Практическое занятие 1

Объектная модель картографической основы.

Модель географических данных , модель геопространственных данных или просто модель данных в контексте географических информационных систем — это математическая и цифровая структура для представления явлений на Земле. Как правило, такие модели данных представляют различные аспекты этих явлений с помощью географических данных , включая пространственное положение , атрибуты, изменения во времени и идентичность. Например, модель векторных данных представляет географию в виде набора точек, линий и полигонов, а модель растровых данных представляет географию в виде матриц ячеек, в которых хранятся числовые значения. [1]Модели данных реализованы во всей экосистеме ГИС, включая программные инструменты для управления данными и пространственного анализа , данные, хранящиеся в различных форматах файлов ГИС , спецификации и стандарты, а также специальные конструкции для установок ГИС.

В то время как уникальная природа пространственной информации привела к собственному набору структур моделей, большая часть процесса моделирования данных аналогична остальным информационным технологиям, включая переход от концептуальных моделей к логическим моделям , а затем к физическим моделям , а также различие между типовые модели и конструкции для конкретных приложений.

Самые ранние компьютерные системы, которые представляли географические явления, были моделями количественного анализа, разработанными во время количественной революции в географии в 1950-х и 1960-х годах; их нельзя было назвать географической информационной системой, потому что они не пытались хранить географические данные в согласованной постоянной структуре, а обычно представляли собой статистические или математические модели. Первое настоящее программное обеспечение ГИС моделировало пространственную информацию с использованием моделей данных, которые впоследствии стали известны как растровые или векторные:

Типы моделей данных

Дополнительная информация: модель данных .

Поскольку мир намного сложнее, чем может быть представлен на компьютере, все геопространственные данные являются неполными приближениями к миру. [9] Таким образом, большинство моделей геопространственных данных кодируют некоторую форму стратегии для сбора конечной выборки из часто бесконечной области и структуру для организации выборки таким образом, чтобы обеспечить возможность интерполяции характера невыборочной части. Например, здание состоит из бесконечного числа точек в пространстве; векторный многоугольник представляет его несколькими упорядоченными точками, которые соединены в замкнутый контур прямыми линиями и предполагают, что все внутренние точки

являются частью здания; кроме того, атрибут «высота» может быть единственным представлением его трехмерного объема.

Процесс проектирования моделей геопространственных данных в целом аналогичен моделированию данных , по крайней мере, в его общей структуре. Например, его можно разделить на три различных уровня абстракции модели: [10]

Концептуальная модель данных , высокоуровневая спецификация того, как информация организована в уме и в корпоративных процессах, без учета ограничений ГИС и других компьютерных систем. Обычно для разработки и визуального представления концептуальной модели используются такие инструменты, как модель объект-связь .

Логическая модель данных , широкая стратегия представления концептуальной модели на компьютере, иногда новая, но часто в рамках существующего программного обеспечения, аппаратного обеспечения и стандартов. Унифицированный язык моделирования (UML), в частности диаграмма классов , обычно используется для визуальной разработки логических и физических моделей.

Физическая модель данных , подробная спецификация того, как данные будут структурированы в памяти или в файлах.

Каждая из этих моделей может быть разработана в одной из двух ситуаций или областей применения :

Общая модель данных предназначена для использования в самых разных приложениях путем обнаружения устойчивых закономерностей в том, как общество в целом концептуализирует информацию и/или структуры, наиболее эффективно работающие в компьютерах. Например, поле — это общая концептуальная модель географических явлений, модель реляционной базы данных и вектор — это общие логические модели, а формат шейп - файла — это общая физическая модель. Эти модели обычно реализуются непосредственно в информационном программном обеспечении и форматах файлов ГИС . В прошлом эти модели разрабатывались академическими исследователями, организациями по стандартизации, такими как Open Geospatial Consortium ., а также поставщиками программного обеспечения, такими как Esri . В то время как академические и стандартные модели являются общедоступными (а иногда и с открытым исходным кодом), компании могут хранить детали своей модели в секрете (как Esri пыталась сделать с покрытием и файловой базой геоданных) или публиковать их открыто (как Esri сделала с шейп-файл). [11]

Конкретная модель данных или ГИС-проект — это спецификация данных, необходимых для конкретного предприятия или проекта ГИС-приложения. Обычно он создается в рамках ограничений выбранных общих моделей данных, чтобы можно было использовать существующее программное обеспечение ГИС. Например, модель данных для города будет включать список слоев данных, которые должны быть включены (например, дороги, здания, участки, зонирование), каждый из которых определяется типом используемой общей пространственной модели данных (например, растровой или векторной).), выбор таких параметров, как система координат и ее столбцы атрибутов.

Концептуальные пространственные модели

Общие геопространственные концептуальные модели пытаются отразить как физическую природу географических явлений, так и то, как люди думают о них и работают с ними. [12] В отличие от стандартного процесса моделирования, описанного выше, модели данных, на которых построена ГИС, изначально не были разработаны на основе общей концептуальной модели географических явлений, а были в значительной степени разработаны в соответствии с технической целесообразностью, вероятно, под влиянием концептуальных представлений здравого смысла, которые еще не были задокументированы.

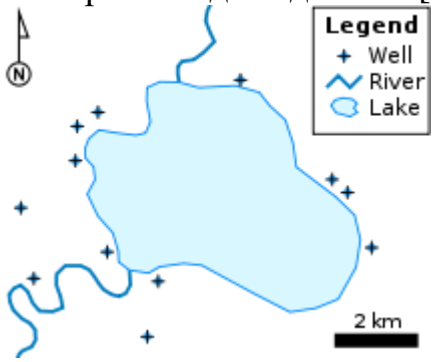
Тем не менее, ранней концептуальной основой, которая оказала большое влияние на раннее развитие ГИС, было признание Брайаном Берри и другими, что географическая информация может быть разложена на описание трех очень разных аспектов каждого явления: пространства, времени и атрибута/свойства/. тема. [13] В качестве дальнейшего развития в 1978 году Дэвид Синтон представил структуру, которая охарактеризовала различные стратегии измерения, данных и картирования как сохранение одного из трех аспектов постоянным, контроль второго и измерение третьего. [14]

Объект (также называемый функцией или сущностью) — это отдельная «вещь», понимаемая как единое целое . Это может быть видимый материальный объект, такой как здание или дорога, или абстрактная сущность, такая как округ или рыночная площадь розничного магазина.

Поле — это свойство, которое изменяется в пространстве , так что оно потенциально имеет отдельное измеримое значение в любом месте в пределах своего экстента. Это может быть физическая, непосредственно измеряемая характеристика материи, родственная интенсивным химическим свойствам, таким как температура или плотность; или это может быть абстрактное понятие, определяемое с помощью математической модели, например вероятность того, что человек, живущий в каждом месте, будет использовать местный парк. [15]

Эти две концептуальные модели не предназначены для представления разных явлений, но часто представляют собой разные способы концептуализации и описания одного и того же явления. Например, озеро — это объект, но температура, прозрачность и степень загрязнения воды в озере — это каждое поле (саму воду можно рассматривать как третье понятие массы , но это не так широко принято) . как объекты и поля). [16]

Векторная модель данных [править]



Простой набор векторных данных с точками, линиями и полигонами, представляющими водные объекты.

Векторная логическая модель представляет каждое географическое положение или явление геометрической формой и набором значений его атрибутов. Каждая геометрическая форма представлена с помощью координатной геометрии , структурированным набором координат (x,y) в географической системе координат , выбранным из набора доступных геометрических примитивов , таких как точки, линии и многоугольники.

Хотя существуют десятки форматов векторных файлов (то есть моделей физических данных), используемых в различных программах ГИС, большинство из них соответствуют спецификации Simple Feature Access (SFA) от Open Geospatial Consortium (OGC). Он был разработан в 1990-х годах путем поиска точек соприкосновения между существующими векторными моделями и теперь закреплен как ISO 19125, эталонный стандарт для векторной модели данных. OGC-SFA включает следующие векторные геометрические примитивы : [17]

Точка : одна координата в двух- или трехмерном пространстве. Многие векторные форматы позволяют одному объекту состоять из нескольких изолированных точек (MultiPoint в OGC-SFA).

Кривая (также называемая ломаной или линией): линия включает в себя бесконечное количество точек, но она представлена конечным упорядоченным набором точек (называемых вершинами), что позволяет программному обеспечению интерполировать промежуточные точки. Традиционно это была линейная интерполяция (OGC-SFA называет этот случай LineString), но некоторые векторные форматы позволяют использовать кривые (обычно дуги окружности или кривые Безье) или одиночный объект, состоящий из нескольких непересекающихся кривых (MultiCurve в OGC). -СФА).

Многоугольник : область также включает в себя бесконечное количество точек, поэтому векторная модель представляет ее границу в виде замкнутой линии (называемой кольцом в OGC-SFA), что позволяет программному обеспечению интерполировать внутреннюю часть. Программное обеспечение ГИС различает внутреннюю и внешнюю части, требуя, чтобы линия располагалась против часовой стрелки, поэтому внутренняя часть всегда находится на левой стороне границы. Почти в каждом формате многоугольник может иметь «дыры» (например, остров в озере) за счет включения внутренних колец, каждое по часовой стрелке (таким образом, внутренняя часть остается слева). Как и в случае с линиями, могут быть разрешены изогнутые границы; обычно один объект может включать в себя несколько полигонов, которые OGC-SFA в совокупности называют поверхностью .

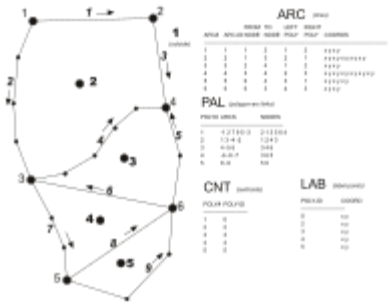
Текст (также называемый аннотациями): меньшинство форматов векторных данных, включая базу геоданных Esri и Autodesk .dwg , поддерживают хранение текста в базе данных. Аннотация обычно представляется в виде точки или кривой (базовой линии) с набором атрибутов, определяющих содержание текста и характеристики дизайна (шрифт, размер, интервал и т. д.).

Геометрическая форма, хранящаяся в наборе векторных данных, представляющих явление, может иметь или не иметь того же размера , что и само явление реального мира. [18] Общепринято представлять объект в более низком измерении, чем его реальная природа, исходя из масштаба и цели представления. Например, город (двухмерная область) может быть представлен в виде точки, а дорога (трехмерная структура) может быть представлена в виде линии. Пока пользователь знает, что последнее является выбором представления, а дорога на самом деле не является линией, это обобщение может быть полезным для таких приложений, как анализ транспортной сети .

Основываясь на этой базовой стратегии геометрических форм и атрибутов, модели векторных данных используют различные структуры для их сбора в единый набор данных (часто называемый слоем), обычно содержащий набор связанных объектов (например, дорог). Их можно разделить на несколько подходов:

Геореляционная модель данных была основой для большинства ранних векторных программ ГИС. [19] Геометрические данные и данные атрибутов хранятся отдельно; изначально это было связано с тем, что для обработки геометрических данных требовался код, специфичный для ГИС, но для управления атрибутами можно было использовать существующее программное обеспечение реляционной базы данных (RDBMS). Например, Esri ARC/INFO (позже ArcInfo) изначально состояла из двух отдельных программ: ARC была написана Esri для пространственного управления и анализа, а INFO была лицензированной коммерческой программой РСУБД. Он был назван «геореляционным», потому что в соответствии с принципами реляционных баз данных геометрия и атрибуты могут быть объединены путем сопоставления каждой формы со строкой в таблице с использованием ключа ., например номер строки или идентификационный номер. [20]

Пространственная база данных (также называемая объектной моделью [20]) впервые появилась в 1990-х годах. Он также использует зрелость систем управления реляционными базами данных , особенно их способность управлять чрезвычайно большими корпоративными базами данных. Вместо того, чтобы хранить геометрические данные отдельно, пространственная база данных определяет тип данных геометрии, что позволяет хранить формы в столбце той же таблицы, что и атрибуты, создавая единый унифицированный набор данных для каждого слоя. Большинство программ РСУБД (как коммерческих, так и с открытым исходным кодом) имеют пространственные расширения, позволяющие хранить и запрашивать геометрические данные, обычно основанные на стандарте Simple Features-SQL от Open Geospatial Consortium . [21]Некоторые форматы данных, не относящиеся к базе данных, также объединяют геометрические и атрибутивные данные для каждого объекта в единую структуру, например GeoJSON .



Изображение модели данных покрытия Arc/INFO, георeляционной топологической векторной модели данных, основанной на ранней модели данных POLYVRT.

Векторные структуры данных также можно классифицировать по тому, как они управляют топологическими отношениями между объектами в наборе данных: [22]

Топологическая модель данных включает в себя топологические отношения как основную часть дизайна модели. [18] : 46 Формат GBF/DIME от Бюро переписи населения США, вероятно, был первой топологической моделью данных; другим ранним примером был POLYVRT, разработанный в Гарвардской лаборатории компьютерной графики и пространственного анализа в 1970-х годах, который в конечном итоге превратился в формат Esri ARC/INFO Coverage. [7] [19] В этой структуре линии ломаются во всех точках пересечения; эти узлызатем может хранить топологическую информацию о том, какие линии там соединяются. Полигоны не хранятся отдельно, а определяются как набор линий, которые все вместе замыкаются. Каждая строка содержит информацию о полигонах справа и слева от нее, тем самым явно сохраняя топологическую смежность. Эта структура была разработана для использования составных линейно-многоугольных структур (например, блока переписи), адресного геокодирования и анализа транспортной сети . Преимущество этого метода заключалось также в повышении эффективности хранения и уменьшении количества ошибок, поскольку общая граница каждой пары смежных полигонов оцифровывалась только один раз. Однако это довольно сложная структура данных. Почти все топологические модели данных также являются георeляционными.

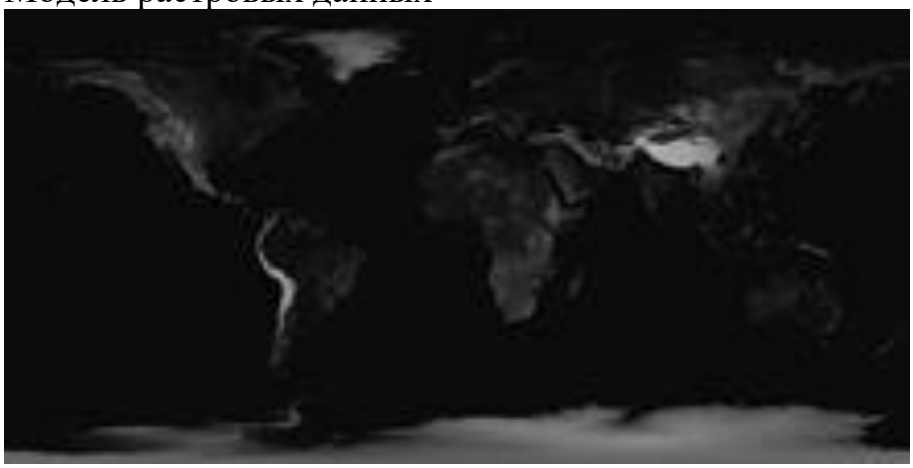
Модель данных спагетти не включает никакой информации о топологии (так называемой, потому что отдельные нити в миске со спагетти могут перекрываться, не соединяясь). [10] : 215 Это было распространено в ранних ГИС-системах, таких как Map Overlay and Statistical System (MOSS), а также в самых последних форматах данных, таких как шейп -файл Esri , язык географической разметки (GML) и почти во всех пространственных базах данных .. В этой модели геометрия каждого объекта кодируется отдельно от любых других в наборе данных, независимо от того, могут ли они быть связаны топологически. Например, общая граница между двумя соседними областями будет дублироваться в каждой форме многоугольника. Несмотря на увеличенный объем данных и вероятность ошибок по сравнению с топологическими данными, эта модель доминировала в ГИС с 2000 года, в

основном благодаря своей концептуальной простоте. Некоторое программное обеспечение ГИС имеет инструменты для проверки правил топологической целостности (например, не позволяющих многоугольникам перекрываться или иметь пробелы) на спагетти-данных для предотвращения и/или исправления топологических ошибок.

Гибридная топологическая модель данных позволяет хранить информацию о топологических отношениях в виде отдельного слоя, построенного поверх набора данных спагетти. Примером может служить набор сетевых данных в базе геоданных Esri . [23]

Векторные данные обычно используются для представления концептуальных объектов (например, деревьев, зданий, округов), но они также могут представлять поля . В качестве примера последнего поле температуры может быть представлено нерегулярной выборкой точек (например, метеостанции) или изотермами , выборкой линий одинаковой температуры. [10] : 89

Модель растровых данных



Растровая сетка высот

См. Также: Растровая графика и форматы файлов ГИС § Растр .

Растровая логическая модель представляет собой поле с использованием тесселяции географического пространства в равномерно распределенный двумерный массив местоположений (каждое из которых называется ячейкой) с одним значением атрибута для каждой ячейки (или более чем одним значением в многоканальном растре).). Как правило, каждая ячейка представляет либо одну центральную точечную выборку (в которой модель измерений для всего растра называется решеткой), либо представляет собой сводку (обычно среднее значение) переменной поля по квадратной области (в которой модель называется сеткой). [9] : 86 Общая модель данных практически такая же, как и для изображений и другой растровой графики ., с добавлением возможностей для географического контекста. Ниже приведен небольшой пример:

Количество осадков в мае 2019 г. (мм)

6	7	10	9	8	6	7	8
6	8	9	10	8	7	7	7

7	8	9	10	9	8	7	6
8	8	9	11	10	9	9	7
8	9	10	11	11	10	10	8
9	9	10	10	11	10	9	8
7	8	9	10	10	9	9	7
7	7	8	9	8	8	7	6

Чтобы представить растровую сетку в компьютерном файле, ее необходимо сериализовать в единый (одномерный) список значений. Хотя существуют различные возможные схемы упорядочения, наиболее часто используемой является row-major , в которой сразу за ячейками в первой строке следуют ячейки во второй строке следующим образом:

6 7 10 9 8 6 7 8 6 8 9 10 8 7 7 7 7 8 9 10 9 8 7 6 8 8 9 11 10 9 9 7 . . .

Для восстановления исходной сетки требуется заголовок с общими параметрами сетки. По крайней мере, ему требуется количество строк в каждом столбце, чтобы он знал, где начинать каждую новую строку, и тип данных каждого значения (т. е. количество битов в каждом значении перед началом следующего значения). [24]

В то время как растровая модель тесно связана с концептуальной моделью поля, объекты также могут быть представлены в растре, по существу, путем преобразования объекта X в дискретное (логическое) поле присутствия/отсутствия X. В качестве альтернативы слой объектов (обычно полигонов) может быть преобразован в дискретное поле идентификаторов объектов. В этом случае некоторые форматы растровых файлов позволяют присоединять к растру векторную таблицу атрибутов путем сопоставления значений ID. [18] Растровые представления объектов часто являются временными, создаются и используются только как часть процедуры моделирования, а не в постоянном хранилище данных. [20] : 135-137

Чтобы быть полезным в ГИС, растровый файл должен иметь географическую привязку , чтобы соответствовать местоположению в реальном мире, поскольку необработанный растр может отображать местоположения только в терминах строк и столбцов. Обычно это делается с помощью набора параметров метаданных либо в заголовке файла (например, в формате GeoTIFF), либо в дополнительном файле (например, в файле привязки). По крайней мере, метаданные географической привязки должны включать местоположение хотя бы одной ячейки в выбранной системе координат и разрешение или размер ячейки , расстояние между каждой ячейкой. Линейное аффинное преобразованиеявляется наиболее распространенным типом пространственной привязки, допускающим вращение и прямоугольные ячейки. [18] : 171 Более сложные схемы пространственной привязки включают полиномиальные и сплайновые преобразования.

Наборы растровых данных могут быть очень большими, поэтому часто используются методы сжатия изображений . Алгоритмы сжатия идентифицируют пространственные шаблоны в данных, затем преобразуют

данные в параметризованные представления шаблонов, из которых могут быть восстановлены исходные данные. В большинстве ГИС-приложений алгоритмы сжатия без потерь (например, Lempel-Ziv) предпочтительнее алгоритмов сжатия с потерями (например, JPEG), поскольку требуются полные исходные данные, а не интерполяция. [10]

Расширения

Начиная с 1990-х годов, когда исходные модели данных и программное обеспечение ГИС совершенствовались, одним из основных направлений исследований в области моделирования данных была разработка расширений традиционных моделей для обработки более сложной географической информации.

Пространственно-временные модели [править]

Время всегда играло важную роль в аналитической географии, по крайней мере, начиная с региональной научной матрицы Брайана Берри (1964) и географии времени Торстена Хегерстранда (1970). [25] [13] На заре эры GIScience в начале 1990-х работа Гейла Ланграна открыла двери для исследования методов явного представления изменений во времени в данных ГИС; [26] это привело к появлению множества концептуальных моделей и моделей данных, появившихся в последующие десятилетия. [27] К 2010 году некоторые формы временных данных стали поддерживаться в готовом программном обеспечении ГИС.

Несколько общих моделей для представления времени в векторных и растровых данных ГИС включают: [28]

Модель моментальных снимков (также известная как слои с отметками времени), в которой весь набор данных привязан к конкретному допустимому времени. То есть это «моментальный снимок» мира того времени.

Объекты с отметками времени , в которых набор данных включает объекты, действительные в разное время, причем каждый объект отмечен временем, в течение которого он был действителен (т. е. в столбцах «дата начала» и «дата окончания» в таблице атрибутов). . Некоторое программное обеспечение ГИС, такое как ArcGIS Pro, изначально поддерживает эту модель с функциями, включая анимацию.

Границы с отметками времени , использующие топологическую модель векторных данных для разложения полигонов на граничные сегменты и отмечающие каждый сегмент временем, в течение которого он был действителен. Впервые этот метод был разработан Исторической ГИС Великобритании .

Факты с отметкой времени , в которых каждый отдельный элемент данных (включая значения атрибутов) может иметь свою собственную отметку времени, что позволяет атрибутам внутри одного объекта изменяться с течением времени или одному объекту (с постоянной идентичностью) иметь разные геометрические формы. в разное время. [29]

Время как измерение , которое рассматривает время как еще одно (3-е или 4-е) пространственное измерение и использует многомерные векторные или растровые структуры для создания геометрии, включающей время. Таким

образом Хэгерстранд визуализировал географию своего времени, и некоторые модели ГИС, основанные на ней, используют этот подход. Формат NetCDF поддерживает управление временными растровыми данными как измерением. [30]

Трехмерные модели [править]

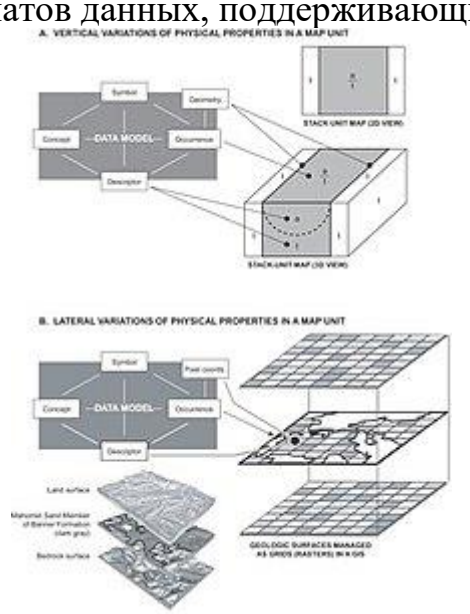
См. Также: 3D компьютерная графика .

Существует несколько подходов к представлению информации о трехмерной карте и к управлению ею в модели данных . Некоторые из них были разработаны специально для ГИС, в то время как другие были заимствованы из трехмерной компьютерной графики или автоматизированного черчения (САПР).

Поля высот (также известные как «пространственные поверхности 2 1/2») моделируют трехмерные явления с помощью одной функциональной поверхности, в которой высота является функцией двухмерного местоположения, что позволяет представлять ее с использованием полевых методов, таких как изолированные точки, горизонтали , растр (цифровая модель рельефа) и триангулированные нерегулярные сети .

Полигональная сетка (связанная с математическим многогранником) является логическим расширением модели векторных данных и, вероятно, является типом трехмерной модели, наиболее широко поддерживаемым в ГИС. Объемный объект сводится к его внешней поверхности, которая представлена набором многоугольников (часто треугольников), которые в совокупности полностью охватывают объем.

Воксельная модель является логическим продолжением растровой модели данных путем тесселяции трехмерного пространства в кубы, называемые вокселями (сочетание объема и пикселя , причем последний сам по себе является сумкой) . NetCDF — один из наиболее распространенных форматов данных, поддерживающих трехмерные ячейки. [30]



Подходы к представлению информации о трехмерной карте и к управлению ею в модели данных. [31]

Векторные карты структурных единиц изображают вертикальную последовательность геологических единиц до заданной глубины (здесь — основание блок-схемы). Этот подход к картированию характеризует вертикальные вариации физических свойств в каждой трехмерной единице карты. В этом примере аллювиальные отложения (пачка «a») перекрывают ледниковый тилл (пачка «t»), и стек-единица, помеченная «a/t», указывает на эту взаимосвязь, тогда как единица «t» указывает, что ледниковый тилл простирается вниз. на указанную глубину. Аналогично тому, как показано на рисунке 11, осуществляется управление вхождением элемента стека (обнажение элемента карты), геометрией (границами элемента карты) и дескрипторами (физическими свойствами геологических единиц, включенных в элемент стека). как для типичной двухмерной геологической карты.

Составные поверхности на основе растров отображают поверхность каждой погребенной геологической единицы и могут содержать данные о латеральных вариациях физических свойств. В этом примере из Soller и др. (1999), [32]верхняя поверхность каждой погребенной геологической единицы была представлена в растровом формате в виде файла ArcInfo Grid. Средняя сетка представляет собой самую верхнюю поверхность экономически важного водоносного горизонта, песка Магомета, который заполняет до- и межледниковую долину, высеченную в поверхности коренных пород. Каждой геологической единицей в растровом формате можно управлять в модели данных таким же образом, как показано для карты стека единиц. Песок Магомета непрерывен в этой области и представляет собой одно появление этой единицы в модели данных. Каждый растр или пиксель на поверхности Магомет Сэнд имеет набор координат карты, записанных в ГИС (в ячейке модели данных, помеченной как «координаты пикселей», которая является растровым следствием ячейки «геометрия» для вектора). данные карты). Каждый пиксель может иметь уникальный набор описательной информации,литология , электропроводность и др.).

Практическое занятие 2

Методы картографического анализа.

ПРИРОДА ПРОСТРАНСТВЕННЫХ ДАННЫХ

Одним из уникальных аспектов пространственных данных является то, что пространственный компонент основан на двух непрерывных измерениях, одном в горизонтальном направлении (на восток) и одном в вертикальном направлении (на север) (14). Другим аспектом является проблема пространственной зависимости, аналогичная временной зависимости. Близлежащие места, вероятно, будут обладать сходные атрибуты; или, другими словами, все связано со всем остальным, и близкие вещи связаны больше, чем отдаленные (16). Эти особенности пространственных данных создают потребность в специальных аналитических методах , и их следует учитывать каждый раз, когда предпринимается попытка реализации проекта, связанного с географией (т.е.

местоположением). Основной проблемой при использовании методов географического и пространственного анализа в эпидемиологии и исследованиях в области здравоохранения является распознавание пространственной структуры процесса, будь то группа событий в области здравоохранения или пространственная картина распространения болезни с течением времени. Связи между людьми, между животными и то , как такие взаимосвязи встроены в комплекс переменных окружающей среды, вполне могут определять или структурировать пространство. Конкретная пространственная структура включает затронутых людей и то, как они связаны в сообществах, а также динамику этих сообществ и их организацию в более крупные единицы. География болезни может дать ценные подсказки для понимания того, как культура, окружающая среда и поведение взаимодействуют со здоровьем и болезнью. Виды пространственные проблемы, которые могут беспокоить исследователей здравоохранения , включают пространственные закономерности заболеваемости и смертности; факторы, связанные с этими моделями; распространение болезней и этиология заболеваний; пространственное распределение, местоположение и регионализация ресурсов здравоохранения; доступ к ресурсам и их использование и факторы, связанные с распределением ресурсов; пространственные аспекты взаимодействия между болезнями и доступом к медицинской помощи; и оценка риска для токсикологии окружающей среды (7).

ПРОСТРАНСТВЕННЫЕ АНАЛИТИЧЕСКИЕ МЕТОДЫ Изучение пространственно связанных объектов или характеристик можно разделить на описание характеристик местоположения , которые различают области (исследовательские аналитические методы) и анализ пространственных взаимосвязей (объяснительные аналитические методы) (17). Некоторые распространенные пространственные методы, используемые в исследованиях в области здравоохранения, включают картирование заболеваний, методы кластеризации, диффузионные исследования, выявление факторов риска с помощью сопоставления карт и регрессионного анализа (7). В этой презентации методов пространственного анализа мы кратко рассмотрим некоторые важные соображения и обзоры, доступные для картирования заболеваний, ряд методов выявления кластеров заболеваний как для точечных, так и для площадных данных, методы "относительных пространств", аспекты исследований диффузии, методы интерполяции и сглаживания пространственных данных, а также некоторые исследования и методы для выявления пространственных факторов риска. В таблице 1 приведен список некоторых методов, обсуждаемых в этой статье, включая методы обнаружения кластеров, методы дисперсии и методы интерполяции; преимущества, недостатки и использование каждого из них; и некоторое компьютерное программное обеспечение (где применимо) для некоторых из методы, описанные в тексте.

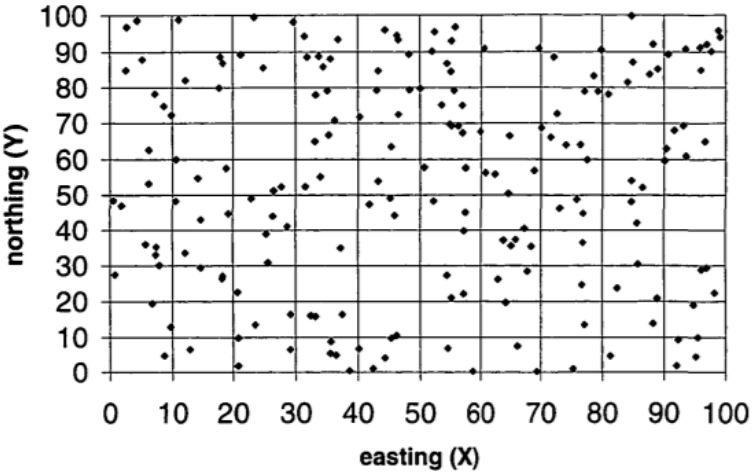
Картографирование заболеваний включает в себя картографирование точечных местоположений случаев заболевания, показателей заболеваемости по районам и стандартизированных показателей. Хотя показатели заболеваемости регулярно наносятся на карту, представление болезней на картах варьируется в

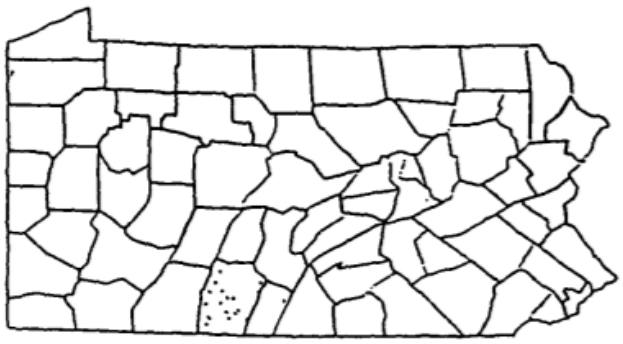
зависимости от целей исследователей. От того, как будет разработана карта, будет зависеть, какую информацию из нее можно извлечь. Источники информации и примеры составления карт заболеваний и представления событий болезни представлены в атласах Клиффа и др. (18), Смоллмана-Рейнора и др. (19) и Пикла и др. (20). Карта является хорошим средством связи, но может также вводит в заблуждение (15, 21). Глаз может обнаружить закономерности в зашумленных данных, отображаемых на картах, но карты плохо отражают сложные взаимосвязи между ответом и объясняющими переменными (21). Например, рисунок 1 представляет собой совокупную карту плотности точек.

