

А.А. Питенко
НЕЙРОСЕТЕВОЙ АНАЛИЗ В
ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ

Красноярск – 2000

Нейросетевой анализ в геоинформационных системах. Питенко А.А.

Представленная в работе технология нейросетевого анализа в ГИС предназначена для решения широкого спектра 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.

Разработанные в рамках технологии методы ориентированы на следующие применения:

- автоматизированное построение нейросетевых блоков для решения задач оценки, диагностики и прогнозирования на основе эмпирических данных в составе существующих геоинформационных систем;
- построение и исследование нейросетевых моделей решения задач анализа данных в ГИС;
- представление и анализ средствами ГИС многомерных данных произвольной природы.

Издание подготовлено при поддержке ФЦП «ИНТЕГРАЦИЯ», проект А0020, направление 2.1.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1. ПРОБЛЕМА АНАЛИЗА ДАННЫХ В ГИС.....	7
1.1. ГИС СРЕДИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ.....	7
1.2. КЛАССИФИКАЦИЯ В ГИС.....	8
1.2.1. Классификация.....	9
1.2.2. Районирование и типология	10
1.2.3. Временные характеристики	10
1.3. ТРУДНОСТИ В КЛАССИФИКАЦИИ ГЕОГРАФИЧЕСКИХ КОМПЛЕКСОВ	10
1.3.1. Местоположение географических явлений и пространственные особенности	11
1.3.2. Плохая формализация.....	11
1.3.3. Лишние данные и различная степень значимости исходных показателей	12
1.3.4. Признаки разной природы.....	12
1.3.5. Несоответствие модели решаемой задаче.....	13
1.3.6. Нормировка исходных показателей.....	13
ИНС и ГИС	13
2. ГИС КАК СРЕДСТВО ВИЗУАЛИЗАЦИИ И АНАЛИЗА ДАННЫХ РАЗЛИЧНОЙ ПРИРОДЫ	15
2.1 ВВЕДЕНИЕ В ГИС.....	15
Данные.....	16
Объекты	16
Слой	16
Объект слоя.....	16
Легенда карты	16
Карта.....	16
2.2 МОДЕЛИ ГИС.....	17
Растровые модели	17
2.3 ОСНОВНЫЕ ИДЕИ МЕТОДА АНАЛИЗА ДАННЫХ В ГИС С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ.....	19
Обобщение задач	19
Формальная постановка.....	20
2.4 АНАЛИТИЧЕСКИЕ ЗАДАЧИ В ГИС, РЕШАЕМЫЕ С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ.....	21
Построение (дополнение) слоя.....	21
Восстановление легенды слоя	22
Районирование и типология	23
Создание моделей поверхностей.....	23
Интерполяция и прогнозное картирование	24
Временной анализ.....	24
Выбор значимых признаков.....	25

2.5 ОСНОВНЫЕ ИДЕИ ВИЗУАЛИЗАЦИИ И АНАЛИЗА ДАННЫХ ПРОИЗВОЛЬНОЙ ПРИРОДЫ	25
<i>Описание задачи</i>	26
<i>Объекты метода</i>	27
3. НЕЙРОСЕТЕВОЙ АНАЛИЗ СВЯЗИ МЕЖДУ СЛОЯМИ	32
3.1. МЕТОДЫ НЕЙРОСЕТЕВОГО АНАЛИЗА СВЯЗЕЙ МЕЖДУ СЛОЯМИ	32
3.1.1. <i>Проблема построения и использования нейросетей в геоинформационных системах</i>	32
3.1.2. <i>Технология построения нейросетевых моделей в составе геоинформационной системы</i>	36
3.1.3. <i>Задачи для нейронных сетей</i>	50
3.2. ПРОГРАММНЫЕ СРЕДСТВА И ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ	52
3.2.1. <i>Реализация программного комплекса для нейросетевого анализа данных в ГИС</i>	52
<i>Программная реализация системы</i>	56
3.2.2. <i>Решение задачи восстановления пропусков</i>	57
4. ТЕХНОЛОГИЯ КАРТИРОВАНИЯ ПРОИЗВОЛЬНЫХ ДАННЫХ	66
4.1. УПРУГИЕ КАРТЫ.....	66
4.1.1. <i>Введение</i>	66
4.1.2. <i>Постановка задачи</i>	68
4.1.3. <i>Построение упругой карты</i>	70
4.1.4. <i>Проектирование многомерных данных на двумерную сетку</i>	71
4.1.5. <i>Использование карты для анализа распределения данных</i>	74
4.2. ПРИМЕНЕНИЕ МЕТОДОВ ВИЗУАЛИЗАЦИИ ПРОИЗВОЛЬНЫХ ДАННЫХ К КАРТОГРАФИРОВАНИЮ ЭКОНОМИЧЕСКИХ ТАБЛИЦ.....	75
4.2.1. <i>Раскраска по признакам</i>	76
4.2.2. <i>Раскраска по плотности данных</i>	77
4.2.3. <i>Восстановление данных, регрессия и прогноз</i>	79
4.2.4. <i>Проблема экстраполяции</i>	82
4.3. ПРИМЕНЕНИЕ ТЕХНОЛОГИИ ДЛЯ ВИЗУАЛИЗАЦИИ И АНАЛИЗА ТАБЛИЦЫ ЭКОЛОГИЧЕСКИХ ИЗМЕРЕНИЙ	82
<i>Атлас №1, №2 информационных раскрасок</i>	84
АТЛАС ИНФОРМАЦИОННЫХ РАСКРАСОК №1	86
АТЛАС ИНФОРМАЦИОННЫХ РАСКРАСОК №2.....	87
ЗАКЛЮЧЕНИЕ.....	88
ЛИТЕРАТУРА.....	89

Введение

Необходимость проанализировать картографические данные, накопленные в Географических информационных системах (ГИС), возникает у представителей различных профессий. Прежде всего, это актуально для управляющих структур, владеющих большими массивами информации, на основе которых принимаются решения. В этом также нуждаются специалисты, оценивающие и прогнозирующие состояние какой-либо области человеческой деятельности, например, рынков сбыта продукции, загрязнения территории и т.п. 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 аналитических систем, которые могут быть составной частью ГИС.

Далеко не все ГИС снабжены возможностями специализированного анализа. Это связано с тем, что четкой схемы проведения таких работ не существует и организации, занимающиеся ими, предпочитают производить анализ по собственным методикам и правилам. Работа со специфическими данными специфическим образом является характерной чертой этого типа анализа. Кроме того, взгляды на приемы его проведения могут меняться с течением времени. Поэтому такие возможности в ГИС представляются средствами создания приложений самими пользователями. Сложность состоит в том, что для каждой специализированной области возникает необходимость создавать отдельное приложение к ГИС и часто даже свою методику обработки. Это не всегда возможно и часто дорого.

Географические комплексы плохо поддаются формализации. Существующий математический аппарат недостаточно приспособлен для решения географических задач. 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 решения таких задач.

ГИС являются хорошей средой для внедрения методов искусственного интеллекта и экспертных систем. Это вызвано, с одной стороны, разнообразием и сложностью данных в ГИС, с другой — наличием большого числа аналитических задач при использовании ГИС. Одновременно с этим большинство проблем и задач в ГИС слабо структурировано и слабо формализовано.

Построение традиционных математических моделей для решения

таких проблем затруднено или сопряжено со значительными затратами, превышающими ожидаемый от модели эффект. Это связано с невозможностью полного исследования внутренних взаимодействий в системе, большим числом влияющих факторов, неполнотой или неточностью описания объектов, динамикой или малой изученностью предметной области. Традиционно такие задачи решаются на неформальном уровне экспертами – специалистами в предметной области. В современных условиях для решения подобных задач используются искусственные нейронные сети.

Нейросетевые модели претендуют на то, чтобы стать универсальным аппаратом, решающим разные специфические задачи из разных проблемных областей в ГИС. Такая универсальность обуславливается тем, что нейросетевые технологии дают стандартный способ решения многих нестандартных задач.

Интеграция основанных на нейросетевых технологиях средств решения слабоформализованных задач и геоинформационных систем позволит существенно повысить качество и скорость обработки информации, расширить их возможности в прикладных, исследовательских, учебных и других задачах.

Представленная в работе технология нейросетевого анализа в ГИС предназначена для решения широкого спектра 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.

Разработанные в рамках технологии методы ориентированы на следующие применения:

- автоматизированное построение нейросетевых блоков для решения задач оценки, диагностики и прогнозирования на основе эмпирических данных в составе существующих геоинформационных систем;
- построение и исследование нейросетевых моделей решения задач анализа данных в ГИС;
- представление и анализ средствами ГИС многомерных данных произвольной природы.

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.

2. ГИС как средство визуализации и анализа данных различной природы

2.1 Введение в ГИС

Геоинформационная система (ГИС) – программно-аппаратный комплекс, предназначенный для сбора, управления, анализа и отображения пространственно распределенной информации.

ГИС – не только и не столько информационные системы для географии, сколько информационные системы с географически организованной информацией. В простейшем варианте геоинформационные системы – сочетание обычных баз данных (атрибутивной информации) с электронными картами, то есть мощными графическими средствами.

Основная идея ГИС – связь данных на карте и в базе данных. ГИС – это и аналитические средства для работы с любой координатно-привязанной информацией. В принципе, ГИС можно рассматривать как некое расширение концепции баз данных. В этом смысле ГИС фактически представляет собой новый уровень и способ интеграции и структурирования информации [32].

ГИС предлагает совершенно новый путь развития картографии. Преодолеваются основные недостатки обычных карт – их статичность и ограниченная емкость как носителя информации. В последние десятилетия бумажные карты из-за перегруженности информацией становятся нечитабельными. ГИС же обеспечивает управление визуализацией информации. Появляется возможность выводить (на экран, на твердую копию) только те объекты или их множества, которые интересуют нас в данный момент. Фактически осуществляется переход от сложных комплексных карт к серии взаимоувязанных частных карт. При этом улучшается структурированность информации, а следовательно, повышается эффективность ее обработки и анализа. В ГИС карта оживает и становится действительно динамическим объектом в смысле:

- изменяемости масштаба;
- преобразования картографических проекций;
- варьирования объектным составом карты;
- возможности опрашивать через карту в режиме реального времени многочисленные базы данных;
- изменения способа отображения объектов (цвет, тип линии и т.п.), в том числе и определения символики через значения атрибутов, то есть синхронизации визуализации с изменениями в базах данных;
- легкости внесения любых изменений.

