

Харьковский национальный университет им. В.Н. Каразина

Кафедра мониторинга окружающей среды и природопользования

Модуль 3. Аналіз, синтез та математична обробка
екологічної інформації за допомогою MapInfo

Тема 5. Розміщення даних на карті

Заняття 19. **Лекция 6** Модель данных и
алгоритмы анализа в ГИС

Учебные вопросы

- **1. Оцифровка карты.**
- **2. Генерализация картографического изображения в ГИС.**

1. ОЦИФРОВКА КАРТЫ

Географические явления и феномены бесконечно сложны и разнообразны. Для того, чтобы абсолютно полно представить их в ГИС, потребовались бы бесконечно большие пространственные базы данных. Чтобы описывать, обрабатывать и анализировать объекты реального мира в ГИС, используют модели данных, в которых абстракцией и агрегацией данных устанавливается необходимый для решения поставленной задачи уровень детальности информации. Это касается как атрибутивной, так и пространственной составляющих геоданных. На географических картах уровень детальности информации определяется масштабом карты. Здесь в качестве примера абстракции данных можно привести процесс камерального создания мелкомасштабных карт на основе крупномасштабных путем их генерализации. Пользователь ГИС смотрит на окружающий мир через модель пространственных данных, которая должна с достаточной полнотой и точностью описывать объекты и явления.

Изображения объектов и явлений на географических картах и космических снимках обычно описываются функцией двух переменных $F(x,y)$, где x,y – географические координаты. Аналитический способ определения функции $F(x,y)$ с целью определения пространственного положения географических объектов практически не применяется. Вместо этого функцию $F(x,y)$ сэмпляют – т.е. представляют выборочными значениями в некоторых точках (их называют выборочными точками).

Способ дискретизации функции $F(x,y)$ определяет модели данных ГИС: растровые модели, DLG-графы векторных ГИС, модели поверхностей TIN и DEM. Представление феноменов и явлений объективной реальности выборочными значениями и кодирование этих значений в пространственной базе данных называется вводом данных в ГИС.

Большая часть характеристик географических объектов и явлений существует на всей территории и изменяется непрерывно (рельеф территории, температура, атмосферное давление и т.п.). Если взять выборочные точки в узлах регулярной решетки, получим растровую модель данных. Размещение точек вдоль трансект либо вдоль границ объектов порождает векторную модель (в виде полилиний и полигонов), а изображение объектов и явлений в изолиниях приводит к 3D-моделям.

Феномены окружающего мира могут быть рассмотрены в трех аспектах: пространственном (вариации значений от одного места на карте до другого), временном (изменения между двумя моментами времени) и тематическом (вариации от одной характеристики до другой). Измеряемые или описываемые свойства объектов попадают в одну из этих категорий, и их полное описание во всех трех аспектах практически невозможно. При построении ГИС-моделей одна из категорий фиксирована, другая – контролируется, а третья – свободно изменяется. Например, при переписи населения фиксируется время (год переписи), контролируется местоположение, используя census track, и измеряется разная тематическая информация (возраст, пол и т.п.).

Ввод данных – наиболее узкое место в геоинформационных технологиях. Затраты на ввод данных часто превосходят 80% от общей стоимости ГИС-проекта. Процесс ввода данных обычно требует участия большого количества операторов, сами операции ввода – утомительны (однообразная, рутинная работа), требуется постоянно контролировать ошибки ввода данных.

Поэтому для снижения трудозатрат и стоимости ввода и для повышения качества ГИС-продуктов нужно автоматизировать процессы ввода данных, насколько это возможно.

В ГИС используется несколько режимов ввода данных. Путем ручного ввода (например, с клавиатуры) могут быть введены табличные данные, элементы оформления и дизайна карты, реже пространственные данные (из-за их большого объема). Ручные устройства определения координат, при помощи которых оператор непосредственно указывает местоположение географического объекта и фиксирует его координаты, позволяют оцифровывать карты, выполнять ручное дешифрирование снимков. Автоматизированные устройства ввода автоматически извлекают геоданные с карт и снимков. В настоящее время производители ГИС пытаются использовать для ввода данных новые технологии, например, голосовой ввод, но в общераспространенном программном обеспечении ГИС эти возможности пока недоступны.

Каждая строка этого перечня содержит сведения об одном человеке. Каждая запись содержит несколько типов сведений о человеке. В терминах баз данных каждая такая строка называется записью. Разные виды сведений называют полями. Поля соответствуют колонкам, так что приведенная выше таблица содержит четыре различных поля (Фамилия, Имя, Адрес и Почтовый индекс).

