов хранится либо в виде широты/долготы, либо в любой картографической проекции.

Одним из самых важных этапов при построении графической базы данных является геометрическое преобразование данных о пространственном местопо-

ложении объектов из одной системы координат в другую в общей картографической проекции. Такие преобразования необходимы для сравнения и анализа объектов, графическая информация о которых хранится в БД ГИС в разных картографических проекциях. Соответствующий математический аппарат широко представлен в стандартном программном обеспечении ГИС.

Растровая модель особенно хорошо подходит для представления явлений реального мира, имеющих непрерывное распределение, например температуры поверхности Земли. Растр представляет собой набор прямоугольных (чаще всего квадратных) ячеек – пикселей – и может быть представлен как прямоугольная матрица чисел подобно двумерным массивам в языках программирования. Для хранения информации в растровой модели можно пользоваться простой файловой структурой с прямой адресацией каждого пикселя или системой управления базой данных.

Растровая модель широко применяется при непосредственной обработке и анализе цифровых изображений, полученных по данным дистанционного зондирования поверхности Земли, а также при решении многих прикладных задач, в частности при мониторинге состояния окружающей среды. Основное преимущество этой модели состоит в слиянии графической и атрибутивной информации в единую регулярную структуру. Однако при моделировании пространства в растровом формате пространственные объекты могут быть представлены с большой точностью только за счет уменьшения размера пикселя, что ведет к увеличению стоимости хранения информации.

При использовании растровой модели операции сопоставления и перекрытия разных пространственных данных об одной территории могут быть проведены без каких-либо сложных геометрических вычислений. При этом в характерном для ГИС послыном представлении информации каждый растровый слой имеет отдельный атрибут. Положение ячейки адресуется номером строки и столбца, поэтому пространственные координаты для каждой ячейки в запоминании не нуждаются, хранятся лишь число колонок и строк и географическое положение первой (угловой) ячейки [3].

Пространственное разрешение растра представляет собой размер пикселя на поверхности Земли. При использовании хотя бы 1 байта на пиксел такое растровое изображение будет занимать достаточно большой объем памяти. Следовательно, требования к памяти возрастают экспоненциально с ростом разрешения, поэтому в растровых моделях применяются алгоритмы сжатия и специальные структуры данных, такие как лексикографический код, квадротомические деревья или код Фримана и т. д.

В растровом формате точечные объекты представляются единичным пикселем, а линии – строкой связанных пикселей. Это не всегда удобно, так как размер пикселя, зафиксированный во время создания растра, может оказаться слишком большим и многие детали будут потеряны. Однако растровая форма хо-

рошо подходит для моделирования пространственной непрерывности, особенно если соответствующий атрибут имеет высокую степень пространственной изменчивости. Такие ситуации часто возникают при обработке спутниковых изображений.

При наличии на растровой карте только части экспериментальных значений данные о поверхности необходимо распространить на всю территорию карты, а также за ее пределы. В настоящее время для этого используется метод простого усреднения.

Чтобы более точно смоделировать пространственную непрерывность методом усреднения, необходимо иметь большое количество точек измерения, которые должны равномерно располагаться на исследуемой территории. При недостатке данных могут быть использованы методы интерполяции поверхности растровых карт. Аргументами функции двух переменных в этом случае будем считать координаты растра (x, y) , а значением функции – количественную характеристику данного пикселя z . Следовательно, возникает задача аппроксимации и ее частные случаи – интерполяция и экстраполяция данных поверхности растровых карт.

Аппроксимация – это процесс получения оценки значений атрибутов в точках, расположенных близко к точкам измерений. Она используется:

- при трансформации растровых изображений;
- преобразовании моделей рельефа местности;
- моделировании непрерывной поверхности с помощью набора отдельных точек.

Аппроксимация, при которой требования точности в узлах исходной сетки предусматривают совпадение соответствующих значений исходной и аппроксимирующей функций, называется *интерполяцией*.

Существует следующие методы интерполяции поверхностей: глобальные, локальные, точные и приближенные.

В глобальных методах ко всем точкам измерений применяется одна и та же функция. Как правило, в качестве интерполяционных функций используются полиномы первой, второй и третьей степени:

$$\begin{aligned} z &= a + bx + cy, \\ z &= a + bx + cy + dx^2 + exy + fy^2, \\ z &= a + bx + cy + dx^2 + exy + fy^2 + \\ &+ gx^3 + hx^2y + icy^2 + fy^2, \end{aligned}$$

где z – искомая высота рельефа в точке с координатами (x, y) ; a, b, c, \dots, i – коэффициенты полиномов, значения которых определяются по известным значениям высот из систем линейных уравнений использованием метода наименьших квадратов.