Рассмотрим основные понятия ГИС, в том или ином виде присутствующие во всех современных геоинформационных системах.

Данные

В ГИС данные делятся на две категории:

- пространственные (местоположение);
- непространственные (атрибуты).

Объекты

Пространственные данные включают географические объекты, представляемые:

- точками;
- линиями;
- полигонами.

Дугами описываются те реальные объекты, которые можно рассматривать как линии. Дуга состоит из отрезков линий и дуг окружностей.

Полигоны – замкнутые области, которые представляют однородные по некоторым критериям участки.

Атрибутивные данные могут включать идентификатор объекта, любую описательную информацию из баз данных, изображение и многое другое.

Слой

Слои в карте подразделяются на два основных вида – растровые и векторные.

Векторные слои – это совокупность простых геометрических объектов (точка, дуга, полигон), которые представляют те или иные объекты на местности. Векторные слои могут также хранить топологию, т.е. информацию о взаимном расположении объектов.

Растровые слои представляют из себя сплошные изображения. Они не могут содержать объекты. Однако они могут служить фоном для векторных слоев

Объект слоя

Каждому объекту векторного слоя может соответствовать запись в базе данных, чем обеспечивается привязка информации к местности. Это соответствие может обеспечиваться в частности назначением каждому объекту соответствующего идентификатора.

Легенда карты

Легенда карты – свод *условных обозначений*, использованных на карте, с текстовыми пояснениями к ним. Обычно, легенды создаются на основе классификаций изображаемых объектов и явлений, они становятся их графической моделью и часто служат для построения классификаторов.

Карта

Представляет собой набор географических слоев, каждый из которых привносит в карту информацию по какой-либо определенной теме. Например, на слой границ некоторой территории может быть нанесен слой рек, затем слой, отображающий количество атмосферных осадков в процентном отношении и т.д.

Электронную карту в ГИС можно рассматривать как многокомпонентную модель реальности. Основными целями ее создания являются:

- графическая коммуникация пространственных отношений и распределений;
- улучшение возможности анализа, обработки и отображения геоинформационных данных;
- визуальное отображение цифровых моделей явлений, невидимых для человеческого глаза;
- автоматизация отображения и картографического анализа в системах управления; исследование объектов, явлений и процессов с учетом динамики их развития и возможного использования;
- получение аналитических решений в графическом виде в режимах реального и разделенного времени и т.д.

2.2 Модели ГИС

Основой визуального представления данных при помощи ГИС-технологий служит так называемая графическая среда. Основу графической среды и соответственно визуализации базы данных ГИС составляют векторные и растровые модели.

В общем случае модели пространственных (координатных) данных могут иметь векторное или растровое (ячеистое) представление, содержать или не содержать топологические характеристики. Этот подход позволяет классифицировать модели по трем типам:

- растровая модель;
- векторная нетопологическая модель;
- векторная топологическая модель.

Все эти модели взаимно преобразуемы. Тем не менее, при получении каждой из них необходимо учитывать их особенности. В ГИС форме представления координатных данных соответствуют два основных подкласса моделей – векторные и растровые (ячеистые или мозаичные). Возможен класс моделей, которые содержат характеристики как векторов, так и мозаик. Они называются гибридными моделями.

В дальнейшем под терминами решетка, мозаика, элемент раstra будем понимать одно и то же. Основу такой классификации составляет атомарная единица (пространства), содержащая представления площадей, линий и точек.

Между векторными и растровыми изображениями имеется различие, характерное именно для ГИС. Растровые изображения отображают поля данных, т.е. носят полевой характер. Векторные изображения в ГИС, как правило, отображают геоинформационные объекты, т.е. носят объектный характер.

Растровые модели

Рассмотрим подробнее растровые модели данных, которые ближе касаются нашей основной задачи, нежели чем векторные. Напомним, что

модель данных представляет собой отображение непрерывных последовательностей реального мира в набор дискретных объектов.

В растровых моделях дискретизация осуществляется наиболее простым способом – весь объект (исследуемая территория) отображается в пространственные ячейки, образующие регулярную сеть. При этом каждой ячейке растровой модели соответствует одинаковый по размерам, но разный по характеристикам (цвет, плотность) участок поверхности объекта. В ячейке модели содержится одно значение, усредняющее характеристику участка поверхности объекта. В теории обработки изображений эта процедура известна под названием пикселизация или растеризация.

Если векторная модель дает информацию о том, где расположен тот или иной объект, то растровая – информацию о том, что расположено в той или иной точке территории. Это определяет основное назначение растровых моделей – непрерывное отображение поверхности.

В растровых моделях в качестве атомарной модели используют двумерный элемент – пиксель (ячейка). Упорядоченная совокупность атомарных моделей образует растр, который, в свою очередь, является моделью карты или геообъекта.

Растровые модели имеют следующие достоинства:

- растр не требует предварительного знакомства с явлениями, данные собираются с равномерно расположенной сети точек, что позволяет в дальнейшем на основе статистических методов обработки получать объективные характеристики исследуемых объектов. Благодаря этому растровые модели могут использоваться для изучения новых явлений, о которых не накоплен материал. В силу простоты этот способ получил наибольшее распространение;
- растровые данные проще для обработки по параллельным алгоритмам и этим обеспечивают более высокое быстродействие по сравнению с векторными;
- некоторые задачи, например создание буферной зоны, много проще решать в растровом виде;
- многие растровые модели позволяют вводить векторные данные, в то время как обратная процедура весьма затруднительна для векторных моделей;
- процессы растеризации много проще алгоритмически, чем процессы векторизации, которые зачастую требуют экспертных решений.

Данные для анализа могут быть получены из векторных слоев, отражающих поля тематических или/и временных характеристик, растеризацией и записаны в таблицу или напрямую занесены туда из отчетов. Таблица, содержащая атрибуты объектов, называется таблицей атрибутов. В таблице каждому объекту соответствует строка таблицы, каждому тематическому признаку – столбец таблицы. Каждая клетка таблицы отражает значение определенного признака для определенного

объекта.

В общем случае ввод информации для задач ГИС осуществляется комплексно: по данным дистанционного зондирования, со снимков спутников, аэроснимков, по материалам дешифрирования снимков, полевым измерениям, по информации с карт.

2.3 Основные идеи метода анализа данных в ГИС с помощью искусственных нейронных сетей

Далеко не все ГИС снабжены возможностями специализированного анализа, например геологического. Связано это с тем, что четкой схемы проведения таких работ, не существует и организации, занимающиеся ими, предпочитают производить анализ по собственным методикам и правилам. Работа со специфическими данными специфическим образом является характерной чертой этого типа анализа. Кроме того, взгляды на приемы его проведения могут меняться с течением времени. Поэтому такие возможности в ГИС представляются средствами создания приложений самими пользователями. Сложность состоит в том, для каждой специализированной области возникает необходимость создавать отдельное приложение к ГИС и часто даже свою методику обработки. Это не всегда возможно и часто дорого.

Нейронные сети претендуют на то, чтобы стать универсальным аппаратом решающим разные специфические задачи из разных проблемных областей в ГИС [33, 34]. Такая универсальность обуславливается тем, что нейросети дают стандартный способ решения многих нестандартных задач [35]. И неважно, что специализированная программа решит лучше один класс задач. Важнее что один нейромимитатор решит и эту задачу и другую и третью и не надо каждый раз создавать специализированные приложения для каждой специфической задачи [36].

Обобщение задач

Как правило, модули, реализующие специализированный анализ для разных проблемных областей, решают одинаковые качественные задачи. Перейдя от специфических частных аналитических задач к общему видению проблемы в целом можно увидеть одно важное обстоятельство. А именно, что большинство аналитических задач сводится к одной проблеме, которая легко формулируется, но сложно решается: к проблеме заполнения пропусков в таблице [35, 37-44].

Учитывая то, что часто методика обработки неизвестна, с этой задачей справляются лучше всего нейронные сети, которые позволяют строить эмпирические зависимости [45, 46] без привлечения дополнительной информации. Проблема заполнения пропусков в таблице тесно связана с задачами, такими как построение отношений на множестве объектов и построение функции по конечному набору значений [35, 47-55]. В такой постановке преследуемая цель – это восстановление пропущенных данных. В нашем случае наиболее общим способом проблема восстановления пропущенных данных формулируется как построение

(дополнение) одного из слоев по информации, имеющейся в других слоях карты. В такой постановке она является решением большинства классификационных задач в ГИС [23, 24].

Методы классификации используются в решении следующих основных задач:

- классификация процессов и явлений;
- районирование, типология;
- выявление определяющих факторов;
- временной анализ;
- интерполяция и создание моделей поверхности
- анализ и прогнозное картирование пространственно распределенных данных и т.д.

Формальная постановка

Пусть, существует набор пространственных данных (сеть мониторинга). Обычно, данные представляются в виде: X , Y – пространственные координаты, Z – зависящая от них переменная. Задачей картирования пространственных данных, как правило, является интерполяция неравномерных данных Z на равномерной координатной сетке.

Как уже отмечалось в первой главе, существует три вида постановки задач относительно расположения географических явлений в пространстве (рис.2.1).

