

1. Проблема анализа данных в ГИС

1.1. ГИС среди информационных технологий

ГИС – это современная компьютерная технология для картирования и анализа объектов реального мира, а также событий, происходящих на нашей планете, в нашей жизни и деятельности [1, 2].

Эта технология объединяет традиционные операции при работе с базами данных, такими как запрос и статистический анализ, с преимуществами полноценной визуализации и географического (пространственного) анализа, которые предоставляет карта. Эти возможности отличают ГИС от других информационных систем и обеспечивают уникальные возможности для ее применения в широком спектре задач, связанных с анализом и прогнозом явлений и событий окружающего мира, с осмыслением и выделением главных факторов и причин, а также их возможных последствий, с планированием стратегических решений и текущих последствий предпринимаемых действий [3, 4, 5].

Пожалуй, главным преимуществом ГИС является наиболее естественное (для человека) представление как собственно пространственной информации, так и любой другой информации, имеющей отношение к объектам, расположенным в пространстве (т.н. *атрибутивной информации*). Пространством в данном случае можно называть не только трехмерное пространство, в котором мы существуем, но и любое *абстрактное пространство произвольной размерности* (см. главу 4). Способы представления атрибутивной информации различны: это может быть числовое значение, таблицы характеристик объекта или базы данных (локальные или удаленные), его фотография, или реальное видеоизображение.

На этапе ввода информации ГИС оказывают большую помощь в наглядном представлении первичной информации, здесь много общего с системами автоматизированного управления производственными и иными объектами (САПР и АСУ). Если сложный объект может быть представлен в виде некоторой схемы, то ГИС может быть удобным интерфейсом для доступа к информации от ее источников. В таком случае с помощью ГИС пользователь может указать курсором на некоторый элемент схемы и получить информацию о характеристиках и состоянии соответствующего ему объекта. Важно также то, что ГИС содержат удобные средства для создания и редактирования таких схем и, естественно, для организации связи с первичными источниками информации. Очевидно и то, что с помощью ГИС может быть организован эффективный доступ к большому объему информации об объектах, имеющих пространственную привязку. Поскольку хранение и поиск больших объемов информации на электронных носителях – задача со своей спецификой, собственно ГИС обычно используют возможности внешних СУБД, и эффективность и

надежность такого взаимодействия – важная характеристика полнофункциональной ГИС.

Обработка пространственных данных в ГИС [6, 7, 8]. Общие аналитические операции с точечными, линейными и площадными объектами: операции переструктуризации данных, трансформация проекций и изменение систем координат, операции вычислительной геометрии, оверлейные операции, операции с рельефом, операции на графах и сетях, интерполяция точечных данных в поверхности. Моделирование в ГИС: статистические модели снижения размерности многомерных массивов пространственных данных, классификации пространственных объектов по комплексу показателей; имитационные модели оценки динамики развития процессов; различного вида оптимизационные и эвристические модели. Многовариантные и диалоговые подходы к моделированию процессов и явлений в ГИС.

Нарастающие информационные потоки в современном обществе, разнообразие информационных технологий, повышение сложности решаемых на компьютере задач увеличивают нагрузку на пользователя этих технологий и ставят задачу переноса проблемы выбора и принятия решений с человека на ЭВМ. Одним из путей решения этой задачи является применение систем анализа данных (аналитических систем), которые могут быть составной частью ГИС [9, 10, 11].

Можно выделить несколько групп задач, требующих применения таких систем в ГИС:

- обработка видеоизображений;
- преобразование растровых изображений в векторные графические модели;
- обработка картографической информации;
- обработка разнородной информации;
- построение моделей объектов или местности;
- анализ моделей ГИС;
- получение новых знаний;
- получение решений на основе геоинформации.

ГИС являются хорошей средой для внедрения методов искусственного интеллекта и экспертных систем [3, 12, 13]. Это вызвано, с одной стороны, разнообразием и сложностью данных в ГИС, с другой — наличием большого числа аналитических задач при использовании ГИС.