Хотя картина распространения болезней очевидна визуально, для понимания сложной природы пространственной тенденции необходимы методы пространственного анализа. Обнаружение кластеров Кластер в эпидемиологии - это ряд событий в области здравоохранения, расположенных близко друг к другу в пространстве и/или времени. Хотя есть те, кто предупреждает о том, чтобы уделять много энергии обнаружению кластеров, большая часть литературы по пространственной эпидемиологии посвящена этой теме (23). Доступно несколько обзоров методов обнаружения кластеров (24-32). Кузик и Эдвардс (33) опишите три общих методологических подхода к обнаружению кластеризации: методы, основанные на количестве ячеек, на автокорреляционных смежностях ячеек с большим количеством и на расстоянии между событиями. В этом обзоре мы рассмотрим методы обнаружения кластеров в точечных данных и площадных данных. Точечные узоры. Кластеры обычно идентифицируются по данным точечного шаблона (34). Цель изучения точечных моделей состоит в том, чтобы определить, когда события систематически организованы или структурированы по сравнению с события распределяются случайным образом (35). Простейший анализ точечных шаблонов - это визуальный осмотр точечной карты, которая отображает географическое распределение событий (рис. 1). Для оценки субъективных впечатлений могут потребоваться статистические инструменты, такие как отношение дисперсии к среднему значению. Обычная нулевая гипотеза при пространственном анализе данных о болезнях заключается в том, что количество случаев заболевания в данной области пропорционально числу людей, подверженных риску (28). При отсутствии кластеризации точки либо случайным образом, либо равномерно распределены в пространстве (33). Анализ обычно предполагает однородный пуассоновский процесс над исследуемой областью, подразумевающий независимость пространственных событий. Анализ ближайших соседей использует расстояния между событиями для получения представления о силе кластеризации точечных данных (35). Тесты значимости, использующие расстояние до ближайших соседей, проверяют полную пространственную случайность. Этот тест может быть использован в качестве предварительной процедуры при анализе данных о событиях и описывает географическое распределение набора точек в соответствии с их расстоянием (36). Анализ ближайшего соседа был

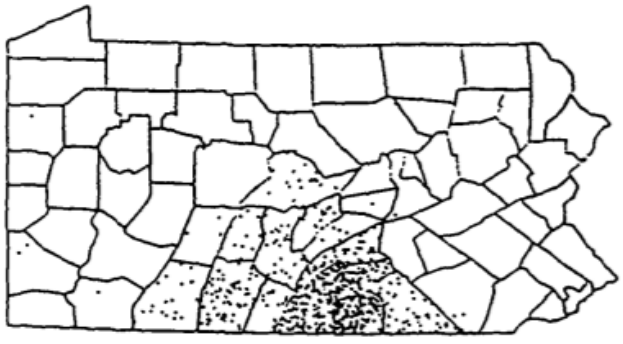
проведен однако он подвергается серьезной критике за то, что в нем не проводится различие между однородными и случайными закономерностями, а также за то, что при анализе областей разного размера с использованием одних и тех же данных будут получены разные результаты (36). Таким образом, масштаб территории играет решающую роль в выявлении кластеров. Может произойти более одного процесса ; например, когда пары точек имеют тенденцию соединяться вместе, в то время как другой процесс может присутствовать, а может и не присутствовать. В результирующей интерпретации может отсутствовать попарная кластеризация или кластеризация более высокого порядка. Этого можно было бы избежать путем изучения, в дополнение к ближайшему соседу, расстояний между первыми ближайшими точками и ближайшей точкой второго или более высокого порядка. Расчет ожидаемого расстояния до k-го ближайшего соседа был получен Томпсоном (37). Чтобы скорректировать неоднородное или непуассоновское распределение населения, подверженного риску, был использован ряд альтернатив методу ближайшего соседа . Бителл (38) оценил функцию относительного риска детской лейкемии в Камбрии, Англия, сравнив расстояния между первым и вторым ближайшими соседями для случаев и случайно выбранных элементов управления. Аналогичный подход был применен Гатреллом и Бейли (39), которые оценили функции k для случайно выбранных случаев и контрольных случаев детской лейкемии в Ланкашире, Англия. Они изучили графики разности этих двух функций в зависимости от расстояния. Значительная кластеризация была выявлена, когда пики превышали аналитический или смоделированный доверительный интервал. Глейзер (40) использовал два альтернативных метода, один из которых учитывал плотность населения алгебраически (41), в то время как второй создавал преобразованную карту, на которой население составляло равномерно распределенный (42) для изучения кластеризации Болезнь Ходжкина в районе залива Сан-Франциско. Общий подход к контролю за распределением населения, подверженного риску, был разработан Кузиком и Эдвардсом (33). Они использовали разновидность подхода k-го ближайшего соседа, при котором случаи рассматриваются с учетом их числа ближайших соседей, которые также являются случаями. Ожидаемое количество ближайших соседей по обращению основано на количестве и доле обращений в выборке "случай-контроль". В отличие от традиционного тест ближайшего соседа, тест Кузика-Эдвардса, учитывает относительное, а не фактическое расстояние между точками. Все эти методы контролируют смещение в сторону кластеризации, которое можно обнаружить с помощью базового теста ближайшего соседа. Квадратичный анализ (или метод подсчета ячеек) - это еще один метод проверки полной пространственной случайности, который решает проблему плотности точек (35, 36). Сетка, обычно квадратная или круговая, размещается на карте случайным образом или в фиксированной последовательности, и в каждой сетке подсчитывается количество событий. Например, на рисунке 2 показано случайное распределение 190 точек, распределенных между 100 ячейками

матрицы 10 х 10. Учитывая , что среднее количество точек на сетке (X) равно 1,9, случайность этого распределения можно проверить, предположив , что точки соответствуют базовому распределению Пуассона, используя критерий хи-квадрат, сравнивающий наблюдаемое и ожидаемое распределение частот количества ячеек. В этом примере есть 14 ячеек, не содержащих точек. Учитывая среднее значение 1,9 и предполагая, что точки следовали распределению Пуассона, ожидаемая вероятность ячейки/

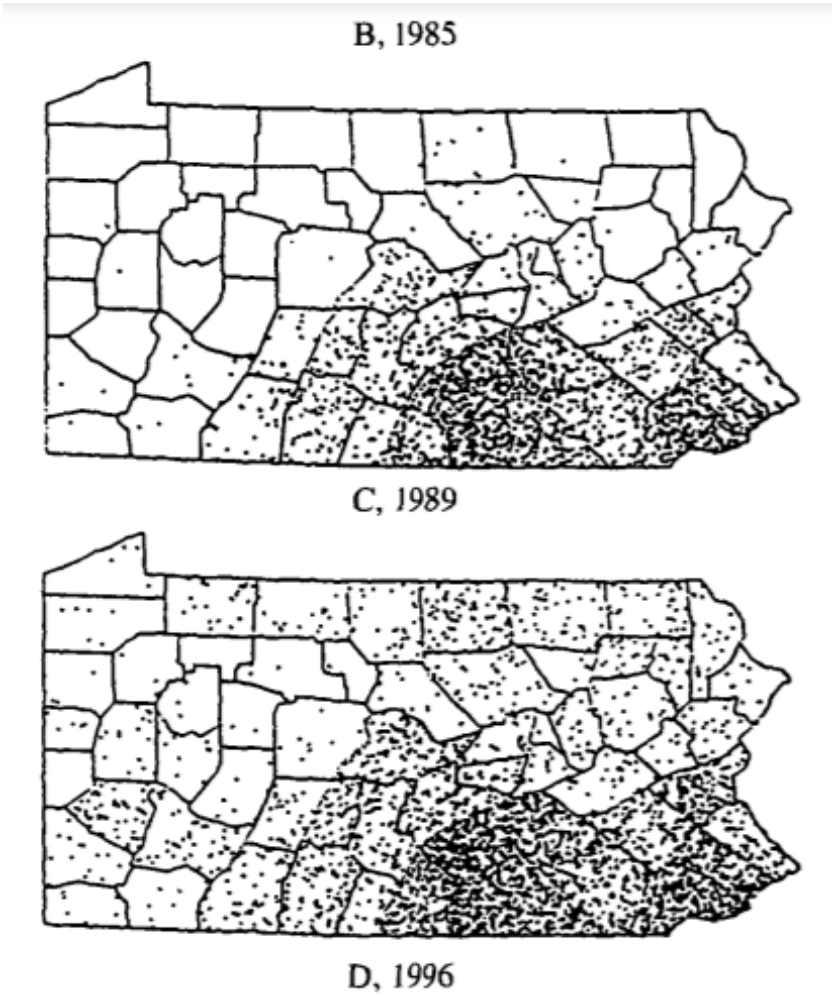




A, 1982



B, 1985



не содержит точек - 0,1496. Следовательно, ожидаемое количество ячеек, не содержащих точек, равно 14,96. Ожидаемое количество ячеек определяется для оставшихся частот точек и вычисляется значение хи-квадрат для определения отклонения от случайности, продемонстрированного этим подходом. В этом примере значение chisquare было рассчитано как 2,841 с помощью/? значение 0,585. Эпидемиологические применения этого подхода включают кластерные исследования стандартизированного коэффициента смертности от болезней сердца (43), системной красной волчанки (44) и смертности от астмы (45). Важный проблема заключается в выборе подходящей формы и размера сетки. Недостатком этого подхода является то, что он не учитывает относительное расположение ячеек. Опеншоу и др. (46) использовали анализ квадратичного типа в машине географического анализа (GAM) для выявления кластеризации случаев детской лейкемии и графического отображения результатов. Они оценили вероятность обнаружения определенного числа случаев лейкемии, учитывая население, подверженное риску в пределах квадрата. Целью этого аналитического инструмента было выявление кластеров, достойных дальнейшего расследования. Бесаг и Ньюэлл (25) представили альтернативу GAM для выявления кластеров редкого заболевания на большой

территории, разделенных на более мелкие единицы. Они исследовали каждый случай, чтобы определить , является ли соответствующий центроид его области центром кластера случаев заданного размера. Результатом было то, что каждый отдельный тест фокусировался только на локальной структуре шаблона без попытки компенсировать очевидный кластер после обнаружения. Эта техника была используется для выявления явных кластеров острых лимфобластных лейкозов, диагностированных у детей в возрасте до 15 лет в нескольких административных округах Англии. Аналогичный метод, тест пространственного сканирования, итеративно выполняет поиск кластеров случаев (26, 47). Этот метод сканирует большую область с круглым окном без предварительного указания местоположения или размера окна. Как только кластер идентифицирован, тест определяет значимость с учетом присущей ему проблемы множественного тестирования. Тест сканирования может идентифицировать как вторичные, так и первичные кластеры и упорядочивать их в соответствии с их коэффициентами вероятности, и был применен к раку молочной железы (48) и кластерные исследования лейкемии (47). Площадные данные. Часто пространственная информация недоступна для точечных данных, данные группируются или суммируются как данные по районам или регионам, или основное внимание уделяется выявлению кластеров в большем масштабе или области. Для выявления площадной кластеризации было разработано несколько тестов. В то время как они традиционно оценивают уровень сходства смежных областей, они различаются по типу данных, которые они анализ: непрерывный, дихотомический, категоричный. Для дихотомических, площадных данных степень кластеризации или дисперсии может быть определена количественно путем измерения количества общих и разнородных "соединений" между областями или методом подсчета соединений (49). Считается, что соединение происходит, если две области примыкают друг к другу. Разнородное соединение возникает, если две смежные области различны, например, черное по сравнению с белым или выше или ниже медианы. В кластеризованных областях относительно меньше разнородных соединений, в рассредоточенных - больше, а в случайных - промежуточное число соединений. Тест joins был использован для выявления схема пространственной кластеризации расположения эндодонтических отделений в Соединенных Штатах (50). Этот метод, однако, страдает от низкой мощности, предположительно из-за потери информации по сравнению с другими областными тестами (51).

Метод был разработан для оценки кластеризации категориальных или ранжированных площадных данных. Первоначальное приложение предназначалось для получения данных о смертности от рака в Японии (52). Области сравниваются с точки зрения соответствия (т.е. идентичны), считается, что смежные области имеют совпадающие значения, если они имеют одинаковую категорию или ранг и в противном случае не совпадают. Подсчитывается количество смежностей , вычисляется количество совпадающих смежностей и сравнивается с ожидаемым числом на основе частотного распределения каждой категории. Уровень значимости разницы

между наблюдаемые и ожидаемые совпадающие смежности рассчитываются для каждой категории и проверяются с помощью теста chisquare. Значение хи-квадрат также может быть рассчитано для всего распределения категорий. Потенциальным недостатком метода Оно является то, что все непохожие соединения обрабатываются одинаково, т.е. смежности с рангами 1 и 2 считаются такими же разными, как и смежности с рангами 1 и 5. Если эти относительные различия важны, метод Оно может оказаться неподходящим. Более подходящий тест для ранжирования данных в смежных областях оценивает то, что называемый непараметрической статистикой смежности рангов, D. Это показатель средней абсолютной разницы в рангах всех смежных областей. Применение этого теста было сделано для выявления ареальной кластеризации рака (53-56). Для сравнения непрерывных данных обычно используются два метода: с Гири и I Морана (57, 58). Эти методы схожи в том, что они сравнивают значения соседних областей для оценки уровня крупномасштабной кластеризации. Кластеризация может быть идентифицирована как результат неожиданно большого числа смежных области, имеющие либо относительно большие, либо небольшие значения. Джумарс и др. (59) рекомендовали применять оба метода, поскольку автокорреляция (кластеризация) может быть обнаружена одним из них, а другой - пропущена. Другие показывают, что тесты, основанные на / Морана, неизменно более эффективны, чем тесты, основанные на с Гири (51, 54, 55). Кроме того, Уолтер (60) обнаружил, что, хотя методы Geary с и Moran / могут реагировать на локализованные кластеры высокого риска, они обладают незначительной способностью обнаруживать сильно локализованные горячие точки. Хотя сообщалось лишь об ограниченном использовании теста Гири (61), Тест Морана часто применялся к различным эпидемиологическим проблемам для изучения локальных кластеров, включая рак (56, 62), показатели смертности от инсульта (63) и болезнь Лайма (64). Был проведен обзор мощности, связанной с несколькими из этих методов кластеризации смежности автор Уолтер (55). Он сообщил, что у Морана / последовательно была более высокая мощность, чем у Гири с, которая превышала мощность теста смежности ранга D. Вывод состоял в том, что мощность теста D была строго ограничена характером его непараметрических данных. Это ограничение является еще более заметно снижение мощности метода подсчета соединений, в котором использовались дихотомические данные (51). Хангерфорд (49) продемонстрировал использование анализа второго порядка с данными о серопревалентности анаплазмоза у крупного рогатого скота в Иллинойсе. Значение в каждой точке (в ее исследовании, центроиды округа) сравнивалось с ожидаемым значением, если все точки и значения были распределены случайным образом. Поскольку измеренное расстояние между точками с аналогичными значениями было меньше, чем ожидалось, для данных была предложена кластеризация. Второй порядок пространственный анализ определяет степень пространственной зависимости между переменными в зависимости от расстояния между точками или областями. Этот метод был использован для анализа пространственной

связи распространенности вируса псевдорабии свиней среди округов Иллинойса (65). Исследователи изучили кластеризацию показателей распространенности вируса псевдорабии свиней в графствах и сравнили эти показатели с географической кластеризацией стад свиней. Округа с высокими показателями распространенности вируса псевдорабии свиней сгруппированы больше, чем наблюдаемая кластеризация округов с большим количеством стад свиней. Пространственный автокорреляционный анализ - это дополнительный метод, используемый для выявления закономерностей заболевания. Он определяется как отношение между значениями одной переменной, которое относится к географическому расположению единиц измерения на карте (36). Хорошее введение в пространственную автокорреляцию дано Гудчайлдом (66). Пространственная автокорреляция является мерой взаимозависимости между значениями переменной в разных географических точках и может использоваться для определения степени пространственной кластеризации (64). Пространственные коррелограммы - это серии Морана / статистики, которые могут быть оценены с большей и большие расстояния от районов, чтобы определить , где пространственные эффекты максимизируются. Этот метод был использован при изучении факторов риска развития анаплазмоза (49). Пространственная коррелограмма - это функция, которая показывает корреляцию между точками выборки (для некоторой переменной). разделены расстоянием h. Корреляция обычно уменьшается с расстоянием, пока не достигнет или не приблизится к нулю. Он описывает автокорреляцию в переменной путем вычисления некоторого индекса ковариации для ряда расстояний с запаздыванием (67). В исследовании географических отношений между графствами лунг показатели смертности от рака, Кеннеди (68) смог продемонстрировать важное влияние соседних округов (местные эффекты) на смертность от рака легких у мужчин и большее региональное влияние на смертность от рака легких у женщин. Показатели смертности соседей первого-пятого порядка были взвешены по географическому соотношению с каждым округом. Остаточные графики указывали на то, что проблемы с автокорреляцией были преодолены с помощью этой авторегрессионной модели. Методы Монте-Карло являются вероятностными методами и могут использоваться для целей моделирования, когда пространственное

данные не являются независимыми, и единицы измерения площади могут быть разных размеров. Методы Монте-Карло - это инструменты, используемые для

решения различных задач путем построения некоторого случайного процесса (69). Иерархические кластеры областей "высокого риска"

могут быть созданы путем ранжирования показателей заболеваемости по территориальным единицам от высокого до низкого. Смежности между

высокопоставленными подразделениями подсчитываются и затем могут быть сопоставлены с результатами моделирования методом Монте-Карло, которое установило бы вероятности возникновения этих смежностей

(70). Относительные пробелы Термин "относительные пространства" относится к событиям или факторам это может быть связано с чем-то иным, чем простое или нетрансформированное географическое пространство. Эти пространства могут быть коммуникационным пространством, пригородным пространством, воздушным пассажирским пространством или любым другим пространством, которое представляется релевантным для анализа (71-73). Многомерное масштабирование - это метод, который использовался для решения проблем с традиционной "географической" проблемой или без нее. Он может быть использован для определения отношений между отдельными людьми в двух, трех или более пространствах. Классический пример многомерного масштабирования был выполнен Клиффом и др. (74, 75) с использованием данные о вспышках кори в Исландии и Соединенных Штатах. В дополнение к исследованиям инфекционных заболеваний, этот метод использовался для определения важных признаков, по которым люди привыкли судить о психиатрических учреждениях (76). Преобразования географического пространства (расстояния между городскими центрами с различной численностью населения) использовались для упрощения сложных иерархических процессов диффузии с использованием картографирования гравитационной модели (77). Процесс Маркова использовался для моделирования передачи синдрома приобретенного иммунодефицита (СПИДа) в Нью-Йорке, штат Нью-Йорк, столичный регион (78). А Марковский процесс - это когда индивиды случайным образом перемещаются между фиксированным набором состояний посредством "перехода". В исследовании Гулда о СПИДе вероятность заражения была связана с "пригородным пространством", или объемом пригородных перевозок, которые могли перевозить инфицированных людей и их вирусный багаж в разные районы и округа столичного региона Нью-Йорка. Из этой процедуры было показано, что структура пространства СПИДа может быть смоделирована в этом "относительном" пространстве и что технология помогла сформировать ход распространения вируса. Диффузионные исследования Распространение можно визуализировать, создав серию карт болезней или событий. Примером может служить доклад Уоллеса о распространении туберкулеза в Нью-Йорке (79) или рисунок 1. Однако для решения сложной пространственно-временной динамики процесса диффузии были разработаны методы моделирования диффузии. Для моделирования диффузионных процессов использовались два общих подхода : стохастический и детерминированный (80). Стохастическая модель имеет элементы, которые включают вероятность; детерминированные модели не допускают случайностей. Там существует три типа процессов распространения: чисто инфекционные, чисто иерархические (когда болезнь или медицинская практика перескакивают с одного места на другое на основе некоторой иерархии, такой как плотность населения) и смешанные иерархические. Важно понимать пространственную "подоплеку", по которой распространяется болезнь , чтобы эффективно моделировать процесс. Для получения руководства по моделированию пространственной диффузии и элементам, характеризующим явления диффузии, читатель обращается к Моррилл и др.

(80). Была выдвинута первичная теория пространственной диффузии автор: Торстен Хагерстранд в 1950-х (80) годах. Метод , который он разработал для моделирования пространственной диффузии, использовал монте- Метод Карло для моделирования процесса диффузии. Первая модель предполагала случайное распространение в пространстве. Вторая модель представила среднее информационное поле, сетку размером 5 x 5, обеспечивающую вероятность принятия при контакте с более ранним усыновителем. Третья модель включала барьеры сопротивления диффузии (80). Фоном или поверхностью, на которой происходит диффузия , может быть плотность популяции людей или животных в любом конкретном месте на карте сетки и может иметь наложил на него географические и другие потенциальные барьеры для распространения. Цель модели состоит в том, чтобы имитировать или имитировать закономерности диффузии. Гилг (34) изучил эпидемию вредителей домашней птицы 1970-1971 годов в Англии и выявил распространение этого инфекционного заболевания с востока на запад. Для каждой ячейки данных он рассчитал временные кривые, которые высветили "местную" эпидемию. Временные кривые графически демонстрировали различия в эпидемии в разных местах, объединяя как пространство, так и время на одной карте. Линейный анализ был использован для сравнения болезни "фронт" движения со случайным блужданием. Фактическое направление движения сравнивается со случайными движениями, чтобы обнаружить какой-либо паттерн (7). Векторы или линии, которые указывают величину и направление, могут быть использованы для описания потока заболеваний через область. После распространения бешенства лис на запад из Польши после второй мировой войны в Соединенном Королевстве была разработана модель пространственного распространения бешенства (72). Эта модель была детерминированной и применялась к зоне, свободной от бешенства. Он предсказал скорость распространения болезни и оценил масштаб вмешательства зона для предотвращения распространения. Акцент в модели делался на скорости распространения, а не обязательно на детерминантах или возможных сдерживающих факторах распространения болезни. Модель распространения бешенства лис была недавно предложена Джелчем и др. (81). Модель включала в себя сетку

Практическое занятие 3

Топологические структуры данных

Обзор модели данных

Топологическая модель данных позволяет работать с данными об узлах, ребрах и гранях в топологии. Например, географические данные переписи населения США представлены в виде узлов, цепочек и полигонов, и эти данные могут быть представлены с использованием модели данных пространственной топологии. Вы можете хранить информацию о топологических элементах и слоях геометрии в таблицах Oracle Spatial и представлениях метаданных. Затем вы можете выполнять определенные пространственные операции, ссылаясь на

топологические элементы, например, находить, какие цепи (например, улицы) имеют какое-либо пространственное взаимодействие с определенным полигональным объектом (например, с парком).

Вы можете использовать модель данных топологии PL/SQL и Java API для обновления топологии (например, для изменения данных о ребре, узле или грани). PL/SQL API для большинства операций редактирования — это пакет SDO_TOPO_MAP, который описан в главе 4 . Java API описан в Разделе 1.8.2 .

1.1.1 Использование топологии, построенной на основе топологических данных

Основные этапы работы с топологией, построенной на основе топологических данных, следующие:

Создайте топологию, используя процедуру SDO_TOPO.CREATE_TOPOLOGY . Это приводит к созданию таблиц <имя-топологии>_EDGE$, <имя-топологии>_NODE$, <имя-топологии>_FACE$ и <имя-топологии>_HISTORY$. (Эти таблицы описаны в Разделе 1.5.1 , Разделе 1.5.2 , Разделе 1.5.3 и Разделе 1.5.5 соответственно.)

Загрузите данные топологии в таблицы node, edge и face, созданные на шаге 1. Обычно это делается с помощью утилиты массовой загрузки, но это можно сделать и с помощью операторов SQL INSERT.

Создайте таблицу объектов для каждого типа геометрического слоя топологии в топологии. Например, в топологии городских данных могут быть отдельные таблицы объектов для земельных участков, улиц и дорожных знаков.

Свяжите таблицы объектов с топологией, используя процедуру SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER для каждой таблицы объектов. Это приводит к созданию таблицы <topology-name>_RELATION$. (Эта таблица описана в Разделе 1.5.4 .)

Инициализируйте метаданные топологии, используя процедуру SDO_TOPO.INITIALIZE_METADATA . (Эта процедура также создает пространственные индексы для таблиц <имя-топологии>_EDGE$, <имя-топологии>_NODE$ и <имя-топологии>_FACE$, а также дополнительные индексы сбалансированного дерева для <имя-топологии>_EDGE$. и таблицы <topology-name>_NODE$.)

Загрузите таблицы объектов с помощью конструктора SDO_TOPO_GEOMETRY. (Этот конструктор описан в Разделе 1.6.2 .)

Запросите данные топологии (например, используя один из топологических операторов, описанных в Разделе 1.8.1).

При необходимости отредактируйте данные топологии с помощью интерфейсов прикладного программирования (API) PL/SQL или Java.

1.1.2 Использование топологии, построенной на основе пространственных геометрий

Чтобы построить топологию из пространственных геометрий, вы должны сначала выполнить стандартные операции по подготовке данных для использования с Oracle Spatial, как описано в Oracle Spatial User's Guide and Reference :

Создайте пространственные таблицы.

Обновите пространственные метаданные (представление USER_SDO_GEOM_METADATA).

Загрузите данные в пространственные таблицы.

Подтвердите пространственные данные.

Создайте пространственные индексы.

Основные этапы работы с топологией, построенной на основе геометрии Oracle Spatial, следующие:

Создайте топологию, используя процедуру SDO_TOPO.CREATE_TOPOLOGY . Это приводит к созданию таблиц <имя-топологии>_EDGE$, <имя-топологии>_NODE$, <имя-топологии>_FACE$ и <имя-топологии>_HISTORY$. (Эти таблицы описаны в Разделе 1.5.1 , Разделе 1.5.2 , Разделе 1.5.3 и Разделе 1.5.5 соответственно.)

Создайте грань юниверса (F0, определено в Разделе 1.2).

Создайте таблицу объектов для каждого типа геометрического слоя топологии в топологии. Например, в топологии городских данных могут быть отдельные таблицы объектов для земельных участков, улиц и дорожных знаков.

Свяжите таблицы объектов с топологией, используя процедуру SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER для каждой таблицы объектов. Это приводит к созданию таблицы <topology-name>_RELATION$. (Эта таблица описана в Разделе 1.5.4 .)

Создайте объект TopoMap и загрузите всю топологию в кеш.

Загрузите таблицы объектов, вставив данные из пространственных таблиц и используя функцию SDO_TOPO_MAP.CREATE_FEATURE .

Инициализируйте метаданные топологии, используя процедуру SDO_TOPO.INITIALIZE_METADATA . (Эта процедура также создает пространственные индексы для таблиц <имя-топологии>_EDGE$, <имя-топологии>_NODE$ и <имя-топологии>_FACE$, а также дополнительные индексы сбалансированного дерева для <имя-топологии>_EDGE$. и таблицы <topology-name>_NODE$.)

Запросите данные топологии (используя один из операторов топологии, описанных в Разделе 1.8.1).

При необходимости отредактируйте данные топологии с помощью интерфейсов прикладного программирования (API) PL/SQL или Java.

Раздел 1.12.2 содержит пример PL/SQL, выполняющий эти основные шаги.

1,2 Основные понятия топологической модели данных

Топология — это раздел математики, изучающий объекты в пространстве. К топологическим отношениям относятся такие отношения, как содержит , внутри , покрывает , покрывает , касается и перекрывается с пересекающимися границами . Топологические отношения остаются постоянными, когда координатное пространство деформируется, например, при скручивании или растяжении. (Примеры отношений, которые не являются топологическими, включают длину , расстояние между и площадь .)

Основными элементами топологии являются ее узлы, ребра и грани.

Узел , представленный точкой , может быть изолирован или может использоваться для связывания ребер. Два или более ребра встречаются в

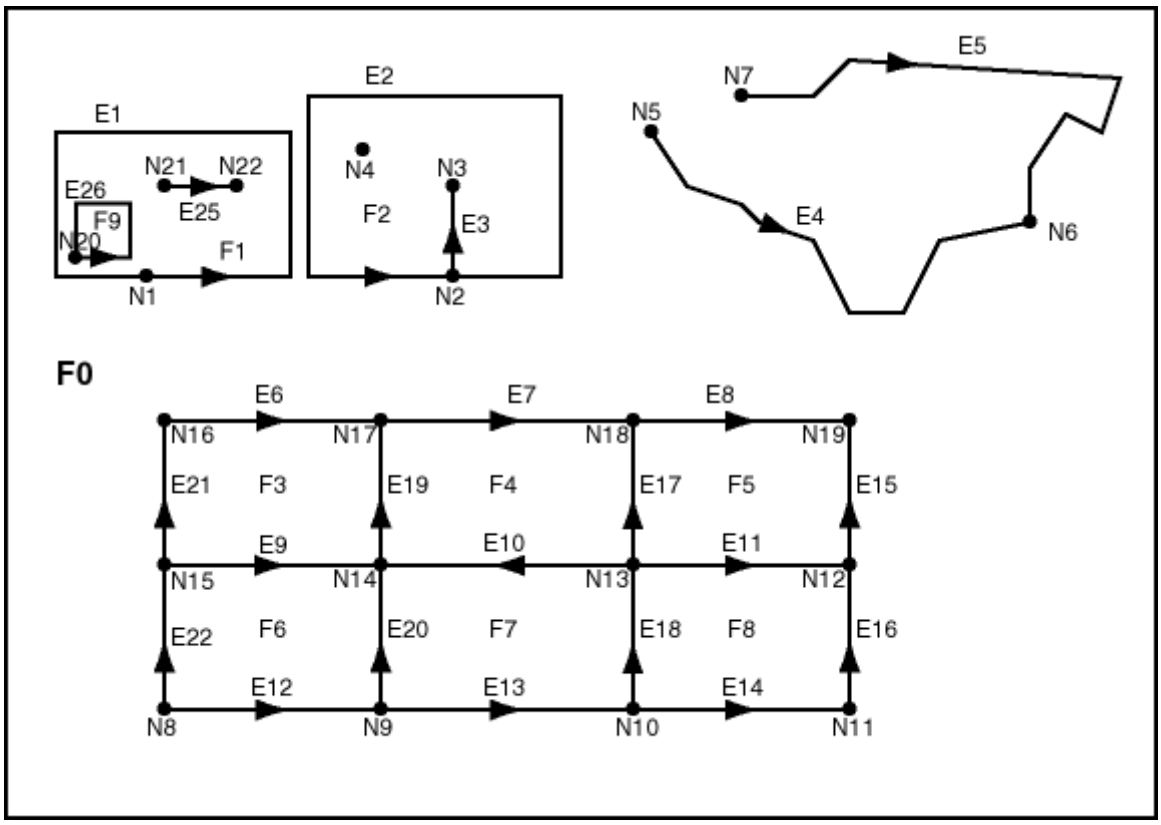
неизолированном узле. С узлом связана пара координат, которая описывает пространственное положение этого узла. Примеры географических объектов, которые могут быть представлены в виде узлов, включают начальные и конечные точки улиц, места, представляющие исторический интерес, и аэропорты (если масштаб карты достаточно велик).

Ребро ограничено двумя узлами: начальным (исходным) узлом и конечным (конечным) узлом . Ребро имеет связанный геометрический объект, обычно строку координат, описывающую пространственное представление ребра. Ребро может иметь несколько вершин, образующих цепочку линий. (Дуги окружности не поддерживаются для топологий.) Примеры географических объектов, которые могут быть представлены в виде ребер, включают сегменты улиц и рек.

Порядок координат задает направление ребра, а направление важно для определения топологических отношений. Положительное направление согласуется с ориентацией нижележащего ребра, а отрицательное направление меняет эту ориентацию на противоположную. Каждая ориентация ребра называется направленным ребром , и каждое направленное ребро является зеркальным отражением своего другого направленного ребра. Начальный узел положительно направленного ребра является конечным узлом отрицательно направленного ребра. Ребро также лежит между двумя гранями и имеет ссылки на обе из них. Каждое направленное ребро содержит ссылку на следующее ребро в непрерывном периметре грани с левой стороны.face , соответствующий многоугольнику, имеет ссылку на одно направленное ребро его внешней границы. Если присутствуют какие-либо островные узлы или островные ребра, грань также имеет ссылку на одно направленное ребро на границе каждого островка. Примеры географических объектов, которые могут быть представлены в виде лиц, включают парки, озера, округа и штаты.

На рис. 1-1 показана упрощенная топология, содержащая узлы, ребра и грани. Стрелки на каждом ребре указывают положительное направление ребра (или, точнее, ориентацию базовой строки линии или геометрии кривой для положительного направления ребра).

Рисунок 1-1 Упрощенная топология



Описание «Рис. 1-1 Упрощенная топология»

Примечания к рисунку 1-1 :

Элементы E (E1, E2 и т. д.) — это ребра, элементы F (F0, F1 и т. д.) — грани, а элементы N (N1, N2 и т. д.) — узлы.

F0 (нулевое лицо) создается для каждой топологии. Это грань вселенной, содержащая все остальное в топологии. Нет никакой геометрии, связанной с гранью вселенной. F0 имеет значение идентификатора лица -1 (отрицательное значение 1).

Для каждой точечной геометрии и для каждого начального и конечного узлов ребра создается узел. Например, грань F1 имеет только ребро (замкнутое ребро) E1, имеющее тот же узел, что и начальный и конечный узлы (N1). F1 также имеет ребро E2 с начальным узлом N21 и конечным узлом N22.

Изолированный узел (также называемыйостровной узел) — узел, изолированный в грани. Например, узел N4 является изолированным узлом на грани F2.

Изолированное ребро (также называемоеостровное ребро) — ребро, изолированное в грани. Например, ребро E25 является изолированным ребром грани F1.

А ребро петли — это ребро, имеющее тот же узел, что и его начальный и конечный узлы. Например, ребро E1 — это ребро петли, начинающееся и заканчивающееся в узле N1.

На ребре не может быть изолированного (островного) узла. Ребро можно разбить на два ребра, добавив узел на ребро. Например, если изначально между узлами N16 и N18 было одно ребро, добавление узла N17 привело к появлению двух ребер: E6 и E7.

Информация о топологических отношениях хранится в специальных информационных таблицах ребер, граней и узлов. Например, таблица информации о ребрах содержит следующую информацию о ребрах E9 и E10. (Обратите внимание на направление стрелок для каждого ребра.) Следующее и предыдущее ребра основаны на левой и правой гранях ребра.

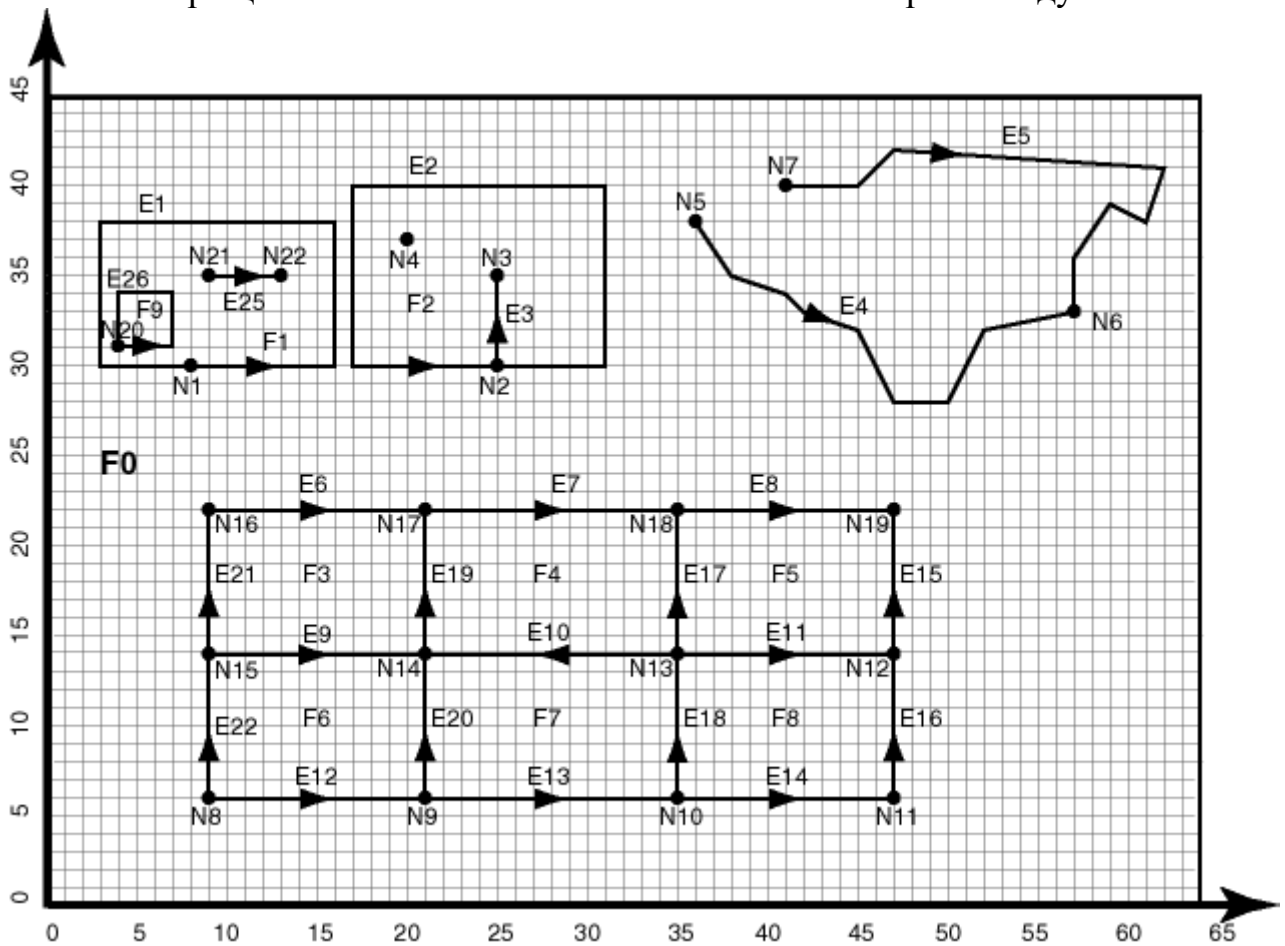
Для ребра E9 начальный узел — N15, а конечный узел — N14, следующее левое ребро — E19, а предыдущее левое ребро — E21, следующее правое ребро — E22, а предыдущее правое ребро — E20, левая грань — F3, а правая грань — F6.