В локальных методах интерполяция применяется последовательно для небольших групп точек исходного набора данных. Оптимальным и широко распространенным локальным методом интерполяции является кригинг. Этот метод легко модифицируется и его можно использовать для интерполяции данных любого типа. Он объединяет целое семейство линейных регрессионных алгоритмов: простой, стандартный, универсальный, блочный кригинг и ко-кригинг.

В точных методах интерполяции результирующая функция принимает в точках измерений те же значения, которые были до интерполяции. К этим методам относятся метод полигонов Тиссена и метод сплайнов.

Метод полигонов Тиссена в силу своей локальной природы глобальные характеристики набора данных не изменяет. Это метод является методом негладкой интерполяции, поскольку на границах полигонов возникают скачки функции. Полигоны Тиссена используются для интерполяции при небольшом числе исходных точек, при этом предполагается, что значение в каждой точке измерений является абсолютно точным.

Метод сплайнов, или сплайн-интерполяция, основывается на применении для интерполяции в окрестностях данного узла кусковых полиномиальных функций, называемых функциями сплайнов. Термин «сплайн» происходит от английского слова *spline* – гибкой линейки, с помощью которой чертежники проводят через заданные точки плавные кривые. Для двумерного случая (на плоскости) функция сплайна математически эквивалентна гибкой линейке и является кубическим полиномом (полиномом третьей степени) – непрерывной функцией, которая имеет непрерывные первую и вторую производные. Для трехмерного случая, когда вместо линии берется интерполированная поверхность, используются бикубические сплайны – полигоны третьей степени двух координат пространства. При сплайн-интерполяции интерполированная линия (двумерный случай) или поверхность (трехмерный случай) в точках измерений совпадает с полученными ранее значениями.

Дополнительно предъявляется требование согласования первых и вторых производных в граничных точках фрагментов и два условия: нулевая или заданная кривизна либо наклон. Условия образуют систему линейных алгебраических уравнений, которая решается с использованием точечных значений переменной, имеющих на каждом фрагменте исследуемой территории, и позволяет найти соответствующие значения коэффициентов полинома.

К достоинствам сплайн-интерполяции следует отнести высокую скорость обработки вычислительного алгоритма. Поскольку сплайн – это кусочно-полиномиальная функция, то при интерполяции одновременно обрабатываются данные по небольшому количеству точек измерений, принадлежащих к рассматриваемому фрагменту. Интерполированная поверхность описывает пространственную изменчивость различного масштаба и в то же время является гладкой. Последнее обстоятельство делает возможным прямой анализ геометрии и топологии поверхности с использованием аналитических процедур.

Однако гладкость интерполированного поверхности обуславливает невозможность корректного отображения резких изменений поверхности-оригинала, что является одним из недостатков метода сплайнов. К другим недостаткам этого метода следует отнести высокую зависимость точности моделирования поверхности от размещения точек измерений (или наблюдений), особенно критическое значение имеет

наличие точек на структурных линиях поверхности-оригинала – водоразделах и тальвегах, если речь идет о топографической поверхности; зависимость результата интерполяции от характера выделения фрагментов; отсутствие методики прямых оценок погрешностей, связанных со сплайн-интерполяцией [4].

В приближенных методах интерполяции исходные и результирующие значения в точках измерений могут различаться.

Требование точности предусматривает незначительное отклонение приближенных, т. е. вычисленных в процессе аппроксимации, значений в узлах исходной сетки от соответствующих исходных значений и, кроме того, малые отклонения вычисленных промежуточных значений от теоретически (или практически) возможных значений исходной функции. В основе этого требования лежит предположение, что аппроксимируемая функция непрерывна и имеет непрерывные производные.

Критерий, по которому отклонения могут считаться допустимо малыми, определяется точностью измерения и представления данных, их полнотой, различиями в колебаниях данных и т. д.

Сохранение непрерывности обеспечивает постепенность изменений, т. е. отсутствие разрывных скачков, аппроксимирующей (приближенной) функции не только при переходе от одной точки внутри промежутка, образованного соседними узлами, к другой близлежащей, но и в самих узлах [5].

Для получения точных результатов при интерполяции данных целесообразно использование методов сплайн-интерполяции. Это обусловлено тем, что исходные данные расположены по поверхности растровой карты не равномерно, а локальными группами, и глобальные методы в этом случае будут давать большие погрешности, ведь точность полиномиальной аппроксимации резко падает при увеличении степени аппроксимирующих полиномов. От этого недостатка можно избавиться с помощью отрезков полиномов невысокой степени, применяемых для представления части узловых точек. Самым известным методом такой аппроксимации является сплайн-аппроксимация на основе применения отрезков кубических полиномов. При этом аппарат сплайн-аппроксимация позволяет получить полиномы, которые дают в узловых точках непрерывность не только представляемой ими функции, но и ее первых и даже вторых производных.