Для исследователя географических комплексов интерес представляют все три модели. В данной работе акцент сделан на первой и второй модели, поскольку конечной целью обработки данных является

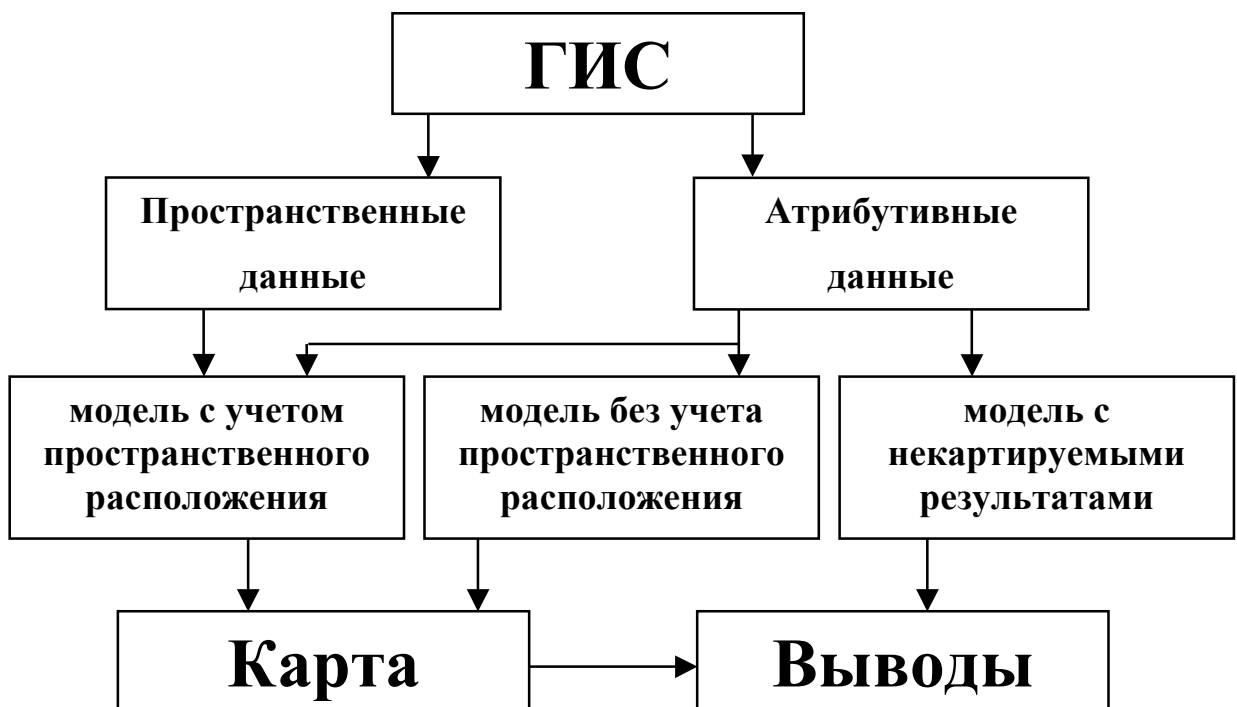


Рис. 2.1. Способы анализа данных ГИС

получение карты.

Рассмотрим более общий случай и введем фундаментальное предположение о фиксированной «вертикальной» связи между слоями. Пусть, как и прежде, существует набор пространственных данных. Предположим что каждую точку сетки с координатами (x, y) характеризует некоторый вектор состояний (z_1, \dots, z_n) . Для всей сетки получаем набор векторов Z_1, \dots, Z_n – параметров в точках сетки мониторинга. Часть параметров – координаты. В общем случае пространственное положение может быть выражено через относительные единицы, например, как обратно пропорциональное квадрату расстояний между объектами.

Данные легко представимы в виде двумерной таблицы, в которой столбцы – это структура параметров Z_1, \dots, Z_n , а строки точки сетки. Показатели состояния Z_1, \dots, Z_n разделяются на входные переменные C_i ($i = 1, \dots, p$), полученные тем или иным способом, и выходные D_j ($j = 1, \dots, q$), ($p + q = n$) – те которые нужно выразить через входные. Т.е. найти функционал: $D_j = F(C_1, \dots, C_p)$, ($j = 1, \dots, q$). Выходные параметры могут быть какими-либо известными классификациями, непрерывными измерениями или другими значениями. Каждый параметр Z_k может быть отдельным атрибутивным слоем в ГИС. Предположение о фиксированной "вертикальной" зависимости между слоями состоит в следующем.

Упростим задачу, сведя ее к классификации на два класса с одним выходным параметром. Если такое разделение возможно, то F является решающим правилом отнесения к одному из классов в зависимости от входных данных. Следовательно, отнесение точки сетки к классу зависит только от параметров самой точки и не зависит напрямую от соседних точек. Все связи между соседями, в том числе пространственное положение, можно закодировать в параметрах Z_1, \dots, Z_n для каждой точки. Для большинства классификационных задач, например, поиска полезных ископаемых по косвенным признакам, прямую пространственную привязку можно исключить. Это позволяет знания о уже разведанной территории переносить на исследуемую. Естественно с учетом некоторой схожести. Пример – заполнение пробелов в данных на исследуемой территории.

Основная задача, которую нужно решить – это задача заполнения пробелов (восстановления, предсказания) в выходных параметрах. Иными словами построение (дополнение) одного или нескольких слоев по информации, имеющейся в других слоях карты. Попутно, возникает ряд проблем, связанных с заполнением пропусков данных во входных параметрах, исключением незначимых для решения основной задачи признаков и других.

2.4 Аналитические задачи в ГИС, решаемые с помощью искусственных нейронных сетей

Опишем круг задач, требующих решения в ГИС, для которых могут быть использованы нейросетевые технологии.

Построение (дополнение) слоя

Основная задача, к которой, так или иначе, относятся остальные,

описанные ниже, это построение слоя. Она означает заполнение его недостающих частей (или построение слоя полностью) по информации, имеющейся в других слоях, на основе нахождения некоторой функциональной зависимости между параметрами, полученными эмпирическим путем, и скрытыми теоретическими параметрами, определяющими существенные характеристики каждой конкретной точки.

Даны слои качественных характеристик одной и той же территории. Слой, который необходимо восстановить, известен частично. Для восстановления слоя при обучении нейросети используется только та информация из слоев, которая покрывает известные участки слоя с пробелами. После обучения можно распространить знания о зависимости между слоями на отсутствующие области карты. Получившиеся знания обладают переносимостью за рамки данной территории. Все описанные ниже задачи можно рассматривать как частный случай данной.

Классификационные задачи. Поскольку при сборе информации для БД приходится иметь дело с результатами измерений, определим по этому показателю три типа задач классификации.

К задачам классификации первого типа относятся те, в которых исходные измерения требуется разделить на устойчивые группы. Их называют задачами классификации без учителя, кластеризации, таксономии, типизации [56-62]. Этот тип классификации применяется для обработки опытных данных.

Задачи классификации второго типа характеризуются тем, что исходные данные уже сгруппированы и требуется оценить их информативность (значимость) относительно совокупности известных эталонов. Такого рода задачи встречаются при распознавании образов [63-65], дешифрировании снимков и т.д.

Задачи классификации третьего типа – задачи разбиения. В них исходные измерения или их функции требуется разбить на устойчивые группы в зависимости от их величины (типичный пример – зонирование) [23, 30, 66].

В ГИС задачи классификации первого типа возникают и решаются при разработке классификаторов, т.е. при организации информационной основы, задачи второго типа – при сборе первичных данных и при использовании ГИС для экспертных решений или оценок. Задачи классификации третьего типа возникают в приложениях ГИС для решения проблем в области экологии, землепользования, статистики и т.п.

Восстановление легенды слоя

Вторая решаемая задача – восстановление легенды. Классификация с учителем – генерация объектов слоя по заданным классификационным правилам. Правила задаются во время обучения нейросети и остаются скрытыми от пользователя. Пользователь имеет возможность задавать, по его мнению, полезные для классификации признаки, выбрав слои, участвующие в обучении. Типичная задача поиск полезных ископаемых по косвенным признакам. Эта задача решается на основе информации об уже

разведанных месторождениях и полевых съемках косвенных признаков. Знания, полученные при обучении, переносимы на другую территорию с известными косвенными признаками.

Районирование и типология

Зонирование. Основное назначение функций этой группы состоит в построении новых объектов – зон до того на карте не существовавших, т.е. участков территорий, однородных в смысле некоторого критерия или группы критериев. Границы зон могут либо совпадать с границами ранее существовавших объектов (задача определения "нарезки" избирательных округов по сетке квартального деления), либо строиться в результате различных видов моделирования (зоны экологического риска). Типичные задачи этого типа: выделение зон градостроительной ценности территорий, зон экологического риска, зонирование урбанизированных территорий по транспортной доступности, построение зон обслуживания поликлиник и т.д. Работа может производиться как с растровыми, так и с векторными изображениями.

В сущности, зонирование это – классификация без учителя. Задан набор объектов, каждому объекту сопоставлен вектор значений признаков (строка таблицы). Требуется разбить эти объекты на классы эквивалентности.

Отнесение объекта к классу проводится путем его сравнения с типичными элементами разных классов и выбора ближайшего. Простейшая мера близости объектов – квадрат евклидоваго расстояния между векторами значений их признаков (чем меньше расстояние, тем ближе объекты). Соответствующее определение признаков типичного объекта – среднее арифметическое значение признаков по выборке, представляющей класс. Другая мера близости, естественно возникающая при обработке сигналов, изображений и т.п. – квадрат коэффициента корреляции (чем он больше, тем ближе объекты). Возможны и иные варианты – все зависит от задачи. Для каждого нового объекта нужно выполнить два действия:

- 1) найти класс, к которому он принадлежит;
- 2) использовать новую информацию, полученную об этом объекте, для исправления (коррекции) правил классификации.

В результате классификации как бы появляются новые имена и правила их присвоения.

Создание моделей поверхностей

Создание моделей поверхностей – это и построение моделей изолинейных изображений по регулярным и нерегулярным сеткам и создание модели трехмерной визуализации, например, построение панорамы города в аксонометрической или иной проекции. Расчет производится по содержащимся в базах данных численным характеристикам. Моделироваться могут, как изображения действительного рельефа или непрерывного поля, современного или с учетом динамических изменений, так и воображаемые поверхности,

построенные по одному или нескольким показателям, например, поверхность цен на землю, плотность дорожной сети или населения и т.п.

Интерполяция и прогнозное картирование

Задача – интерполяция пространственно распределенных данных. Сводится к задаче построения функции по конечному набору значений и как следствие к задаче заполнения пробелов. Цель – извлечение максимума информации из набора данных, учитывая возможные ошибки измерений, неравномерную плотность сетки мониторинга, и прочие помехи, встречающиеся при реальных измерениях. Данные по окружающей среде обладают неоднородностью как на крупных, так и на мелких масштабах, что затрудняет анализ. Нейросетевая обработка обладает рядом преимуществ перед детерминистическими моделями.

Временной анализ

Временной анализ растровых изображений. В качестве таких изображений в ГИС обычно выступают снимки или растеризованные векторные изображения. Преимущество снимков – в их современности и достоверности, поэтому часто встречающийся вид анализа в этой группе – временной. Сравняются и ищутся различия между снимками различной давности, таким образом, оценивается динамика произошедших изменений. Не менее часто анализируются пространственные взаимосвязи двух или нескольких явлений.