Для ребра E10 начальный узел — N13, а конечный узел — N14, следующее левое ребро — E20, а предыдущее левое ребро — E18, следующее правое ребро — E17, а предыдущее правое ребро — E19, левая грань — F7 и правое лицо F4.

Дополнительные примеры данных, связанных с границами, включая иллюстрацию и пояснения, см . в Разделе 1.5.1 .

На рис. 1-2 показана та же топология, что и на рис. 1-1 , но добавлена сетка и номера единиц по осям x и y. Рисунок 1-2 полезен для понимания вывода некоторых примеров в главах 3 и 4 .

Рис. 1-2 Упрощенная топология с линиями сетки и номерами модулей



Описание «Рис. 1-2 Упрощенная топология с линиями сетки и номерами модулей»

1,3 Геометрия топологии и слои

Геометрия топологии (также называемая функцией) представляет собой пространственное представление объекта реального мира. Например, Main Street и Walden State Park могут быть именами топологических геометрий. Геометрия хранится в виде набора топологических элементов (узлов, ребер и граней), которые иногда также называют примитивами . Каждая геометрия топологии имеет уникальный идентификатор (назначаемый Spatial при импорте или загрузке записей), связанный с ней.

Уровень геометрии топологии состоит из геометрий топологии, обычно определенного типа геометрии топологии, хотя он может быть набором нескольких типов (см. Раздел 1.3.2 для получения информации о слоях коллекции). Например, Streets может быть слоем геометрии топологии, включающим геометрию топологии Main Street , а State Parks может быть слоем геометрии топологии, включающим геометрию топологии Walden State Park . Каждый слой геометрии топологии имеет уникальный идентификатор (назначаемый Spatial), связанный с ним. Данные для каждого слоя геометрии топологии хранятся втаблица характеристик . Например, таблица объектов с именем CITY_STREETS может содержать информацию обо всех геометриях топологии (отдельных дорогах или улицах) в геометрическом слое топологии Streets .

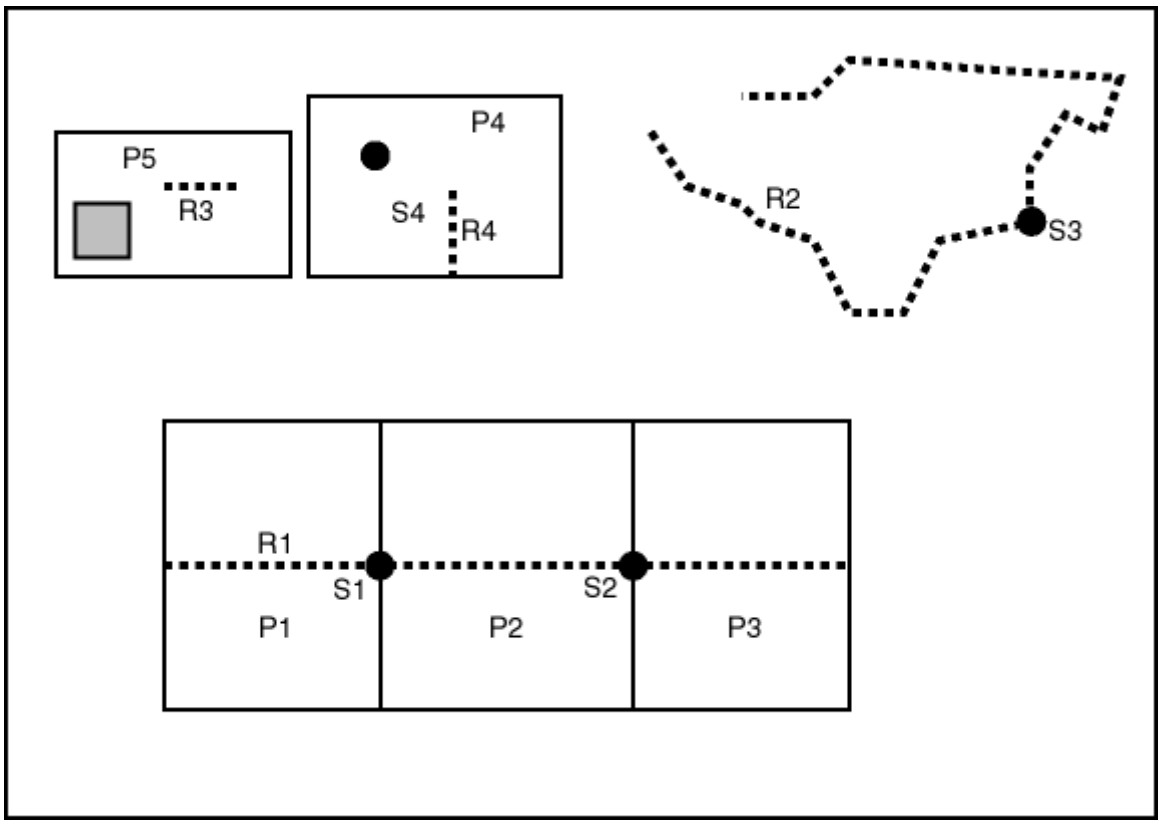
Каждая геометрия топологии (функция) определяется как объект типа SDO_TOPO_GEOMETRY (описанный в разделе 1.6.1), который идентифицирует тип геометрии топологии, идентификатор геометрии топологии, идентификатор слоя геометрии топологии и идентификатор топологии для топологии.

Метаданные топологии автоматически поддерживаются Spatial в представлениях USER_SDO_TOPO_METADATA и ALL_SDO_TOPO_METADATA, которые описаны в Разделе 1.7.2 . Представления USER_SDO_TOPO_INFO и ALL_SDO_TOPO_INFO (описанные в Разделе 1.7.1) содержат подмножество метаданных этой топологии.

1.3.1 Особенности

Часто в топологии меньше признаков, чем топологических элементов (узлов, ребер и граней). Например, дорожный объект может состоять из множества ребер, площадной объект, такой как парк, может состоять из множества граней, а некоторые узлы могут не быть связаны с точечными объектами. На рис. 1-3 показаны точечные, линейные и площадные объекты, связанные с топологией, показанной на рис. 1-1 в разделе 1.2 .

Рисунок 1-3 Элементы в топологии



Описание «Рис. 1-3 Элементы топологии»

На рис. 1-3 показаны следующие типы объектов топологии:

Точечные объекты (дорожные знаки), показанные темными кружками: S1, S2, S3, иS4

Линейные объекты (дороги или улицы), показанные пунктирными линиями: R1, R2, R3иR4

Объекты площади (земельные участки), показанные в виде прямоугольников: P1, P2, P3, P4, иP5

Земельный участок P5не включает в свою площадь заштрихованную территорию. (В частности, P5включает грань , F1но не грань F9. Эти грани показаны на рис. 1-1 в разделе 1.2 .)

Пример 1-12 в разделе 1.12.1 определяет эти функции.

1.3.2 Слои коллекции

Слой коллекции — это слой геометрии топологии, который может содержать топологические элементы различных типов геометрии топологии. Например, используя CITY_DATAтопологию из примеров в Разделе 1.12 , вы можете создать слой коллекции, содержащий определенные элементы земельных участков, городских улиц и дорожных знаков.

Чтобы создать слой-коллекцию, выполните практически те же действия, что и для создания других типов слоев. Создайте таблицу объектов для слоя, как показано в следующем примере:

```
CREATE TABLE collect_features ( -- Выбранные гетерогенные функции
  feature_name VARCHAR2(30) ПЕРВИЧНЫЙ КЛЮЧ,
  функция SDO_TOPO_GEOMETRY);
```

Свяжите таблицу объектов с топологией, указав COLLECTIONпараметр topo_geometry_layer_typeв вызове процедуры SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER , как в следующем примере:

ВЫПОЛНИТЬ SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER('CITY_DATA', COLLECTED_FEATURES', 'FEATURE', 'COLLECTION');

Чтобы загрузить таблицу объектов для слоя коллекции, вставьте необходимые строки, как показано в примере 1-1 .

Пример 1-1 Загрузка таблицы объектов для слоя коллекции

```
-- Возьмите R5 из слоя CITY_STREETS.
ВСТАВЬТЕ В ЗНАЧЕНИЯ collect_features(
 'C_R5',
 SDO_TOPO_GEOMETRY('CITY_DATA',
   2, -- tg_type = линия/многострочная
   4, -- tg_layer_id
   SDO_TOPO_OBJECT_ARRAY(
     SDO_TOPO_OBJECT(20, 2),
     SDO_TOPO_OBJECT(-9, 2)))
);

-- Возьмите S3 из слоя TRAFFIC_SIGNS.
ВСТАВЬТЕ В ЗНАЧЕНИЯ collect_features(
 'C_S3',
 SDO_TOPO_GEOMETRY('CITY_DATA',
   1, -- tg_type = точка/многоточка
   4, -- идентификатор топографического слоя
   SDO_TOPO_OBJECT_ARRAY(
     SDO_TOPO_OBJECT(6, 1)))
);

-- Возьмите P3 из слоя LAND_PARCELS.
ВСТАВЬТЕ В ЗНАЧЕНИЯ collect_features(
 'C_P3',
 SDO_TOPO_GEOMETRY('CITY_DATA',
   3, -- tg_type = (мульти)полигон
   4,
   SDO_TOPO_OBJECT_ARRAY(
     SDO_TOPO_OBJECT(5, 3),
     SDO_TOPO_OBJECT(8, 3)))
);

-- Создать коллекцию из полигона и точки.
ВСТАВЬТЕ В ЗНАЧЕНИЯ collect_features(
```

```
  'C1',
  SDO_TOPO_GEOMETRY('CITY_DATA',
    4, -- tg_type = коллекция
    4,
    SDO_TOPO_OBJECT_ARRAY(
      SDO_TOPO_OBJECT(5, 3),
      SDO_TOPO_OBJECT(6, 1)))
);

-- Создать коллекцию из полигона и линии.
ВСТАВЬТЕ В ЗНАЧЕНИЯ collect_features(
  'C2',
  SDO_TOPO_GEOMETRY('CITY_DATA',
    4, -- tg_type = коллекция
    4,
    SDO_TOPO_OBJECT_ARRAY(
      SDO_TOPO_OBJECT(8, 3),
      SDO_TOPO_OBJECT(10, 2)))
);

-- Создать коллекцию из линии и точки.
ВСТАВЬТЕ В ЗНАЧЕНИЯ collect_features(
  'C3',
  SDO_TOPO_GEOMETRY('CITY_DATA',
     4, -- tg_type = коллекция
     4,
     SDO_TOPO_OBJECT_ARRAY(
       SDO_TOPO_OBJECT(-5, 2),
       SDO_TOPO_OBJECT(10, 1)))
);
```

1,4 Иерархия слоев геометрии топологии

В некоторых топологиях геометрические слои топологии (элементарные слои) имеют одно или несколько отношений родитель-потомок в иерархии топологии . То есть слой на самом верхнем уровне состоит из объектов в его дочернем слое на следующем уровне в иерархии; дочерний слой может состоять из объектов своего дочернего слоя на следующем нижележащем слое; и так далее. Например, топология землепользования может иметь следующие геометрические слои топологии на разных уровнях иерархии:

Штаты на самом высоком уровне, который состоит из объектов из его дочернего слоя, округов.

Округа на следующем уровне ниже, который состоит из объектов дочернего слоя Tracts.

Участки на следующем уровне ниже, которые состоят из объектов из его дочернего слоя, групп блоков.

Группы блоков на следующем уровне ниже, которые состоят из объектов из дочернего слоя Land Parcels.

Земельные участки на самом низком уровне иерархии

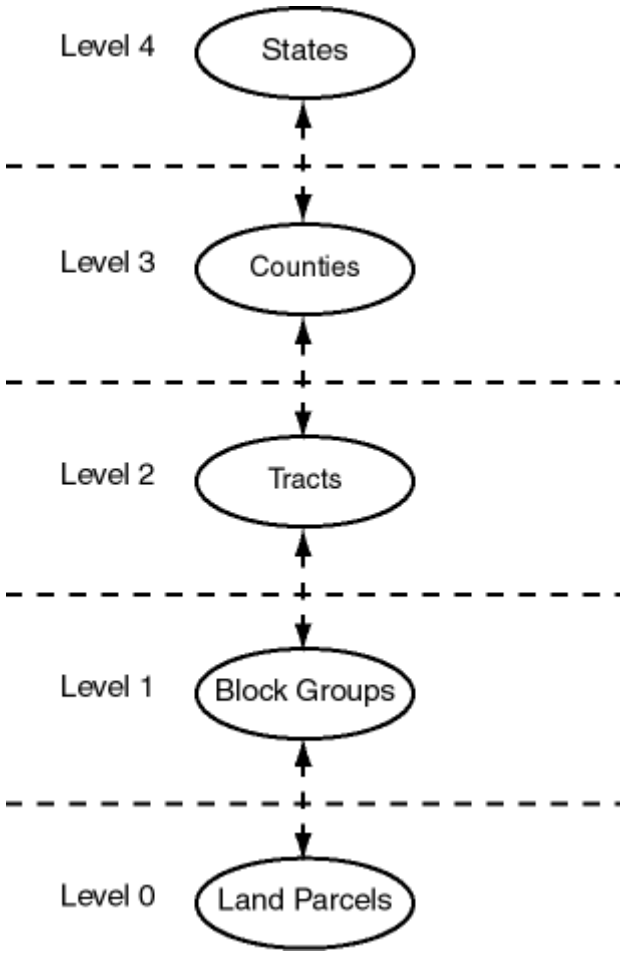
Если слои геометрии топологии в топологии имеют такое иерархическое отношение, намного эффективнее моделировать слои как иерархические, чем если бы вы задавали все слои геометрии топологии на одном уровне (то есть без иерархии). Например, более эффективно строить объекты SDO_TOPO_GEOMETRY для округов, указывая только участки в округе, чем задавая все земельные участки во всех группах кварталов во всех участках округа.

Самый низкий уровень (для геометрического слоя топологии, содержащего наименьшие виды объектов) в иерархии — это уровень 0, а последующие более высокие уровни нумеруются 1, 2 и т. д. Геометрические слои топологии на смежных уровнях иерархии имеют отношение родитель-потомок. Каждый слой геометрии топологии на более высоком уровне является родительским слоем для одного слоя на более низком уровне, который является его дочерним слоем . У родительского слоя может быть только один дочерний слой, а у дочернего слоя может быть один или несколько родительских слоев. В предыдущем примере слой Counties может иметь только один дочерний слой Tracts; однако слой Tracts может иметь родительские слои с именами Counties и Water Districts.Примечание:

Иерархия геометрического слоя топологии чем-то похожа на сетевую иерархию, которая описана в Разделе 5.5 ; однако есть существенные различия, и их не следует путать. Например, самый низкий уровень иерархии геометрического слоя топологии — 0, а самый низкий уровень сетевой иерархии — 1; и в иерархии геометрического слоя топологии у каждого родителя должен быть один дочерний элемент, и у каждого дочернего элемента может быть много родительских элементов, тогда как в сетевой иерархии у каждого родительского элемента может быть много дочерних элементов, и у каждого дочернего элемента должен быть один родительский элемент.

На рис. 1-4 показан предыдущий пример иерархии геометрического слоя топологии. Каждый уровень иерархии показывает номер уровня и слой геометрии топологии на этом уровне.

Рис. 1-4 Иерархия геометрического слоя топологии



Чтобы смоделировать слои геометрии топологии как иерархические, укажите дочерний слой в child_layer_idпараметре при вызове процедуры SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER для добавления в топологию родительского слоя геометрии топологии. Сначала добавьте слой геометрии топологии самого низкого уровня (уровень 0); затем добавьте слой уровня 1, указав слой уровня 0 в качестве его дочернего элемента; затем добавьте слой уровня 2, указав слой уровня 1 в качестве его дочернего элемента; и так далее. Пример 1-2 показывает добавление пяти слоев геометрии топологии, чтобы установить 5-уровневую иерархию.

Пример 1-2 Моделирование иерархии геометрического слоя топологии

```
-- Создайте топологию. (Нулевой SRID в этом примере.)
ВЫПОЛНИТЬ SDO_TOPO.CREATE_TOPOLOGY('LAND_USE_HIER', 0.00005);

-- Создание таблиц признаков.
CREATE TABLE land_parcels ( -- Земельные участки (выбранные грани)
  feature_name VARCHAR2(30) ПЕРВИЧНЫЙ КЛЮЧ,
  функция SDO_TOPO_GEOMETRY);

СОЗДАТЬ ТАБЛИЦУ block_groups (
  feature_name VARCHAR2(30) ПЕРВИЧНЫЙ КЛЮЧ,
```

```
  функция SDO_TOPO_GEOMETRY);

CREATE TABLE тракты (
  feature_name VARCHAR2(30) ПЕРВИЧНЫЙ КЛЮЧ,
  функция SDO_TOPO_GEOMETRY);

СОЗДАТЬ ТАБЛИЦУ округов (
  feature_name VARCHAR2(30) ПЕРВИЧНЫЙ КЛЮЧ,
  функция SDO_TOPO_GEOMETRY);

СОЗДАТЬ ТАБЛИЦУ состояний (
  feature_name VARCHAR2(30) ПЕРВИЧНЫЙ КЛЮЧ,
  функция SDO_TOPO_GEOMETRY);

-- (Другие шаги, не показанные здесь, такие как заполнение таблиц функций
-- и инициализация метаданных.)
. . .
-- Связать таблицы объектов с топологией; включать информацию об иерархии.

ЗАЯВИТЬ
  land_parcels_id ЧИСЛО;
  block_groups_id ЧИСЛО;
  tracts_id ЧИСЛО;
  counties_id ЧИСЛО;
НАЧИНАТЬ
SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER('LAND_USE_HIER',
'LAND_PARCELS',
  'ФУНКЦИЯ','ПОЛИГОН');
ВЫБЕРИТЕ tg_layer_id INTO land_parcels_id ИЗ user_sdo_topo_info
  ГДЕ топология = 'LAND_USE_HIER' AND table_name =
'LAND_PARCELS';
SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER('LAND_USE_HIER',
'BLOCK_GROUPS',
  'ФУНКЦИЯ','ПОЛИГОН', NULL, land_parcels_id);
ВЫБЕРИТЕ tg_layer_id В block_groups_id ИЗ user_sdo_topo_info
  ГДЕ топология = 'LAND_USE_HIER' AND table_name =
'BLOCK_GROUPS';
SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER('ЗЕМЛЯ_USE_HIER',
'ТРАКТЫ',
  'ФУНКЦИЯ','ПОЛИГОН', NULL, block_groups_id);
ВЫБЕРИТЕ tg_layer_id В tracts_id ИЗ user_sdo_topo_info
  ГДЕ топология = 'LAND_USE_HIER' AND table_name = 'ТРАКТЫ';
```

```
SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER('ЗЕМЛЯ_USE_HIER',
'ГРАНЫ',
  'ФУНКЦИЯ','ПОЛИГОН', NULL, tracts_id);
ВЫБЕРИТЕ tg_layer_id INTO counties_id ИЗ user_sdo_topo_info
  ГДЕ топология = 'LAND_USE_HIER' AND table_name = 'COUNTIES';
SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER('ЗЕМЛЯ_USE_HIER',
'СОСТОЯНИЯ',
  'ФУНКЦИЯ','ПОЛИГОН', NULL, counties_id);
КОНЕЦ;/
```

На каждом уровне выше уровня 0 каждый слой может содержать объекты, построенные из объектов следующего более низкого уровня (как это сделано в примере 1-2), объекты, построенные из топологических элементов (граней, узлов, ребер) или их комбинации. Например, слой участков может содержать участки, построенные из групп блоков, или участки, построенные из граней, или и то, и другое. Однако каждый объект внутри слоя должен быть построен либо из объектов следующего более низкого уровня, либо из топологических элементов. Например, конкретный участок может состоять из групп блоков или граней, но не может состоять из комбинации групп блоков и граней.

Чтобы вставить или обновить объекты геометрии топологии в таблицах объектов для уровней в иерархии, используйте соответствующие формы конструктора SDO_TOPO_GEOMETRY. Таблицы объектов описаны в разделе 1.3 , а конструкторы SDO_TOPO_GEOMETRY описаны в разделе 1.6.2 .

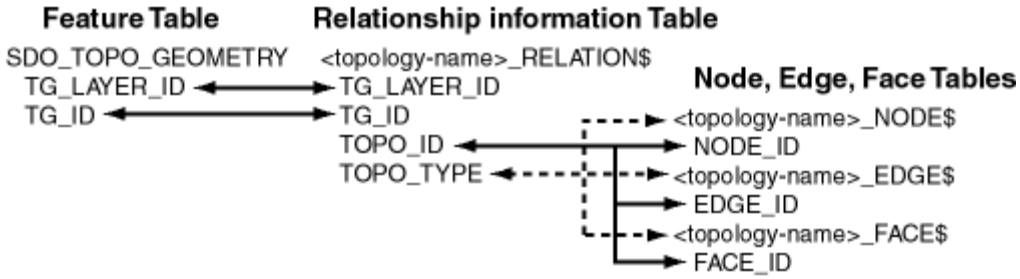
1.5 Таблицы топологической модели данных

Чтобы использовать возможности пространственной топологии, вы должны сначала вставить данные в специальные таблицы ребер, узлов и граней, которые создаются Spatial при создании топологии. Таблицы ребер, узлов и граней описаны в Разделе 1.5.1 , Разделе 1.5.2 и Разделе 1.5.3 соответственно.

Spatial автоматически поддерживает таблицу информации об отношениях (<topology-name>_RELATION$) для каждой топологии, которая создается при первом связывании таблицы объектов с топологией (то есть при первом вызове процедуры SDO_TOPO.ADD_TOPO_GEOMETRY_LAYER , которая определяет топологию). Информационная таблица отношений описана в Разделе 1.5.4 .

На рис. 1-5 показана роль таблицы информации об отношениях в соединении информации в таблице признаков с информацией в связанных с ней узлах, ребрах или таблицах граней.

Рисунок 1-5 Сопоставление между таблицами признаков и таблицами топологии



Как показано на рис. 1-5 , сопоставление между таблицами объектов и таблицами узлов топологии, ребер и граней происходит через таблицу <topology-name>_RELATION$. Особенно:

Каждая таблица объектов включает столбец типа SDO_TOPO_GEOMETRY. Этот тип включает атрибут TG_LAYER_ID (уникальный идентификатор, назначаемый Oracle Spatial при создании слоя), а также атрибут TG_ID (уникальный идентификатор, назначаемый каждому объекту в слое). Значения в этих двух столбцах имеют соответствующие значения в столбцах TG_LAYER_ID и TG_ID в таблице <topology-name>_RELATION$.

У каждого объекта есть одна или несколько строк в таблице <topology-name>_RELATION$.

Учитывая значения TG_LAYER_ID и TG_ID для объекта, набор узлов, граней и ребер, связанных с объектом, можно определить путем сопоставления значения TOPO_ID (идентификатор узла, ребра или лица) в <topology-name>_RELATION$ с соответствующим значением идентификатора в таблице <имя-топологии>_NODE$, <имя-топологии>_EDGE$ или <имя-топологии>_FACE$.

Следующие соображения относятся к именам схем, таблиц и столбцов, которые хранятся в любых представлениях метаданных Oracle Spatial. Например, эти соображения относятся к именам таблиц ребер, узлов, граней, взаимосвязей и исторических данных, а также к именам любых столбцов в этих таблицах и схемам для этих таблиц, которые хранятся в представлениях метаданных топологии, описанных в Разделе 1.7. .

Имя должно содержать только буквы, цифры и символы подчеркивания. Например, имя не может содержать пробел ( ), апостроф ('), кавычку (") или запятую (,).

Все буквы в именах преобразуются в верхний регистр перед сохранением имен в представлениях метаданных или перед доступом к таблицам. Это преобразование также применяется к любому имени схе

Практическое занятие 4

Пространственные данные и знания.

Многие области географических исследований носят скорее наблюдательный, чем экспериментальный характер, поскольку пространственный масштаб зачастую слишком велик, а географические проблемы слишком сложны для экспериментирования. Исследователи

приобретают новые знания путем поиска закономерностей, формулирования теорий и проверки гипотез наблюдениями. Благодаря постоянным усилиям научных проектов, государственных учреждений и частного сектора были и продолжают собираться обширные географические данные. Теперь мы можем получать гораздо более разнообразные, динамичные и подробные данные, чем когда-либо прежде, с помощью современных методов сбора данных, таких как глобальные системы позиционирования (GPS), дистанционное зондирование с высоким разрешением, услуги и опросы с учетом местоположения, а также добровольное участие в Интернет. географическая информация (Гудчайлд , 2007 г.). Вообще говоря, география и связанные с ней пространственные науки перешли из эпохи бедных данных в эпоху богатых данных (Miller & Han, 2009). Доступность обширных пространственных и пространственно-временных данных с высоким разрешением предоставляет возможности для получения новых знаний и лучшего понимания сложных географических явлений, таких как взаимодействие человека и окружающей среды и социально-экономическая динамика, а также для решения неотложных реальных проблем, таких как глобальный климат. изменение и распространение пандемического гриппа.

Однако традиционные методы пространственного анализа были разработаны в эпоху, когда данных было относительно мало, а вычислительная мощность не была такой мощной, как сегодня (Miller & Han, 2009) .). Столкнувшись со все более доступными массивными данными и сложными вопросами анализа, на которые они потенциально могут ответить, традиционные методы анализа часто имеют одно или несколько из следующих трех ограничений. Во-первых, большинство существующих методов ориентированы на ограниченную перспективу (например, одномерная пространственная автокорреляция) или на определенный тип модели отношений (например, линейная регрессия). Если выбранная точка зрения или предполагаемая модель не подходит для анализируемого явления, анализ может в лучшем случае показать, что данные не показывают интересных взаимосвязей, но не может предложить никаких лучших альтернатив. Во-вторых, многие традиционные методы не могут обрабатывать очень большие объемы данных. В-третьих, недавно появившиеся типы данных (такие как траектории движущихся объектов, географическая информация, встроенная в веб-страницы,

Существует острая потребность в эффективных и действенных методах извлечения неизвестной и неожиданной информации из наборов данных беспрецедентно большого размера (например, миллионы наблюдений), высокой размерности (например, сотни переменных) и сложности (например, разнородные источники данных, пространственные – временная динамика, многомерные связи, явные и неявные пространственные отношения и взаимодействия). Для решения этих проблем *интеллектуальный анализ пространственных данных и обнаружение географических знаний* стали активной областью исследований, сосредоточенной на развитии теории, методологии и практики извлечения полезной информации и знаний из массивных и сложных пространственных баз данных

Интеллектуальный анализ пространственных данных имеет глубокие корни как в традиционных областях пространственного анализа (таких как пространственная статистика, аналитическая картография, исследовательский анализ данных), так и в различных областях интеллектуального анализа данных в статистике и информатике (таких как кластеризация, классификация, анализ ассоциативных правил, визуализация информации и т. д.). визуальная аналитика). Его целью является интеграция и дальнейшее развитие методов анализа больших и сложных пространственных данных в различных областях. Неудивительно, что исследовательские усилия по интеллектуальному анализу пространственных данных часто относятся к разным направлениям, таким как пространственная статистика, геовычисления, геовизуализация и интеллектуальный анализ пространственных данных, в зависимости от типа методов, на которых сосредоточено исследование.

Интеллектуальный анализ данных и обнаружение знаний — это итеративный процесс, включающий несколько этапов, включая выбор данных, очистку, предварительную обработку и преобразование; включение предшествующих знаний; анализ с помощью вычислительных алгоритмов и/или визуальных подходов, интерпретация и оценка результатов; формулирование или модификация гипотез и теорий; приспособление к данным и методу анализа; повторная оценка результата; и т.д. Интеллектуальный анализ данных и обнаружение знаний носят исследовательский характер и являются более индуктивными, чем традиционные статистические методы. Он естественным образом вписывается в начальную стадию процесса дедуктивного открытия, когда исследователи разрабатывают и модифицируют теории на основе информации, полученной из данных наблюдений.

В литературе обнаружение знаний относится к вышеупомянутому многоэтапному процессу, в то время как интеллектуальный анализ данных в узком смысле определяется как применение вычислительных, статистических или визуальных методов. Однако на практике применение любого метода интеллектуального анализа данных должно осуществляться в соответствии с описанным выше процессом, чтобы обеспечить значимые и полезные результаты. В этой статье термины «интеллектуальный анализ пространственных данных» и «обнаружение географических знаний» используются взаимозаменяемо, и оба относятся к общему процессу обнаружения знаний.

2 . Общие задачи интеллектуального анализа пространственных данных

Интеллектуальный анализ пространственных данных — это растущая область исследований, которая все еще находится на очень ранней стадии. За последнее десятилетие благодаря широкому применению технологии GPS, совместному использованию и картографированию пространственных данных через Интернет, дистанционному зондированию с высоким разрешением и услугам на основе определения местоположения все больше и больше областей исследований создали или получили доступ к высококачественным географическим данным. данные для включения пространственной информации и анализа в различные исследования, такие как социальный анализ и бизнес-

приложения . Помимо области исследований, частные предприятия и широкая общественность также проявляют огромный интерес как к предоставлению географических данных, так и к использованию обширных ресурсов данных для различных прикладных нужд. Таким образом, вполне ожидаемо, что в ближайшие годы будет разрабатываться все больше и больше новых способов использования пространственных данных и новых подходов к интеллектуальному анализу пространственных данных. Хотя в этом разделе мы попытаемся представить обзор распространенных методов интеллектуального анализа пространственных данных, читатели должны знать, что интеллектуальный анализ пространственных данных — это новая и захватывающая область, границы и возможности которой еще предстоит определить.

Интеллектуальный анализ пространственных данных включает в себя различные задачи, и для каждой задачи часто доступен ряд различных методов, будь то вычислительные, статистические, визуальные или их комбинация. Здесь мы лишь кратко представляем выбранный набор задач и связанных с ними методов, включая классификацию (классификация с учителем), анализ правил ассоциации, кластеризацию (классификация без учителя) и многомерную геовизуализацию.

2.1 . Пространственная классификация и предсказание

Классификация заключается в группировании элементов данных в классы (категории) в соответствии с их свойствами (значениями атрибутов). Классификацию также называют контролируемой классификацией, в отличие от неконтролируемой классификации (кластеризации). Для «контролируемой» классификации требуется обучающий набор данных для обучения (или настройки) модели классификации, проверочный набор данных для проверки (или оптимизации) конфигурации и тестовый набор данных для оценки производительности обученной модели. Методы классификации включают, например, деревья решений, искусственные нейронные сети (ANN), оценку максимального правдоподобия (MLE), линейную дискриминантную функцию (LDF), методы опорных векторов (SVM), методы ближайших соседей и рассуждения на основе прецедентов (CBR).

Методы пространственной классификации расширяют методы классификации общего назначения, чтобы учитывать не только атрибуты объекта, который необходимо классифицировать, но также атрибуты соседних объектов и их пространственные отношения (Ester et al., 1997 , Koperski et al., 1998). Визуальный подход к пространственной классификации был представлен в (Андриенко и Андриенко, 1999), где дерево решений, полученное с помощью традиционного алгоритма C4.5 (Куинлан, 1993), сочетается с визуализацией карты для выявления пространственных закономерностей правил классификации. Индукция дерева решений также использовалась для анализа и прогнозирования поведения при пространственном выборе (Thill & Wheelerm, 2000).). Искусственные нейронные сети (ИНС) использовались для решения широкого круга задач пространственного анализа . Дистанционное зондирование является одной из основных областей, в которых обычно

используются методы классификации для классификации пикселей изображения по маркированным категориям .

Модели пространственной регрессии или прогнозирования составляют особую группу регрессионного анализа, которая учитывает независимую и/или зависимую переменную ближайших соседей при прогнозировании зависимой переменной в определенном месте, например пространственные авторегрессионные модели (SAR) (Anselin et al., 2006). , Cressie, 1983 , Pace et al., 1998). Однако методы пространственной регрессии, такие как SAR, часто включают манипуляции с матрицей пространственных весов размером n на n , что требует значительных вычислительных ресурсов, если n велико. Поэтому более поздние исследования были направлены на разработку подходов к поиску приблизительных решений для SAR, чтобы он мог обрабатывать очень большие наборы данных (Griffith, 2004)., Kazar et al., 2004, Smirnov and Anselin, 2001).

2.2 . Анализ правил пространственной ассоциации

Анализ ассоциативных правил изначально предназначался для обнаружения закономерностей между элементами в больших базах данных транзакций (Agraval, Imielinski, & Swami, 1993). Пусть $I = \{ i1, i2, \ldots, im \}$ будет набором предметов (т. е. предметов, купленных в транзакциях, таких как компьютер, молоко, велосипед и т. д.) . Пусть D — набор транзакций, где каждая транзакция T — это набор таких элементов, чтоT⊆I. Пусть X — набор элементов, и говорят, что транзакция T содержит X тогда и только тогда, когдаИкс⊆T. Правило ассоциации имеет вид:Икс⇒А ТАКЖЕ, кудаИкс⊂I,А ТАКЖЕ⊂Iа такжеИкс∩А ТАКЖЕ=∅. ПравилоИкс⇒А ТАКЖЕвыполняется в наборе транзакций D с *достоверностью* c , если c % всех транзакций в D , которые содержат X , также содержат Y . ПравилоИкс⇒А ТАКЖЕимеет *поддержку* s в наборе транзакций D , если s % транзакций в D содержатИкс∪А ТАКЖЕ. Уверенность обозначает силу, а поддержка указывает на частоту применения правила. Часто желательно обратить внимание на те правила, которые имеют достаточно большую поддержку (Agraval et al., 1993).

Подобно добыче правил ассоциации в транзакционных или реляционных базах данных, правила пространственной ассоциации могут быть получены в пространственных базах данных с учетом пространственных свойств и предикатов (Appice et al., 2003 , Han and Kamber, 2001 , Koperski and Han, 1995 , Mennis and Лю, 2005). Правило пространственной ассоциации выражается в видеА⇒Б[s %, c %], где А и В — наборы пространственных или непространственных предикатов, s % — поддержка правила, а c % — достоверность правила.

Очевидно, что многие возможные пространственные предикаты (например, близко_к, далеко_далеко, пересекаются, перекрываются и т. д.) могут использоваться в правилах пространственной ассоциации. Учет различных пространственных предикатов при выводе правил ассоциации из больших наборов пространственных данных требует значительных вычислительных ресурсов. Другая потенциальная проблема с анализом правил пространственной ассоциации заключается в том, что может быть сгенерировано большое

количество правил, и многие из них очевидны или общеизвестны. Знание предметной области необходимо, чтобы отфильтровать тривиальные правила и сосредоточиться только на новых и интересных находках.

Анализ моделей пространственного совместного расположения духовно похож на анализ правил ассоциации, но технически сильно отличается от него (Shekhar & Huang, 2001). При заданном наборе данных о пространственных объектах и их местоположениях шаблон совместного размещения представляет собой подмножества объектов, часто расположенных вместе, например, определенный вид птиц имеет тенденцию обитать рядом с определенным типом деревьев. Конечно, местоположение — это не транзакция, и два объекта редко существуют в одном и том же месте. Следовательно, указанное пользователем соседство необходимо в качестве контейнера для проверки того, какие объекты находятся в одном и том же соседстве. Были предложены меры и алгоритмы для анализа моделей пространственного совместного размещения (Huang et al., 2006 , Lu and Thill, 2008 , Shekhar and Huang, 2001) .).