Поля базы данных упорядочены (первое, второе, третье и т.д.), и обычно первое поле показывается в виде первой колонки базы данных. Второе поле показывается справа от первой и т.д., последнее поле показывается в самой правой колонке.

В примере записи приведены не в алфавитном порядке, как это обычно бывает в справочниках. Для быстрой обработки запросов (поиска по базе) базы данных используют механизм индексации. Без индексов было бы сложно найти что-либо в базе из сотен записей, не говоря уже о базах из сотен тысяч записей.

Индекс в базе данных подобен указателю в книге. Индекс представляет собой Набор ссылок (как и в книге, где каждому названию соответствуют номера страниц). В книгах указатели, как правило, используют алфавитный порядок, чтобы легче было найти интересующую тему.

Так, MapInfo требует, чтобы были проиндексированы те поля, к которым Вы применяете команду **Запрос > Найти**. Индексы также используются для оптимизации процедур объединения таблиц и SQL-запросов.

Запрос – это по сути то же самое, что и вопрос. Данные, которые отбираются в результате выполнения запроса, являются частью базы данных. Например, если надо узнать, сколько клиентов проживает в радиусе двух километров от Вашего сервис-центра (и информация об адресах проживания клиентов имеется в базе), то надо сделать соответствующий запрос к такой базе. Результатом будут данные запроса. Можно считать данные запроса подмножеством Ваших данных, как в приведенном примере – список клиентов, проживающих не далее 2 километров от сервис-центра. MapInfo Professional имеет встроенные модули, помогающие делать запросы и показывать их на карте.

Данные для ГИС-проекта могут быть также получены из других ГИС и CAD/CAM – систем; в этом случае ввод данных сводится к преобразованиям между различными ГИС-форматами. Существуют специальные обменные форматы для передачи пространственных данных между разными ГИС (DFX/DBF, MIF/MID и т.п.). Современные попытки разработки стандартов на пространственные данные и на процедуры обмена базируются на XML.

Одним из способов снижения затрат на ввод данных является также разделение цифровых данных (data sharing), когда несколько коллективов, занятых разработкой ГИС, совместно создают, владеют и используют банк геопространственных данных, что позволяет избежать ситуаций, когда две организации тратят ресурсы на ввод одинаковых карт. Кроме того, сейчас в сети Интернет появляется все больше свободно доступных пространственных данных (космические снимки и цифровые карты GoogleEarth, цифровые модели местности NASA и многое другое).

Для ввода данных в ГИС необходимы технические и программные средства преобразования пространственных данных различных типов в цифровую форму. Пространственные данные кодируются в виде списка координат, а атрибутные данные чаще всего представлены в виде таблиц. Для пространственных данных часто требуются преобразования между различными проекциями и системами координат цифровой карты, космического снимка и т.п.

Первыми устройствами для аналого-цифрового преобразования картографической информации в ГИС были дигитайзеры – устройства ручной оцифровки карт, схем и планов в виде последовательности точек, положение которых описывается прямоугольными декартовыми координатами плоскости. Дигитайзер состоит из плоского стола и указующего устройства – съемника информации в виде курсора или пера. Рабочее поле стола может быть выполнено из прозрачного материала и иметь подсветку. Дигитайзеры различаются размером рабочего поля (форматы А4–А0), точностными характеристиками: точностью (accuracy), контролируемой погрешностью курсора, точностью поля дигитайзера, конструктивным разрешением (величиной минимального шага – инкремента). Интегральная точность современных дигитайзеров обычно лежит в пределах сотых или десятых долей миллиметра.

Растровая модель пространственных данных описывает характер исследуемого географического явления на всей территории. Растровая модель представляет окружающий мир в виде регулярной сети ячеек и является самой простой из распространенных моделей пространственных данных. Растр – набор данных, имеющих географический характер, значения которых организованы в прямоугольный массив объектов.



Разрешение растрового слоя можно охарактеризовать как минимальный линейный размер самой мелкой части географического пространства, для которого в слое записываются данные. В общем случае эта часть прямоугольная, но чаще всего – квадратная. Ориентация слоя – угол между действительным географическим севером и направлением, заданным линией столбцов раstra. Зона растрового слоя – множество соприкасающихся клеток раstra, имеющих одинаковое значение. Не все растровые карты содержат зоны – если клетки слоя содержат значения непрерывно изменяющегося в пространстве явления, этот слой не будет содержать зон. Значения ячеек – хранящаяся в слое информация о географическом явлении по строкам и столбцам. Ячейки, принадлежащие одной зоне, имеют одинаковые значения.