Из всех разновидностей и направлений в ГИС наиболее близко к теме данной работы математико-картографическое моделирование и геоинформационное картографирование – одна из основ ГИС-технологий, применяемых при принятии решений, управлении, проведении экспертиз, составлении прогнозов и т.п. [5]

1.2. Классификация в ГИС

Многие пространственные математические модели создаются в картографической форме: это морфометрические карты (расчленения поверхностей, уклонов, градиентов и др.), карты полей плотности и

интенсивности явлений, фоновые (трендовые) и остаточные поверхности, поля пространственных корреляций и взаимных соответствий, анизотропии явлений, синтетические карты распределения главных факторов и факторных нагрузок, интегрального районирования и многие другие [14 – 19]. Собственно математическое моделирование предполагает более углубленный анализ, нежели просто вычисление количественных показателей [20, 21]. Имеется в виду построение пространственно-временных моделей структуры, динамики, взаимосвязей объектов и явлений и на этой основе – создание более сложных моделей и прогноза их дальнейшего развития. Большинство аналитических задач такого рода можно сформулировать как задачу классификации географических комплексов. В данной работе рассматриваются задачи в ГИС так или иначе связанные с восстановлением функции по конечному набору значений или с разбиением конечного множества объектов на классы. Рассмотрим подробнее задачу классификации географических комплексов и связанные с ней сложности.

1.2.1. Классификация

Классификация изучаемых явлений в той или иной мере присуща всем наукам, а ее использование в географии имеет давнюю историю. Географический подход к изучению явлений природы и общественной жизни предполагает территориальную изменчивость этих явлений и ее изучение с помощью методов классификации. Районирование территории, типология и оценка комплексов зачастую с представлением на карте полученных результатов являются не только методами, но и целями исследований. Можно сказать, что в географии относительно чаще, чем в других науках, классификация выступает не только (и даже не столько) методом, но и целью научного исследования [22, 23, 24].

При изучении географических комплексов часто ставится задача предварительной классификации показателей и факторов, описывающих эти комплексы. Обычно она выступает как метод исследования, в то время как классификации комплексов могут выступать и в роли методов, и в роли целей. Содержательная интерпретируемость классификаций показателей и факторов, необходимая для их эффективного использования в исследованиях, требует получения таких результатов, когда показатели и факторы, отнесенные к одной группе, имеют близкий содержательный смысл или описывают близкие стороны явлений. Различные алгоритмы корреляционного или факторного анализа, которые используются или могут быть использованы для решения подобных задач, не всегда отвечают поставленным условиям. Их применение иногда требует проведения большого объема работы по корректировке полученных результатов и приведению к схеме, элементы которой имеются у исследователя а priori.

Наличие у исследователя априорной схемы (построенной часто по аналогии с ранее проведенными классификациями) играет важную, а часто и определяющую роль. Именно с ее позиций обычно оценивается качество

полученных результатов. Методы многомерной автоматической классификации бывают полезны тем, что они подтверждают существующую схему и позволяют уточнять отдельные ее элементы.

При классификациях комплексов может решаться как задача разделения исследуемой совокупности явлений на классы, так и отнесения одного или нескольких явлений к уже существующим. Эти классы могут быть описаны перечислением списка явлений, ими охватываемых, указанием общих свойств явлений, включенных в них, либо характеристикой отдельных их представителей, рассматриваемых как типичные. Каждый из этих случаев имеет свои сложности при реализации.

Наконец, помимо классификаций комплексов возникает задача классификации их взаимодействий (в том числе и территориальных). Она оказывается гораздо труднее из-за неясностей с описаниями процессов взаимодействий. Особенно это проявляется при классификациях территориальных систем как целостных совокупностей явлений вместе с их процессами взаимодействий.