2.3 . Пространственная кластеризация, регионализация и анализ точечных шаблонов

Кластерный анализ широко используется для анализа данных, который организует набор элементов данных в группы (или кластеры) таким образом, чтобы элементы в одной группе были похожи друг на друга и отличались от элементов в других группах (Gordon, 1996 , Jain and Dubes, 1988 , Джейн и др., 1999). В различных областях исследований, таких как статистика, распознавание образов, интеллектуальный анализ данных, машинное обучение и пространственный анализ, было разработано множество различных методов кластеризации.

Методы кластеризации можно разделить на две группы: кластеризация с разделением и иерархическая кластеризация. Методы сегментации кластеров, такие как K-средние и самоорганизующаяся карта (SOM) (Kohonen, 2001), делят набор элементов данных на несколько непересекающихся кластеров. Элемент данных относится к «ближайшему» кластеру на основе меры близости или несходства. Иерархическая кластеризация, с другой стороны, организует элементы данных в иерархию с последовательностью вложенных разделов или групп (Jain & Dubes, 1988). Обычно используемые иерархические методы кластеризации включают метод Уорда (Ward, 1963), кластеризацию с одной связью, кластеризацию со средней связью и кластеризацию с полной связью (Gordon, 1996 ,Джейн и Дубес, 1988 г.).

Чтобы учесть пространственную информацию при кластеризации, были изучены три типа кластерного анализа, в том числе пространственная кластеризация (т. е. кластеризация пространственных точек), регионализация (т. статистика пространственного сканирования). Для первого типа, пространственной кластеризации, сходство между точками данных или кластерами определяется пространственными свойствами (такими как местоположения и расстояния). Методы пространственной кластеризации могут быть секционированными или иерархическими, основанными на плотности или

сетке. Читатели могут обратиться к (Han, Kamber, & Tung, 2001) за всесторонним обзором различных методов пространственной кластеризации.

Регионализация — это особая форма кластеризации, которая направлена на группировку пространственных объектов в пространственно смежные кластеры (т. е. регионы) при оптимизации целевой функции. Многие географические приложения, такие как климатическое зонирование, анализ ландшафта, сегментация изображений дистанционного зондирования, часто требуют, чтобы кластеры были географически непрерывными. Существующие методы районирования, основанные на концепции кластеризации, можно разделить на три группы: (1) многомерная (непространственная) кластеризация с последующей пространственной обработкой для перегруппировки кластеров в регионы (Fovell & Fovell, 1993); (2) кластеризация с пространственно взвешенной мерой несходства, которая рассматривает пространственные свойства как фактор формирования кластеров (Wise, Haining, & Ma, 1997) .) и (3) кластеризация с ограничением смежности, которая обеспечивает пространственную смежность во время процесса кластеризации (Guo, 2008).

Анализ точечных паттернов, также известный как анализ «горячих точек» (Brimicombe, 2007), фокусируется на обнаружении необычных концентраций событий в пространстве, таких как географические скопления болезней, преступлений или дорожно-транспортных происшествий. Общая задача исследования состоит в том, чтобы определить, существует ли избыток наблюдаемых событийных точек (например, случаев заболевания) для области (например, в пределах определенного расстояния до места). Для поиска таких пространственных кластеров было разработано несколько статистических данных сканирования, таких как машина географического анализа (GAM) Openshaw et al., 1987 , Openshaw et al., 1990 и семейство статистики сканирования пространства-времени Kulldorff, 1997 , Kulldorff et al . др., 2005 г.. Статистические данные для обнаружения пространственных кластеров становятся все более доступными для неевклидовых пространств, особенно для сетевых пространств (Шиоде и Шиоде, 2009 г., Се и Ян, 2008 г., Ямада и Тил, 2007 г.).

Тестовая статистика, используемая в GAM, представляет собой подсчет точек (например, случаев заболевания) в пределах области (т. е. круглой области вокруг точки решетки). Чтобы определить, является ли количество точек в области значимым, используется процедура Монте-Карло для генерации большого количества (например, 500) наборов случайных данных, каждый из которых представляет реализацию нулевой гипотезы в одной и той же области. Значение тестовой статистики вычисляется для каждого набора случайных данных, и таким образом выводится распределение значений тестовой статистики при нулевой гипотезе. Путем сравнения фактического значения тестовой статистики (т. е. количества точек) и полученного распределения получается уровень значимости тестовой статистики в области. Потенциальная проблема с GAM, как отмечено в (Rogerson & Yamada, 2009), заключается в том, что его трудно приспособить к проблеме множественного тестирования. Его вычислительная нагрузка также является

недостатком, но более или менее вся статистика сканирования требует значительной вычислительной мощности для поиска и тестирования локальных кластеров.

Статистика пространственного сканирования, разработанная Kulldorff, 1997 , Kulldorff et al., 2005.рассчитывает отношение правдоподобия для каждой локальной области. Чтобы преодолеть проблему многократного тестирования, статистика сканирования использует максимальное отношение правдоподобия (которое является максимальным отношением правдоподобия среди всех локальных областей) в качестве тестовой статистики. Таким образом, метод статистики сканирования сообщает о наиболее вероятном кластере, хотя также предоставляется набор вторичных кластеров. Сначала он вычисляет отношение правдоподобия для каждой из набора зон и находит максимум. Чтобы получить уровень значимости, репликации набора данных генерируются при нулевой гипотезе с учетом общего количества точек. Для каждой повторности снова вычисляется тестовое статистическое значение (т. е. максимальное отношение правдоподобия находится по всем перечисленным локальным областям).

2.4 . Геовизуализация

Геовизуализация касается развития теории и методов, облегчающих построение знаний посредством визуального исследования и анализа геопространственных данных, а также внедрения визуальных инструментов для последующего поиска, синтеза, передачи и использования знаний (MacEachren & Kraak, 2001). Как развивающаяся область геовизуализация привлекла интерес из различных родственных областей и развивалась в различных направлениях исследований, как видно из недавно отредактированного тома о геовизуализации Дайкса , МакИхрена и Краака (2005) .. Основное различие между традиционной картографией и геовизуализацией заключается в том, что первая фокусируется на разработке и использовании карт для передачи информации и общественного потребления, а вторая делает упор на разработку интерактивных карт и связанных с ними инструментов для исследования данных, генерации гипотез и построения знаний. MacEachren, 1994 , MacEachren и Kraak, 1997).

Геовизуализация также тесно связана с исследовательским анализом данных (EDA) и исследовательским анализом пространственных данных (ESDA) (Анселин, 1999 г., Бейли и Гатрелл, 1995 г., Тьюки, 1977 г.) .), который связывает статистические графики и карты и полагается на человека-эксперта для взаимодействия с данными, визуального выявления закономерностей и формулирования гипотез/моделей. Однако, чтобы справиться с современными большими и разнообразными наборами пространственных данных и облегчить обнаружение и понимание сложной информации, геовизуализация должна решать несколько основных задач, включая (1) эффективную и действенную обработку очень больших наборов данных; (2) одновременная работа с несколькими точками зрения и многими переменными для обнаружения сложных закономерностей и (3) разработка эффективного пользовательского интерфейса и интерактивной стратегии для облегчения процесса обнаружения.

Для обработки больших наборов данных и визуализации общих закономерностей визуальные подходы часто сочетаются с вычислительными

методами (такими как кластеризация, классификация и анализ ассоциативных правил), чтобы суммировать данные, выделять структуры и помогать пользователям исследовать и понимать закономерности (Андриенко и Андриенко, 1999 , Guo et al., 2005 , Ward, 2004). Чтобы визуализировать несколько точек зрения и множество переменных, нам часто необходимо сочетать визуализацию с методами уменьшения размеров, такими как многомерное масштабирование, анализ основных компонентов (PCA), самоорганизующиеся карты (SOM) (Agarwal and Skupin, 2008 , Kohonen , 2001) . или другие методы преследования проекций (Cook, Buja, Cabrera, & Hurley, 1995).). Многомерное картирование уже давно является интересной исследовательской проблемой, для решения которой были разработаны многочисленные подходы, такие как специально разработанные символы (Чернофф и Ризви, 1975 , Чжан и Пазнер, 2004), множественные связанные представления (Дайкс, 1998 , МакИхрен и др., 1999 , Monmonier, 1989 , Yan and Thill, 2009) и подходы, основанные на кластеризации (Guo et al., 2003 , Guo et al., 2005). Усилия по исследованию третьей проблемы превратились в активную подобласть, называемую визуальной аналитикой (Thomas & Cook, 2005).

3 . Обзор статей

Здесь мы предлагаем обзор статей в специальном выпуске. Эти статьи вносят свой вклад в литературу по интеллектуальному анализу пространственных данных различными способами. Некоторые статьи расширяют существующие методы, такие как искусственные нейронные сети (ИНС) и пространственная кластеризация, для учета проблем пространственной зависимости и пространственного масштаба. Другие разрабатывают новые методы для типов пространственных данных, которые только недавно стали широко доступными, таких как данные о пути и траектории, описывающие движущиеся объекты. Вклад других статей касается новых приложений методов интеллектуального анализа данных.

Как отмечалось ранее, классификация и прогнозирование являются фундаментальной задачей интеллектуального анализа данных, и ИНС входят в число широко используемых методов классификации. Однако обычная ИНС не учитывает пространственную зависимость и ассоциации между соседними объектами. Ченг и Ван (2009)попытаться решить эту проблему при разработке ИНС для пространственно-временного прогнозирования. Их подход включает пространственные ассоциации между наблюдениями в динамические рекурсивные нейронные сети (DRNN), подход ANN, который включает обратную связь от предыдущих итераций входных и выходных данных модели. Такие обратные связи делают DRNN хорошим кандидатом для моделирования данных временных рядов. В настоящей статье авторы предлагают улучшить предсказание цели, включив не только значение цели в предыдущий интервал времени, но и значения ближайших наблюдений. Три тематических исследования служат для демонстрации этого подхода с использованием различных типов данных с различными пространственными и временными записями — прогнозирование лесных пожаров, экономический валовой внутренний продукт и температура.

Одна из основных проблем интеллектуального анализа пространственных данных связана с обработкой новых типов данных. Недавние достижения в области внедрения GPS для создания устройств с определением местоположения привели к созданию огромного объема данных о движущихся объектах. Обнаружение закономерностей в этих данных является сложной задачей как из-за огромного объема, так и из-за временного характера данных. Додж, Вайбель и Форутан (2009)решить эту проблему в своей статье, посвященной классификации траекторий движения. Авторы представляют способ характеристики траекторий движущихся объектов как с глобальной точки зрения, т. е. тех свойств, которые характеризуют всю траекторию объекта, так и с локальной точки зрения, т. е. тех свойств, которые характеризуют части траектории объекта. Свойства включают такие характеристики, как длина пути и прямолинейность, а также скорость и ускорение. С помощью этих извлеченных характеристик SVM применяется для классификации траекторий по категориям. Два типа данных, данные о транспортировке движущихся транспортных средств, а также данные отслеживания глаз, используются для демонстрации предлагаемого подхода.

Третья статья Пей, Чжу, Чжоу, Ли и Цинь (2009 г.)фокусируется на разработке нового метода анализа точечных паттернов. Авторы отмечают, что установленные методы пространственной кластеризации часто чувствительны к параметризации алгоритма кластеризации, особенно к масштабу, в котором теоретически происходит кластеризация, поскольку такое предположение часто должно быть сделано априори при применении метода кластеризации. Следовательно, результаты кластеризации могут быть весьма субъективными. Для решения этой проблемы авторы представляют новый метод кластеризации, который они называют методом коллективного ближайшего соседа (CLNN). В основе CLNN лежит различие между точками, распределение которых можно объяснить причинно-следственным механизмом, и точками, распределение которых можно объяснить случайным «шумом», где отличительной характеристикой между двумя процессами является интенсивность кластеризации. CLNN расширяет предыдущее исследование, разрабатывая процедуру повторения различных шкал измерения для оценки интенсивности. Авторы демонстрируют CLNN, используя как синтетические данные, так и тематическое исследование, посвященное выявлению кластеров землетрясений в Китае по сейсмическим данным.

Лу, Чен и Хэнкок (2009)также сосредоточены на интеллектуальном анализе данных о движущихся объектах, но они заинтересованы в кластеризации и обнаружении аномалий пути, т. Е. Идентификации выбросов в наборе путей движущегося объекта. Эти авторы предлагают несколько новых метрик, которые можно использовать для определения степени сходства между путями, включая метрики, отражающие глобальный характер пути, например периметр области, которую занимает путь. Другие показатели характеризуют меньшие участки пути, которые могут быть сегментированы на основе ребер, общих с другим путем. Подход к обнаружению аномалий пути был протестирован в двух сценариях на наборе синтетических путей, нанесенных на карту в уличной сети

Ольденбурга, Германия. В первом сценарии генерировался набор кратчайших путей между случайно определенными начальными и конечными узлами в городе, а также меньшее множество путей, которые были вынуждены посетить вершину не на кратчайшем пути. Алгоритм стремился выделить те пути, которые не были кратчайшими путями, из большего набора. Во втором сценарии набор «обычных» путей определяется таким образом, что все пути в наборе начинаются в определенном районе города и заканчиваются в определенном районе города, где каждый нормальный путь является кратчайшим путем. Регионы представляют места, куда люди обычно едут, например, в торговый центр или жилой комплекс. Определены два исключительных набора путей. В первом наборе пути между кластерами не являются кратчайшим путем. Во втором наборе пути имеют нормальную длину, но не заканчиваются кластером. Результаты свидетельствуют об эффективности этих показателей для выявления аномалий пути,

Генетический алгоритм — это подход, вдохновленный биологическими системами, для определения оптимальных или почти оптимальных решений задач оптимизации. Классификация данных пространственного изображения является одним из примеров такого рода задач, к которым могут быть применены эволюционные алгоритмы, и служит предметом исследования, представленного Моммом, Иссоном и Кушмаулом (2009).. Авторы рассматривают классификацию мультиспектральных изображений дистанционного зондирования с использованием нелинейной комбинации как спектральной информации, так и текстурных метрик. Проблема здесь заключается в выборе используемой метрики текстуры изображения, поскольку наиболее информативная метрика текстуры различается в разных предметных областях. Выбор и сочетание текстурной метрики со спектральной информацией для наиболее точной классификации изображений можно рассматривать как задачу оптимизации, решаемую с помощью генетического программирования. Подход генетического программирования сравнивался с обычным классификатором изображений K-Means с использованием мультиспектрального изображения Quickbird для Оксфорда, штат Миссисипи, США, с упором на дифференциацию классов с похожими спектральными, но разными текстурными характеристиками.

Шад, Месгари и Абкар (2009)также используют генетические алгоритмы в качестве подхода к оптимизации. Однако в данном случае целью является интерполяция данных о загрязнении воздуха, особенно твердыми частицами, с использованием метода нечеткого индикатора принадлежности Кригинга. Здесь нечеткая логика используется для моделирования неопределенности в процессе прогнозирования, где степень предсказанного членства может быть представлена с использованием нечеткой логики. Задача оптимизации заключается в определении оптимальной параметризации функции нечеткой логики для определения степени принадлежности. Для этой цели авторы используют генетические алгоритмы, в которых различные нечеткие функции принадлежности могут конкурировать, чтобы максимизировать точность оценки загрязнения. Этот подход был применен к набору данных о концентрации

твердых частиц в воздухе, собранных на станциях мониторинга загрязнения воздуха в Тегеране, Иран.

4 . Заключение

Благодаря широкому применению географических информационных систем (ГИС) и технологии GPS, а также все более развитой инфраструктуре для сбора, обмена и интеграции данных все больше и больше областей исследований получают доступ к высококачественным географическим данным и создают новые способы включения пространственных данных. информацию и анализ в различных исследованиях. Частные предприятия и широкая общественность также проявляют все больший интерес как к предоставлению, так и к использованию географических данных. Эти данные стали более разнообразными, сложными, динамичными и намного большими, чем когда-либо прежде, и поэтому их труднее анализировать и понимать. Интеллектуальный анализ пространственных данных и обнаружение знаний превратились в активную область исследований, которая фокусируется на развитии теории, методологии, и практика извлечения полезной информации и знаний из массивных и сложных пространственных баз данных. Статьи в этом специальном выпуске освещают избранный набор подходов и приложений в интеллектуальном анализе пространственных данных. Как отмечалось ранее, интеллектуальный анализ пространственных данных все еще находится на очень ранней стадии, и его границы и возможности еще предстоит определить. Существуют как возможности, так и проблемы, стоящие перед исследованиями в области интеллектуального анализа пространственных данных.

Интеллектуальный анализ пространственных данных — это не задача, выполняемая нажатием одной кнопки. Мы часто заявляем, что «пусть данные говорят сами за себя». Однако данные не могут рассказать историю, если мы не сформулируем соответствующие вопросы и не используем соответствующие методы для получения ответов из данных. Интеллектуальный анализ данных управляется данными, но также, что более важно, ориентирован на человека, когда пользователь контролирует выбор и интеграцию данных, очистку и преобразование данных, выбор методов анализа и интерпретацию результатов. Это итеративный и индуктивный процесс обучения, встроенный в общую дедуктивную структуру.

Обилие пространственных данных предоставляет захватывающие возможности для новых направлений исследований, но также требует осторожности при использовании этих данных. Данные часто поступают из разных источников и собираются для разных целей при различных условиях, таких как неопределенность измерения, предвзятая выборка, различные единицы площади и ограничение конфиденциальности. Важно понимать качество и характеристики выбранных данных.

Тщательный отбор, предварительная обработка и преобразование данных необходимы для обеспечения значимого анализа и результатов. Какие переменные следует выбрать? Какую систему измерения, такую как евклидово пространство или неметрическое сетевое пространство, следует

использовать? Какие пространственные отношения или контекстную информацию следует учитывать? Могут ли выбранные данные адекватно отражать сложность и характер проблемы?

Новые типы данных и новые области применения (такие как анализ движущихся объектов и траекторий, пространственно встроенные социальные сети, пространственная информация в веб-документах, геокодированные мультимедиа и т. д.) значительно расширили границы исследований интеллектуального анализа пространственных данных. Новые типы данных и приложения часто требуют разработки новых методов интеллектуального анализа данных и открытия новых типов шаблонов.

Обработка очень больших объемов и понимание сложной структуры пространственных данных — еще две основные проблемы интеллектуального анализа пространственных данных, которые требуют как эффективных вычислительных алгоритмов для обработки больших наборов данных, так и эффективных подходов к визуализации для представления и изучения сложных закономерностей.

Практическое занятие 5

НЕ-факторы представления пространственных объектов и явлений

Мониторинг характеристик окружающей среды часто затруднен из-за того, что их определение размыто. Эта проблема возникает во многих приложениях, таких как мониторинг естественной растительности и лесных массивов, развитие землепользования, развитие побережья и т. д. Такие процессы мониторинга часто основаны на использовании данных дистанционного зондирования, из которых извлекается информация о соответствующих особенностях. быть извлечены. Нечеткость понятий и определений в таких приложениях имеет два последствия:

1.Особенности могут быть идентифицированы только с ограниченным уровнем уверенности, так что их пространственное описание имеет существенную нечеткость.

2.Если процессы, подлежащие мониторингу, выражаются через изменения состояний этих признаков, то этот мониторинг также может осуществляться с ограниченной достоверностью.

Эти два аспекта мониторинга будут обсуждаться в этой статье. Сначала будут обсуждаться концептуальные аспекты идентификации нечетких пространственных объектов, затем следует объяснение того, как неопределенность состояний объектов влияет на анализ переходов состояний.

1.1 . Идентификация нечетких пространственных объектов

Пространственная протяженность геообъектов обычно определяется через границы, точнее, через положение граничных точек. Поэтому анализ геометрической неопределенности объектов часто основывается на моделях точности координат этих точек. В этом контексте хорошо известен метод эпсилон-диапазона (Dunn et al., 1990) . Однако решения этой проблемы, найденные в литературе, неудовлетворительны. Причина в том, что

геометрическая неопределенность геообъектов связана не только с точностью координат; это скорее проблема определения объекта и тематической неопределенности (см. также обсуждение этой темы, например, у Chrisman, 1991 ; Burrough, 1986 ; Burrough and Frank, 1995 ;Берроу и Макдоннелл, 1998 г.; Гудчайлд и др., 1992). С этим последним аспектом нельзя справиться только с помощью геометрического подхода.

Это становится очевидным, когда картографирование основано на извлечении признаков из цифровых изображений с растровой структурой, а не с векторной структурой, обычно используемой в геодезии и фотограмметрии. Неопределенность классификации изображений дистанционного зондирования в первую очередь считается тематической, и уверенность в том, что пиксель принадлежит к тематическому классу, может быть выражена через функцию правдоподобия, которая оценивается в процессе классификации . 1993 . Затем сегменты изображения могут быть сформированы из смежных наборов пикселей, относящихся к одному и тому же классу. Если эти сегменты представляют собой пространственные размеры объектов, то неопределенность геометрии этих объектов связана с тем, что значение функции правдоподобия варьируется в зависимости от пикселя.Canters, 1997 , Fisher, 1996 , Wickham et al., 1997 , Usery, 1996 , Brown, 1998 , Gahegan and Ehlers, 1997 . Мы будем использовать подход Molenaar, 1994 , Molenaar, 1996 , Molenaar, 1998, чтобы объяснить, как тематическая неопределенность распространяется на геометрию объектов, когда объекты извлекаются из изображений или растровых данных.

1.2 . Динамика объекта

Для наблюдения за явлениями окружающей среды мы должны иметь возможность сравнивать состояния пространственных объектов в разные эпохи. Литература по этой теме ограничена, и лишь немногие публикации обсуждают динамику объектов, особенно пространственные изменения, в общих чертах Yuan, 1996 , Hornsby and Egenhofer, 1997 . Еще меньше литературы о динамическом поведении нечетких объектов. Обнаружение динамики нечетких объектов является вторым пунктом, который будет рассмотрен в этой статье. Мы будем следовать подходу Ченга (1998) и Ченга и Моленаара (1997) .на основе оценки перекрытия пространственных размеров объектов в последующие годы. Несколько параметров будут объяснены для идентификации типа процесса, имевшего место между двумя эпохами. Этот подход будет проиллюстрирован на примере динамики отложений вдоль побережья Нидерландов.

2 . Аспекты неопределенности нечетких пространственных объектов

Пусть $U M = \{\ldots, O i, \ldots\}$ будет универсумом карты M , где термин «карта» относится к пространственной базе данных , содержащей описание местности. $U M$ — это совокупность всех объектов местности, представленных в этой базе данных. Синтаксический подход к обработке информации о пространственных объектах, разработанный в Molenaar, 1994 , Molenaar, 1996 , Molenaar, 1998, позволяет различать три типа утверждений относительно существования этих пространственных объектов:

1.Экзистенциальное утверждение, утверждающее, что существуют пространственные и тематические условия, подразумевающие существование объекта *O*.

2.Экстенсиональное выражение, идентифицирующее геометрические элементы, которые описывают пространственную протяженность объекта.

3.Геометрическое утверждение, определяющее фактическую форму, размер и положение объекта в метрическом смысле.

Эти три типа утверждений тесно связаны. Экстенсиональное и геометрическое утверждения подразумевают экзистенциальное утверждение — если объект не существует, он не может иметь пространственной протяженности и геометрии. Геометрическое утверждение также подразумевает экстенсиональное утверждение. Все три типа утверждений могут иметь определенную степень неопределенности, и хотя эти утверждения связаны между собой, они дают нам разные точки зрения, подчеркивая различные аспекты неопределенности в отношении описания пространственных объектов.

2.1 . Экзистенциальная неопределенность

Неопределенность существования объекта O может быть выражена функцией *:* Exist(*O*)∈[0,1]. Если эта функция имеет значение 1, мы уверены, что объект существует, если она имеет значение 0, мы уверены, что он не существует. Этот последний случай приводит к философской проблеме, потому что как мы можем делать экзистенциальные утверждения об объектах, которые не существуют, или, скорее, как мы можем идентифицировать несуществующие объекты и ссылаться на них в качестве аргумента этой функции? Эта проблема не будет здесь подробно рассматриваться, но мы будем следовать прагматичному подходу, ограничивая область значений функции значением Exist(*O*)∈(0,1), что означает, что функция может принимать любое значение, большее 0 и меньшее или равное 1. Неопределенность экзистенциального утверждения связана с тем, что процедуры наблюдения, такие как интерпретация фотографий или анализ спутниковых изображений , может идентифицировать условия наблюдения, предполагающие, что объект может существовать в каком-то месте, не давая определенной уверенности в том, что он действительно существует как независимый объект. Затем «наблюдаемый объект» получает идентификатор объекта, но на самом деле он может быть частью другого объекта. Функция «существовать» выражает в данном случае неопределенность фактического реального состояния наблюдаемого объекта. Проблема возникает из-за того, что во многих приложениях ГИС можно только косвенно ссылаться на объекты реального мира через описания, предоставляемые системами наблюдения. Эванс, 1982 , Нил, 1990 , Куайн, 1960 . Если существование объектов неизвестно, то вселенная карты становится нечеткой вселенной:UM={…,{Оя,Существует(Оя)},…}

Членами этой нечеткой вселенной являются объекты, функция которых выражает неопределенность их существования. Эта ситуация принципиально отличается от ситуации, рассматриваемой теорией нечетких (под)множеств Кауфмана, 1975 , Клира и Фолгера, 1988 , Клира и Юана, 1995 , Циммермана, 1985 . Там существование членов универсального

множества не является неопределенным, нечеткими являются только подмножества универсального множества. Поэтому понятия теории нечетких подмножеств и нечетких рассуждений следует применять в нашей ситуации с осторожностью. Если вселенная карты не определена, то что?

Попытаемся сформулировать ответ на этот вопрос, рассматривая полное множество граней векторной карты (или множество растровых элементов) как универсальное множество, из которого могут быть сгенерированы подмножества путем присвоения граней пространственному экстенту объектов. Это означает, что случай, когда *UM состоит из нечетких* объектов, интерпретируется в том смысле, что пространственная протяженность этих объектов является нечеткими подмножествами универсального множества граней. Формализм, разработанный в Molenaar, 1994 , Molenaar, 1996 , Molenaar, 1998 , поможет развить этот подход в следующих разделах этой статьи.

2.2 . Экстенсиональная неопределенность

Предположим, что геометрия объектов некоторого *UM представлена* в векторном формате, т. е. геометрия описана в узлах, ребрах и гранях (или 0-, 1- и 2-ячейках). Пусть Geom(M) — геометрическая составляющая карты, т. е. совокупность всех геометрических элементов, описывающих геометрию всех объектов вселенной. Пусть Face(M) будет набором всех граней в Geom(M), и пусть функция Part 22 [f , O] выражает отношение между гранью $f \in$Face(M) и объектом $O \in U\ M$. Если эта функция имеет значение 1, то лицо принадлежит пространственному экстенту объекта, если оно имеет значение 0, то это не так. Определяем множество:Лицо(O)={f|Часть22[ф, O] = 1}

Тогда это экстенсиональное утверждение объекта в том смысле, что Face(O) определяет пространственную протяженность O . В этих обозначениях геометрическое описание объектов организовано пообъектно. Эта формулировка верна для четких объектов, т. е. объектов с идентифицируемыми границами (см. конец раздела 2.4).

Для нечетких объектов отношение между лицом и объектом не может быть установлено с уверенностью, так что мы имеем Part 22 [f , O]$\in$[0,1] . Таким образом, пространственная протяженность нечеткого объекта неопределенна и определяется как:Лицо(O)={f|Часть22[ф,O]>0}На рис. 1 представлен пример этого случая.

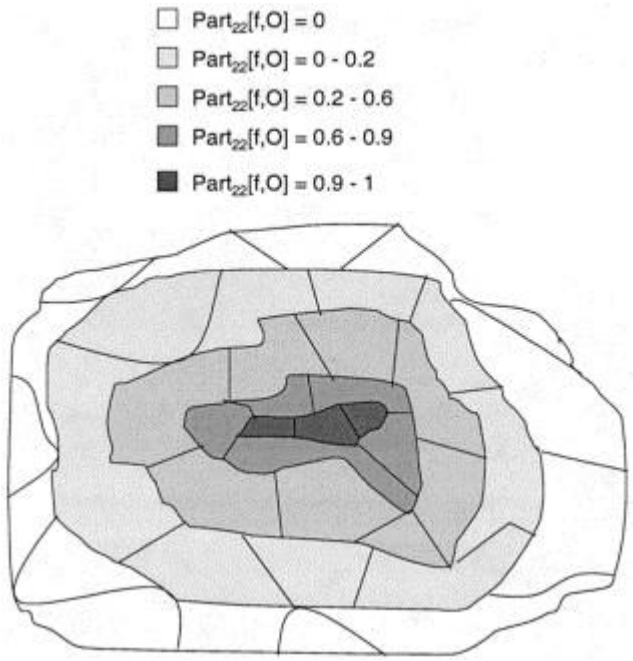


Рис. 1 . Объект с нечеткой пространственной протяженностью .

Синтаксический подход, разработанный Molenaar , 1994 , Molenaar, 1996 , Molenaar, 1998, показывает, что векторная и растровая геометрия обладают сходной выразительной силой . Это означает, что обработка пространственной неопределенности в принципе должна быть одинаковой для обеих геометрических структур. Следовательно, должна быть возможность комбинировать или даже унифицировать векторный и растровый подходы, описанные в литературе. В случае, если геометрия карты представлена в растровом формате, функция Часть 22 [ ] вычисляется для каждой ячейки растра.

Пусть множество граней, связанных с объектом *O* а с уровнем достоверности *c* , равно:Лицо(Oa|c)={eДж|Часть22[фДж, Oa]≥c}

Это условная пространственная протяженность объекта, сравнимая с (сильным) α-разрезом в Клире и Юане (1995) . С помощью этого набора мы можем определить условные функции:Часть22[фДж, Oa|c]=1⇒fДжеЛицо(Oa|c)Часть22[фДж, Oa|c]=0⇒fДж∉Лицо(Oa|c)Отношения между гранями и условными пространственными размерами объекта четкие.

Molenaar (1998) определил нотацию, в которой геометрическое описание площадных объектов было организовано по граням. Это обозначение также может быть изменено для обработки неопределенности, и тогда набор площадных объектов, которые имеют нечеткую связь с лицом, будет следующим:AO(e)={O|Часть22[ф,O]>0}

С помощью этого выражения можно интерпретировать объект нечеткой области как нечеткое поле. Затем для каждой грани *f* функция Part 22 [*f* , *O*] оценивается для каждого объекта карты. Если карта имеет растровую структуру, то для ячеек будет оцениваться набор AO( ).

Затем объекты можно объединить в один слой путем наложения их нечетких полей. Предположим, что нечеткие поля n объектов были представлены в растровом формате и эти *n* растровых слоев должны быть

объединены операцией наложения. Атрибуты от *A* 1 до *An* растра будут представлять функции Part 22 [cell, *O*] для *n объектов, m* . е. для каждой ячейки значение атрибута *k* представляет значение функции Part *22* [cell, *O k*] .

Полевое представление нечетких объектов представляется вполне естественным. Этот факт может объяснить, почему многие картографические дисциплины так сильно склоняются к полевому подходу, а не к объектному. Этот подход часто используется из-за кажущейся синтаксической простоты. Следствием этого может быть то, что объектная структура некоторых описаний местности остается скрытой, а вместе с ней и большая часть семантического содержания этих описаний.

2.3 . Пространственно непересекающиеся объекты

Предположим, что объекты определены так, что они образуют пространственное разделение картируемой области, тогда описание местности имеет структуру однозначной векторной карты Molenaar, 1989 , Molenaar, 1998 . Это означает, что объекты образуют полное покрытие картируемой области и что они не перекрываются, т. е. они пространственно исключают друг друга. Эти требования можно интерпретировать следующим образом для объектов с нечеткой пространственной протяженностью:

Каждая грань принадлежит объекту вселенной карты:(∀f∈Лицо(М))⇒Часть22[ф, уМ]=1

Тогда функция NPart 22 [*f* , *O*]=1−Part 22 [*f* , *O*] выражает уверенность в том, что грань *f* не принадлежит объекту *O*. Тот факт, что объекты вселенной являются пространственно исключительными, может быть выражен тем фактом, что дополнением пространственной протяженности объекта *O* а является совокупная протяженность всех других объектов вселенной *U M − O* а :(∀f∈FM)(∀OaeUM)⇒NPart22[ф, Оазнак равноЧасть22[ф, уМ−Оа]так что в нечеткой однозначной векторной карте мы имеем:Часть22[ф, Оа] = 1-Часть22[ф, уМ−Оа]

Пусть XPart 22 [*f , O*] выражает уверенность в том, что грань f принадлежит исключительно объекту *O* , а не какому-либо другому объекту. Для грани *f k* и объекта *O i* значение этой функции равно:XPart22[фк, Оязнак равноМИН.(Часть22[фк, Оя],МИН.дж ≠ я(NPart22[фк, ОДж]))

Минимальные операторы были применены, потому что мы требуем, чтобы грань принадлежала *O i , а не O j* для любого *j ≠ i* , это подразумевает *и* условие.

Для однозначных векторных карт мы должны потребовать, чтобы расширение:АО(фДж) = {Оа|МАХPart22[фДж, Оазнак равноМАКСИМУМОя(XPart22[фДж, Оя])}содержит только один элемент, т. е. существует только один объект *O* а, для которого функция XPart 22 имеет максимальное значение для грани *f j* . Если объектов с одинаковым максимальным значением больше, то для выбора уникального объекта требуется дополнительное свидетельство.

2.4 . Объекты с выпуклой нечеткой пространственной протяженностью

Граф смежности может быть определен для пространственной протяженности каждого объекта. Граф смежности объекта *O* состоит из:

1.Узлы, представляющие грани, принадлежащие пространственному экстенту O .

2.Ребра, чтобы каждое ребро выражало смежность граней, представленных узлами, которые оно соединяет.

Затем элементарные площадные объекты могут быть определены как объекты, которые имеют пространственную протяженность, состоящую из непрерывного набора граней, так что граф смежности этих граней является связным. Это определение позволяет использовать объекты элементарной площади (или даже грани) с отверстиями. Это определение справедливо для четких объектов.

Это определение элементарных объектов нуждается в модификации, прежде чем его можно будет применить к нечетким объектам, но исходное намерение может быть сохранено. Здесь можно использовать концепцию выпуклых нечетких множеств, описанную в Клире и Юане (1995) , но согласно Моленаару (1998) она будет сформулирована по-другому . Во-первых, некоторые вспомогательные определения будут сформулированы до того, как можно будет дать определение элементарных объектов нечеткой области.

Определение 1

Нечеткий объект O имеет вложенные условные пространственные экстенты, если: $(\forall\, c\,i\,,\, c\,j \mid c\,i > c\,j\,) \Rightarrow$ (Лицо ($O \mid c\,i\,$) $\subset$ Лицо ($O \mid c\,j\,$)).

Это означает, что набор граней, представляющий пространственную протяженность объекта для высокого уровня достоверности, должен содержаться в наборе граней для более низкого уровня достоверности.