Местоположение определяется упорядоченной парой ординат (номером строки и столбца ячейки). Обычно известны реальные географические координаты нескольких углов растрового изображения.

В большинстве случаев с ячейкой раstra связано только одно значение. Совокупность этих ячеек со связанными значениями образуют растровый слой. База данных может содержать несколько таких слоев.

Оцифровку содержимого карты можно производить в точечном режиме, когда оператор устанавливает точки явно нажатием кнопки на курсоре, или в потоковом режиме, при котором положение точек снимается автоматически через некоторый промежуток времени (обычно 0,1–1 сек.) или при движении курсора на заданное количество позиций. В точечном режиме точки выбираются субъективно; два оператора никогда не оцифруют одну и ту же ломаную одинаково. Но оператор в этом режиме может размещать точки более правильно, т.е. чаще ставить точки на изгибах линий и реже – на прямых участках. В потоковом режиме точки размещаются независимо от формы кривой, поэтому создается много избыточных точек, которые нужно отфильтровать.

Большая часть существующих бумажных карт создается без учета того, что когда-нибудь их будут оцифровывать, поэтому при оцифровке карт неизбежно возникают проблемы.

Часто на картах в целях более наглядного изображения географических объектов жертвуют их позиционной точностью. Например, по узкой береговой полосе проходят ЛЭП, автомобильная и железная дороги. Если вынести эти объекты на карту точно по тем координатам, где они расположены, три линии могут слиться в одну. В этом случае эти линии рисуют на небольшом расстоянии друг от друга, пренебрегая позиционной точностью и сохраняя топологию объектов. Если граница территорий проходит по естественным объектам (по реке, не выраженной в масштабе карты по ширине), линию границы изображают поочередно на разных берегах этой реки.





2. ГЕНЕРАЛИЗАЦИЯ КАРТОГРАФИЧЕСКОГО ИЗОБРАЖЕНИЯ В ГИС

Сводка сегментов ГИС-проекта

Планшет топографической карты, основной источник данных для создания цифровых карт, является самостоятельным картографическим производением. В геоинформационных системах планшеты цифровых карт часто используются совместно с целью создания топографических основ какой-либо территории, например, административного района или края в целом.

Для решения этой задачи необходимо совмещение тополого-метрической информации на границах планшетов.

Совместное использование планшетов цифровых топографических карт в системе координат Гаусса–Крюгера требует пересчета координат в одну зону. Но даже после этого метрика объекта, пересекающего границы планшета, будет разорвана. Величина расхождения в метрике чаще всего лежит в пределах картографической точности, поэтому корректной будет операция по усреднению метрики объектов на стыках листов. Эта операция называется "сводкой" планшетов топографических карт.

Сводка планшетов сводится к выполнению ряда простых операций. В зависимости от расположения сводимых листов задается широта или долгота сведения. Задается критерий близости объектов – минимальное расстояние между вертексами линейных объектов на соседних листах, при котором они считаются различными. Алгоритм сводки состоит в последовательном сравнении с критерием близости расстояний между вертексами всех элементах декартового произведения множеств объектов двух сводимых листов.

Чтобы не перебирать все пары объектов, по заданной ширине полосы сведения выбираются объекты, имеющие выход на границу со сводимым листом. В результате получим множество пар объектов, вертексы которых находятся достаточно близко друг от друга, чтобы считать эти объекты изображением одного предмета. Если расстояние между вертексами объектов не превосходит точности карты, их можно свести без ошибок в одну точку.

а) – выделение объектов, имеющих выход на границу листа; б) – поиск пар, представляющих один объект на местности; в) – усреднение

Генерализация картографического изображения в ГИС

Под картографической генерализацией понимают процесс обработки картографического изображения, осуществляемый посредством абстрагирования и содержательного обобщения объекта в соответствии с масштабом и назначением карт. Цифровые карты крупных масштабов являются источником данных для создания мелкомасштабных карт, применяемых для печати обзорных карт небольших форматов. Мелкомасштабные карты создаются путем генерализации исходных цифровых карт.