Анализ временных рядов содержит комплекс задач, которые сводятся к построению функций по конечным наборам значений и заполнению пробелов в таблицах. Временные ряды представляют собой специальный вид таблиц и заслуживают отдельного рассмотрения. Для каждого типа объектов выделяется набор постоянных признаков (констант) и множество свойств, меняющихся со временем (переменных признаков). Предполагается, что в любой момент времени для каждого объекта существуют свои значения переменных признаков. Вот, например, три задачи, специфичные для обработки временных рядов:

а) определение констант (всех или части) по известным значениям переменных в разные моменты времени;

б) предсказание значений части переменных в некоторый момент времени по известным значениям констант, переменных в нескольких предшествовавших моментах времени и части переменных в текущий момент;

в) определение объема данных о прошлом, достаточных для предсказания будущего на конкретное время и с заданной точностью.

Обычная задача при временном анализе – получение прогноза. Легко заметить, что решение такой задачи немногим отличается от решения задачи по восполнению пробелов в слое на основе информации, заключенной в других слоях. Единственное концептуальное отличие состоит в том, что слои вместо разных пространственных признаков содержат изменение во времени одного и того же слоя. Для примера, упростим описание анализа временных рядов. Возьмем за основу

известные слои одной и той же территории в количестве $N-1$ (без последнего), а в качестве восстанавливаемого слой с номером N . Произведем обучение нейросети. Для прогноза в качестве входных параметров возьмем слои в количестве $N-1$ (без первого) и подадим на вход нейросети. На выходе получим прогнозируемый слой $N+1$.

Выбор значимых признаков

Анализ значимости. Как уже было сказано, основной задачей является восполнение пробелов в данных и решается она применительно к данной области построением слоя по слою (или нескольким слоям). При этом исследуется вопрос, какие из входных сигналов являются доминирующими, (значимыми) при принятии нейросетью решения, а какие нет. Другими словами, насколько каждый слой участвующий в построении влияет на восполнение пробелов. Такая информация дает знание, например, о том какие признаки можно убрать из рассмотрения, а какие оставить. То есть решается задача нахождения оптимального набора исходных показателей, которые полностью описывают изучаемые явления. Это может помочь в понимании сущности географического комплекса.

Значимость по слою складывается из значимости точек сетки. Благодаря такой информации можно видеть, какие области из слоев участвующих в качестве входов были значимы при построении. Таким образом, получаем представление о территориальном распределении значимости.

Нейросетевые технологии анализа данных решают такие задачи и позволяют помочь в оптимальном выборе системы исходных показателей, исследовать признаки на дублирование, выяснить значимость исходных признаков для решения основной задачи.

Подробнее все поставленные задачи и методы их решения в нейросетевом базисе описаны в третьей главе.

2.5 Основные идеи визуализации и анализа данных произвольной природы

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 можно называть не только трехмерное пространство, в котором мы существуем, но и любое абстрактное пространство произвольной размерности.

Это свойство ГИС является определяющим для использования предлагаемого подхода визуализации данных, поскольку основное качество ГИС – это наглядность. Кроме того, современные ГИС имеют множество мощных инструментов для анализа. Отображение в готовых ГИС произвольных данных позволяет подключить для визуализации и анализа весь накопленный арсенал средств обработки пространственной информации.

Предлагаемый новый подход позволяет отображать многомерные

данные в общем случае различной природы и не обязательно числовые [19, 67, 68]. Например, можно представлять патенты, статьи, курсы акций, временные ряды, ... Из данных создается карта, куда они будут нанесены. Карта это трансформируемый объект – атлас и на этом атласе можно отразить существенные детали данных, дающие представление об их структуре.

Описание задачи

Обычно подразумевается, что данные при нанесении на карту отображаются на какую-либо подложку. Например, какие-либо характеристики накладываются на территорию. Посмотрим на это с новой стороны и предположим, что подложка порождается с использованием самих данных. Отобразив данные специальным образом, мы получим их визуальное представление в виде некоторого многообразия, например, в виде пленки натянутой на многомерные данные. Пленку затем спроецируем на поверхность. Данные же отображаются ближайшей точкой пленки. В результате и получим подложку для данных, функций над данными, производных, показателей значимости, классификаций, отношений данных, различных тематик и др. [69, 70]

Такая визуализация многомерных данных осуществляется понижением размерности с сохранением некоторых специфических особенностей исходного пространства данных.

Картографирование является прямым следствием визуализации данных [71]. Объекты, с которыми оперирует метод, отображаются в слои, если угодно одной территории данных, с которыми можно проводить аналитические операции, принятые в ГИС. Если существует географическая привязка, то ничто не мешает отображать результат анализа дополнительно ко всему прочему на реальные территории.

Перейдем к формальному описанию задачи визуализации и картографирования данных. Она заключается в отображении многомерных данных в представимую человеком размерность, например, на плоскость так, чтобы точки данных, близкие на плоскости (на карте), были близки и в исходном пространстве (обратное в общем случае неверно).

Понятно, что, визуализируя данные, мы можем получать большое количество информации о них без какой-либо обработки. Становятся видимыми области группировки данных и разреженные области. Например, упрощается решение задач классификации. Видно количество кластеров, их форма, взаимное расположение и т.д. Обратим внимание, что это естественная классификация данных.

Заметим, однако, что все это видно когда данные отображаются на многообразия малой размерности. Размерности один, два максимум три. Типичные же данные при решении серьезных задач это, например, 100 мерное пространство и 100 – 100000 точек в нем. Даже если размерность или объем выборки меньше, то все равно осмысленно ее представить человек не в состоянии за исключением двух- или трехмерных.

Для дальнейшего изложения необходимо описать сущности, с

которыми оперирует метод.

Объекты метода

Данные

Данные, которые можно картографировать могут быть любые, т.е. все. Слово "все" выступает здесь в трех значениях.

Все данные как любая информация о мире. Предлагаемый подход позволяет строить отображение многомерных данных, заключенных в таблицах, в "человеческом" виде. Отсюда следует, что если данные можно представить в таблице, то они могут быть картографированы. А такой информации в мире большинство.

Все для одной задачи, без изъятия, целиком, в полном составе. Нейросетевые технологии позволяют решать такую задачу как определение значимости входной информации для решения задачи. Поэтому можно давать все собранные данные по задаче и получать сокращенный набор признаков, факторов необходимых для ее решения.

Все в смысле, какие есть, с пробелами и неполные. Нейросетевые методы позволяют заполнять пробелы в данных [35, 72]. Для заполнения пропусков, как правило, решается или задача построения функции по конечному набору значений или задача построения отношений на множестве объектов. Для этого могут использоваться разные методы, например, линейная регрессия, транспонированная регрессия, нейросетевая нелинейная регрессия, линейный и квазилинейный факторный анализ, мозаичная регрессия [47, 48, 73-76].

В некотором смысле любая обработка данных заполняет области незнания. Считается, что нейросети делают это "хорошо". И большое преимущество имеют в области плохо формализуемых и нестандартных задач, а также в тех случаях, когда плотность пробелов высока, расположены они нерегулярно, а данных немного, например, число объектов (строк) примерно таково же, как и число признаков (столбцов).

Предложенный алгоритм картографирования данных большой размерности не требует предварительного априорного заполнения пробелов. В общем случае может быть несколько вариантов работы с неполными данными. Отображать с пропусками – в этом случае пропущенные данные не влияют на построение поверхности (карты). Заполнить пропуски перед отображением. Заполнить во время и путем отображения многомерного пространства данных.

Чаще всего данные должны быть предварительно нормированы (обезразмерены) – переходом в каждом столбце таблицы к "естественной" единице измерения. Обычно нормировка производится на единичное среднеквадратичное отклонение в столбцах или на единичный разброс данных в каждом столбце (если нет каких-либо специфических ограничений, связанных со смыслом задачи).

Графически облака данных представляются точками на одной из координатных плоскостей базового пространства. Об информативном

отображении данных будет сказано дальше.

Многообразия

Существует многомерное облако данных. Многообразия это построенные в этом облаке поверхности малой размерности, приближающие его.

Некоторое представление может дать описание самоорганизующейся карты (Self-Organizing Map – SOM). В 1982 году финский ученый Тойво Кохонен [77] предложил ввести в базовое правило обучения нейросети информацию о пространстве. Построение топографических карт (карт Кохонена) является методом, дающим оптимальное представление информации в виде координат двумерной сетки.

В многомерное пространство данных погружается двумерная сетка. Эта сетка изменяет свою форму таким образом, чтобы по возможности точнее аппроксимировать облако данных. Каждой точке данных ставится в соответствие ближайший к ней узел сетки. Таким образом, каждая точка данных получает некоторую координату на сетке.

Такое отображение локально непрерывно: близким точкам на карте соответствуют близкие точки в исходном пространстве (обратное, вообще говоря, не верно: близким точкам в исходном пространстве могут соответствовать далекие точки на карте). Таким образом, распределение данных на двумерной карте позволяет судить о локальной структуре многомерных данных.

Такая топографическая самоорганизующаяся карта дает наглядное представление о структуре данных в многомерном входном пространстве, геометрию которого мы не в состоянии представить себе иным способом. Визуализация многомерной информации является главным применением SOM.

Достоинства SOM начинают проявляться после нанесения на нее какой-либо графической информации. Различные раскраски топографической карты являются удобным средством для выявления взаимосвязей различных факторов. В принципе, любая характеристика порождает свою раскраску карты. Вместе подобные раскраски дают исчерпывающую и наглядную картину. Здесь имеется полная аналогия с географическими картами различных типов на одной и той же географической сетке, которые в совокупности дают полное представление о данной местности.

При построении многообразий можно пользоваться классическим методом главных компонент. Для определенности возьмем двумерный случай. При этом плоскости над данными строятся по двум главным компонентам. Также построения могут вестись по комбинациям пар компонент получающимся в результате дальнейшей обработки.

Предложенная технология моделирует данные (в общем случае – с пробелами) многообразиями (линейными и нелинейными) малой размерности. Для построения многообразий используется линейный метод

главных компонент, квазилинейный метод, надстраиваемый над линейным и использующий его результаты, существенно нелинейный метод, построенный с помощью формализма самоорганизующихся кривых [75]. Разработан метод построения *упругой карты*, моделирующей данные [69-71].