Второе определение требует, чтобы для каждого уровня определенности пространственная протяженность объекта была связной, т. е. множество граней имело связный граф смежности.

Определение 2

Нечеткий объект является связным, если: $(\forall\, c > 0) \Rightarrow$Face($O \mid c\,$) *связен.*

Когда объект соответствует этому определению, тогда каждый условный пространственный экстент соответствует определению четких элементарных площадных объектов. С помощью этих двух определений можно определить объекты с выпуклыми нечеткими пространственными размерами:

Определение 3

Нечеткий объект имеет выпуклую нечеткую пространственную протяженность, если он связан и если его условные протяженности вложены друг в друга.

Теперь мы готовы определить элементарные нечеткие объекты:

Определение 4

Объект нечеткой области является элементарным, если он имеет выпуклую нечеткую пространственную протяженность.

Это определение действительно похоже на определение элементарных четких объектов, которое теперь является частным случаем для ситуации, когда зона между $c = 0$ и $c = 1$ схлопывается до ширины, равной 0, т. е. мы получаем∀ся,сДж∈(0,1]⇒Лицо(О|сязнак равноЛицо(О|сДж).

Это означает, что все условные размеры четких объектов идентичны.

3 . Идентификация нечетких объектов

Департамент исследований Министерства общественных работ Нидерландов ежегодно проводит наблюдения за профилями высот вдоль побережья Нидерландов, чтобы отслеживать изменения пляжей и дюн . Геоморфологические единицы должны быть извлечены из данных о высоте с помощью трехэтапной процедуры.

1.Полноразмерный растр получается из профилей посредством интерполяции.

2.Ячейкам растра присвоены классы высот, относящиеся к береговой полосе, пляжу и авандюнам.

3.Будут образованы смежные области, состоящие из ячеек, принадлежащих этим трем классам.

Проблема в том, что это можно сделать только с ограниченной уверенностью, потому что нет фиксированного значения высоты, где кончается береговая полоса и начинается пляж, а также нет случая перехода от пляжа к передним дюнам (Cheng et al., 1997) . Это означает, что классы высот, относящиеся к трем типам регионов, могут быть определены только приблизительно, поскольку они нечеткие.

Здесь мы обсудим, как неопределенность распространяется от классификации ячеек растра до формирования сегментов растра. Пример берегового мониторинга будет использован для пояснения более общей ситуации, когда в процессе идентификации объектов неопределенность тематических данных влияет на определение пространственных размеров объектов.

В этом примере есть три класса высоты, так что для каждой растровой ячейки *P* будет оцениваться вектор:[Л(п,С1),л(п,С2),л(п,С3)]Т(0≤L(п,Ск)≤1)где *L* (*P* , *C*) представляет значение функции принадлежности ячейки сетки *P* , принадлежащей классу *C*. Для каждого класса *C* можно выделить *k регионов, состоящих из ячеек с L* (*P* , *Ck*)> *Threshold k* . Затем каждую область можно интерпретировать как нечеткую протяженность пространственного объекта, принадлежащего классу *C k* . Во многих приложениях разрешены нечеткие пространственные перекрытия между объектами, так что объекты имеют нечеткие переходные зоны Burrough, 1996 , Usery, 1996.. В переходных зонах пиксели могут принадлежать нескольким объектам.

Геоморфологические единицы в этом примере были определены таким образом, что они должны формировать раздел нанесенной на карту области, и это может быть только в том случае, если классы являются пространственно исключительными. Таким образом, каждая ячейка сетки должна принадлежать только одному классу высоты и, следовательно, только одному объекту. Поэтому пусть N *L [* P , Ck *] = 1 − L* [*P , Ck] представляет непринадлежность* , т. е. уверенность в том, что *P не принадлежит* классу *Ck* , и пусть *XL* [*pij , Ck*] выражают уверенность в том, что ячейка сетки принадлежит классу *C k* , а не какому-либо другому классу *C l* с *l ≠ k* . Последний может быть

выражен с помощью минимального оператора(1)ИксЛ[п,Скзнак равноМИН.(Л[п,Ск],МИН.л≠к(НЛ[п,Сл]))

P должен принадлежать только одному классу, требуя только, чтобы существовал один класс, для которого функция X *L* [ ] имеет максимальное значение для *P* . Если классов с одинаковыми максимальными значениями больше, то для выбора уникального класса требуются дополнительные доказательства. Его можно представить как(2)ЕслиИксЛ[п,Скзнак равноМАКСИМУМсл(ИксЛ[п,Ся])(l=1,…,N)тогдаД[п,Ск]=1,иначеД[п,Ск]=0.

После отнесения ячеек к классам будет сформирована область *S* а типа класса *C k* по следующим двум правилам, Molenaar, 1996 , Molenaar, 1998 ,длявсесеткаклеткипijeSa,D[Pij,Ск]=1а также(3)еслипклеSaa такжеСОСЕДНИЙ[Пкл,Пij]=1а такжеД[Пij,Ск]=1тогдапijeSa

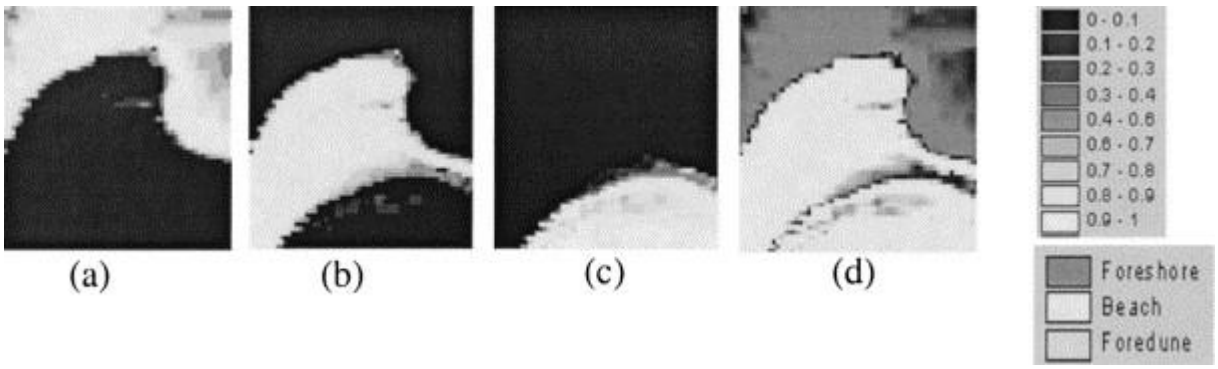
ADJACENT[*P kl* , *P ij*] выражает отношение смежности между ячейками сетки *P kl* и *P ij* , и его значение равно 0 или 1. *P ij* будет присвоено *S* а только в том случае, если *D[P ij ,C k]* =1 . Уверенность в том, что это назначение правильное, зависит от уверенности в том, что ячейка была правильно назначена *C k* . Следовательно, связь между *P ij* и *S* а можно записать в виде(4)Часть[Пij,Сазнак равноМИН.(Д[Пij,Ск],ИксЛ[Пij,Ск])

уравнение (4) выражает, как тематическая неопределенность (неопределенная классификация ячеек сетки) распространяется в экстенсиональную неопределенность (т. е. нечеткую пространственную протяженность) объекта. Например, пусть ячейка сетки имеет значения членства для трех классов:Л[п,С]т1знак равно0,20,70,0такэтоНЛ[п,С]т1знак равно1−0,21−0,71−0,0знак равно0,80,31,0и согласно уравнению (1) ,ИксЛ[п,С]т1знак равноМИН.0,2,0,3,1,00,7,0,8,1,00,0,8,0,0,2знак равно0,20,70,0

С этими результатами мы находим, чтоИксЛ[п,С2знак равноМАКСИМУМСя(ИксЛ[р,Ся])(я=1,2,3)такэтоД[п,С2]=1.

Это означает, что данная ячейка относится к классу *C* 2 с достоверностью 0,7.

На рис. 2(a), (b) и (c) показаны пространственные размеры береговой полосы, пляжа и авандюны в исследовании прибрежной геоморфологии (Cheng et al., 1997) . Каждый из этих рисунков показывает нечеткую пространственную протяженность одного из объектов. На рис. 2(d) они объединены в один слой . На этом рисунке показано, что нельзя определить четких границ, но есть переходные зоны, где оттенки серого выражают неопределенность привязки растровых ячеек к объектам.



Нечеткая классификация и нечеткие объекты в 1989 году.

Практическое занятие 6

Модели знаний и логического вывода в пространственных ситуациях

Карта опасностей, более точно называемая картой предрасположенности к оползням (LSM), предназначена для прогнозирования наиболее вероятных мест обрушения склонов. В этом контексте «восприимчивость» определяется как вероятность возникновения оползня, если не учитывать временные факторы или триггеры, такие как дожди и землетрясения. Оценка восприимчивости к оползням — это процесс оценки этой вероятности на основе физических характеристик местности, таких как уклон, землепользование и литология. Кроме того, исследователи часто рассматривают пространственные корреляции между важными характеристиками местности и распространением оползней в прошлом .

Наиболее важными атрибутами производительности LSM являются надежность и точность будущих вероятностей. Степень точности в основном зависит от количества и качества имеющихся данных, рабочего масштаба карты и методологии анализа и моделирования

Было предложено несколько моделей и методов для создания LSM с использованием географических информационных систем (ГИС). Однако до недавнего времени не было единого мнения о том, какой класс методов наиболее подходит. Процедуры генерации LSM можно классифицировать как прямые или косвенные (Carrara and Guzzetti, 1995), а также как качественные (Barredo et al., 2000) или количественные (Agnesi et al., 2003).

Прямые методы основаны на подробной геоморфологической карте и наносят непосредственно на это поле степень опасности. Хотя прямые методы имеют ряд преимуществ, они требуют много времени и в значительной степени зависят от опыта геоморфолога (Barredo et al., 2000).

Все непрямые методы, согласно Clerici et al. (2002) и Su¨zen and Doyuran (2004) , имеют некоторые общие этапы:

Картирование прошлых оползней в целевом регионе.

Картирование набора геологических/геоморфологических факторов, которые, как предполагается, прямо или косвенно связаны с неустойчивостью склонов.

Оценка корреляций этих факторов с неустойчивостью склона.

Разделение территории на участки различной оползнеопасности на основе этих соотношений (районирование опасности).

Количественные подходы используют математические методологии для оценки восприимчивости при строгих ограничениях. Качественные подходы, с другой стороны, полагаются на экспертное мнение человека или группы (Neaupane and Piantanakulchai, 2006). Качественные методы субъективны и определяют зоны опасности в объяснительных терминах. Количественные методы оценивают вероятность возникновения оползня численно в каждой точке региона, поэтому определяют опасные зоны на непрерывной шкале (Guzzetti et al., 1999).). Для точной оценки этих вероятностей требуется актуальная карта инвентаризации оползней с полной информацией о прошлых массовых перемещениях. Количественные методы также менее субъективны, чем любые качественные подходы, и совсем недавно были предприняты попытки академическими и исследовательскими учреждениями (Ermini et al., 2005).

В литературе по оценке подверженности оползням предлагается множество количественных подходов. Двумерные статистические модели Su¨zen and Doyuran (2004) и Thiery et al. (2007) см. широкое использование. Другие исследователи предпочитают многомерные статистические методы, такие как дискриминантный анализ (Carrara et al., 2003) или линейную и логистическую регрессию (Dai and Lee, 2002 , Dai and Lee, 2003 , Ayalew and Yamagishi, 2005 , Yesilnacar and Topal, 2005 , Greco et al . al., 2007) или нелинейные методы, такие как искусственные нейронные сети (Lee et al., 2004 , Arora et al., 2004 ,Гомес и Кавзоглу, 2005 г., Эрмини и др., 2005 г., Есилнакар и Топал, 2005 г., Канунго и др., 2006 г.). Примеры качественных или полуколичественных подходов включают теорию нечеткой логики Zadeh (1965) , используемую Saboya et al. (2006) ; процесс аналитической иерархии (AHP) Саати (1980) , используемый Ялчином (2008) ; аналитический сетевой процесс (ANP) Saaty (1999) , используемый Neaupane and Piantanakulchai (2006) ; и метод взвешенной линейной комбинации (WLC) Ayalew et al. (2004) .

Нейро-нечеткая модель Kanungo et al. (2005) представляет собой косвенный, полностью количественный метод оценки подверженности оползням. Этот метод, хотя и не связанный напрямую с нейро-нечеткой системой, предложенной в этой статье, тем не менее дал замечательные результаты. Он присваивает степени членства ландшафтным факторам на основе их вклада в возникновение оползней. Метод использует ИНС для решения проблемы регрессии и классификации.

Согласно литературным данным, картирование подверженности оползням обычно страдает от двух проблем. Если используется качественная методика, то немоделированные явления или неполные знания ослабляют экспертные решения. Количественные методы также страдают от неточных или малоточных данных. Метод, сочетающий экспертные знания с объективными данными или,

иначе говоря, сочетание количественного и качественного подходов, может привести к более надежному результату. Экспертные знания могут компенсировать недостатки физических данных, в то время как количественный анализ устраняет некоторую субъективность индивидуальных мнений.

Система нечеткого вывода (FIS) представляет собой гибкую и нелинейную модель, которая включает в себя экспертные знания в стиле человеческого мышления. Это подходит для построения основы простых выводов. С другой стороны, возможности обучения и результаты нейронных сетей делают их естественным дополнением к нечетким системам. ИНС могут автоматизировать или поддерживать процесс разработки нечеткой системы для данной задачи.

В этой статье предлагается новая стратегия интеграции экспертных знаний и существующих данных об оползнях в надежный LSM. Кроме того, мы разрабатываем расширение на основе ГИС с использованием программного обеспечения ArcGIS®, реализующего эту процедуру. На первом этапе выходные данные системы нечеткого вывода интегрируются с фактическими значениями интенсивности оползня. (Следует отметить, что существует несколько возможных определений интенсивности, основанных на физических параметрах, таких как размеры оползня, объем перемещенного материала и глубина. В этом исследовании используется глубина, нормированная на интервал [0, 1].) Второй , нейронная сеть обучается решать задачу регрессии между восприимчивостью к оползням и выбранными характеристиками физического ландшафта. Наконец, модель обобщается и тестируется для всей изучаемой территории.

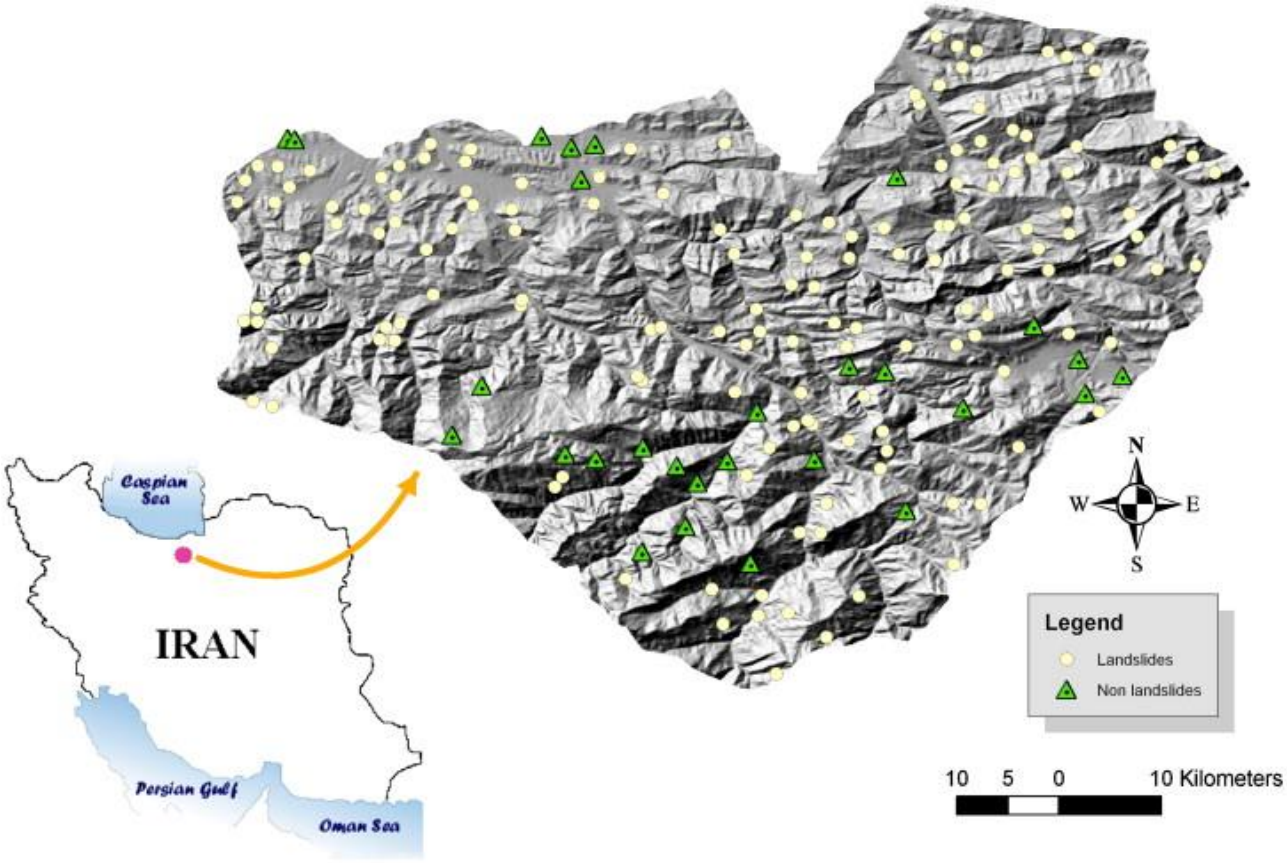
В этой статье не делается попыток изменить известные процедуры построения базы нечетких правил. Скорее, он использует возможности нейронных сетей для выявления и обобщения ассоциаций между физическими входными данными (параметрами ландшафта) и скорректированными выходными данными системы нечеткого вывода.

Область исследования и используемые материалы

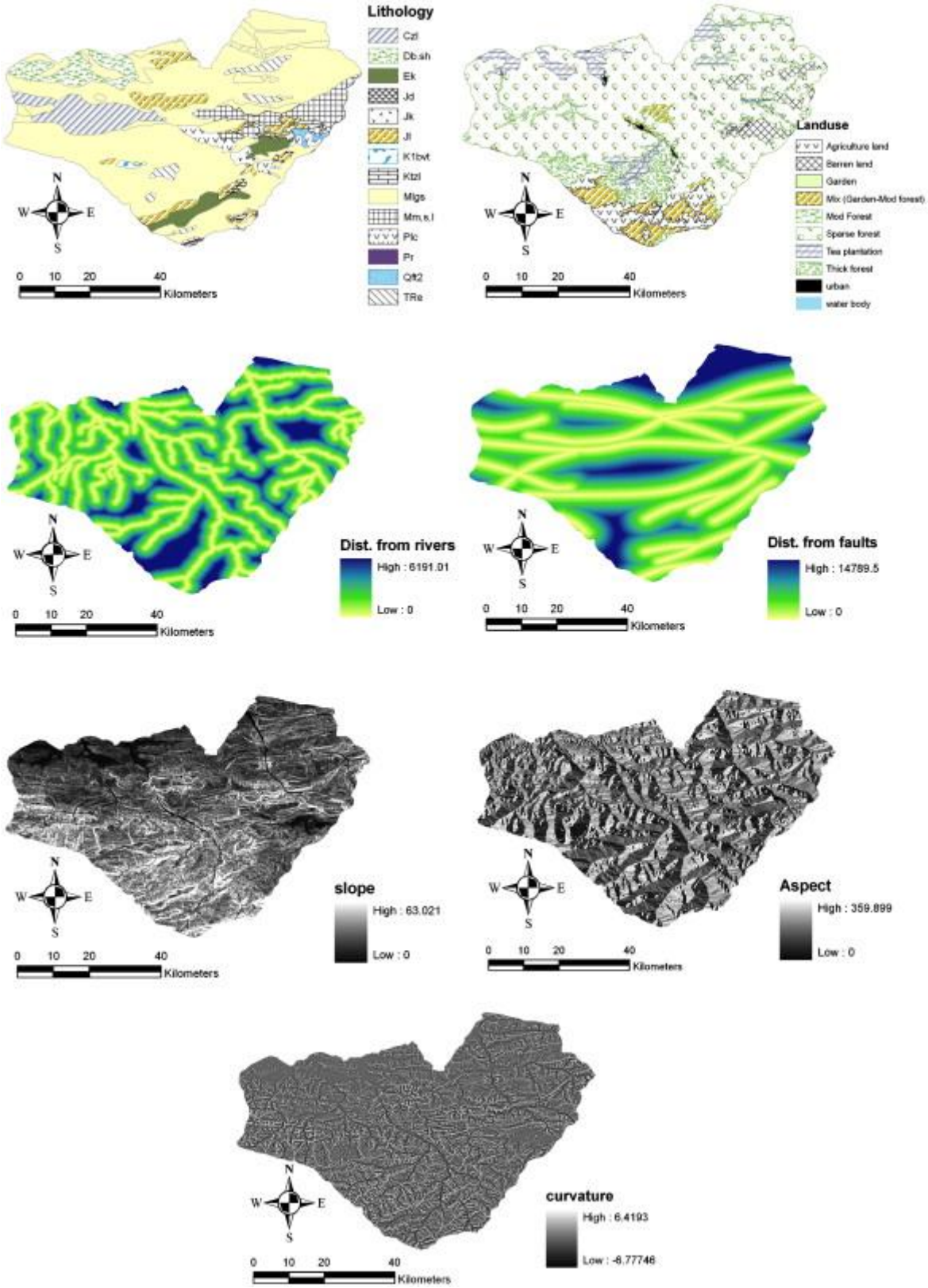
Провинция Мазандаран представляет собой горный и лесной регион, расположенный в центральной части гор Альборз, которые являются причиной большинства оползней, происходящих в Иране. Район исследований находится в пределах 52°31′ и 53°27′ восточной долготы, 35°52′ и 36°30′ северной широты и имеет площадь около 3440 км 2 . Климат региона средиземноморский, с относительно большим годовым количеством осадков (около 1000 мм/год). Район исследований находится вблизи Фирузкоухского района, который отличается относительно высокой сейсмичностью. Кроме того, через исследуемую территорию проходит активная ветвь зоны разломов гор Альборз (AMFZ). Чрезмерное количество осадков, разветвленная дренажная сеть и в целом низкая устойчивость почвы и горных пород к давлению и колебаниям уклона широко признаны наиболее важными факторами, ответственными за оползни и обвалы склонов в этом регионе (www.ngdir.ir).

Для целей настоящего исследования наиболее важными геоморфологическими единицами являются склоновые зоны, способствующие возникновению оползней. Эти зоны являются прерывистыми, поэтому мы начали с подготовки точной инвентаризационной карты оползней. В общей

сложности был собран 151 активный и спящий оползень (рис. 1), а также параметры, связанные с нестабильностью, такие как уклон, экспозиция, кривизна, землепользование, литология, расстояние от рек и расстояние от разломов



Район исследования и его оползневая инвентаризация. На карте также отмечены несколько мест, где маловероятно возникновение оползней.



Карты семи причинных факторов подверженности оползням.

Для этого исследования используются только поступательные оползни, наиболее распространенный тип. Эти оползни хорошо распределены по исследуемому региону. Наклон, экспозиция и кривизна получаются из цифровой модели рельефа (ЦМР). Наклон считается одним из наиболее важных факторов, особенно когда угол большой. Экспозиция оказывает косвенное влияние на неустойчивость склона через его связь с преобладающим ветром. Литология

(состав и структура) связана с подверженностью оползням, поскольку более прочные породы более устойчивы к движущим силам. Землепользование также играет важную роль, поскольку частота оползней обратно пропорциональна густоте растительности (бесплодные земли более подвержены оползням, чем густые леса). Близость к разломам способствует неустойчивости склона и его восприимчивости, влияет не только на структуру поверхности, но и на проницаемость местности. Наконец, эрозия, связанная с реками в холмистых районах, является непосредственной причиной многих оползней. Хотя в данных могут присутствовать и другие причинные факторы, эти семь были выбраны из-за силы их корреляции с окраинами, их репрезентативности на изучаемой территории, их широкой пространственной изменчивости, простоты измерения и отсутствия избыточности.Ялчин, 2008).

Все предварительные тематические карты (инвентаризация оползней, цифровая модель рельефа (ЦМР) с точностью до 1 м по высоте и разрешением 80 м, реки, разломы, литология и землепользование) были получены в мелком масштабе (1:1 000 000) от National Geosciences. База данных Ирана (NGDIR) и Организации по управлению лесными массивами и водоразделами (FRW).

Нечеткая система вывода

FIS обычно рассуждают неопределенной и неточной информацией. Знания, которые они воплощают, часто неточны, точно так же, как человеческое знание несовершенно. В отличие от других качественных методов, таких как многокритериальное принятие решений (Vahidnia et al., 2009a), FIS представляет собой гибкую нелинейную модель неопределенных правил, лежащих в основе данных. Существует несколько исследований пригодности нечетких концепций для картирования подверженности оползням. Сабойя и др. (2006)использовали нечеткий вывод для определения областей, где вероятны оползни, в Итаперуне, Бразилия. Они собрали экспертные оценки, используя лингвистические переменные, и рассчитали индекс потенциального отказа (FPI). Они утверждали, что «интеллектуальные модели», использующие нечеткую логику, приемлемы в геотехнической инженерии, если в модели присутствуют суждения, индукция и дедукция. Низкая стоимость и быстрое время отклика, связанные с моделированием лингвистических переменных в форме нечетких логических правил, делают эту методологию очень привлекательной для картирования подверженности оползням.

Однако модель Сабойи остается по существу линейной и не поддерживает динамический вывод для различных ситуаций нестабильности. Кроме того, как эта модель, так и модель нечеткого взвешивания Wang et al. (2009) полностью полагаются на знания экспертов. Напротив, Ercanoglu и Gokceoglu (2004) создали LSM, используя нечеткие отношения на основе данных, без каких-либо экспертных знаний. В следующем разделе мы опишем систему нечеткого вывода, которая включает в себя как знания, так и данные. В этом разделе мы кратко опишем необходимые понятия нечетких множеств, нечетких операций и нечетких отношений (Zadeh, 1965 , Cox, 1995 , Chen and Pham, 2000).

Нечеткая логика

В теории нечетких множеств *функция принадлежности* отображает каждому элементу дискурсивной вселенной *значение принадлежности* от 0 до 1. В этом исследовании используется треугольная функция принадлежности, самый простой и наиболее широко используемый вариант. В *нечеткой логике* наиболее распространенными математическими операциями, связанными с выводом, являются пересечение, объединение и дополнение. *Нечеткие отношения* отображают элементы одной вселенной в элементы другой посредством декартова произведения.

В отличие от классической (не-нечеткой) логики, нечеткая логика позволяет утверждениям быть частично истинными и частично ложными. Чтобы описать частичные значения истинности и приблизительное человеческое мышление, Заде установил бесконечнозначную логическую систему. Он определил первичные логические операторы – И (∧), ИЛИ (∨), НЕТ (¬), импликацию (→) и эквивалентность (↔) – для любых двух неточных предложений p и q (Chen and Pham, 2000). Нечеткие множества и нечеткие операторы являются «субъектами» и «глаголами» нечеткой логики, соответственно, и могут использоваться для получения значений частичной истинности для утверждений в форме «если предшествующее, то последующее» (Hudec and Vujosevic, 2004). Набор операторов «если-то» (база правил) и импликации (→) используются для формулировки условных операторов, составляющих нечеткую логику (Saboya et al., 2006), например:(1)ЕслиИкс1являетсяAi,1а также/илиИкс2являетсяAi,2а также/или,…,ИксрявляетсяAi,ртогдаа такжеявляетсяБi,i=1,…,k

Здесь y — определяемая переменная (в нашем случае — восприимчивость), а $x\,1$, $x\,2$,…, $x\,p$ — входные переменные (в нашем случае извлеченные из причинных факторов). $A\,i,j$ — нечеткие множества, представляющие лингвистические переменные («низкий», «умеренный», «высокий» и т. д.), а i — индекс правила. $B\,i$ — еще одна лингвистическая переменная, предполагаемая выходными данными.

Нечеткий вывод

Нечеткий вывод — это процесс сопоставления заданного набора входных данных с выходными данными с использованием нечеткой логики (Siler and Buckley, 2005). Метод нечеткого вывода Мамдани (Mamdani and Assilian, 1975) является наиболее широко используемой методологией. Наш собственный подход к выводу несколько отличается от метода, описанного в исходной статье, особенно в отношении отношения импликации. Мы используем импликационное отношение Zadeh (1973) , которое можно интерпретировать с помощью нечетких функций принадлежности и нечетких правил if-then следующим образом:(2)v(p→д)⇔v(p)→v(д)⇔(¬v(p))∨(v(p)∧v(д))

Здесь p и q — неточные утверждения, а v — степень истинности, точно такая же, как и у нечеткой функции принадлежности.

Нечеткий вывод, использованный в этом исследовании, основан на правилах вывода Мамдани и Ассиляна (1975) и импликации Заде (уравнение 2). Следовательно, используется уравнение(3)мюБ˜(а также)=∨Икс∈Икс{[1−мюА˜(Икс)]∨[мюА˜(Икс)∧мюр˜(Икс,а также)]|Икс∈Икс,а

также∈A ТАКЖЕ}A˜a такжеБ˜являются антецедентом и консеквентом, определенными во вселенных X и Y соответственно.р˜— нечеткое отношение между входами (x) и выходами (y) системы, а μ — функция принадлежности.

Шаги процесса логического вывода, показанные на рис. 3 , резюмируются ниже:

Нечеткие входные данные: возьмите каждый входной сигнал и определите степень, в которой он принадлежит каждому из соответствующих нечетких множеств (треугольные числа, представляющие лингвистические переменные). То есть каждый ввод заменяется набором значений членства.

Применение нечетких операторов: если антецедент данного правила состоит из более чем одной части, нечеткие операторы (в этом исследовании используются только MAX, MIN и NOT) применяются для получения одного нечеткого числа из нечетких входных переменных.

Применение метода импликации: чтобы определить результат каждого правила, консеквент преобразуется по отношению к антецеденту с использованием функции импликации (импликация Заде).

Агрегировать все выходные данные: поскольку решения должны основываться на всех правилах FIS, результаты отдельных правил должны быть каким-то образом агрегированы (в этом исследовании мы используем оператор MAX). В результате получается одно нечеткое множество.

Дефаззификация: агрегированный нечеткий набор охватывает диапазон выходных значений, поэтому его необходимо дефаззифицировать, чтобы получить один числовой вывод. Используем метод центроидов:

(4)а также0=∫а также×мюI˜Agg(а также)га также∫мюI˜Agg(а также)кудамюI˜Agg(а также)- функция принадлежности, связанная с шагом агрегации, иа также0является дефаззифицированным значением.

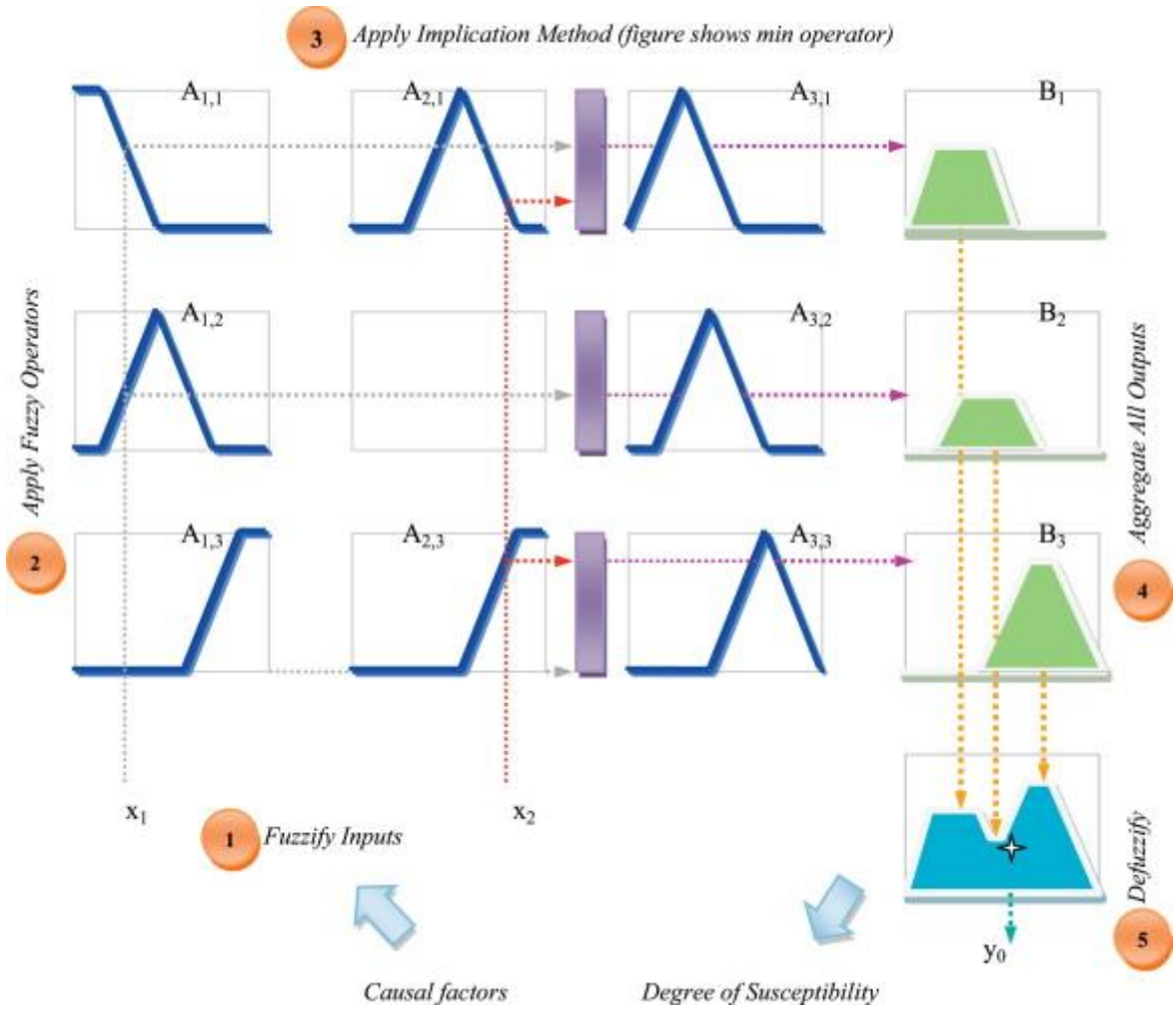


Рис. 3 . Стратегия нечеткого вывода на основе Мамдани (используя 2 входа и 3 правила, получается 1 выход).

Практическое занятие 7

Построение образов ситуаций.