В качестве аппроксимирующей функции – сплайна – строится цепочка полиномов или других сравнительно простых типов функций. Соседние по цепочке полиномы склеиваются так, что и их значения, и значения их производных на стыках совпадают, при этом все исходные значения попадают на некоторые стыки. Таким образом, внутри промежутков, образованных узлами исходной сетки, оказывается по одному (реже по два или три) полиному из указанной цепочки. Дополнительно выполняются значения производных на концах всего интервала. Внутренние стыки, не лежащие в узлах исходной сетки, рассматриваются как дополнительные узлы. Описанные сплайны называются основными. Параметры всех полиномов

основного сплайна связаны друг с другом и определяются с помощью решения системы линейных уравнений.

Другим вариантом сплайн-аппроксимации является обобщение метода скользящей средней. Каждое значение аппроксимирующей функции (сплайна) складывается из близлежащих исходных значений с коэффициентами в виде некоторых полиномов. А поскольку одновременному согласованию подлежат не все используемые полиномы, а только близлежащие, то такого рода сплайны называют локальными, или В-сплайнами.

Признанными достоинствами основных сплайнов и В-сплайнов считаются высокая точность, достаточная гладкость и технологическая отработанность. Но эти достоинства в полной мере проявляются лишь на равномерной сетке и при невысоких перепадах градиентов. Поэтому для восполнения данных и обеспечения их сопоставимости желательно сохранить достоинства сплайнов, но избавиться от отмеченных выше недостатков [5].

После вычисления значений промежуточных точках растровой карты необходимо распространить данные карты за ее пределы с целью прогнозирования.

Экстраполяция – это, собственно, сам процесс прогнозирования с помощью формулы или алгоритма, полученного из аппроксимации. При кажущейся простоте подстановки значений в уже найденную функцию необходимо убедиться в том, что область определения полученной функции на прогнозируемом интервале вписывается в допустимые значения реальной физической величины, которая аппроксимируется.

Кубическая сплайн-экстраполяция представляет собой быстрый и устойчивый способ экстраполяции функций и является альтернативой полиномиальной экстраполяции. В основе кубической сплайн-экстраполяции лежит заимствованный из сплайн-интерполяции принцип разбиения заданного экстраполируемого интервала на небольшие отрезки, на каждом из которых функция задается полиномом третьей степени.

Основными достоинствами кубической сплайн-экстраполяции являются ее устойчивость и малая трудоемкость, что позволяет получать коэффициенты кубического полинома с высокой точностью. В про-

цессе экстраполирования находится ближайшая точка, в которой уже вычислены коэффициенты сплайна, и строится полином, который дает значение искомой функции в текущей точке.

Однако метод сплайн-экстраполяции имеет и серьезный недостаток: значение ошибки будет увеличиваться с отдалением экстраполируемой точки от заданного начального интервала в случае линейности исходной функции.

При реализации алгоритмов аппроксимации предлагается объединить методы кубической и квадратичной (линейной) экстраполяции на границах экстраполируемого интервала, что поможет уменьшить ошибку, если отсутствует информация о характере поведения функции с одной из сторон исследуемой точки [6].

Таким образом, наиболее эффективным методом аппроксимации данных поверхности растровых карт является метод сплайнов.

Библиографические ссылки

1. Сайт группы компаний «Центр пространственных исследований» [Электронный ресурс]. URL: <http://www.geointellect.ru/?id=137> (дата обращения: 20.12.2011).
2. Принципы географического анализа экологической информации [Электронный ресурс]. URL: <http://loi.sssc.ru/gis/ecoinf/C4-1.htm> (дата обращения: 20.12.2011).
3. Пространственная информация и ее свойства [Электронный ресурс]. URL: <http://www.agiks.ru/data/articles/ddzz/pi.htm> (дата обращения: 20.12.2011).
4. Локально-детерминированные методы интерполяции. Метод сплайнов [Электронный ресурс] // Ты и я : науч.-попул. журнал. 2010. URL: <http://www.tyia.ru/lokalno-determinirovannyye-metody-interpolyacii-metod-splajnov> (дата обращения: 20.12.2011).
5. Математическая обработка данных: изометрические задачи [Электронный ресурс]. URL: <http://www.sati.archaeology.nsc.ru/Home/pub/Data/?html=Chap8.htm&id=210> (дата обращения: 20.12.2011).
6. Шалагинов А. В. Кубическая сплайн-экстраполяция временных рядов // Системный анализ и информационные технологии : материалы Междунар. науч.-техн. конф. Киев, 2011. С. 397.

V. V. Kukarcev, O. A. Antamoshkin

APPROXIMATION OF THE SURFACE DATA OF RASTER TYPE MAPS IN THE GEOINFORMATION SYSTEM

The authors consider a bitmap representation of data in geographic information systems, and the problem of approximation of the available data on a raster type map and substantiate the choice of the spline approximation method for a geographic map.

Keywords: approximation, raster, surface, geoinformation system.

© Кукарцев В. В., Антамошкин О. А., 2012