Экстраполяция и интерполяция получаемых зависимостей производится линейно и с помощью формул Карлемана. Метод решает следующие задачи:

- 1) заполнение пробелов в данных;
- 2) ремонт данных – корректировка значений исходных данных так, чтобы наилучшим образом работали построенные модели;
- 3) построение вычислителя, заполняющего пробелы в поступающей на вход строке данных (в предположении, что данные в ней связаны теми же соотношениями, что и в исходной таблице).

Существует еще одна техника получения многообразий малой размерности моделирующих данные называемая "метод узкого горла". От нейросети требуется выдать те же вектора данных, которые были получены на входе, т.е. быть для них прозрачной. Эффект заключается в сокращении числа нейронов среднего слоя нейросети после ее обучения. Сеть таким образом можно разделить пополам по ее узкой части. Это будет напоминать кодирование-декодирование данных. Сократив до возможного минимума средний слой нейросети, получим на ее выходе внутренние координаты данных. По ним можно строить многообразия.

Применение многообразий малой размерности требует постановка задачи, а именно, визуально представить данные в естественном для человека виде. Иначе картография данных не имеет смысла. Поскольку не только отобразить, но и представить многомерное пространство данных в реальных задачах не представляется возможным.

Подведем итог. Многообразиями малой размерности могут быть в простейшем случае прямые, ломаные и надстройки над ними типа кривых и более сложные плоскости, пленки и упругие карты. Эти объекты располагаются в облаке данных, аппроксимируя их. Необходимо понимать, что вряд ли будет одно универсальное многообразие, поскольку решаемая задача в каждом конкретном случае накладывает условия на приближение. Поэтому их может быть несколько для одного и того же набора данных, и они могут составлять даже сообщества многообразий.

Проекции

С каждым *многообразием* связан проектор на него, с помощью которого данные отображаются на многообразии. Построение этого проектора может вестись различными способами [70].

"Образцовый" путь построения проектора – метод максимума правдоподобия (максимума вероятности, максимума энтропии...). Он предполагает, что плотность вероятности в точке на многообразии больше, чем в тех точках, которые в нее проецируются.

Также и с каждым *отображением данных* в меньшей размерности связано многообразие, погруженное в пространство данных и строящееся из тех же соображений. Эта двойственность "многообразие-проектор" является основой многих преобразований карт данных.

Построенный проектор дает знание об отображении точки данных из исходного пространства данных на многообразие малой размерности. Например, из стомерного пространства на плоскость или кривую. Каждая проекция – это новая подложка для данных и топологические свойства данных меняются от проекции к проекции. Зная правила проектирования, можно отображать дополнительные точки данных уже после построения карты.

Вновь поступившие данные занимают свое место в многомерном пространстве, проектор определяет их место на плоскости. Для прикладных задач это мгновенная классификация. Более того, при изменении некоторых характеристик новой точки данных проектор помогает отследить траекторию движения точки по плоскости. Также может решаться обратная задача, какие свойства и как нужно изменить, чтобы попасть в определенный класс.

Многообразие и проекция – это две взаимосвязанных вещи. У многообразия есть проектор, у каждого проектора многообразие. Иными словами, есть случаи, когда в облаке данных строится многообразие малой размерности, например SOM или *упругая карта*, а затем определяется проекция, отображающая туда данные, и также каждый проектор определяет свое многообразие в данных.

Иньекции

Операция, обратная проекции. Объект, сопутствующий проекции. Оператор, отображающий точки с плоскости в многообразие R^n . Позволяет, выбрав точки на плоскости, узнать, где они находятся в пространстве. Например, определить пространственное расположение точек класса, точек разделяющей поверхности или выбрав на плоскости область, определить характеристики точки которые удовлетворяют условию выбора.

Развертки

Проекция многообразий малой размерности на стандартные многообразия. Многообразия в данных могут быть различной формы от плоских до сферических. Развертки – это отображения многообразий, которые уже имеют "хорошую" размерность, на некоторый набор стандартных, например, на прямую, плоскость, сферу (глобус), тор (глобус в форме бублика) ...

Слои

Теперь, опишем, пожалуй, главную сущность, в которой и заключается смысл картографирования данных. Слои имеют тот же смысл

что и слои в ГИС более того это они и есть. Существенная разница в том, что они отображают. Отображаться в слоях могут как все вышеописанные сущности, так и дополнительные характеристики данных которые и составляют основное информационное содержание карты [19].

Данные представляют собой при отображении точки. Они образуют в карте точечный слой. Многообразия отображаются сетками и образуют слои сеток. Проекции и инъекции выглядят как прямые, соединяющие точки данных с соответствующими узлами на многообразиях. Развертки образуют топографическую основу карты.

Удобным инструментом визуализации данных является раскраска описанных объектов аналогично тому, как это делают на обычных географических картах. Порождать свою раскраску ячеек сетки, проекций, данных и др. объектов могут различные характеристики данных. Это могут быть известные классификационные признаки, значимости, зависимости, производные. Любые функции над данными могут служить основой для раскраски.

Собрав воедино карты всех интересующих нас признаков, получаем топографический атлас, дающий интегральное представление о структуре многомерных данных.

3. Нейросетевой анализ связи между слоями

3.1. Методы нейросетевого анализа связей между слоями

3.1.1. Проблема построения и использования нейросетей в геоинформационных системах

Общая постановка задачи

Опираясь на описание проблемы восстановления пропусков в таблицах, предлагается соединить методы позволяющие делать это наиболее эффективно с программно-инструментальным комплексом, визуализирующим результаты работы. Здесь и далее речь идет об использовании нейросетевых методов обработки информации применительно к географическим информационным системам.

Исходя из предпосылки, что часто даже большое количество информации не может помочь решить проблему, пока она не будет визуализирована на карте, мы приходим к необходимости широкого использования ГИС для обработки и анализа картографических данных. Круг ее возможных потребителей чрезвычайно широк. Прежде всего, это, конечно, управляющие структуры, владеющие большими массивами информации, на основе которых принимаются решения. В картографических данных также нуждаются специалисты, оценивающие и прогнозирующие состояние какой-либо области человеческой деятельности, например, рынков сбыта продукции, загрязнения территории и т.п. Хотя хранящаяся в ГИС информация и представляет собой основную ценность, она приносит практическую пользу только при решении прикладных задач.

В данной работе рассматриваются нейросетевые модели, предназначенные для решения задач относящихся к классу проблем построения функции по конечному набору значений или построение отношений на множестве объектов [35, 78]. К этому классу относятся распространенные и актуальные в ГИС задачи классификации процессов и явлений, районирования и типологии, временной анализ географических комплексов, интерполяция и создание моделей поверхности, анализ и прогнозное картирование пространственно распределенных данных [17, 24]. Такие задачи формализуются как задачи построения действительной функции по конечному набору значений, классификации, анализа временных рядов, выявления зависимостей в данных.

Общая постановка таких задач может быть представлена в следующем виде. Существует набор переменных, описывающих состояние объекта или явления (входных переменных), необходимо найти значения некоторых целевых параметров (выходных переменных). То есть формально. Дано: объект или процесс, который является предметом

исследования. Требуется: Получить значение некоторого зависимого параметра, который характеризует состояние объекта с точки зрения целей исследования.

В [35] подобный тип задач определяется как задача *заполнения пробелов в таблицах данных*. Такое определение предполагает, что постановка задачи может быть представлена в виде таблицы с неизвестными значениями некоторых целевых параметров. Обычно поля таблицы соответствуют выделенным признакам, описывающим объект или процесс, а записи – конкретным примерам проявления этих признаков.

Задача заполнения пробелов в данных в свою очередь порождает задачу выбора метода восстановления. В работе рассматривается нейросетевой метод восстановления информации [11, 35, 48, 76, 79-82].

Для обеспечения эффективного использования нейронных сетей в ГИС необходимо рассмотреть вопросы интеграции и взаимодействия нейросетевых моделей и существующих ГИС.

Проблема построения и использования нейросетевых моделей в ГИС

Проблему применения нейросетевых компонент в ГИС, так же как и в любой прикладной информационной системе можно рассматривать как совокупность следующих проблем:

- Проблема *программной интеграции искусственных нейронных сетей и геоинформационных систем*. Определяет вопросы, связанные с разработкой методов и схем взаимодействия нейросетевых компонент и ГИС, организацией обмена данными и системы запросов между компонентами.

- Проблема *создания нейросетевых моделей в составе геоинформационной системы*. Включает разработку технологии построения нейросетевых моделей, разработку методов автоматизации процесса построения нейросети.

- Проблема *использования нейросетевых моделей в составе геоинформационной системы*. К этой проблеме можно отнести обеспечение устойчивого функционирования, повышение «прозрачности» работы нейросети, получение дополнительной информации о модели, оценку качества работы сети.

- Проблема *технической реализации нейросетевых компонент*. Проблема технической реализации состоит в определении средств построения нейросетевой компоненты, разработке программной системы и обеспечении информационного, программного и технологического соответствия систем.

Проблема интеграции НС и ГИС

ГИС давно уже перестали быть чисто научными инструментами исследователя. Геоинформатика – наука прикладная, решающая проблемы других, тематических областей. Даже самый полнофункциональный ГИС

не может учесть потребности всех и каждого. Реальные же применения имеют свою специфику, которая может сильно расходиться с тем, что предполагал разработчик системы. *Наращиваемая функциональная часть* – одна из важнейших черт современных геоинформационных систем. Можно создать свою, новую функцию, соединяющую сотню уже существующих в ГИС.

Проблема интеграции ИНС и ГИС может быть решена, по крайней мере, тремя способами (рис.3.1):

- 1) интеграция НС моделей в ГИС;
- 2) развитие интерфейса между ИНС и ГИС, как самостоятельными системами;
- 3) создание НС систем с включением интерфейса взаимодействия с ГИС.

Особенности и различия интеграции определяются возможностями ГИС такими как: встроенный язык программирования; средства DDE и OLE; функциональные DLL. Все эти способы требуют написания ГИС-приложения.