В ГИС понятие генерализации задается в более обобщенном виде – как редукция "информационной плотности" в пространственной базе данных, при которой сохраняется ее общая структура и семантика. Процесс генерализации в ГИС сложен и требует больших вычислительных затрат. Высокая вычислительная сложность машинной генерализации объясняется тем, что в этом процессе учитываются как геометрические аспекты географических сущностей – форма, структура, детализация, так и негеометрические – роль и важность объекта в контексте карты (рисунок 4).



Достижение полной автоматизации этого процесса невозможно по причине бесконечного многообразия отношений между объектами и явлениями. Наиболее изученная часть проблемы – упрощение геометрии объектов. Здесь самым известным и часто используемым алгоритмом возможно является эвристический алгоритм Дугласа и Пеке для простой открытой полигональной цепи.

На мелкомасштабных картах отображается обзорное представление картографируемого объекта или явления. Для этой задачи мелкие детали, имеющиеся на картах более крупного масштаба, только препятствуют комплексному взгляду на предмет. Поэтому при создании карт мелких масштабов на исходной карте выделяются наиболее общие классы объектов: реки, озера, крупные населенные пункты, автомобильные и железные дороги. Содержательная разгрузка карты выполняется также путем удаления мелких объектов. С исходной карты удаляются линейные объекты, длина которых меньше некоторой величины, определяемой масштабом карты. Таким же способом – выборкой по площади – удаляются мелкие полигональные объекты.

При уменьшении масштаба объекты могут менять свою пространственную локализацию. Площадные объекты, которые в новом масштабе не выражаются по одному из направлений, заменяются линейными объектами.

Если площадной объект перестает выражаться по всем направлениям, то он заменяется точечным. Линейный объект заменяется точечным, если его длина перестает выражаться в масштабе карты.

Агрегация предполагает переход от частных понятий к общим понятиям и контролируется в основном негеометрическими правилами. Группа близко расположенных объектов при уменьшении масштаба может быть выражена единым условным знаком, обозначающим некоторое собирательное понятие.

Геометрическая сторона картографической генерализации алгоритмически наиболее сложна. Для разных типов объектов решение задачи генерализации будет своим. Однако, можно выделить общие операции, которые применяются при генерализации картографических объектов: замена ломаной отрезком при заданном предельном угле излома, удаление точек, лежащих на одной прямой, объединение нескольких условных знаков один, замена масштабного изображения внемасштабным.

а) замена ломаной линии прямой при заданном предельном угле излома; б) замена масштабного изображения внемасштабным знаком; в) удаление точек, лежащих между близкими параллельными прямыми; г) объединение нескольких условных знаков

Существует несколько алгоритмов генерализации полилиний. Алгоритмы, независимые от формы линии, наиболее просты – в них просто удаляется заданное количество точек (например, удаление каждой точки), но генерализируют форму линии очень грубо и могут порождать топологические ошибки.

В локальных алгоритмах чтобы определить, оставить текущую точку или нет, используются характеристики соседних точек. Линейные и площадные объекты цифровой карты описываются последовательностью точек, которые должны быть расположены достаточно близко, чтобы обеспечить плавность линии при визуализации в исходном масштабе карты. В мелком масштабе плавность линий обеспечивается меньшим числом точек. Поэтому ломаная на участках, имеющих небольшой угол излома, заменяется прямой линией. Если несколько идущих подряд точек лежат на одной прямой, они не несут информации и удаляются.

В глобальных алгоритмах генерализации полилиния рассматривается целиком. Типичным примером глобальной генерализации, нашедшим применение во многих современных ГИС-пакетах, является алгоритм Дугласа-Пекье. Рассмотрим его работу (рисунок 6).

Сначала введем параметр толерантности ε , устанавливающий степень генерализации карты, зависимый от масштаба и выбираемый эмпирически.

Пусть V_i – начальная точка, а V_j – конечная точка полилинии. Вычислим расстояния от всех остальных точек полилинии до прямой $V_j V_i$, найдем из них максимальное d_{\max} и обозначим наиболее удаленную от $V_j V_i$ точку V_f . Если $d_{\max} < \varepsilon$, то все точки между V_i и V_j удаляются, а $V_j V_i$ представляет генерализованную линию. Если $d_{\max} \geq \varepsilon$, то полилиния $V_j V_i$ расщепляется на две части – $V_j V_f$ и $V_f V_i$ для которых рекурсивно выполняется данный алгоритм.

Генерализация полилинии алгоритмом Дугласа-Пекье: а) – определение самой дальней точки от прямой; б) – рекурсия