Система поддержки принятия решений (DSS) представляет собой компьютеризированную информационную систему, предназначенную для поддержки деловых, организационных и управленческих действий по принятию решений [1] . DSS обслуживают уровни управления, операций и планирования организации и помогают принимать решения. Многие пространственные проблемы полуструктурированы или плохо определены [2] , потому что все их аспекты не могут быть измерены или смоделированы. Для эффективной поддержки принятия решений для полуструктурированных пространственных задач были созданы системы поддержки пространственных решений (SDSS). Подобно DSS, SDSS обеспечивает основу для интеграции систем управления базами данных с аналитическими моделями, возможностями графического отображения и табличной отчетности, а также знаниями лиц, принимающих решения [3] .. SDSS применяется во многих областях, таких как сельское хозяйство [4] , бизнес [5] , энергетика [6] , противопожарная

защита [7] , анализ пригодности земель [8] , транспорт [9] , коммунальные услуги [10] и водоснабжение. управление ресурсами [11] .

Проблемы с принятием решений по множеству критериев часто возникают в повседневной жизни, в деловых ситуациях и при инженерных задачах. Для принятия обоснованного решения был разработан ряд методов многокритериального принятия решений (MCDM) [12] .. Подходы MCDM могут оценивать большое количество критериев принятия решений и возможных решений с помощью сравнений и определения приоритетов. Возможность расчета приоритетов делает методы MCDM подходящим инструментом для измерения влияния критериев на среду принятия решений. Сочетание анализа на основе геопространственной информационной системы (ГИС) с методами MCDM позволило создать SDSS на основе нескольких критериев (MC-SDSS), целью которого является определение и сравнение решений задачи пространственного решения на основе комбинации множества факторов, которые могут быть хотя бы частично представлены картами [13] .

MC-SDSS использовался во многих различных областях, таких как выбор места для медицинских учреждений [14] , управление водными ресурсами [15] , управление стихийными бедствиями [16] , экологические исследования [17] , [18] , оценка пригодности земли [19] , городское планирование [20] , выбор места жительства [21] и т.д. В настоящее время существует хорошо зарекомендовавшая себя литература по интеграции ГИС-MCDM [22] , а методы и приложения обсуждались в некоторых исследованиях, таких как [13] и [23] .

В этом документе представлена MC-SDSS для оценки производительности транспортной сети (TNP) в чрезвычайных ситуациях. Транспортная сеть имеет тесную взаимосвязь со стихийными бедствиями и управлением ими. С одной стороны, транспортные сети представляют собой крупномасштабные, пространственно распределенные и сложные системы, что делает их уязвимыми для стихийных бедствий. С другой стороны, транспортная сеть облегчает обмен товарами и услугами, особенно в ситуациях управления стихийными бедствиями. Поскольку большинство бедствий требует значительного материально-технического обеспечения для оказания помощи пострадавшим и транспортировки оборудования и гуманитарных грузов [24] , [25] , можно сказать, что транспортные сети являются основой служб управления чрезвычайными ситуациями [26] .. Растущее понимание этих вопросов привело к растущему научному и практическому интересу к предмету TNP в ситуациях бедствий [27] , [28] , и для оценки TNP был разработан ряд систем поддержки принятия решений, таких как [29] , [30] .

Кроме того, для оценки и анализа TNP в некоторой литературе были предложены различные показатели эффективности (таблица 1).

Таблица 1 . Общие определения показателей эффективности.

Мера	Общее определение
Уязвимость	Уязвимость — это восприимчивость системы

Мера	Общее определение
	к угрозам и инцидентам, вызывающим снижение производительности.
Надежность	Надежность — это степень, в которой при заранее определенных обстоятельствах сеть способна поддерживать функцию, для которой она была изначально разработана.
Надежность	Надежность – это вероятность того, что система остается работоспособной на удовлетворительном уровне.
Риск	Риск — это сочетание вероятности события и его последствий с точки зрения производительности системы.
Контакт	Воздействие — это вероятность и вероятная серьезность того, что конкретный элемент системы подвергнется удару или подвергнется воздействию угрозы.
Важность	Важность заключается в том, что снижение производительности того или иного элемента инфраструктуры влияет на пользователей.
Критичность	Если дорога и слабая, и важная, компонент критический

Мера	Общее определение
устойчивость	Устойчивость определяется как способность транспортной сети приспосабливаться к возмущениям/катастрофам и возвращаться к нормальному функционированию в «разумные» сроки.

Хотя в этой области использовались различные термины (всеобъемлющий обзор был сделан Faturechi и Miller-Hooks [57]), как Sullivan and et al. [51] и Knoop и соавт. [58] , каждый исследователь пролил свет на определенный аспект TNP. Насколько нам известно и согласно обзору Mattsson и Jenelius [59], предыдущие работы редко представляли комплексный подход к оценке TNP в чрезвычайных ситуациях. В этой статье используется термин «стратегическая дорога» как относительно новое понятие для оценки TNP. Стратегическая дорога определяется как «дорога, которая обеспечивает адекватное реагирование на чрезвычайные ситуации в случае стихийного бедствия, а ее выход из строя имеет серьезные негативные последствия». Это определение вводит несколько показателей эффективности для оценки TNP в чрезвычайных ситуациях:

-

Пропускная способность дороги: внутренние характеристики дороги (такие как ширина, тип и топографическое состояние) влияют на использование дороги в ситуациях управления стихийными бедствиями.

Уязвимость дорог: некоторые из дорог имеют несколько важных объектов инфраструктуры (например, туннели и мосты), что делает их более уязвимыми для стихийных бедствий.

Доступ к центрам экстренной помощи: в случае стихийного бедствия требуется доступ к службам экстренной помощи (таким как полиция, пожарные, спасательные и медицинские центры).

Важность дорог: некоторые дороги играют значительную роль в геометрической структуре сети, и их выход из строя оказывает огромное влияние на TNP.

Вышеупомянутые индикаторы учитывают различные аспекты TNP в ситуациях стихийных бедствий и требуют разработки полностью интегрированного MC-SDSS для оценки TNP.

Оставшаяся часть статьи организована следующим образом: Модель решения проблемы вместе с ее применением в реальном примере представлена в Разделе 2 . Предварительные определения, необходимые для описания MC-SDSS, представлены в разделе 3 . Затем раздел 4 посвящен описанию реализации предложенной системы и достигнутых результатов исследования. В этом разделе также представлены некоторые обсуждения, а статья заканчивается некоторыми выводами и предложениями для будущей работы.

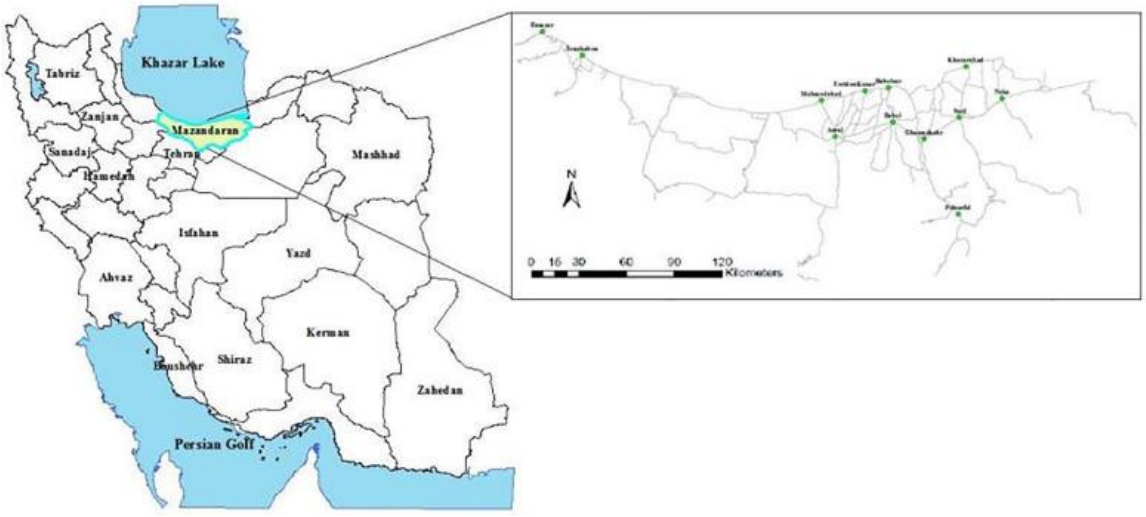
2 . Модель и применение

2.1 . Структура модели

Общая цель исследования состоит в том, чтобы предложить MC-SDSS для определения стратегически важных дорог в условиях стихийных бедствий. Для достижения этой цели необходимо предпринять следующие шаги, показанные на , после определения требуемых критериев решения задачи собираются соответствующие данные. Затем выполняются два параллельных режима действий: один для создания карт критериев, а другой для определения весов критериев. Путем введения различных моделей пространственного анализа генерируются четыре требуемые карты критериев, и с использованием методов MCDM веса критериев рассчитываются на основе предпочтений экспертов. Затем, следуя предыдущим шагам и используя взвешенную линейную комбинацию (WLC), создается карта стратегических дорог. Наконец, выполняется анализ чувствительности, чтобы оценить устойчивость результатов из-за вариаций весов.

2.2 . Описание области исследования

Разработанный MC-SDSS применяется в провинции Мазандаран для определения стратегических дорог. Провинция Мазандаран расположена в северной части Ирана и окружена горами Альборз на юге и Каспийским морем на севере (рис. 6). Эта ситуация ограничивает его доступ к другим соседним провинциям в случае стихийных бедствий. Таким образом, внутренние дороги Мазандарана играют важную роль в чрезвычайных ситуациях. Мазандаран имеет различную топографию и сильную дождливую погоду. Дорожная сетьрайона является очень востребованным и привлекает множество людей для отдыха в течение всего года. Он имеет 141,5 км шоссе и 388,7 км основных дорог. Наличие разветвленной дорожной сети, нестабильных погодных условий, высокого рельефа рельефа и высокого спроса делает эту дорожную сеть уязвимой для стихийных бедствий. Местоположение исследуемого участка показано на рис. 2 . По собранным нами данным, он состоит из 220 участков дорог.



Область исследования: провинция Мазандаран.

3 . Предварительные определения

3.1 . Определение критериев

Как упоминалось выше во вводном разделе, определение стратегических дорог приводит к различным показателям эффективности для оценки TNP в чрезвычайных ситуациях: показатели уязвимости, важности, пропускной способности и доступности. Эти показатели эффективности более подробно описаны в следующих разделах.

3.1.1 . Критерий уязвимости

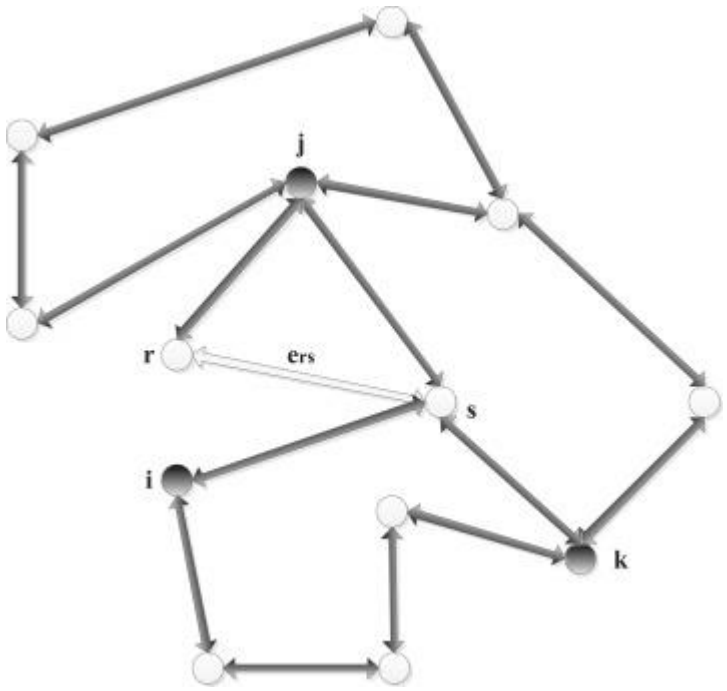
Оценка TNP включает в себя оценку важнейших инфраструктур, включая мосты и туннели [60] . Пан и др. [61] объяснили, что национальная дорожная инфраструктура, особенно автомобильные мосты и туннели, подвержены структурным дефектам . Шринивас и др. [62] при условии, что дорожное разрушение обычно связано с разрушением моста или туннеля. В текущем исследовании, основанном на идее о том, что чем больше мостов и туннелей на дороге, тем она более уязвима, предлагается следующий критерий уязвимости:(1)CV=∑Bci.viГде *C V* — значение уязвимости для каждого участка дороги (звена), *v i* — количество уязвимых инфраструктур, содержащих мосты и туннели в пределах дороги (*i*), а *wc i* — их относительный вес. Шаги для расчета этой критериальной карты с использованием инструментов ГИС

Как показано на рис. 3 , сначала подсчитывается количество существующих туннелей и мостов путем пересечения слоя дороги со слоем туннеля и моста соответственно, а затем с помощью простого метода MCDM (например, WLC) генерируется карта критериев уязвимости.

3.1.2 . Критерий важности

Анализ сетевых нарушений — это методологический подход, который успешно применялся к задачам технического обслуживания и планирования транспорта с целью выявления и ранжирования наиболее важных звеньев в сети, а также для оценки надежности сети в целом [32] , [38] , [63]

Для ранжирования дорог по их важности в геометрической структуре сети в данной статье применяется подход, аналогичный описанному в [32] . Рассмотрим дорожно-транспортную сеть как граф (G), содержащий пару узлов и ребер (V, N), где узлы представляют города, а ребра, соединяющие два города. Базовая модель может быть использована для измерения критерия важности с точки зрения изменения общей стоимости проезда между двумя городами в случае отказа данного соединения. Хотя в качестве обобщенных затрат могут рассматриваться различные параметры, такие как расстояние, время, деньги и т. д., в исследовании в качестве обобщенных затрат выбран параметр времени. Схематическое изображение этой проблемы показано на рис. 4 , а критерий важности ссылки может быть записан какуравнение (2) :(2)Ci=∑i=1н∑Дж=1igij−∑i=1н∑Дж=1igджрсГде *C i* — это разница во времени как обобщенная стоимость проезда между двумя городами при отказе данной линии связи, *d ij* обозначает наименьшую стоимость проезда от узла *i* до узла *j* , а *d ijrs* обозначает наименьшую стоимость проезда от узла *i* до узла j. узел *j* , если сетевой канал *e rs* выходит из строя (является каналом, соединяющим узлы *r* и *s*).



Как показано, чем больше увеличивается стоимость проезда от узла *i* к узлу *j* *при отказе* сетевого звена , тем более важным *является звено* . Для расчета критерия важности связи сначала создается граф, содержащий города (кружки на рис. 4) и соединяющие их дороги (стрелки). Затем некоторые из важных узлов (городов) рассматриваются как базовые узлы (черный кружок), а затем вычисляется совокупная сумма наименьших взаимных затрат на перемещение между этими узлами. Наконец, ссылка удаляется из сети (белая стрелка), и влияние ее отказа измеряется с помощью уравнения. (2). Этот процесс выполняется для всех звеньев сети.

3.1.3 . Критерий производительности

Реагирование на чрезвычайные ситуации включает в себя междугородние перевозки товаров, таких как медицинские материалы, персонал, специальное оборудование, а также проведение спасательных работ, продуктов питания и других товаров, используемых в операциях по оказанию помощи [64] . Поэтому дорожно-транспортная сеть должна быть способна выдерживать нагрузку транспортных средств различных типов в аварийных ситуациях. Большая нагрузка на дорожную инфраструктуру может привести к ее выходу из строя. В данном исследовании используется эмпирический метод оценки грузоподъемности различных дорог. Этот метод показан в таблице 2 .

Как показано в Таблице 2 , если дорога расположена во внутренней части страны и ее тип слоя - высококачественный асфальт, то ее первичный дневной тоннаж составляет 60 000 тонн. Также, если дорога узкая, расположена в местности с низким рельефом и неподходящими погодными условиями, ее дневной тоннаж будет уменьшен на 25%, 30% и 20% соответственно. Рассчитывая ежедневный тоннаж различных дорог, можно классифицировать TNP на основе пропускной способности дорог.

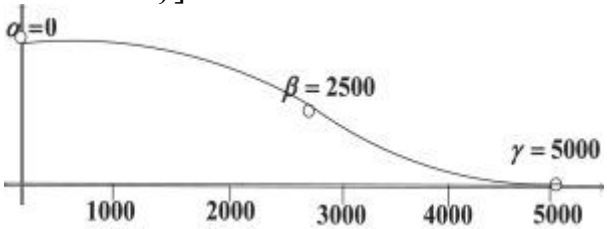
3.1.4 . Критерий доступности

По данным Федерального агентства по управлению в чрезвычайных ситуациях (FEMA) США, когда происходит стихийное бедствие, эвакуация людей и обеспечение доступа к объектам аварийно-спасательных служб являются главными приоритетами, за которыми следует доступ к инфраструктурам аварийных операций, таким как центры аварийных операций и центры распределения снабжения . 66] ; и структура дороги может сильно повлиять на доступность этих услуг [67] . Существует значительный объем работ, в которых исследуется применение различных мер доступности в контексте дорожных сетей [68] , [69] . Тейлор [70]определяет доступность как простоту доступа к услугам и объектам при использовании дорожной сети.

В этом исследовании предлагаемый критерий доступности ориентирован на звено и может использоваться для определения точных звеньев проезжей части в сети, которые являются наиболее важными с точки зрения облегчения доступа к службам экстренной помощи. Дороги, расположенные на близком расстоянии от поставщиков аварийно-спасательных служб (таких как центры здравоохранения, больницы и полицейские участки), играют важную роль в экстренном реагировании на стихийные бедствия [67] . Простая линейная агрегация в уравнении. (3) используется для расчета доступности к центрам аварийных поставщиков.(3)CA=∑i=1н(Bai.г˜i)Где *C A* — значение доступности каждой дороги для поставщиков аварийно-спасательных служб, *wa i* — относительный вес каждого поставщика аварийно-спасательных служб, полученный на основе суждений экспертов иг˜i– нормированное расстояние каждой дороги до места расположения аварийных поставщиков. Этапы расчета этого критерия показаны на рис. 5 .

Как показано на рис. 5 , для расчета критерия доступности сначала создается 5-километровый буфер (на основе рекомендаций экспертов) для каждой дороги. Затем, накладывая буферный слой на слой аварийных поставщиков, подсчитывают количество аварийных поставщиков, находящихся в буферных зонах . Далее, поставщики экстренных служб должны быть оценены в зависимости от их близости к дороге. Для этой цели используется сплайн (на основе рекомендаций экспертов) в качестве функции уменьшения расстояния для нормализации оценки расстояния до поставщиков аварийных служб. Затем вес каждого аварийного поставщика умножается на нормализованную оценку расстояния до дороги, и, наконец, результаты суммируются.

Математическая форма рис. 6 показана в уравнении. (4) :(4)[s(u;a,б,c)={1→u≤a1−2[(u−c)/(c−a)]2→a<u≤б2[(u−a)/(c−a)]2 →б<u≤c0→u>c}]



3.2 . Аналитический иерархический процесс (AHP)

AHP как известный метод MCDM используется для получения приоритетов и предпочтений лиц, принимающих решения, относительно

критериев. Она была разработана Саати [71] в 1970-х годах и, согласно обзорам Хо и Ма [72] и Сипахи и Тимора [73] , является наиболее часто используемой процедурой для оценки весов критериев [74] . Что касается простоты, AHP также имеет возможность систематически декомпозировать сложную проблему принятия решений с использованием иерархических уровней [75] .. Он рассматривает набор критериев оценки и набор альтернативных вариантов, среди которых выбирается наилучшее разрешение. Наилучший вариант — это не тот, который оптимизирует каждый отдельный критерий, а тот, который обеспечивает наиболее подходящий компромисс между различными критериями [76] . Методология AHP включает четыре этапа [77] :

Структурирование иерархии: определение цели анализа и построение модели иерархической структуры. Модель структурируется сверху по цели, по критериям и подкритериям на промежуточных уровнях, до альтернатив, которые ставятся на низший уровень иерархии;

Построить матрицу парных сравнений : элементы определенного уровня сравниваются попарно относительно определенного элемента верхнего уровня. Цель такого анализа состоит в том, чтобы определить степень относительной важности элементов.

Состав приоритетов: как правило, веса определяются с помощью матричной алгебры для определения главного собственного вектора . B=(B1,B1,…,Bн)из ПКМ, где w i >0 и$\sum$i=1нBi=1. Главный собственный вектор для каждой матрицы при нормализации становится вектором приоритетов (т.е. весов) для этой матрицы [78] .

Оценка непротиворечивости: непротиворечивость — это мера, позволяющая оценить, является ли относительное суждение, данное респондентом, последовательным или нет. Суждение называется непротиворечивым, если оно соответствует логике предпочтения транзитивного свойства (т.aij.aДжейк=ая) [71] . Каждая запись a ij матрицы PCM представляет важность *i* -го критерия по отношению к *j* -му критерию. Следовательно, этот шаг обеспечивает логическую последовательность суждений заинтересованных сторон [78] .

По двум основным причинам метод AHP применялся для оценки риска в исследованиях по защите критической инфраструктуры ; Во-первых, это легко реализуемый подход; ценная функция, редко встречающаяся в других методах. Во-вторых, метод AHP может легко унифицировать результаты оценки для альтернативных схем и может быть легко проведен, когда к оценке привлекаются эксперты с разным опытом [79] . Берньери и др. [80] предложили систему поддержки принятия решений, основанную на AHP, чтобы предложить наилучший способ восстановления безопасного состояния критической инфраструктуры в случае аварии. На основании AHP, Klein и соавт. [81]разработала новый модифицированный метод - MAHP - для оценки и сравнения мер защиты критической инфраструктуры по качественным критериям. Используя метод AHP, Deshmukh и соавт. [82] оценили воздействие наводнений как стихийного бедствия на инфраструктуру и связанные с ней отрасли и сообщества с точки зрения критичности и уязвимости

инфраструктуры, а также серьезности социальных и экономических последствий, если будет затронута критически важная инфраструктура. Хуанг и др. [83] предложили новый метод, учитывающий эту взаимозависимость .и эффекты обратной связи между различными типами критически важных инфраструктур с использованием гибридной модели, которая представляет собой комбинацию Лаборатории испытаний и оценки принятия решений (DEMATEL) и Аналитического сетевого процесса (ANP).

Практическое занятие 8

Методы использования опыта в принятии решений

Планирование и внедрение крупномасштабных устойчивых энергетических систем создают серьезные проблемы для заинтересованных сторон из-за сложности. Благодаря быстрому росту доступности данных о зданиях появляются возможности для анализа существующих данных о зданиях и разработки стратегического и эффективного планирования энергопотребления. Однако для интеграции имеющихся данных об энергии и планировании требуются систематические подходы. Одним из возможных решений для крупномасштабного энергетического анализа зданий является пространственный анализ энергетических данных с использованием моделирования географической информационной системы (ГИС) [4] . Этот подход широко используется для регионального, городского и национального планирования [5] .и является одним из основных инструментов для представления крупномасштабных географических данных в визуальном формате. ГИС обеспечивает основу для сбора, управления и анализа крупномасштабных данных в географическом контексте. Визуальное представление данных в системе ГИС может помочь заинтересованным сторонам выполнить качественный и количественный анализ для поддержки принятия решений [6] .

Энергетическое планирование на основе ГИС требует обширных данных для принятия решения по энергетической политике [4] . Анализ отдельных зданий часто затруднен в больших масштабах из-за ограниченной доступности данных и проблем с конфиденциальностью пользователей [7] . Одно из наиболее многообещающих решений для энергетического анализа зданий с ограниченной информацией может быть выполнено путем моделирования запасов зданий [8] . Тем не менее, большинство исследований сосредоточено на разработке моделей строительных фондов без учета аспектов, которые интегрируют пространственную информацию для процессов принятия решений [9] .

Как правило, крупномасштабное моделирование строительных фондов использует два подхода, а именно инженерное моделирование и моделирование на основе данных [10] . Инженерно-ориентированные подходы используют архетипы зданий, которые представляют собой различные типы жилья из строительного фонда, для расчета энергопотребления с использованием

численных имитационных моделей [11] . Тем не менее, существующие исследования по моделированию городской энергии часто полагаются на агрегированные данные о зданиях и впредь не учитывают детальный анализ характеристик зданий [8] . Модели, управляемые данными, используют исторические данные о строительных фондах для построения взаимосвязей между входными и выходными данными с использованием статистических методов или методов машинного обучения [12] .. Этот подход полезен, когда доступны ограниченные исторические данные. Однако существующие исследования сосредоточены на традиционных статистических методах, и лишь в ограниченном числе исследований широко применяются методы машинного обучения с использованием пространственных признаков.

Градостроителям, местным органам власти и лицам, определяющим политику в области энергетики, часто приходится проводить энергетическое планирование и анализ в масштабе района или квартала. В то время как органы власти на национальном уровне часто сталкиваются с трудностями при координации крупных разрозненных источников индивидуальной информации, местные органы власти не имеют доступа к данным о фондах зданий за пределами соответствующей сферы полномочий. Поскольку предыдущие исследования по энергетическому моделированию зданий в основном сосредоточены на анализе национального или городского масштаба для планирования энергетической политики [13]Таким образом, эти стратегии не рассматриваются должным образом в рамках детального анализа на местном или региональном уровне. Таким образом, местные органы власти не полностью информированы при принятии решений по энергетической политике в своей местности, поскольку планирование энергетики часто не рассматривается должным образом в рамках структур планирования на местном или региональном уровне. [6] . Кроме того, в существующих подходах к моделированию энергопотребления отсутствует пространственная информация для детального анализа на основе ГИС в различных масштабах.

Существует несколько проблем, связанных с внедрением многомасштабного моделирования энергопотребления зданий на основе ГИС, которые включают: (1) доступность данных, (2) несогласованность данных, (3) масштабируемость данных (4) интеграцию данных (5) геокодирование и, (6)) вопросы конфиденциальности данных. Данные об энергоэффективности здания обычно недоступны для всей пространственной области. Более того, из-за несоответствий в имеющихся крупномасштабных данных об энергопотреблении и отсутствия масштабируемых подходов к картированию энергопотребления зданий сохраняется разрыв между моделированием энергопотребления зданий и традиционными методами планирования [14] .. Заинтересованные стороны сталкиваются с проблемами масштабируемости из-за требования, чтобы энергетическое планирование осуществлялось на национальном уровне. Точно так же проблемы с интеграцией существуют при крупномасштабном картографировании ГИС для планирования и анализа, поскольку доступные данные разрежены, непоследовательны, разнообразны и неоднородны [15] . Имеющиеся данные не обеспечивают полного охвата и

имеют неизвестное качество. К сожалению, большинство данных обследований фонда зданий не имеют геокодирования для картографирования ГИС. Кроме того, конфиденциальность данных также является серьезной проблемой для детального картирования результатов ГИС [6] . Следовательно, требуется надежный подход к моделированию на основе ГИС, который помог бы прогнозировать энергетические характеристики всего фонда зданий с использованием ограниченных ресурсов для комплексного анализа решений.

В этом материале представлен обобщающий восходящий подход, основанный на данных, для многомасштабного картирования энергоэффективности жилых зданий на основе ГИС. В предыдущих исследованиях часто разрабатывались немасштабируемые фреймворки, подходящие для конкретного приложения. Методология, описанная в этом исследовании, является обобщаемой и масштабируемой и впредь может применяться к существующим доступным данным о фонде зданий. Кроме того, разработанная методология интегрирована с новым решением, основанным на данных, для поддержки геокодирования данных о строительных фондах для картографирования ГИС. Восходящий подход, основанный на данных, прогнозирует энергоэффективность здания, используя доступные ограниченные данные о фонде зданий. Кроме того, методология сравнивает различные алгоритмы контролируемого машинного обучения, чтобы определить оптимальную модель энергопотребления здания на основе данных для крупномасштабной реализации.

Материал включает в себя реализацию функции разработки и определяет оптимальные функции для разработки модели, управляемой данными, что значительно повышает точность модели. Кроме того, предлагаемый подход реализует подход пространственной агрегации для определения энергоэффективности на уровне микрорайона, района, города и округа. Методология дополнительно связывает прогнозируемые результаты с доступными пространственными ресурсами (социальными, экономическими или экологическими данными) для планирования и принятия решений с использованием подхода многокритериального анализа решений (MCDA). В целом, это исследование выводит новый комплексный подход, чтобы помочь местным органам власти анализировать потребление энергии жилым сектором и выбросы CO 2 .выбросы в различных географических масштабах, от местного до национального уровня. Это исследование демонстрирует применение методологии для жилищного фонда Ирландии.

Модели строительных фондов на основе ГИС можно эффективно использовать для разработки и оптимизации планирования устойчивой энергетики в масштабах городов. Моделирование на основе ГИС предполагает использование данных из различных источников. Связанные подходы к моделированию ГИС различаются в зависимости от доступных и необходимых данных, как описано в следующих разделах.

Моделирование данных на основе ГИС

Данные о здании, необходимые для энергетического моделирования, включают три основные категории, а именно смоделированные, эталонные и

измеренные данные. Смоделированные данные генерируются с помощью инженерных инструментов моделирования энергопотребления зданий, Эталонные данные можно получить из общедоступных наборов данных, доступных исследователям для сравнения результатов моделирования и проверки производительности моделей. Реальные данные собираются с помощью переписи, опроса, выставления счетов, счетчиков энергии и датчиков окружающей среды [8].. Моделирование на основе данных часто использует реальные данные. В категории реальных данных данные переписи включают статистические данные о фонде зданий в различных масштабах (местных, национальных и международных), в то время как данные обследований включают дополнительные выборочные исследования отдельных зданий в пределах определенной территории населения. Данные об электричестве здания и счетчиках могут быть доступны с разной степенью детализации временных рядов (частот измерения), например, поминутно, ежечасно, ежемесячно и ежегодно. Использование этих данных зависит от приложений, таких как прогнозирование, предсказание и оценка интенсивности использования энергии (EUI) [14] .

Методологии, используемые для сбора данных о реальном фонде зданий, различаются в зависимости от страны. Например, Министерство энергетики Соединенных Штатов поддерживает одну из крупнейших баз данных фонда зданий, Базу данных эффективности зданий (BPD), которая включает информацию о фонде жилых и коммерческих зданий [19] . Точно так же каждое государство-член в ЕС поддерживает свою собственную базу данных EPC, содержащую важную информацию об энергоэффективности здания о его фонде зданий [20] . Тем не менее, использование имеющихся данных для принятия решений часто является сложной задачей для акционеров (градостроителей, местных органов власти и лиц, определяющих политику в области энергетики), поскольку данные противоречивы, разнообразны, скудны и неоднородны [15] .

Имеющиеся данные для энергетического моделирования обычно имеют неполный охват и неадекватное качество. Например, любой набор данных Сертификата энергоэффективности (EPC) представляет только часть всего фонда здания. К сожалению, большинство данных опросов не геокодированы, а геокодирование — это процесс преобразования данных в формат, основанный на местоположении. Пользователи часто не следуют стандартному формату при сборе адресов зданий. Этот неструктурированный формат адреса вносит несоответствия в картографирование ГИС [21] .

Существует два различных способа геокодирования существующих наборов данных, а именно геокодирование интерфейса прикладного программирования (API) и подход, основанный на данных. API геокодирования — это коммерческая услуга, предоставляемая некоторыми ведущими картографическими компаниями, такими как Google, Институт экономических и социальных исследований (ESRI) и Bing. Однако эти службы плохо работают, если данные неструктурированы и неформатированы. Такие службы сопоставляют адрес на основе предопределенных описательных данных. Кроме того, эти услуги могут быть дорогостоящими для геокодирования

крупномасштабных наборов данных. С другой стороны, подход, основанный на данных, реализует алгоритмы сопоставления нечетких строк для геокодирования. Этот процесс полезен для основанных на опросе и противоречивых данных. Этот подход эффективно работает со сложными адресами и приоритетами для конкретных случаев. Например, даже если адрес не совпадает правильно, [21].

Существует ограниченное количество исследований, в которых используется геокодирование с использованием подхода, основанного на данных. Среди них большая часть исследований связана с повышением эффективности сопоставления адресов и, таким образом, не предоставляет обобщенного, масштабируемого решения для различных сценариев [22] . Существует несколько возможностей для расширения существующей работы по предварительной обработке адресов наряду с сопоставлением адресов [21] .

Энергетическое моделирование зданий на основе ГИС

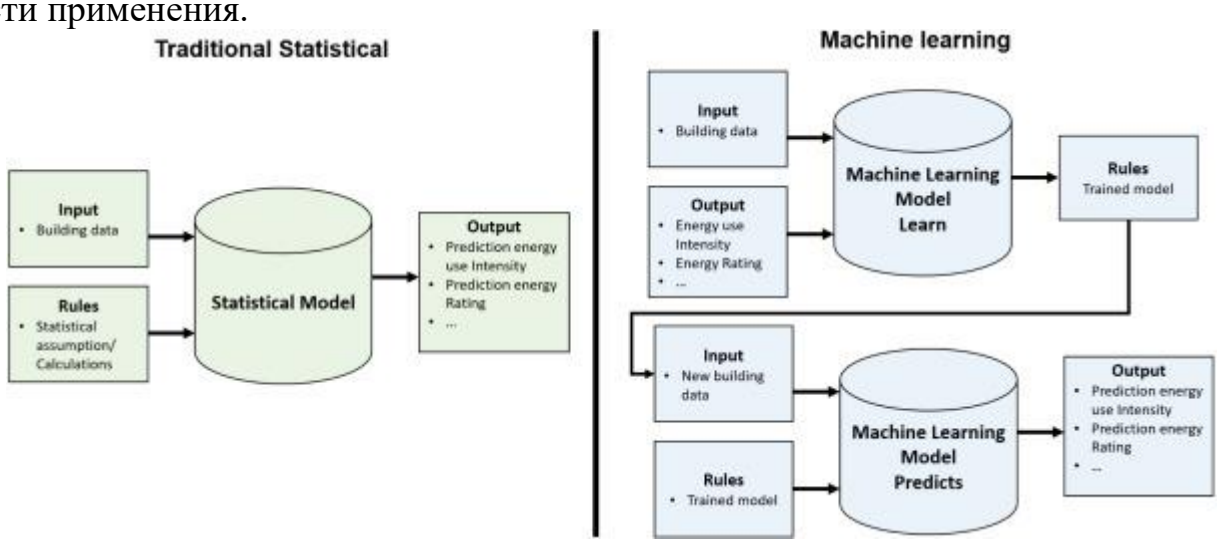
В крупномасштабном моделировании строительного фонда обычно используются два подхода, а именно инженерный подход и подход, основанный на данных [10] . Инженерный подход использует подробную физику здания для определения энергоэффективности. Эти инструменты часто требуют подробных данных о геометрических и негеометрических свойствах зданий; неспособность обеспечить точные входные данные может привести к неправильным результатам. Отныне для моделирования всего района потребуется огромное количество данных. Использование архетипов зданий упрощает этот подход за счет классификации строительного фонда с использованием репрезентативных зданий. В нескольких недавних проектах, основанных на моделировании городской энергии, использовался инженерный подход.

В этих исследованиях в основном используются инженерные методы с синтетическими экспериментальными данными (таблица 1). Поскольку инженерные методы, использующие архетипы, реализуют ограниченное количество типологий, при моделировании энергопотребления заложено множество допущений и неопределенностей. Эти допущения напрямую влияют на точность результатов и, следовательно, ограничивают надежность принятия решений в больших масштабах [8] .