1.2.2. Районирование и типология

Важным понятием является районирование. Районирование определяется как процедура вычленения целостных территориальных систем, когда внимание исследователей концентрируется на различиях между ними, а при типологии и оценке основной критерий – однородность выделяемых таксонов. В этом, главным образом, заключается отличие районирования. Однако, если рассматривать район как некоторое единство (не сводящееся лишь к однородности) слагающих систему элементов с их процессами взаимодействий, то само районирование можно считать специфической формой классификаций в широком понимании. В отличие от районирования, типология и оценка могут приводить к образованию территориально расчлененных таксонов, свойства которых определяются содержательной сущностью решаемых задач.

1.2.3. Временные характеристики

Очень важным основанием географических классификаций являются временные, генетические характеристики изучаемых комплексов. Во всех географических исследованиях комплексы должны рассматриваться как пространственно-временные образования. Временные и генетические характеристики играют важную роль и в классификациях, встречающихся в других естественных и общественных науках (биологии, геологии, экономике, истории и др.).

1.3. Трудности в классификации географических комплексов

Алгоритмы методов автоматической классификации должны входить в математическое обеспечение ЭВМ, используемых географами при проведении исследований. Доступность соответствующих стандартных программ и умение пользоваться ими необходимо для прогресса в области использования ЭВМ в научных исследованиях. Однако в настоящее время эти алгоритмы создаются чаще всего без конкретного учета специфики тех наук, в которых они могут

использоваться. Такая их универсальность, конечно, имеет положительное значение, позволяя использовать алгоритмы, уже разработанные математиками-прикладниками или другими специалистами.

С другой стороны, она чревата и возможностью возникновения ряда сложностей, связанных с недоучетом специфики географических задач и подходов. Это вызывает необходимость анализа современного состояния классификаций в географии, специфических черт и перспектив их использования. Остановимся на некоторых аспектах этих проблем.

1.3.1. Местоположение географических явлений и пространственные особенности

При классификации географических комплексов приходится сталкиваться с разными трудностями, некоторые из них имеют место в различных науках, другие характерны именно для географии. Часть их связана с проблемой адекватности применяемых методов существу и уровню строгости поставленных задач. В частности, важным показателем, характеризующим географические явления, является их местоположение. Существующие алгоритмы обычно работают со статистическими характеристиками, лишь косвенно или вовсе не отражающими расположение явлений. При этом количественная оценка относительной значимости географического положения в настоящее время не ясна. Это вызывает необходимость дополнительного учета территориального аспекта моделируемых явлений и здесь, прежде всего, обращаются к картографическому методу. Имеются попытки модификации методов статистической обработки показателей для учета пространственного положения [25, 26, 27, 28].

Как уже отмечалось, одним из факторов, определяющих наличие специфики применения классификаций в географии, является, прежде всего, привязка объектов к земной поверхности, хотя это не обязательное условие. Вообще, среди таких моделей можно выделить три разновидности:

- 1) результаты реализации моделей не подлежат пространственному анализу и не наносятся на карту,
- 2) результаты поддаются картографированию, но пространственный аспект не учитывается на этапе реализации математических алгоритмов,
- 3) без учета пространственного положения явлений невозможно реализовать математические расчеты [29].

1.3.2. Плохая формализация

Географические явления плохо поддаются формализации. Существующий математический аппарат недостаточно приспособлен для решения географических задач [24].

Формулировки географических задач, описания явлений допускают некоторый произвол или двоякое толкование, по крайней мере, на современном этапе исследований. Строгие алгоритмы многомерной классификации могут не соответствовать уровню строгости и точности

самих задач. Это иногда приводит к результатам, не отвечающим существу и содержательному смыслу. В плане решения данной проблемы внимание исследователей (географов и не географов) привлекает теория искусственного интеллекта и попытки разработки на ее основе методов классификации. Одному из эффективных примеров их использования в географических исследованиях посвящена третья глава данной работы. В частности речь пойдет об искусственных нейронных сетях.