ГИС-приложения – специально разработанные для решения каких-то

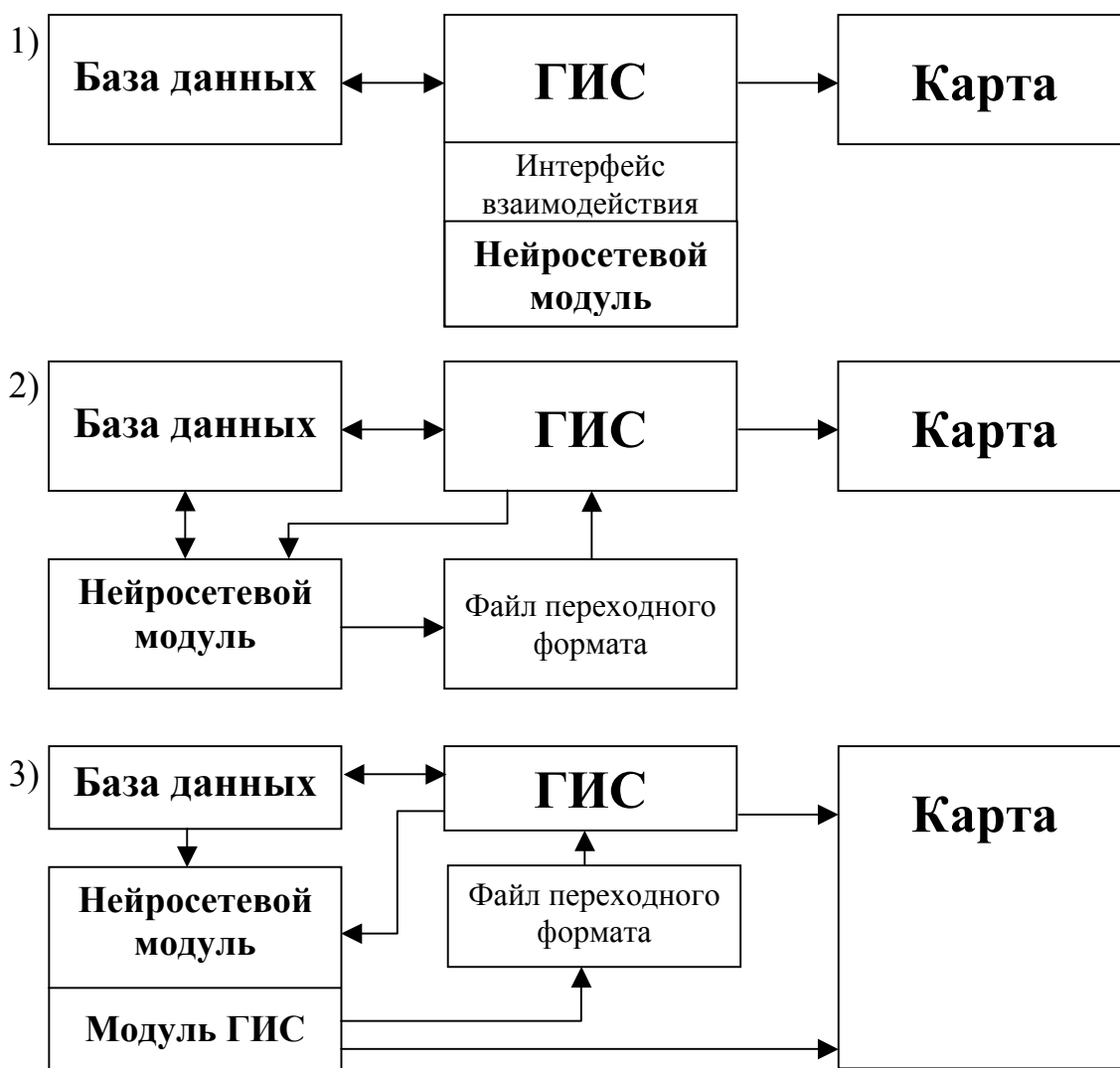


Рис. 3.1. Способы интеграции ИНС и ГИС

конкретных задач алгоритмы обработки данных. Если обобщить известные средства *создания приложений*, их можно четко разделить на две неравные группы.

Первая – преобладающая – располагает собственной, встроенной средой разработки, имеет свой оригинальный язык программирования. Это ArcView, MapInfo, Sinteks. Другая (меньшая) часть ГИС только помогает разработчику создать геоинформационное приложение, а среды разработки в себе не несет. GIS Component (Геоконструктор) GeoGraph –

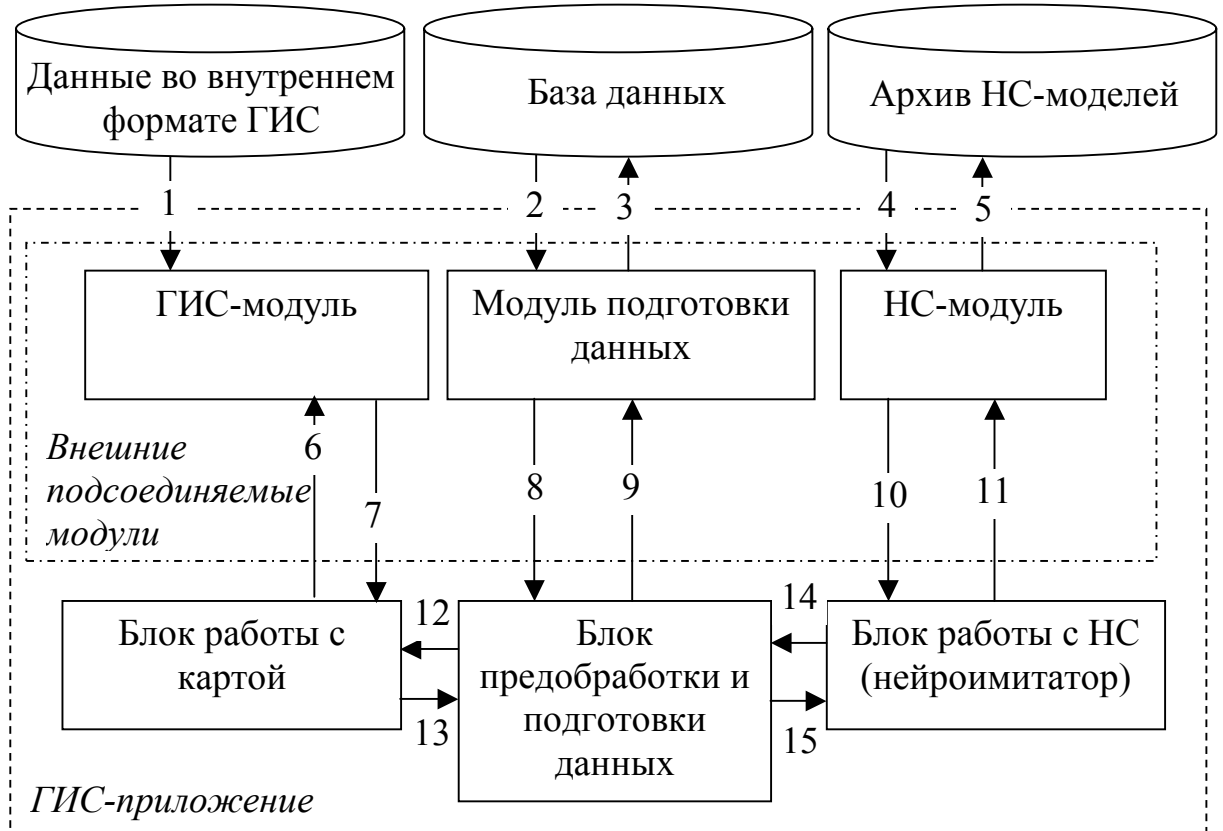


Рис. 3.2. Схема интеграции ИС и ГИС.

это библиотека для Visual-сред программирования (Visual C, Delphi); WinGIS работает в режиме DDE-сервера, обслуживая запросы внешних программ. При этом относительно просто достигнуть высокой производительности, программа ориентирована на конкретную архитектуру системы.

В обеих перечисленных группах в последние годы наметились следующие тенденции: переход к объектно-ориентированному методу, ориентация на непрофессионала даже в плане создания приложений, объединение максимума потенциальных возможностей (а именно – обмена с GPS, доступа к "большим" БД, средств импорта из форматов множества других систем) в рамках единой, интегрированной среды.

В работе используется подход, реализующий создание ГИС-приложения на основе библиотеки компонент ГИС (рис.3.2). Программная система созданная по такой технологии может осуществлять стандартный для ГИС набор операций с картой и не имеет ограничения на расширяемость модулей для решения аналитических задач.

Потоки данных между блоками:

1. Файлы карты во внутреннем формате ГИС-модуля.
- 2-3. Таблицы исходных данных.
- 4-5. Файлы описания нейросетей.
6. Данные для тематических слоев, запросы к карте, географические операции, манипулирование слоями, задание свойств.
7. Данные о карте, слоях, объектах слоя, связях с таблицами атрибутов.
- 8-9. Таблицы исходных данных во внутреннем формате системы, параметры и типы предобработки.
- 10-11. Параметры нейросетей (формирование, обучения, ...).
12. Данные для создания тематических слоев.
13. Данные об объектах слоя и связях с таблицами атрибутов (внутренние идентификаторы).
14. Данные обработанные НС. Результаты в терминах поставленной задачи.
15. Предобработанные географические и табличные данные в формате пригодном для обработки нейросетевой моделью.

Интересным также представляется подход интеграции модулей нейросетевой обработки данных в ГИС средствами самой полнофункциональной ГИС. Расчетные модули представляют собой программные динамически подсоединяемые библиотеки DLL. Механизм обмена данными и интерфейс взаимодействия с нейросетевыми моделями разрабатывается с использованием внутреннего языка геоинформационной системы.

3.1.2. Технология построения нейросетевых моделей в составе геоинформационной системы

Одним из эффективных средств решения слабо формализованных задач на основе примеров являются искусственные нейронные сети (ИНС). Распространение нейронных сетей объясняется следующими достоинствами нейросетевого подхода:

1. Автоматическая настройка параметров нейросетевой модели для решения задачи на примерах. Не требуется участие эксперта для построения модели, решающей задачу.
2. Универсальность. Нейросети позволяют стандартным образом, без учета семантики, решать любые задачи, которые допускают представление в виде набора примеров, содержащих входные и выходные данные [35, 83-86].
3. Устойчивость при работе с зашумленными и недостоверными данными [35, 85, 87, 88].
4. Возможность адаптации (дообучения) к новым условиям.
5. Устойчивость к сбоям и разрушениям элементов.
6. Высокий параллелизм, присущий нейросетевым моделям.
7. Способность эффективно обрабатывать данные высокой размерности, разнотипные данные.