Подходы, основанные на данных, с другой стороны, не требуют подробных знаний о здании, поскольку эти подходы оценивают энергоэффективность здания на основе исторических данных либо с использованием статистических моделей, либо моделей машинного обучения [50] . В то время как статистические модели используют выборочные данные о зданиях для построения математической зависимости между энергопотреблением здания и характеристиками [11] , модели машинного обучения реализуют алгоритмы, которые изучают данные для прогнозирования энергоэффективности здания с минимальными предположениями [51] .. Традиционная статистическая модель использует входные данные (данные о здании) и заранее определенные правила (статистические предположения/расчеты) для прогнозирования выходных данных, таких как

интенсивность использования энергии и рейтинг энергопотребления. С другой стороны, модель машинного обучения состоит из двух этапов. На первом этапе используются входные данные (данные о здании) и выходные данные (энергоемкость использования энергии и характеристики энергопотребления) для обучения модели обучения (обученная модель). На втором этапе эти правила (обученная модель) и входные данные модели (данные о новом здании) используются для прогнозирования выходных данных (рис. 1) [52] . Поскольку модели машинного обучения могут прогнозировать энергетическую эффективность с ограниченной информацией, эти подходы привлекли большое внимание в энергетическом секторе за последние несколько лет [12] .. Кроме того, эти подходы часто обеспечивают высочайший уровень точности с использованием доступных данных об энергопотреблении зданий [14] . Однако лишь в ограниченном числе исследований применяются подходы, основанные на данных, в различных масштабах с использованием моделей машинного обучения (таблица 1).

Существующие исследования по моделированию энергопотребления зданий на основе ГИС для сравнения масштабов, подходов, применения и области применения.



Методологические различия между традиционными методами статистического моделирования и моделирования машинного обучения.

Как правило, модели машинного обучения реализуют либо алгоритмы регрессии, либо алгоритмы классификации [12] . Алгоритмы регрессии оценивают реальные значения (числовые или непрерывные) выходных переменных, таких как потребление энергии. Наиболее распространенные алгоритмы регрессии включают линейную регрессию, деревья решений, случайный лес, глубокое обучение, обобщенные линейные модели, деревья с градиентным усилением и регрессию опорных векторов (SVR) [53] .. Алгоритмы классификации эффективны, когда выходная переменная представляет собой назначенную метку (дискретную или категориальную), такую как рейтинг энергопотребления или тип здания. Обычно используемые алгоритмы классификации включают ближайший сосед, наивный байесовский подход, обобщенную линейную модель, логистическую регрессию, глубокое

обучение, деревья решений, случайный лес, деревья с градиентным усилением и машину опорных векторов (SVM), индукцию правил и нейронные сети [54] .

В этом исследовании реализуется подход, основанный на данных, с использованием моделей машинного обучения для моделирования энергопотребления зданий в различных масштабах. Подход, основанный на данных, обеспечивает надежные результаты моделирования энергопотребления при наличии данных о запасах зданий. В основном исследования, основанные на данных об энергопотреблении зданий, сосредоточены либо на прогнозировании энергопотребления одного здания, либо на строительстве кластеров ограниченных типологий [43] . В этих исследованиях реализованы традиционные статистические модели, а именно линейная регрессия, множественная линейная регрессия, нелинейная регрессия и анализ условного спроса [55].. Большинство этих моделей полагаются на характер данных; предположения модели слишком строги и не отражают реальность. Чтобы преодолеть эти ограничения, модели машинного обучения используют такие методы, как предварительная обработка данных, выбор функций и перекрестная проверка, чтобы улучшить качество данных перед созданием системных моделей.

В нескольких исследованиях реализуется крупномасштабное моделирование энергопотребления зданий на основе ГИС с использованием моделей машинного обучения (таблица 1). Например, Ма и Ченг разработали систему для оценки энергоемкости здания в городском масштабе путем интеграции ГИС и технологии больших данных [26] . Точно так же в другом исследовании Контокоста и Талл сформулировали прогнозную модель, основанную на данных, для оценки энергопотребления зданий в масштабе города [49] .. Стоит отметить, что существующие исследования в основном сосредоточены на формулировании структуры городского масштаба, которая использует синтетические данные для создания моделей с ограниченным акцентом на моделировании ГИС. Например, Nutkiewicz, Yang и Jain разработали основу для интеграции инженерного моделирования (синтетических данных) и методов машинного обучения в многомасштабный рабочий процесс моделирования городской энергии [13] . Точно так же Аббасабади и Азари предложили структуру моделирования использования энергии в городах (UEUM) для моделирования энергии городского строительства и транспорта с использованием машинного обучения [43] . Существует несколько возможностей расширить предыдущую литературу путем введения обобщенной методологии многомасштабного моделирования.

многомасштабных энергетических характеристик жилых зданий с использованием подходов, основанных на данных.

Методология

Одной из серьезных задач для градостроителей и политиков является анализ и визуализация больших наборов данных и извлечение значимой информации из данных [6] . Моделирование на основе ГИС обеспечивает основу для сбора, управления и анализа крупномасштабных данных в

географическом контексте. Таким образом, моделирование и планирование энергопотребления зданий на основе ГИС помогает собирать, хранить и визуализировать подробную информацию [6] . Следовательно, обобщенная методология на основе ГИС позволила бы проводить широкий спектр анализов, тем самым помогая заинтересованным сторонам максимально использовать аналитические возможности методов энергетического планирования и моделирования [4] .

Разработанный подход учитывает энергоэффективность здания на основе ГИС в нескольких масштабах. Картографирование энергоэффективности многомасштабных жилых зданий на основе ГИС состоит из семи этапов (рис. 2).

Начальный этап включает сбор данных из различных источников (строительный фонд, перепись населения, ГИС и географические данные);

Следующий шаг посвящен геокодированию данных о фондах зданий;

Этап предварительной обработки и выбора объектов следует процедуре геокодирования и использует подходы, основанные на данных, для улучшения качества данных о фондах зданий;

На следующем этапе, разработке архетипов зданий, используются предварительно обработанные данные фонда зданий для определения архетипов, представляющих фонд зданий;

Этап разработки модели, основанной на данных, прогнозирует энергетическую эффективность здания в больших масштабах с использованием восходящего подхода;

Шаг многомасштабного картографирования ГИС отображает результаты энергоэффективности здания; а также

Наконец, на этапе энергетического планирования анализируются результаты моделирования для планирования или принятия решений. На этом этапе анализируются и определяются приоритетные области для реализации долгосрочных и устойчивых решений, связанных с энергетикой.

В следующих разделах более подробно описываются отдельные этапы методологии.

Сбор информации

В процессе сбора данных собираются наборы данных, необходимые для картирования ГИС. Эти наборы данных включают перепись населения, географические данные, геометрию зданий и информацию, не связанную с геометрией. Процесс сбора данных дополнительно объединяет данные из этих ресурсов, которые могут быть представлены в качестве вспомогательного средства визуализации для обоснования решений в области энергетической политики. В больших масштабах существующие базы данных зданий часто являются основным источником информации о фонде зданий. Эти существующие базы данных запасов зданий могут быть представлены в форме базы данных энергетических паспортов зданий, которая содержит как геометрическую, так и негеометрическую информацию. Геометрические данные состоят из информации о форме здания, типе здания, структуре здания, количестве этажей и соотношениях окон и стен. Негеометрические данные

здания включают в себя значения коэффициента теплопередачи ограждающих конструкций, строительные узлы, и свойства систем отопления, вентиляции и кондиционирования воздуха (HVAC). Кроме того, прогнозирование энергопотребления также зависит от показателей энергоэффективности здания. Данные EPC обычно предоставляют обзор геометрической и негеометрической информации в дополнение к показателям эффективности здания. Кроме того, набор данных включает количественные данные о зданиях (данные национальной статистики или переписи населения), необходимые для определения количества зданий, присутствующих в определенной области.

Точно так же для моделирования трехмерных данных ГИС требуются данные о площади здания и высоте здания. Наиболее подходящей моделью стандартного формата является формат геопространственных векторных данных, также известный как шейп-файл. Данные контура здания и границ обычно доступны в формате шейп-файла, который содержит точки, линии и полигоны. Эти данные могут быть собраны для желаемой области из OpenStreetMap или национального географического обзора, которые содержат географические данные достаточного качества. Высота здания может быть сформулирована как произведение количества этажей и средней высоты здания для конкретной местности [31] . Данные светового обнаружения и определения дальности (ЛИДАР) также можно использовать для определения высоты здания. Однако данные LIDAR часто недоступны на районном уровне [31] .. Наконец, для процесса геокодирования требуется национальная географическая база данных, содержащая адрес здания с пространственной информацией.

Геокодирование

Эта процедура следует за сбором данных и включает геокодирование данных о фондах зданий. В этом исследовании реализован подход, основанный на данных, который использует базовые базы данных и национальные географические базы данных (рис. 3). Частично геокодированные данные фонда зданий дополняются набором географических данных, включающим геокодированные адреса фонда жилых зданий. Чтобы эффективно сократить пространство поиска, этот процесс сегментирует набор данных на основе городов и округов. Сегментация повышает точность поиска, гарантируя, что алгоритмы нечетких строк используют только области поиска, в которых находится адрес.

Поскольку данные, собранные с помощью опросов, таких как данные EPC, обычно содержат нерелевантную, неполную, зашумленную, избыточную и противоречивую информацию, этот процесс реализует процедуру предварительной обработки адреса. Расхождения в основном возникают, когда пользователь не следует стандартной процедуре сообщения адресов. Предварительная обработка адресов устраняет эти несоответствия, используя очистку данных и преобразование данных перед реализацией алгоритмов сопоставления адресов. Очистка данных удаляет и заменяет неправильные, неполные, повторяющиеся и неструктурированные адреса соответствующими словами, что повышает производительность алгоритма

сопоставления нечетких строк. Процесс очистки данных касается орфографических ошибок, пропущенных пробелов, неправильных типов, аббревиатур, синонимов, лишних слов, звуков и переставленных букв (Таблица 2).). В процессе очистки адресов используется словарь очистки данных, который содержит список предопределенных неполных или нерелевантных слов вместе с возможностью их замены. Наконец, в процессе преобразования данных из адресов извлекается такая информация, как улица, район, город и почтовый индекс. Этот процесс дополнительно помогает процессу фильтрации адресов.

Процесс геокодирования дополнительно реализует фильтрацию адресов на нескольких уровнях, а именно, номер дома/квартиры, улица и небольшой район (скопление зданий рядом с улицей). В процессе геокодирования используются алгоритмы сопоставления нечетких строк для сравнения адресов с существующими доступными национальными базами данных геокодированных адресов. В этом исследовании сравниваются четыре различных алгоритма нечеткого сопоставления, включая Яро, Яро-Винклера, Левенштейна и Жаккара, на основе оценки совпадения. Все эти алгоритмы сопоставления строк хорошо зарекомендовали себя при сопоставлении сложных строк, основываясь на существующей литературе [56] , [57] . Эти алгоритмы могут быть математически сформулированы с использованием уравнений. (1) , (2) , (3) соответственно.(1)simДж=0еслиm=013m|s1|+m|s2|+m−т m,иначекуда|si|длина строкиsi;mколичество совпадающих символов;тсоставляет половину числа транспозиций. Два персонажа изs1a такжеs2соответственно, считаются совпадающими только в том случае, если они совпадают и не дальше, чемМаксимум(|s1|,|s2|)2−1. Количество транспозиций определяется как половина совпадающих символов, которые не находятся в одном и том же индексе.(2)simДжВ=simДж+ℓp1−simДжкудаsimДжподобие Яро для строкs1a такжеs2;ℓдлина общего префикса в начале строки до четырех символов;рявляется постоянным коэффициентом масштабирования того, насколько оценка корректируется вверх из-за наличия общих префиксов.рне должно превышать 0,25, иначе сходство может стать больше 1. Стандартное значение этой константы в работе Винклера равнор=0.1.

Примечание. Приведенные выше записи представляют собой фиктивные адреса. Исходные адреса не используются из соображений конфиденциальности.

Расстояние Яро – ВинклерадВопределяется какдВ=1−simДжВ.(3)яА такжеva,бi,Дж=maИксi,Джеслиміні,Дж=0міняА такжеva,бi−1,Дж+1яА такжеva,бi,Дж−1+1яА такжеva,бi−1,Дж−1+kиначекудаk=0, еслиаi=бi, 1 иначе.яА такжеva,бэто расстояние между первымиперсонажиаи первыйДжперсонажиб.

Jaccard — это алгоритм сопоставления строк на основе токенов. Расчет заключается в том, чтобы найти количество общих строковых токенов и разделить его на общее количество уникальных строковых токенов. Это выражается в математических терминах в формуле. (4) .(4)Дж(А,Б)=|А∩Б||А∪Б|=|А∩Б||А|+|Б|−|А∩Б|где числитель — это пересечение (общие строковые токены), а знаменатель — объединение (уникальные строковые токены).

Процесс сопоставления нечетких строк сопоставляет адрес на основе двух уровней. Первоначально этот процесс сопоставляет адреса на основе номеров домов/квартир на уровне отдельных зданий. При отсутствии номеров домов/квартир процесс сравнивает адреса на основе названий улиц на уровне района. Процесс сопоставления присваивает баллы от 0 до 1 различным алгоритмам сопоставления строк. Затем эти оценки определяют критерии наименьшего совпадения, которые следует рассматривать как геокодированный адрес. Наименее совпадающие критерии можно определить вручную, используя образец набора данных. Эти геокодированные адреса затем сохраняются в базе данных фонда жилых домов.

При выборе пространственной проекции часто используются различные опорные координаты, определяющие расположение отдельных зданий на складе. Системы отсчета географических координат (CRS) определяют опорные точки пространственной проекции x, y на поверхности земли, такие как значения долготы и широты. Общие картографические проекции, используемые в настоящее время, включают универсальную поперечную меркаторскую (UTM) и военную сеточную справочную систему (MGRS). Национальная географическая база данных обычно содержит ссылки на пространственную проекцию (координаты x, y) для адресов. Поэтому в этом исследовании считается, что система отсчета координат аналогична той, которая используется в национальной географической базе данных при геокодировании адресов [21] .

Предварительная обработка строительного материала

Предварительная обработка строительного фонда включает четыре последовательных этапа, а именно: статистический анализ, предварительную обработку данных, обнаружение выбросов и выбор признаков (рис. 4). Статистический анализ помогает сделать первоначальные выводы и выводы из данных. Этот анализ включает в себя выполнение арифметических операций (среднее, медиана и мода) и последующие визуальные представления (гистограммы, графики плотности, диаграммы). Предварительная обработка данных включает преобразование реальных или необработанных данных в понятный формат. Во время предварительной обработки данные проходят ряд операций, таких как очистка данных, интеграция данных, сокращение данных, преобразование данных и дискретизация данных [14] .

Обнаружение выбросов или аномалий является важным шагом перед реализацией алгоритма обучения. Выбросы — это точки наблюдения, находящиеся на ненормальном расстоянии от большинства других значений в пространстве выборки данных. Как правило, процедура обнаружения выбросов реализует методы на основе расстояния, плотности и локального фактора выбросов (LOF) [14] .

Процесс выбора признаков определяет подмножество наиболее релевантных переменных или атрибутов для представления архетипа и разработки модели обучения. Этот процесс удаляет нерелевантные, избыточные и менее важные функции, которые не влияют на производительность модели обучения, и, таким образом, уменьшает входную размерность, сложность и вычислительную нагрузку модели обучения [14] .

Выбор функций, рассматриваемый как одна из основных концепций машинного обучения, которая сильно влияет на точность обучения, обычно использует инженерные методы или методы, основанные на данных. Инженерные методы используют инженерную оценку и существующие практики в литературе [58] . Методы, управляемые данными, используют различные статистические подходы для разработки моделей обучения [14] . Как правило, методы отбора на основе данных идентифицируют и ранжируют функции на основе нескольких статистических тестов, таких как прирост информации, порог дисперсии / стандартного отклонения, коэффициент корреляции и тесты хи-квадрат [59].. Например, коэффициент корреляции отфильтровывает те признаки, которые точно отражают целевой признак. Точно так же пороговое значение дисперсии/стандартного отклонения фильтрует функции, которые имеют самые или очень разные значения. В этом исследовании используются как инженерные методы, так и методы, основанные на данных, для определения подмножества наиболее важных функций. На первом этапе инженерный метод определяет оптимальные характеристики на основе существующих исследований. На следующем этапе метод выбора на основе данных идентифицирует функции с помощью нескольких статистических тестов. Однако тип выбора признаков зависит от общего количества признаков и качества данных.

Разработка архетипов зданий

Создание архетипов требует двух основных подэтапов, таких как сегментация и характеристика. Процесс сегментации определяет количество архетипов зданий, необходимых для представления фонда жилых зданий в различных масштабах. Существуют различные критерии сегментации строительного фонда, например, тип здания, год постройки, климатическая зона или пространственная информация [7] .

Процесс характеристики определяет физические свойства каждого архетипа здания, такие как строительная ткань, система отопления, освещение и оборудование для горячего водоснабжения [8] . Этот процесс оценивает значения характеристик архетипа здания на основе критериев сегментации с использованием подхода, основанного на данных. Критерии сегментации группируют данные, а затем выполняют операцию агрегирования для каждого кластера, чтобы получить свойства каждого архетипа. Агрегирование может быть выполнено путем применения арифметических или геометрических математических операций (среднее, медиана или мода). Полученное агрегированное значение представляет характеристики одного архетипа здания.

Это исследование создает архетипы на местном уровне, а не на уровне страны или города для детального анализа с использованием подхода среднего виртуального здания на основе статистических данных о здании. Таким образом, на локальном уровне сформулированные архетипы представляют весь кластер зданий в этой местности, что помогает в формулировании всех данных о фонде зданий. Локальные архетипы дают двойное преимущество. Во-первых, эти архетипы решают проблему доступности данных для моделирования. Во-вторых, архетипы локального уровня косвенно решают вопросы

конфиденциальности данных, используя небольшие области для картографирования ГИС и разработки моделей.

Разработка энергетической модели здания на основе данных

Разработка крупномасштабной модели машинного обучения для повышения энергоэффективности здания с использованием подхода, основанного на данных, требует нескольких этапов (рис. 5). Процесс разработки модели использует предварительно обработанные данные о строительных фондах и начинается с разделения данных в целях обучения и тестирования, после чего следует реализация алгоритмов обучения. Затем процесс анализирует эффективность разработанной модели обучения.

Разделение данных — это процесс разделения набора данных на наборы для обучения и тестирования. Набор обучающих данных — это подмножество данных, которое используется для разработки обученной модели. Набор данных тестирования — это подмножество, которое оценивает модель для оценки беспристрастной конечной производительности моделей. Наиболее распространенным подходом к разделению данных является использование случайной выборки данных, которая случайным образом разбивает данные на 80–20% для обучения и тестирования соответственно [14] .

Процесс разработки модели машинного обучения реализует алгоритмы классификации для формулировки модели обучения. Классификация является частью контролируемого алгоритма машинного обучения, который предсказывает класс данного набора точек данных; классы также известны как метки или категории. В этом исследовании используются восемь различных алгоритмов классификации для прогнозирования энергии, а именно: Наивный Байес, Обобщенная линейная модель, Логистическая регрессия, Глубокое обучение, Деревья решений, Случайный лес, Деревья с градиентным усилением и Машина опорных векторов. Эти восемь алгоритмов обеспечивают превосходную производительность при использовании для классификации, прогнозирования или прогнозирования энергопотребления, как видно из предыдущих исследований [12] , [53] .

Оценка модели проверяет эффективность моделей классификации. Некоторые из показателей оценки включают ACCuracy (ACC), точность, полноту (уравнения (7) , (5) , (6)) и время выполнения [53] .(5)ppA
такжеcisiОн=TPTP+ФР(6)pA
такжесаяя=TPTP+ФН(7)ACC=TP+THTP+ФР+ФН+THгде Истинные Положительные (TP) - это случаи, которые предсказаны положительно и на самом деле верны. Правда отрицательные (TH) - это случаи, которые предсказаны отрицательно, но верны. Ложные срабатывания (ФР) — это случаи, которые прогнозируются положительными, но на самом деле являются ложными. Ложноотрицательные (ФН) - это случаи, которые предсказаны как отрицательные, но ложные

Как правило, производительность алгоритма классификации можно оценить с помощью точности. Точность представляет собой отношение правильно предсказанных наблюдений к общему количеству наблюдений. Большие значения точности означают лучшую

производительность модели. Однако точность дает неправильные результаты для несбалансированных выходных меток. Поэтому в этом исследовании в дополнение к точности рассматриваются альтернативные показатели производительности, такие как показатели точности и полноты. Точность представляет количество положительных предсказаний класса/метки, которые фактически принадлежат положительному классу/метке. Точно так же отзыв — это число положительных классов, предсказанное из всех положительных результатов в наборе данных. Точность или отзыв можно использовать в виде матрицы путаницы, которая показывает общую сводку производительности результатов прогнозирования класса. Матрица путаницы представляет собой определенный макет таблицы, который обеспечивает визуализацию производительности алгоритма классификации. В этом исследовании также учитывается время вычислений для разработки модели обучения, поскольку процесс разработки модели должен быть эффективным для крупномасштабных данных о строительном фонде. Наконец, наиболее обученная модель обучения, основанная на показателях эффективности, в дальнейшем используется для прогнозирования энергоэффективности здания для всего фонда здания.

Многомасштабное ГИС-моделирование

Этот процесс отображает результаты энергоэффективности здания в нескольких масштабах ГИС, от уровня отдельного здания до национального уровня. Из-за ограниченной доступности данных по отдельным зданиям в национальном масштабе модель обучения прогнозирует энергетические характеристики здания с использованием архетипов зданий. Поэтому в модели обучения используются входные признаки из разработанных архетипов зданий на локальном уровне в разделе 3.4.. Эти архетипы зданий помогают генерировать данные об отдельных зданиях в национальном масштабе. Это исследование разрабатывает архетипы зданий в масштабе небольшой области/района для проведения мелкозернистого анализа. Процесс многомасштабного ГИС-моделирования на основе данных использует концепцию восходящего подхода для моделирования всего фонда зданий. Процесс моделирования состоит из двух основных этапов, а именно моделирования здания и многомасштабного моделирования (рис. 6).

На первом этапе реализуется моделирование ГИС на уровне здания с использованием разработанной модели обучения (описанной в разделе 3.5).). Процесс начинается со сбора входных данных объектов для всего фонда зданий. Этот процесс извлекает входные характеристики из нескольких источников, таких как архетипы зданий, географические данные и данные переписи, и передает эти функции в наиболее подготовленную модель обучения для прогнозирования энергоэффективности здания. Данные архетипа здания помогают собрать значения входных признаков для всего фонда здания. Входные функции включают исходные функции, используемые для создания лучшей обучающей модели. Географические данные или данные переписи включают количественные данные о зданиях (количество зданий) в каждом географическом масштабе. На следующем этапе лучшая модель обучения прогнозирует энергетические характеристики здания для всего фонда

здания. Наконец, спрогнозированные результаты используются для 2D- или 3D-моделирования ГИС каждого жилого дома. 2D-моделирование ГИС требует, чтобы площадь здания отображала энергетические характеристики здания. Трехмерное ГИС-моделирование выполняется путем выдавливания контура здания через данные о высоте здания.

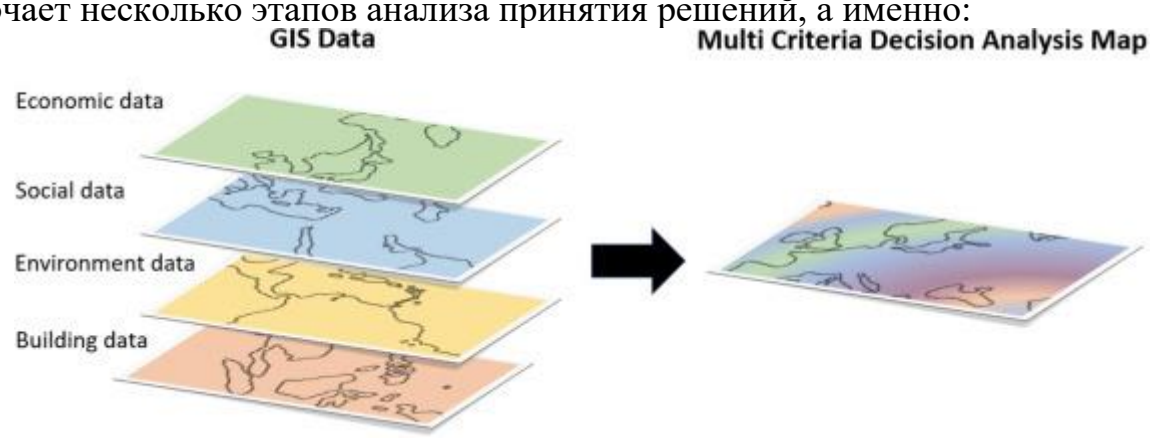
На втором этапе результаты прогнозирования энергоэффективности в масштабе здания могут быть дополнительно распространены на несколько масштабов, таких как небольшие районы, районы, города и округа. В процессе используется концепция «снизу вверх» для агрегирования результатов на уровне здания на более высокий географический уровень. Таким образом, подход пространственного объединения или агрегации используется для моделирования ГИС с несколькими масштабами. Пространственное объединение — это ГИС-операция, которая агрегирует данные из одного географического слоя в другой с пространственной точки зрения. Для отображения пространственного соединения в большом масштабе требуется файл форм для каждого масштаба. Процесс пространственного объединения включает в себя отдельные здания, небольшие участки и кварталы. Агрегированные отдельные строения представляют собой небольшие участки. Все постройки на небольшой территории объединены в районы. На аналогичной ноте прогнозы на районном уровне можно использовать для моделирования городов или округов. Наконец, процесс объединяет спрогнозированные 2D- и 3D-слои здания, чтобы сформулировать полную карту для планирования и анализа энергопотребления здания.

Энергетическое планирование и анализ зданий

Этот процесс реализует сопоставление результатов энергетического планирования. Созданные карты помогают заинтересованным сторонам анализировать и определять приоритетные области для реализации энергоэффективных стратегий. Результаты могут быть дополнительно использованы для определения областей, в которых лица, ответственные за энергетическую политику, могут проводить целевые общественные мероприятия/кампании для увеличения активности по модернизации. По сравнению с широкомасштабными массовыми кампаниями, целевые кампании по модернизации на базе местных сообществ с большей вероятностью будут успешными в увеличении активности по модернизации в районе [60] .

Поскольку интеграция данных из различных ресурсов представляет собой серьезную проблему для крупномасштабного картографирования ГИС, в этом процессе также реализуется основанный на ГИС подход к многокритериальному анализу решений (MCDA) для поддержки принятия сложных решений с использованием нескольких источников данных анализа решений [9] . Подход MCDA помогает лицам, принимающим решения, принять наилучшее возможное решение с учетом множества критериев. Кроме того, этот подход полезен, когда различные заинтересованные стороны имеют конфликтующие цели, задачи и интересы. Подход MCDA на основе ГИС обычно используется для оценки потенциала возобновляемых источников энергии, управления отходами, лесного хозяйства, сельского хозяйства и сектора окружающей среды [62] .. В этом

исследовании результаты прогнозирования энергоэффективности здания на основе ГИС могут быть интегрированы с данными анализа решений, такими как социальные, экономические или экологические данные, для принятия комплексных решений в больших масштабах (рис. 7). MCDA на основе ГИС включает несколько этапов анализа принятия решений, а именно:



1.Определите проблему и поставьте цель или задачу для анализа принятия решения. Цель может быть связана со стратегической политикой или результатами проекта более высокого уровня, такими как экономическое, социальное или устойчивое развитие;

2.Собирайте данные для принятия решений и прогнозируйте результаты в формате слоев ГИС на основе цели MCDA. Данные анализа пространственного решения могут быть получены из национальной переписи населения или пространственной базы данных;

3.Определите соответствующие пороговые значения для критериев принятия решений или факторов для каждого пространственного слоя. Это может быть получено из мнений экспертов, заинтересованных сторон или из существующей литературы по соответствующим областям. Критерий каждого уровня должен поддаваться измерению, чтобы отражать выполнение отдельных задач;

4.Стандартизируйте или преобразуйте слои критериев в относительную шкалу. Этот процесс позволяет сравнивать каждый из уровней критериев и экспертные знания с значимыми оценками;

5.Определите вес (в процентах) каждого критерия на основе его приоритета, важности и цели. Как правило, метод аналитической иерархии (AHP) используется для определения веса каждого слоя [63] . AHP — это метод попарного сравнения, в котором для оценки веса используется опыт экспертов или заинтересованных сторон. Кроме того, такой метод дает как экспертам, так и заинтересованным сторонам равные возможности внести свой вклад в определение качественной и количественной значимости каждого слоя;

6.Агрегируйте или объедините сгенерированные слои на основе критериев с определенными весами. Окончательная агрегированная карта по многим критериям, разработанная с использованием метода взвешенной линейной комбинации (WLC), может использоваться для получения индекса пригодности (приоритета) или оценки.Sакаждой областиаследующим образом (уравнение (8)) [64] :(8)Sa=∑i=1нBiИксікудаBiрасчетный вес критериевікак

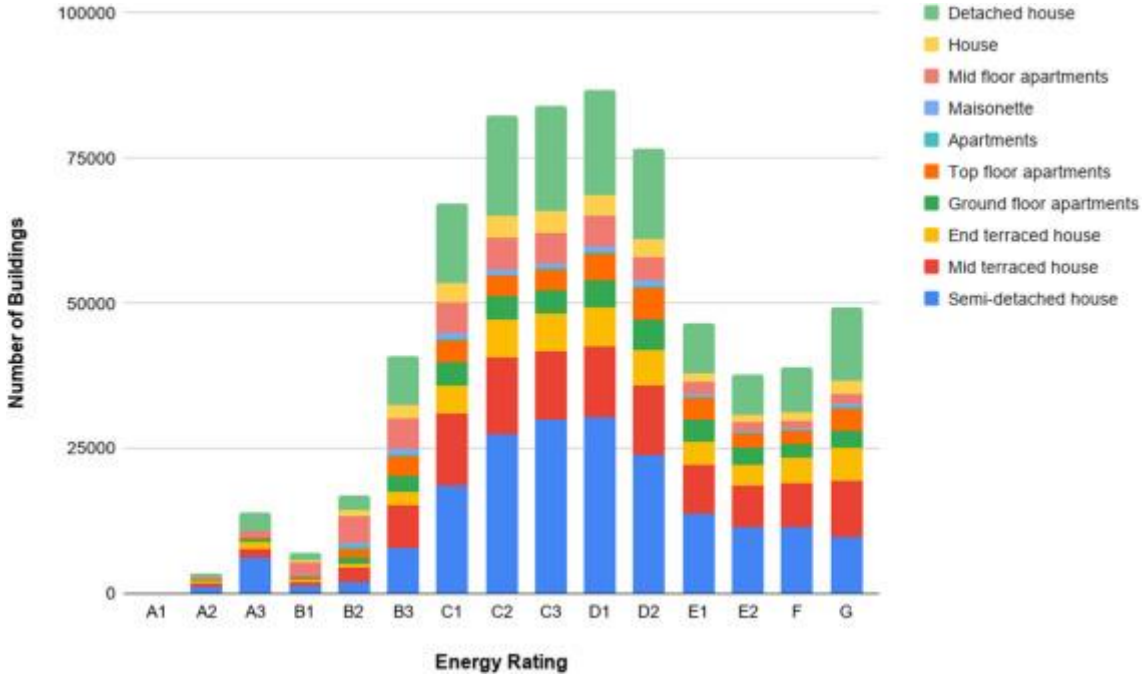
определено на шаге 5;Иксiэто оценка области,ав отношенииiкритерии, определенные на шаге 4, иi=1,2,…,нгде n — общее количество взвешенных критериев; а также

7.Проверьте и проанализируйте окончательную карту ГИС.

Таблица 3 . Требования к формированию данных и связанные с ними источники данных для примера из Ирландии.

Тематическое исследование

Основная цель данного тематического исследования – разработать методологию расчета энергоэффективности зданий на основе ГИС для всего фонда зданий Ирландии. Методология объединяет подход, основанный на данных, с восходящим моделированием для прогнозирования (оценки) энергетических характеристик здания в различных масштабах с использованием пространственной информации. Это исследование демонстрирует применение разработанного подхода на примере фонда жилых домов Ирландии. Географический масштаб рассматривается в Ирландии на нескольких уровнях, включая округа, города, избирательные округа и небольшие районы. Это позволяет проводить анализ энергоэффективности здания при различных пространственных разрешениях. В этом исследовании предлагается основанная на ГИС структура для многомасштабного картографирования энергоэффективности жилых зданий, которая может выступать в качестве инструмента визуального анализа для разработчиков политики в области энергетики.



Скачать : Скачать изображение в высоком разрешении (319 КБ)

Скачать : Скачать полноразмерное изображение

Рис. 8 . Распределение данных EPC в Ирландии указывает на то, что значительная часть рейтингов энергопотребления зданий находится в диапазоне от C1 до D2, при этом сдвоенные и отдельно стоящие дома представляют собой самый высокий процент типов домов.

Сбор информации

Сбор данных о фонде зданий городского масштаба является довольно сложной задачей, поскольку информация об отдельных зданиях часто недоступна. Процесс сбора данных включает в себя получение необработанных данных о фонде зданий из различных источников, а именно, набора данных EPC, набора данных переписи зданий, данных о площади здания, географических данных здания, данных ГИС (файлы формы небольших областей, районов, городов) и данных из энергетики. программы эффективности, администрируемые Управлением по устойчивой энергетике Ирландии (таблица 3).

Поддерживаемый Управлением по устойчивой энергетике Ирландии (SEAI), набор данных EPC (также называемый сертификатом Building Energy Rating (BER)) жилого фонда Ирландии представляет измеренный фонд зданий и включает более 200 характеристик зданий, включая строительную ткань, системы отопления, оценка конечного использования CO2выбросы, предполагаемые поставки и предполагаемое потребление первичной энергии. Ирландский набор данных EPC содержит энергетический рейтинг для каждого здания, который ранжирует энергетические характеристики здания по градуированной шкале от G до A1 на основе расчетного потребления энергии на квадратный метр в год [65] . Ирландский набор данных EPC содержал около 695 000 жилых зданий (на конец 2019 г.), при этом основная доля рейтингов зданий находится в пределах C1 и D2, при этом самый высокий процент типов зданий приходится на сблокированные и отдельно стоящие дома (рис. 8).

Ирландская перепись, которая проводится каждые четыре года Центральным статистическим управлением (ЦСУ), собирает ряд точек данных о здании, в котором живет респондент. Таким образом, перепись обеспечивает количество зданий в каждом географическом районе [68] . Согласно набору данных CSO 2016, в Ирландии насчитывается около 1 983 715 жилых домов, в отличие от набора данных EPC, который состоит из 695 000 жилых домов. Это говорит о том, что данные EPC доступны только для≈39% жилого фонда [70] . В этом исследовании используются алгоритмы машинного обучения для прогнозирования энергетического рейтинга оставшихся 61% акций с использованием ограниченных переменных.

База данных GeoDirectory содержит географическую информацию обо всем строительном фонде Ирландии [66] . Поскольку процесс картирования ГИС требует геокодирования зданий, метод геокодирования преобразует базу данных зданий EPC с использованием базы данных GeoDirectory. Эта база данных, опубликованная An Post (Ирландская почтовая служба) и Ordnance Survey Ireland, содержит геокодированные адреса 2 014 357 жилых зданий.