1.3.3. Лишние данные и различная степень значимости исходных показателей

При многих классификациях встает проблема оптимального выбора системы исходных показателей, которая должна всесторонне (в той мере, в которой это требуется существом задачи) описывать изучаемые явления. При этом возникает соблазн включить в рассмотрение все доступные данные. Однако это может привести к их избыточности. Они не должны дублировать друг друга, быть производными один от другого и т. д. В противном случае они могут затушевывать наиболее значимые признаки и привести к искажению конечного результата. Трудно найти тот критерий, который позволяет оценить, необходим ли тот или иной показатель как индикатор характеристики географического комплекса. Самую существенную помощь здесь может оказать глубокое познание сущности комплекса, что позволяет установить круг показателей, его отображающих. Другой путь – экспериментальная проверка степени их влияния на конечный результат.

Еще одна трудность — различная степень значимости, важности используемых показателей для характеристики комплексов. Некоторые из них столь важны, что их исключение не позволяет моделировать явления, другие же лишь дополняют, уточняют основную систему. Это требует "взвешивания" показателей, ведущего к дифференциации степени их влияния на конечный результат. Однако, определение "весов" - самостоятельная, сложная, во многом не решенная задача. В третьей главе описываются подходы решающие перечисленные выше проблемы:

- оптимальный выбор системы исходных показателей;
- дублирование исходных признаков;
- значимость исходных признаков для решения основной задачи.

1.3.4. Признаки разной природы

Большинству классификационных задач в географии свойственны показатели различной природы: имеющие количественное выражение, оценивающие значения какого-либо признака без его количественного выражения, а также носящие чисто качественный характер (например, пришедшие из какой-нибудь другой классификации). Это накладывает определенные ограничения на возможности использования всего многообразия методов многомерной классификации. Из-за того, что значительная часть данных, которые учитываются при классификациях, имеет качественный характер, используемые алгоритмы должны уметь работать с характеристиками нечисловой природы. Решение этой

проблемы также рассмотрено в главе 3.

1.3.5. Несоответствие модели решаемой задаче

Существенны трудности при выборе наиболее подходящей модели. В настоящее время существует огромное количество различных алгоритмов, пригодных для решения классификационных задач, но не отвечающих в полной мере специфике отображаемых комплексов. Это ведет к разработкам математических моделей самими географами, при этом стараются с помощью математических алгоритмов отобразить наиболее яркие типичные стороны географических комплексов. В одном случае требуется выявить ядра районообразования, установить основные районообразующие связи и формировать районы как целостные с содержательных позиций и территориально нерасчлененные совокупности исходных единиц; в другом при оценках комплексов ставится условие гомогенности исходных единиц по комплексу показателей, а также ранжирования выделяемых однородных таксонов и т. д.

1.3.6. Нормировка исходных показателей

Модификации математических алгоритмов с целью приближения их содержательной сути начинается уже с нормировки исходных показателей. Например, при оценках географических комплексов необходимо привести систему исходных данных к логически сопоставимому виду, когда используемые показатели должны описывать отклонения характеристик комплекса от оптимальных оценок. Это позволяет правильно задать ориентацию показателей между логическими полюсами наилучших и наихудших условий для каждого из них и тем самым правильно соизмерить их между собой [30]. Структура вычислительного алгоритма и на других этапах исследования также должна быть согласована с требованиями адекватного моделирования содержательной сущности явлений.

ИНС и ГИС

Анализ литературы показал фактически полное отсутствие использования искусственных нейронных сетей (ИНС) среди огромного количества работ, связанных с ГИС-тематикой. Исключения представляют работы, затрагивающие решение частных задач. Например, в работах, ведущихся в институте проблем безопасного развития атомной энергетики, нейросети используются для решения задачи интерполяции [17]. Есть работы, посвященные использованию искусственных нейронных сетей в географических информационных системах для оценки устойчивости сельскохозяйственных земель [31]. Есть попытки описать взаимодействие систем искусственного интеллекта с ГИС [12].

Несмотря на то, что существует некоторое количество работ, в которых высказывается пожелание использования нейросетевых технологий [3, 6, 12, 13], общей методологии использования нейросетей в ГИС до сих пор не создано. Также нет общего описания и классификации задач, для которых возможно использование нейронных сетей.

По-видимому, такое положение вызвано отсутствием удобного

средства для решения задач ГИС нейронными сетями.