В отличие от «традиционных» статистических методов, нейросети выдают не статистически достоверное, а правдоподобное решение задачи и могут применяться при недостатке эмпирических данных для статистического исследования [35]. В качестве достоинств ИНС по сравнению со статистическим подходом можно назвать универсальность и автоматизированный режим настройки в условиях сильной априорной неопределенности, что позволяет быстро получить приемлемый результат. Нейросетевые модели налагают слабые ограничения на возможные функции распределения переменных и позволяют избегать априорных предположений о виде функций распределения переменных и структуре модели [35, 80].

Нейронные сети особенно продуктивны в решении слабоструктурированных задач, так как обычно эксперт может легко структурировать задачу до уровня «черного ящика» или системы данных на основе методов системного анализа [45, 46, 89, 90], то есть указать входные и выходные параметры системы, не указывая метода решения. Как отмечают исследователи, например [35, 77, 79, 91], для решения реальных задач нейросетевой подход является часто более эффективным.

Получение данных из ГИС

Пусть существует набор пространственных данных (сеть мониторинга). Данные представляются в виде: X , Y – пространственные координаты, Z – зависящая от них переменная. Данные, снятые в узлах сетки, нас интересуют по причине того, что такие данные легко представимы в виде двумерной таблицы. Из которой наиболее просто формируется обучающая выборка.

Обычно, первоначально информация существует в виде сетки мониторинга или отбора проб, поэтому предварительная специальная предобработка не требуется. Если все же данные в ГИС представлены не в сеточном виде (фактически это своеобразный растр) а в векторном то их необходимо предварительно растривать. Растривать поля данных можно по-разному. В зависимости от сути поставленной задачи. После растривания возможно проведение фильтрации для сглаживания, равномеризации или для контраста границ и учета пространственного положения. Алгоритмы получения данных из ГИС описаны в разделе 3.2.1.

В данном разделе остановимся на описании построения нейросетевых моделей для решения задач в ГИС. Основы современной нейроинформатики были заложены в работах [92-94], Теоретические и практические вопросы применения ИНС нашли отражение в работах [35, 36, 75, 78, 80, 85, 95-102]. Наиболее развернуто и последовательно вопрос стандартизации нейросетевых моделей рассмотрен в монографии Е.М. Миркеса [36].

Проект стандарта описан в форме разработки компонент «идеального» нейрокомпьютера. Нейрокомпьютер представлен в виде нескольких модулей, реализующих определенные задачи и взаимодействующих через фиксированный набор запросов. Выделяются следующие компоненты нейрокомпьютера (рис.3.3): задачник, предобработчик, сеть, учитель, интерпретатор, оценка, контрастер, исполнитель.

Остановимся на описании блоков нейросетевой модели необходимых для решения задачи анализа данных в ГИС.

Подготовка обучающей выборки и предобработка данных

При обучении сетей всех видов с использованием любых алгоритмов обучения сети необходимо предъявлять примеры, на которых она обучается решению задачи. Источником данных для сети является задачник.

Формируемый для нейросети задачник представляет собой прямоугольную таблицу, поля которой содержат информацию о входных данных примеров задачи, правильные ответы и другую информацию. Важным вопросом является разделение всех доступных данных на обучающую и тестовую выборки таким образом, чтобы обеспечить их независимость и представительность. Эта проблема решается для каждой конкретной задачи отдельно.

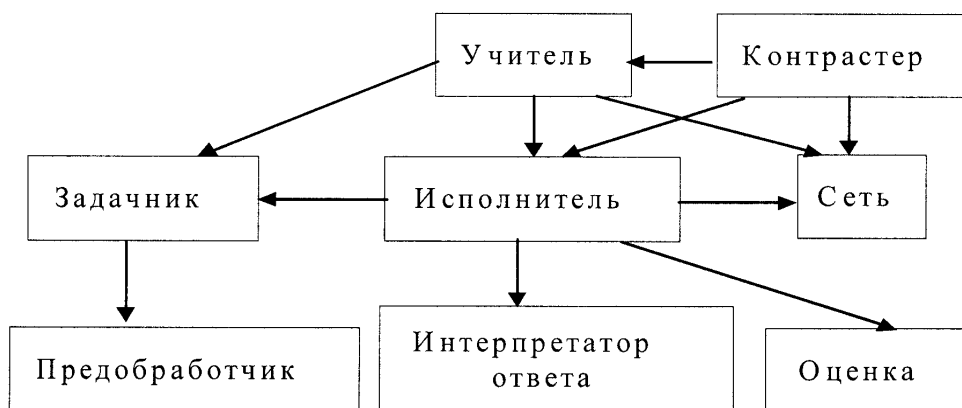


Рис. 3.3. Компоненты нейрокомпьютера

Информация поступает к нейронной сети в виде набора ответов на некоторый список вопросов. Можно выделить три основных типа признаков:

- бинарный признак (возможен только один из ответов – истина или ложь);

- качественный признак (принимает конечное число значений). Для него нельзя ввести осмысленное расстояние между состояниями;
- число.

В процессе обучения параметры подаются на вход сети, значения целевого параметра сравниваются с ее выходом. Однако в большинстве задач использование непосредственных значений из выборки нецелесообразно, т.к. различные переменные имеют различные диапазоны изменения значений, что приводит к затруднениям в процессе настройки весов нейросети. Кроме этого, выборка может содержать как количественные переменные, так и качественные, измеренные в номинальной шкале. Для обеспечения эффективности процесса обучения применяется предварительная обработка данных.

Цель предобработки входных данных – преобразование входных сигналов таким образом, чтобы обеспечить эффективную работу нейронной сети. Для количественных признаков стандартными процедурами предобработки являются нормировка и центрирование, которые обеспечивают универсальность нейронной сети при работе с произвольными данными и позволяют сохранять параметры сети в оптимальном для функционирования диапазоне.

Простейшая предобработка числовых признаков

Числовые сигналы рекомендуется масштабировать и сдвигать так, чтобы весь диапазон значений попадал в диапазон приемлемых входных сигналов. Эта предобработка проста и задается следующей формулой:

$$c' = \frac{(c - c_{min})(b - a)}{(c_{max} - c_{min})} + a, \quad (3.1)$$

где $[a, b]$ - диапазон приемлемых входных сигналов, $[c_{min}, c_{max}]$ – диапазон значений признака c , c' – предобработанный сигнал, который будет подан на вход сети.

Структура нейронной сети

Нейросети могут рассматриваться как разновидность вычислительных моделей. Возможна общематематическая (или алгоритмическая) запись нейронной сети, отражающая функционирование модели, но нейронная сеть может представлять собой довольно сложную модель, число параметров нейронных сетей, применяемых на практике, может составить порядка 1000 – 10000. «Прямое» описание подобной функции, очевидно, неэффективно. Поэтому для описания сетей выработана специальная «схемотехника», выделены элементы и структуры [35, 36], использование которых позволяет сделать описание нейросети более структурированным, компактным и прозрачным.

Нейронная сеть представляется как конструкция, состоящая из элементарных блоков – синапсов, сумматоров, умножителей, нелинейных элементов и точек ветвления.

На рис.3.4 приведены все элементы, необходимые для построения нейронных сетей. Возможно расширение списка нелинейных преобразователей. Вертикальными стрелками обозначены входы параметров (для синапса – синаптических весов или весов связей), а горизонтальными – входные сигналы элементов. С точки зрения функционирования элементов сети сигналы и входные параметры элементов равнозначны. Различие между этими двумя видами параметров

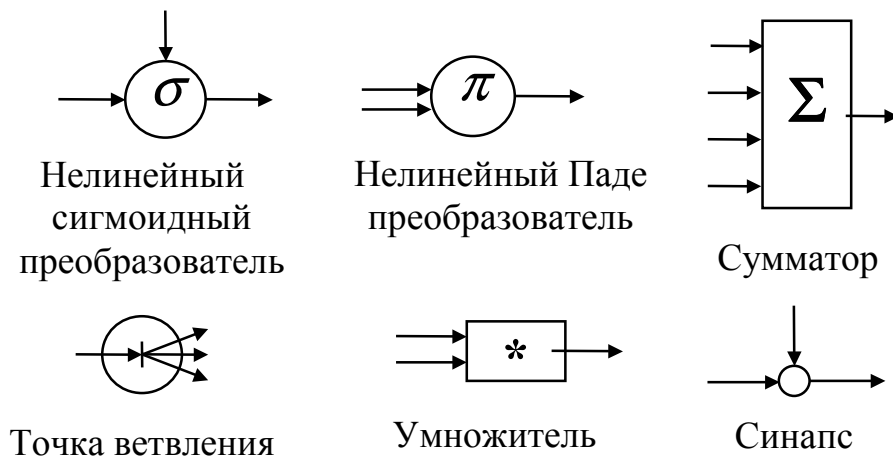


Рис. 3.4. Простейшие элементы сети

относятся к способу их использования в обучении. Также считается, что параметры каждого элемента являются его свойствами и хранятся при нем. Совокупность параметров всех элементов сети называют вектором параметров сети. Совокупность параметров всех синапсов называют вектором обучаемых параметров сети, картой весов связей или синаптической картой. Кроме этого, нужно различать входные сигналы элементов и входные сигналы сети. Они совпадают только для элементов входного слоя сети.

Из приведенных на рис. 3.4 элементов можно построить практически любую нейронную сеть.

Выделим характерные признаки большинства нейросетевых моделей. Прежде всего, нейронные сети являются распределенными структурами и могут быть представлены в виде направленного графа. Узлы представляют процессорные элементы, называемые также нейронами, которые преобразуют информацию в соответствии с некоторой (обычно не сложной) функцией. Ребра задают структуру обмена сигналами между нейронами. Ребрам обычно приписан скалярный параметр – вес связи или синаптический вес. Определены также входные синапсы, по которым подаются компоненты входного сигнала, и выходные, с которых снимается преобразованный сетью сигнал.