Набор данных схемы модернизации жилья в Ирландии содержит количественные данные для жилых зданий, которые завершили энергетическую модернизацию в рамках одной из программ SEAI. Домовладельцы обращаются в SEAI за грантами, которые субсидируют стоимость их обновлений. Набор данных, поддерживаемый SEAI, включает 265 182 модернизированных здания и включает дома, которые были модернизированы в рамках одной из программ

повышения энергоэффективности SEAI, таких как Better Energy Homes, Warmer Homes, Better Energy Communities и пилотная программа Deep Retrofit [69] .

В этом исследовании используется многомасштабная концепция для картографирования ГИС для отдельных ирландских зданий, небольших территорий, районов, городов и округов. Каждый небольшой район представляет собой группу зданий, а группа небольших районов составляет один район. Процесс сопоставления сопоставляет прогнозируемый энергетический рейтинг здания с фондом здания. Согласно информации, полученной от CSO, Ирландия состоит из 26 административных округов, 5 городов, 139 муниципальных районов и 18 641 небольшого района с более чем двумя миллионами жилых зданий. Площадь застройки и границы небольших территорий, районов, городов и округов получены из Ordnance Survey Ireland [67] .. Данные о высоте отдельных зданий, опубликованные Школой географии Дублинского университетского колледжа, доступны только для жилого фонда Дублина. Стоит уточнить различия в структурах картографирования 2D и 3D ГИС, реализованных в этом примере. 2D-карта представляет энергетическую эффективность здания на всех мультимасштабных уровнях для Ирландии. С другой стороны, трехмерная карта представляет только энергетические характеристики здания на уровне города Дублин.

Таблица 4 . Пример алгоритма сопоставления нечетких строк для объяснения критериев геокодирования для ирландских адресов.

Примечание. Приведенные выше записи представляют собой фиктивные адреса. Исходные адреса не используются из соображений конфиденциальности.

Геокодирование

Отсутствие геокодированных данных создает серьезную проблему при реализации картографирования ГИС. Поскольку ирландский набор данных EPC не включает геокодированные адреса, в процессе геокодирования каждому жилому зданию в наборе данных EPC присваивается геокод с использованием современного алгоритма программирования на основе Java (алгоритм 1). Поскольку обработка всего ирландского фонда зданий требует огромных вычислительных ресурсов, в этом исследовании применяется метод параллельного программирования, который использует несколько процессов для сокращения времени вычислений. В процессе геокодирования используются два набора данных, а именно: набор данных EPC жилых зданий Ирландии (содержит адреса для геокодирования) и база данных Irish GeoDirectory (содержит геокодированные адреса жилых зданий). В этом тематическом исследовании База данных GeoDirectory содержит три разные географические системы координат, которые назначают уникальные опорные проекции каждого здания, а именно: ирландскую сетку (восток, север), ирландскую поперечную меркаторскую (восток, север) и ETRS89 (долгота, широта). Затем процедура геокодирования сегментирует данные по городам и округам. Крайне важно внедрить метод предварительной обработки вышеупомянутых наборов данных, поскольку процесс сбора данных EPC выполняется вручную, а в данных отсутствуют геокодированные характеристики (долгота и широта). Кроме того, оценщики EPC могут следовать или не следовать стандартной процедуре для

заполнения географической информации, такой как адрес, почтовый индекс. Предварительная обработка адресов устраняет эти несоответствия до того, как данные можно будет использовать для реализации любых алгоритмов сопоставления адресов.

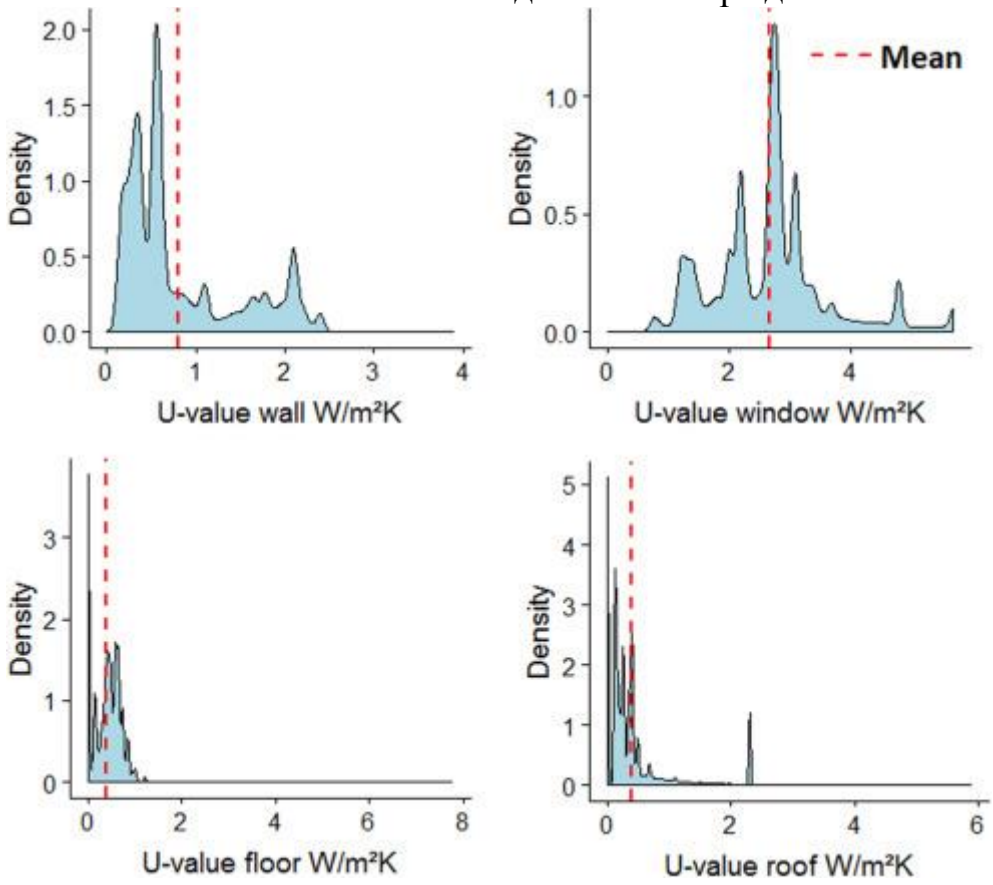
Процесс геокодирования использует обработанный набор данных EPC для реализации алгоритмов нечеткого сопоставления для сопоставления строк. В этом исследовании сравниваются четыре различных алгоритма нечеткого сопоставления, а именно: Яро, Яро-Винклера, Левенштейна и Жаккара. Процесс сопоставления строк фильтрует и сравнивает адреса на двух уровнях. Первый уровень сравнивает адреса EPC, которые содержат номер дома или квартиры, со всеми адресами в базе данных GeoDirectory на уровне отдельного здания. Второй уровень сравнивает те EPC-адреса, которые не содержат номеров домов или квартир, с ближайшими небольшими областями в базе данных GeoDirectory. Затем процесс сопоставления строк сравнивает алгоритмы на основе оценки совпадения, которая определяет критерии наименьшего совпадения для адресов геокодирования в наборе данных EPC (Таблица 4).). Результаты показывают, что минимальные оценки соответствия для алгоритмов Яро-Винклера, Яро, Жаккара и Левенштейна составляют 0,90, 0,80, 0,50 и 0,50 соответственно для сравнений на уровне зданий (таблица 5). Аналогичным образом, минимальные оценки соответствия для алгоритмов Яро-Винклера, Яро, Жаккара и Левенштейна составляют 0,85, 0,75, 0,40 и 0,40 соответственно для сравнений на уровне небольших областей (таблица 5). Наконец, результаты сопоставления адресов сохраняются в базе данных для картографирования ГИС.

Предварительная обработка материала

Процедура предварительной обработки запасов зданий извлекает характеристики зданий в Ирландии и связанное с ними энергопотребление с использованием методов, управляемых данными. Эти методы включают начальный статистический анализ, предварительную обработку данных, обнаружение выбросов и методы выбора признаков.

Первоначальный статистический анализ графиков плотности для значений U крыши и пола показывает, что весь спектр значений U содержит значительное количество нулей (рис. 9). Этап предварительной обработки данных устраняет эти несоответствия; средние значения для объектов (определенные с помощью кластеризации типа здания) отсутствующие значения в наборе данных. Предварительная обработка данных также включает фильтрацию и преобразование данных. В то время как фильтрация данных удаляет ненужные экземпляры данных, преобразование данных преобразует все категориальные и номинальные значения в числовые значения, поскольку методы, управляемые данными, обычно обрабатывают числовые значения. Кроме того, метод преобразования данных уменьшает несколько комбинаций рейтинговых классификаторов (например, A, B, C, D и EFG) из существующих рейтинговых меток (A1, A2, …, E, F, G). [48] . Классификаторы генерируют кластеры смежных рейтингов энергии. Например, классификатор с меткой EFG включает в себя отдельные рейтинговые метки E, F и G. Это делается для того, чтобы

определить, повлияет ли уменьшение количества классификаторов на эффективность модели обучения, используемой при прогнозировании энергетического рейтинга здания. Кроме того, десять типов жилых зданий в наборе данных EPC объединены в пять основных, а именно: квартиры (верхние, средние, цокольные, мезонеты), двухквартирные дома, отдельные дома, террасные (средние или торцевые) дома и бунгало (дома). Это важно для обеспечения согласованности типов зданий в наборе данных GeoDirectory.



U-значения (Вт/м2K) для города Дублин, состоящего из более чем 250 000 жилых зданий, со средним значением для иллюстрации исходного статистического анализа.

Таблица 6 . Выбранные объекты и их типы образуют набор данных EPC Ирландии для разработки модели на основе данных.

В этом исследовании используется алгоритм LOF для удаления выбросов из набора данных EPC, поскольку этот алгоритм подходит для больших наборов данных [14] . Алгоритм LOF использует функцию расстояния для измерения плотности объектов друг относительно друга. Евклидова мера расстояния используется с алгоритмом LOF для этого тематического исследования. Нижняя и верхняя границы для минимальных точек для меры расстояния установлены равными 10 и 20 соответственно. Результаты показывают, что набор данных EPC содержит значительное количество выбросов; например, в значениях u окна здания, стены, крыши и пола в наборе данных EPC (рис. 9).

Хотя набор данных EPC содержит более 200 переменных, в этом исследовании рассматриваются только те переменные, которые влияют на разработку архетипа и модели обучения. В тематическом исследовании

используются как инженерные методы, так и методы, основанные на данных, для определения подмножества наиболее важных функций. На первом этапе инженерный метод определяет 63 признака из более чем 200 признаков, основанных на существующих исследованиях [14] , [71] .. На следующем этапе метод выбора на основе данных идентифицирует 43 функции из выбранных 63 функций на основе нескольких статистических тестов, таких как порог дисперсии / стандартного отклонения и коэффициент корреляции. Коэффициент корреляции удаляет те признаки, которые точно отражают выходной признак. Выходной характеристикой данных EPC является маркировка энергетического рейтинга здания, выраженная в единицах первичной энергии (кВтч/(м2·а такжер)). Корреляция менее 0,01% и более 50% свидетельствует о том, что эта функция не оказывает существенного влияния на энергетический рейтинг здания. Метод порога стандартного отклонения исключает слишком похожие или непохожие признаки. К удаленным объектам относятся либо те, у которых более 90% всех значений идентичны, либо объекты с большим количеством отсутствующих значений. Например, элементы уровня пола, такие как U-значения ткани пола, не соответствуют установленным критериям, поскольку почти все значения идентичны. Оценщики EPC при проведении опросов часто предоставляют значения по умолчанию для элементов уровня пола из-за отсутствия точных данных. Другие исключенные функции включают дату, идентификатор и целевое значение.

В процессе выбора функций перечислены 43 важные функции из первоначальных 200 функций. Эти 43 функции могут быть классифицированы на основе ограждающих конструкций здания, системы обогрева строительной ткани, горячего водоснабжения, пространственных и выходных меток (Таблица 6). Окончательные обработанные данные содержат только улучшенную количественную информацию о строительных запасах, которая используется для формулирования архетипов и разработки моделей обучения.

Разработка архетипов зданий

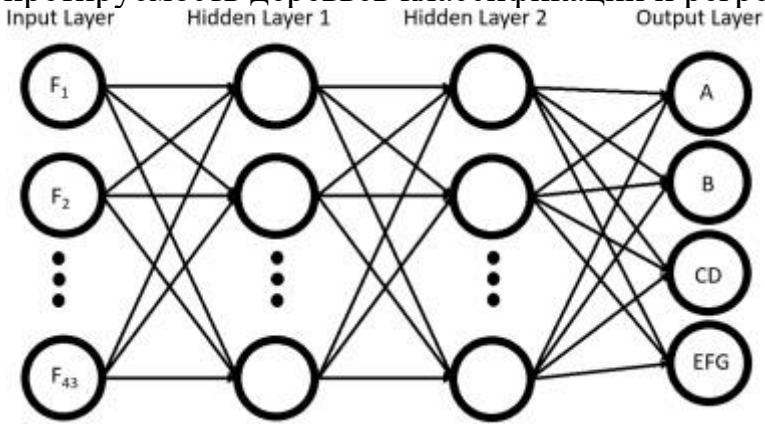
Процедура построения архетипов включает в себя два основных процесса, а именно сегментацию и характеристику. Для рассматриваемого фонда зданий сегментация с использованием критериев типа здания определяет пять типов зданий. Поскольку отдельные здания в структурированном наборе данных содержат собственный набор значений различных переменных (признаков), процесс характеристики объединяет (медиана) эти значения для зданий, принадлежащих к одному конкретному сегменту (архетипу). Таким образом, в результате агрегирования получается единый набор значений связанных переменных. Эти агрегированные значения для каждого типа здания представляют характеристики отдельного архетипа здания. Это исследование реализует агрегирование на уровне небольшой площади с использованием сегментации типа здания для детального анализа. По анализам их 18. 641 небольшой район в Ирландии. В базе данных GeoDirectory существует пять различных типов зданий, а именно: квартиры, таунхаусы, отдельные дома, двухквартирные дома и бунгало. В результате этого процесса было идентифицировано 93 205 архетипов зданий небольших районов на основе

сегментации типов зданий, которые представляют более двух миллионов жилых домов в Ирландии.

Разработка энергетической модели здания на основе данных

Этот процесс включает в себя формулировку модели классификации машинного обучения энергоэффективности здания. Процесс начинается с разделения данных, при котором ирландский набор данных EPC случайным образом разбивается на две группы для создания наборов данных для обучения и тестирования.

В этом исследовании используются и сравниваются восемь алгоритмов для разработки модели классификации, а именно: наивный байесовский алгоритм, обобщенная линейная модель, логистическая регрессия, глубокое обучение, дерево решений, случайный лес, деревья с градиентным усилением и машина опорных векторов. Алгоритмы сравниваются на основе различных классификаций энергетических рейтингов. В тренировочном процессе рассматриваются семь классификаций энергетических рейтингов. Результаты показывают, что алгоритмы глубокого обучения могут эффективно обрабатывать сложные и многомерные данные с большим количеством входных функций (таких как набор данных Irish EPC, используемый в этом тематическом исследовании). Аналогичным образом, алгоритм GBT эффективно обрабатывает наборы данных как с категориальными, так и с числовыми значениями. Кроме того, оба алгоритма могут обрабатывать отсутствующие записи в наборе данных, тем самым повышая точность модели. Хотя исследования показали, что SVM часто является оптимальным выбором для прогнозирования энергопотребления зданий, этот алгоритм, безусловно, не подходит для обработки больших наборов данных. Алгоритмы глубокого обучения обеспечивают высокую точность для значительного числа сценариев классификации по сравнению с другими алгоритмами. Хотя интерпретируемость алгоритмов глубокого обучения меньше по сравнению с такими алгоритмами, как деревья классификации и регрессии, большое количество входных признаков значительно снижает интерпретируемость деревьев классификации и регрессии.

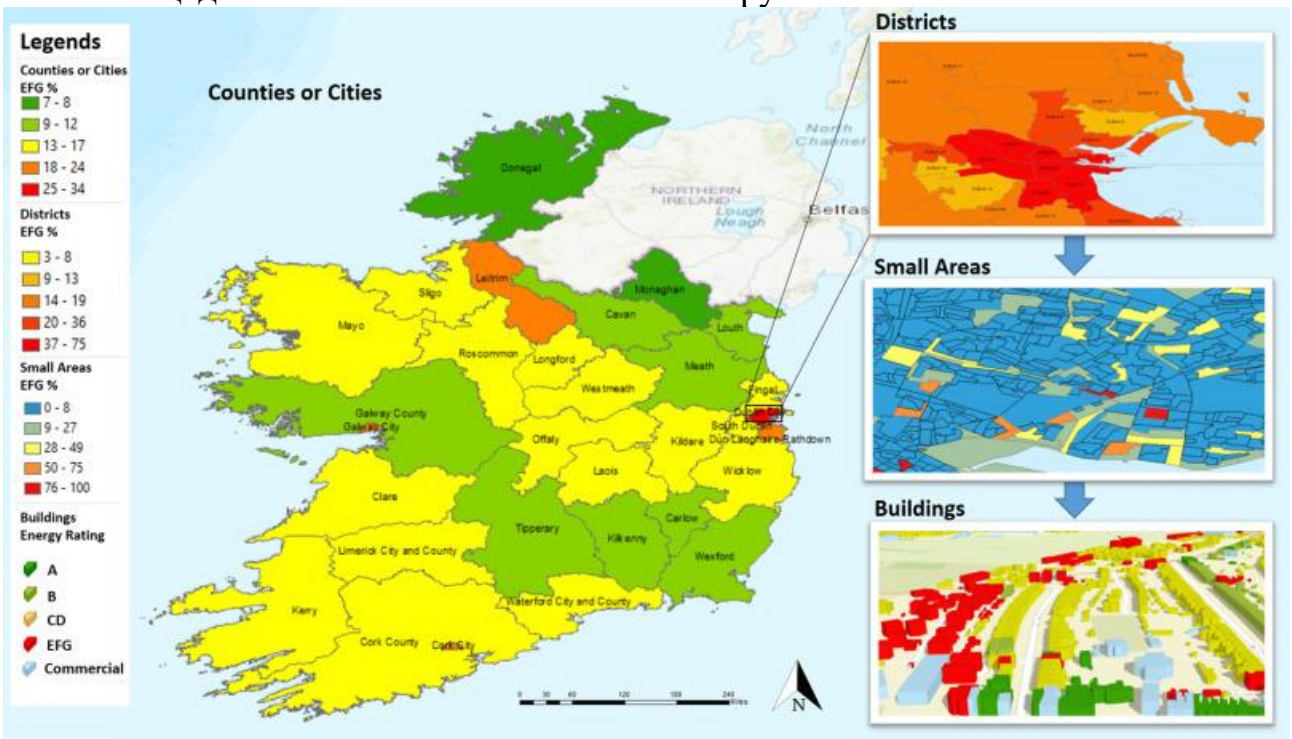


входном, 2 скрытых и 1 выходном слое для прогнозирования энергетического рейтинга здания с использованием данных EPC Ирландии.

Классификация A, B, CD и EFG дает наивысшую точность 88% при использовании алгоритма глубокого обучения (рис. 10). Стоит отметить, что эта классификация приемлема для заинтересованных сторон; цель часто состоит

в том, чтобы определить здания со значительно низкими характеристиками. Эти данные указывают на два важных вывода. Разработанная модель, управляемая данными, может рассчитать энергетический рейтинг здания, используя ограниченное количество входных данных с высочайшей точностью. Кроме того, точность модели может быть повышена за счет агрегирования меток с более низким классом энергопотребления. Например, максимальная производительность модели с фактической классификацией энергопотребления (A1, A2,…, E, F, G) составляет 76 %, а точность модели увеличивается на 12 % с агрегированной классификацией (A, B, CD и EFG).) с использованием алгоритма глубокого обучения (Рис. 10).

3D-карта Дублина показывает прогнозируемый энергетический рейтинг различных зданий; каждое здание представляет собой архетип здания небольшой площади с использованием онлайн-инструмента ArcGIS.



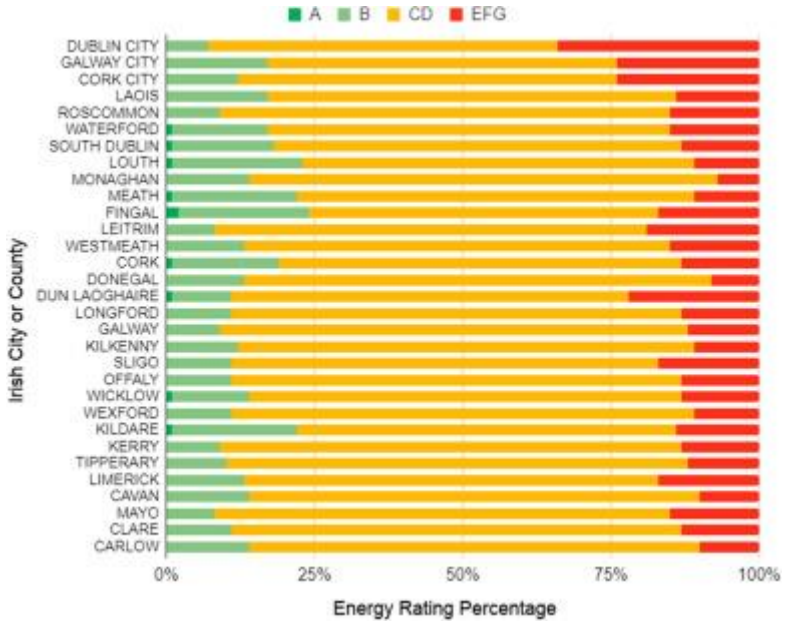
Многомасштабная карта Ирландии показывает процент результатов прогнозирования рейтинга энергопотребления зданий EFG, что помогает оценить распределение спроса на энергию в нескольких масштабах.

Выбранная модель обучения включает в себя четыре классификатора энергетического рейтинга, а именно, A, B, CD, EFG. Модель обеспечивает высочайшую точность 88% с использованием алгоритма глубокого обучения, который использует 43 входных блока с 2 скрытыми слоями, каждый из которых имеет размер 50 единиц и 4 выходных блока (рис. 11).). Более глубокое исследование модели с использованием матрицы путаницы показывает, что точность четырех выходных классов составляет от 75% до 95%. Матрица путаницы — это таблица, описывающая эффективность каждой метки в модели классификации. Точно так же значения отзыва составляют от 76% до 99% для четырех выходных меток. Значения точности являются самыми высокими для класса EFG (95%), что указывает на то, что модель правильно предсказывает 14 522 из всех 15 254 результатов прогнозирования. Значения отзыва являются самыми высокими для A-класса (99%), показывая, что модель точно предсказывает 1899 из всех 1920 фактических результатов (Таблица 7).

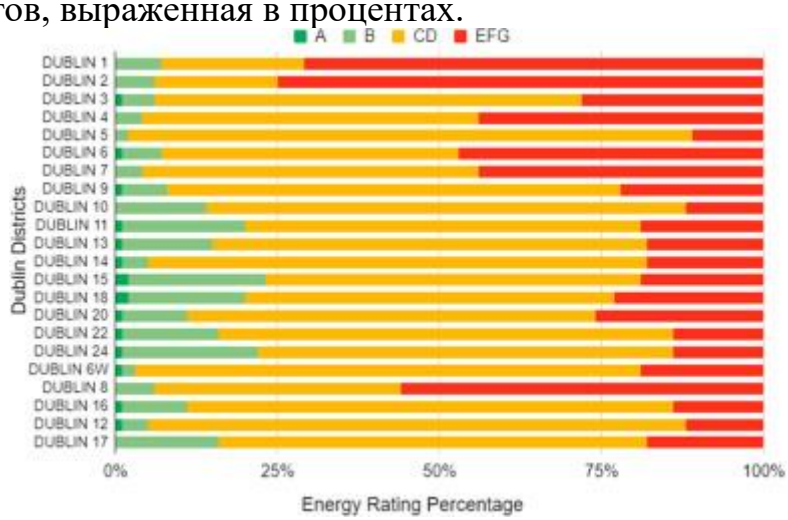
Многомасштабное ГИС-моделирование

Этот процесс включает в себя сопоставление результатов прогнозирования EPC в нескольких масштабах, от уровня отдельного здания до национального уровня. Разработанные архетипы зданий используют 43 исходных объекта для представления уникальных зданий на небольших территориях. Эти архетипы представляют весь ирландский фонд зданий с использованием набора данных GeoDirectory. Сформулированная модель глубокого обучения использует значения 43 входных признаков для оценки энергоэффективности всего фонда зданий. В процессе картирования эти результаты моделирования сопоставляются для получения 2D- и 3D-карт ГИС с помощью инструмента ArcGIS. Как упоминалось ранее, 3D-карты ГИС рассматривают только фонд зданий города Дублина как площадь здания, а данные о высоте доступны только для этого конкретного региона (рис. 12).

Наконец, результаты прогнозирования энергоэффективности зданий на основе 2D-ГИС распространяются на малые территории, районы, города и страны с использованием подхода пространственного объединения или агрегации. В процессе используется восходящий подход для агрегирования результатов прогнозирования энергоэффективности от здания до национального масштаба. Процесс начинается с пространственного агрегирования результатов энергетического рейтинга зданий на уровне небольшой площади. На следующем этапе все прогнозы для малых районов объединяются на уровне района, города и округа. Разработанную карту можно использовать для визуализации распределения рейтингов энергопотребления в нескольких масштабах. Например, результаты могут помочь определить процент неэффективных зданий (рейтинг EFG) для нескольких уровней (рис. 13).



Доля прогнозов энергетического рейтинга для ирландских городов / округов, выраженная в процентах.

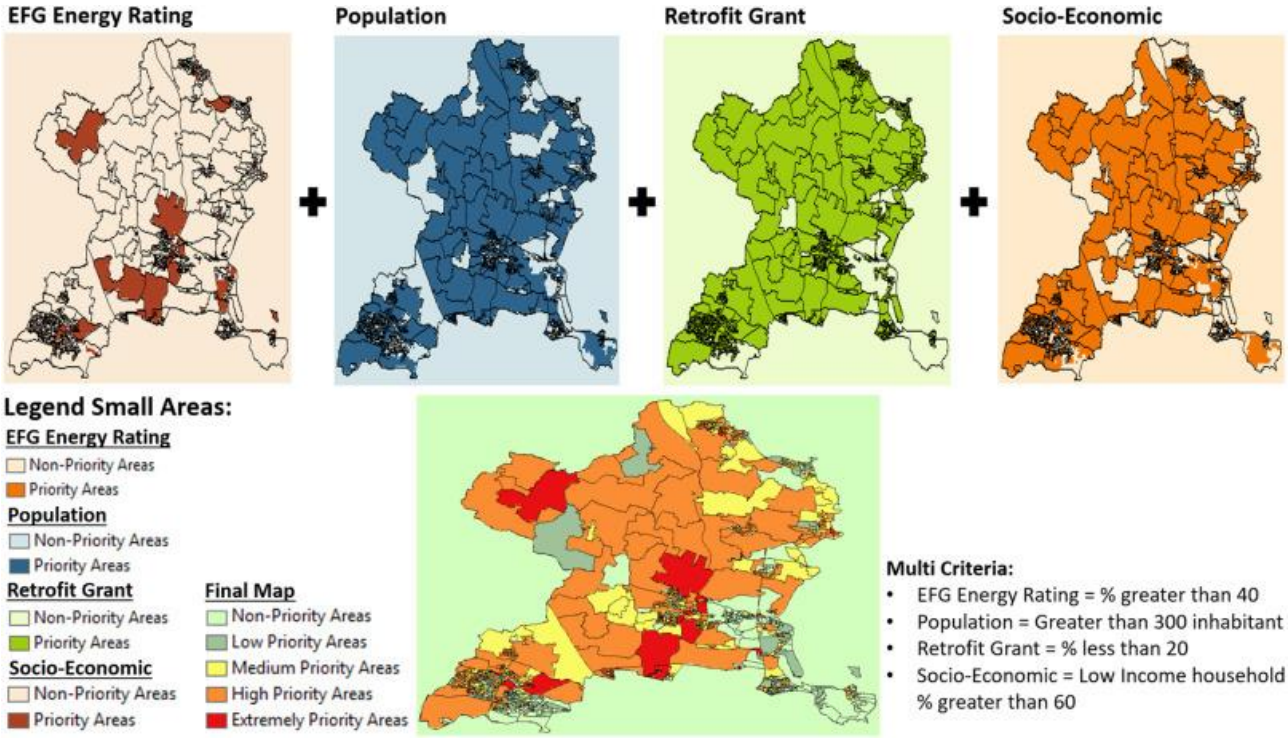


Доля прогнозов энергетического рейтинга для различных районов Дублина и округов, выраженная в процентах.

Результаты показывают, что самый высокий процент прогнозируемых рейтингов EFG принадлежит городским советам Дублина, Корка и Голуэя, которые составляют 34%, 24% и 24% от общего числа жилых зданий соответственно. (рис. 14). Более того, при проведении анализа на уровне районов Дублина результаты показывают, что районы в центре города (Дублин 1 и Дублин 2) имеют наибольшее количество рейтингов EFG (рис. 15). Точно так же заинтересованные стороны могут определить распределение энергетических рейтингов небольших территорий в конкретных районах, таких как Дублин 1 или Дублин 2. Кроме того, для каждой небольшой области результаты трехмерного моделирования зданий помогают выполнять детальный анализ.

Разработанная многомасштабная карта помогает лицам, принимающим решения, выявлять районы, где находится большое количество энергоэффективных зданий. Затем эту информацию можно использовать для проведения целевого социального маркетинга на уровне сообщества, что может

увеличить скорость модернизации в этом районе. На карте также обозначены кластеры ирландских жилых домов с низкими энергетическими характеристиками, что также указывает на то, в каких районах плохой уровень изоляции и производительности систем отопления. Результаты также показывают потребность в отоплении или электричестве в данной области для целей энергетического планирования. Например, на карте можно указать области, в которых проекты централизованного теплоснабжения могут быть эффективными.



Карты небольших районов округа Дублин (Фингал) для многокритериального анализа решений, чтобы помочь заинтересованным сторонам проанализировать и определить приоритетные области для реализации схем устойчивой энергетики.

Энергетическое планирование и анализ зданий

Этот процесс включает в себя применение карт ГИС для энергетического планирования и принятия решений. Для этого конкретного тематического исследования основная цель состоит в том, чтобы определить области в ирландских округах, где лица, ответственные за энергетическую политику, могут проводить целевые общественные мероприятия/кампании для увеличения активности по модернизации. Следовательно, этот процесс реализует подход многокритериального анализа решений (MCDA) для формулирования решений по планированию. Стоит отметить, что результаты прогнозирования энергоэффективности зданий на основе ГИС могут быть интегрированы с данными анализа решений.

Три разных уровня данных анализа решений, такие как субсидия на модернизацию, население и социально-экономический (доход домохозяйства), рассматриваются для реализации комплексного принятия решений и анализа решений. Ирландские гранты на модернизацию относятся к финансовой

поддержке, доступной домовладельцам для модернизации своих жилых зданий. Эта поддержка субсидирует затраты на повышение энергоэффективности. Население и социально-экономические данные собираются из базы данных переписи населения Ирландии. Этот анализ поможет в реализации мероприятий по модернизации, которые улучшат энергетические характеристики зданий и, таким образом, снизят потребление энергии. Кроме того, результаты также могут помочь градостроителям в планировании с учетом энергопотребления и анализе решений.

В этом тематическом исследовании рассматриваются небольшие районы округа Фингал в Дублине, насчитывающие более 114 000 жилых зданий, для демонстрации применения анализа MCDA на основе ГИС с использованием инструмента ArcGIS. Каждый слой четко указывает неприоритетные области, а также приоритетные области на основе определенных критериев. Первый слой, прогнозируемый энергетический рейтинг EPC, представляет собой небольшие районы, в которых более 40% зданий имеют рейтинг EFG. Общая приоритетная территория земли для уровня рейтинга EFG оценивается в 66.km2, на долю которого приходится около 14% всей площади Дублинского Фингала. Второй слой, население, представляет собой количество жителей зданий с населением более 300 человек на небольшой территории. Общая приоритетная земельная площадь для слоя населения оценивается в 373km2, на долю которого приходится около 77% всей площади Дублинского Фингала. Третий слой представляет собой распределение грантов на модернизацию по площади, которая была получена с максимальным охватом 20%. Общая приоритетная земельная площадь для слоя грантов на модернизацию оценивается в 385km2, на долю которого приходится около 80% всей площади Дублинского Фингала. Наконец, четвертый социально-экономический слой представляет собой критерий наличия более 60% малообеспеченных домохозяйств на малой территории. Общая приоритетная земельная площадь для социально-экономического слоя оценивается в 433km2, на долю которого приходится около 90% всей площади Дублинского Фингала (Таблица 8).

Окончательная карта показывает агрегированную работу карт взвешенных критериев (социально-экономические, население, субсидия на модернизацию), включая карту прогнозируемого энергетического рейтинга. Веса назначаются каждому слою с использованием метода аналитического иерархического процесса (AHP). Самый высокий вес 0,58 присвоен социально-экономическому слою, за которым следует вес 0,25, присвоенный слою рейтинга EFG. Вес 0,11 присвоен гранту на модернизацию, а вес 0,06 присвоен слою населения. Окончательная сводная карта распределяет приоритетные области по пяти классам: неприоритетные, низкоприоритетные, среднеприоритетные, высокоприоритетные и чрезвычайно приоритетные. Площадь земли для общей чрезвычайно приоритетной категории оценивается в 41km2, что составляет около 9% от общей площади Дублинского Фингала. Земельный участок высокого приоритета оценивается в 274km2, на долю которого приходится около 57% всей площади Дублинского Фингала. Земельные участки со средним и

низким приоритетом оцениваются в 88 и 35km2, что составляет около 18% и 7% от общей площади Фингала Дублина соответственно.

Окончательно разработанные карты из этого анализа помогают заинтересованным сторонам анализировать и определять приоритетные области для реализации решений в области устойчивой энергетики. На карте показаны области, которые наиболее выгодны для целевых кампаний на уровне сообществ, направленных на увеличение активности по модернизации (рис. 16). Кроме того, при рассмотрении ирландского фонда зданий заинтересованные стороны будут заинтересованы в выявлении потенциальных областей, в которых будет высока потребность в схемах модернизации с большими субсидиями. Например, такой подход мог бы помочь SEAI определить потенциальные цели для расширения охвата программы «Теплые дома».

Реальное применение анализа MCDA может быть расширено для поддержки принятия решений в городах и облегчения энергетического планирования и анализа в городских районах за счет минимизации выбросов CO.2и использование энергии. Например, в зоне высокого приоритета определено 4918 жилых зданий с рейтингом EPC. Модернизация зданий в здании с рейтингом A уменьшит≈1751 МВтч/м2/год экономия на ПНВ и≈407 т/м2/год в CO2снижение. Эти оценки помогут градостроителям планировать ремонт в районах, представляющих особый интерес. Кроме того, политики будут следить за строительным сектором с точки зрения энергоэффективности и выбросов углерода.