В структуре сети можно выделить группы нейронов, функционирующих одновременно и параллельно (слои). В многообразии нейронных сетей можно выделить две базовые архитектуры — слоистые и полносвязные сети. В данной работе используются сети слоистой архитектуры (рис. 3.5).

Нейроны первого слоя получают входные сигналы, преобразуют их и через точки ветвления передают нейронам второго слоя. Далее срабатывает второй слой и т.д. до k -го, который выдает выходные сигналы для интерпретатора и пользователя. Если не оговорено противное, то *каждый* выходной сигнал i -го слоя подается на вход *всех* нейронов $i + 1$ -го. Число нейронов в каждом слое может быть любым и никак заранее не связано с количеством нейронов в других слоях. Стандартный способ подачи входных сигналов: каждый нейрон первого слоя получает все входные сигналы. С такой структурой нейросеть можно представить как сложную вектор-функцию:

$$F^l(\mathbf{a}, \mathbf{b}, \mathbf{x}) = \sum_{i=1}^{n_p} (a_{i,l}^{p+1} \cdot f_i^p(\mathbf{a}^{p-1}, \mathbf{b}^{p-1}, \mathbf{f}^{p-1}(\dots(\dots, \mathbf{f}^2(\mathbf{a}^1, \mathbf{b}^1, \mathbf{f}^1(\mathbf{a}^0, \mathbf{b}^0, \mathbf{x})))))), \quad (3.2)$$

где l – номер компоненты выходного вектора, \mathbf{a}, \mathbf{b} – векторы параметров или весов связей, \mathbf{x} – вектор входных данных или переменных, p – число слоев сети, n_p – число нейронов в p -м слое, $f_i^k(\mathbf{a}, \mathbf{b}, \mathbf{f})$ – функция поведения Паде-нейрона:

$$f_i^{k+1}(\mathbf{a}^k, \mathbf{b}^k, \mathbf{f}^k) = \frac{\alpha_0^k + \sum_{j=1}^{m_k} \alpha_j^k f_j^k}{c + \beta_0^k + \sum_{j=1}^{m_k} \beta_j^k f_j^k} \quad (3.3)$$

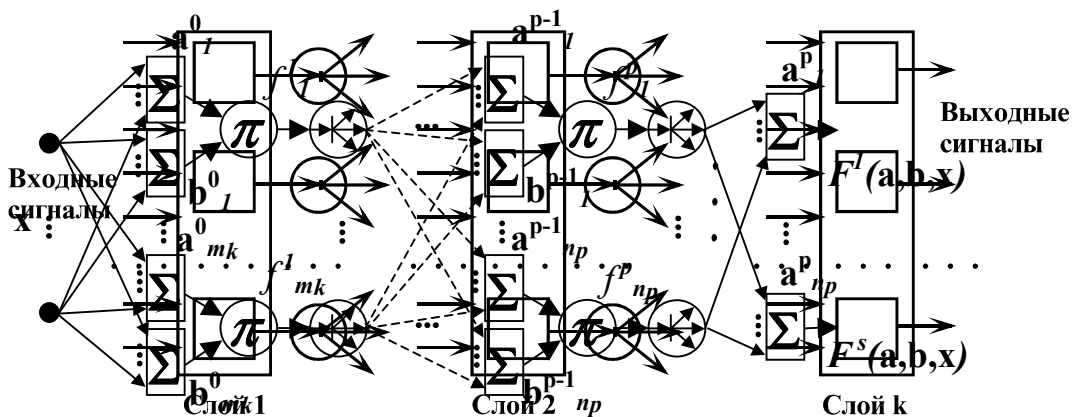


Рис. 3.6. Нейросеть с нелинейными Паде-преобразователями
 Рис. 3.5. Сеть слоистой архитектуры

где $c > 0$, $k = 1 \dots p$ – номер слоя сети, i – номер нейрона в k -м слое. Пример нейросети вида (3.2) представлен на рис.3.6.

Функционирование сети

В работе нейронной сети можно различить два процесса: обучения сети и использования обученной сети. При использовании обученной сети происходит только решение сетью определенной задачи. При этом синаптическая карта сети остается неизменной. Работа сети при решении задачи называется *прямым функционированием*.

При обучении нейронных сетей методом обратного распространения ошибки нейронная сеть (и каждый составляющий ее элемент) должна уметь выполнять обратное функционирование. **Обратным функционированием** называется процесс работы сети, когда на *выход* сети подаются определенные сигналы, которые далее распространяются по тем же связям, что и при прямом функционировании до входа сети. При прохождении сигналов обратного функционирования через элемент с обучаемыми параметрами вычисляются поправки к параметрам этого элемента. Если на выход сети с непрерывными элементами подается производная некоторой функции F от выходных сигналов сети, то вычисляемые сетью поправки являются элементами градиента функции F по обучаемым параметрам сети. Исходя из этого требования, определяются правила обратного функционирования для элементов сети.

Покажем это для одного элемента сети: Паде нейрона.

Нелинейный Паде преобразователь

Нелинейный Паде преобразователь или Паде элемент имеет два входных сигнала и один выходной. Обозначим входные сигналы через A, B . Тогда выходной сигнал Паде элемента равен A/B (рис. 3.7).

При обратном функционировании на выход Паде элемента подается сигнал $\partial F / \partial(A/B)$. На входах сигналов A и B должны быть получены сигналы обратного функционирования, равные

$$\frac{\partial F}{\partial A} = \frac{\partial F}{\partial(A/B)} \frac{\partial(A/B)}{\partial A} = \frac{1}{B} \frac{\partial F}{\partial(A/B)} \quad (\text{рис 3.7a}) \quad \text{и} \quad (3.4)$$

$$\frac{\partial F}{\partial B} = \frac{\partial F}{\partial(A/B)} \frac{\partial(A/B)}{\partial B} = -\frac{A}{B^2} \frac{\partial F}{\partial(A/B)}, \quad \text{соответственно (рис. 3.7б)}. \quad (3.5)$$

Различные нелинейные преобразователи могут иметь различное количество параметров и различную структуру их описания. Так, например, при n входных сигналах, стандартный адаптивный сумматор

имеет $n + 1$ параметр – $y = \alpha_0 + \sum_{i=1}^n \alpha_i x_i$, Паде-нейрон – $2n + 2$ параметра –

$$y = \frac{\alpha_0 + \sum_{i=1}^n \alpha_i x_i}{c + \beta_0 + \sum_{i=1}^n \beta_i x_i}. \quad \text{Соответственно для сети из } k \text{ слоев, } m \text{ нейронов в}$$

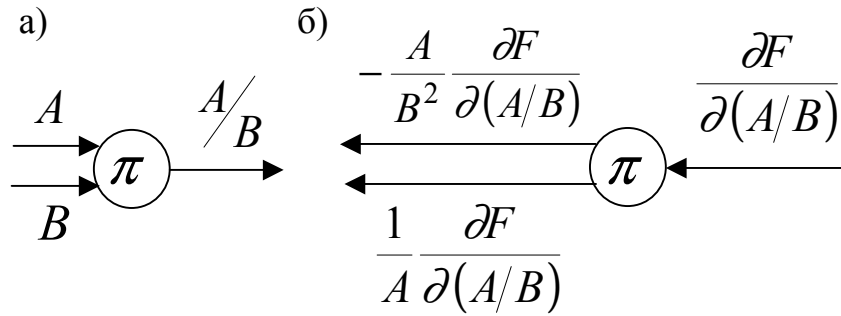


Рис. 3.7. Прямое (а) и обратное (б) функционирование нелинейного Паде преобразователя

каждом слое и n входных сигналах количество параметров без учета выходного сумматора равно $2m((n+1)+(m+1)(k-1))$.

Исследуем вопрос о степени полиномов в числителе и знаменателе отношения в зависимости от числа слоев в нейросети.

На выходе первого слоя $n^{(1)} = \frac{(a_i, x)}{(b_i, x)}$ имеем отношение полиномов

первой степени. На выходе второго слоя

$$n^{(2)} = \frac{(A, n_i^{(1)})}{(B, n_i^{(1)})} = \frac{\sum_i^m A_i \frac{(a_i, x)}{(b_i, x)}}{\sum_i^m B_i \frac{(a_i, x)}{(b_i, x)}} = \frac{\sum_i^m A_i (a_i, x) \prod_{j \neq i} (b_j, x)}{\sum_i^m B_i (a_i, x) \prod_{j \neq i} (b_j, x)} \quad (3.6)$$

отношение полиномов степени m . На выходе третьего слоя полиномы будут степени m^2 .

Оптимизационное обучение нейронных сетей

Представить обучение нейронных сетей как задачу оптимизации можно тогда, когда удастся *оценить* работу сети. Это означает, что можно указать, хорошо или плохо сеть решает поставленные ей задачи и оценить это количественно. Для этого строится *функция оценки*. Задание функции оценки определенного вида позволяет вычислить для каждого примера обучающей выборки оценку работы нейросети и определить направление изменения ее параметров для уменьшения значения оценки H . Процесс итерационного изменения параметров нейросетевой функции называется обучением. Выбор нейросетевого решателя в виде функции (3.2) позволяет применить градиентные методы оптимизации для поиска такого вектора параметров, который доставляет минимум выбранному функционалу оценки.

Функция оценки, как правило, явно зависит от выходных сигналов сети и неявно (через функционирование) – от всех ее параметров. Простейший и самый распространенный пример оценки – сумма квадратов расстояний от выходных сигналов сети до их требуемых значений:

$$H = \frac{1}{S} \sum_{s=1}^S \left[\frac{1}{2} \sum_{x \in P_{out}} (Y_s^p(a, b, x_s) - Y^*(x_s))^2 \right], \quad (3.7)$$

где H – оценка работы нейросети, $Y_s^p(a, b, x_s)$ – значение p -ой компоненты вектора выходного сигнала нейросети, $Y^*(x)$ – требуемое значение выходного сигнала, S – число примеров.

Устройство, вычисляющее оценку, надстраивается над нейронной сетью, и градиент оценки может быть вычислен с использованием описанного принципа двойственности.

В тех случаях, когда оценка является суммой квадратов ошибок,

$$H = \frac{1}{2} \sum_{x \in P_{out}} (Y(x) - Y^*(x))^2 \quad (3.8)$$

значения независимых переменных двойственного функционирования $\mu(x)$ для нейронов выходного слоя P_{out} при вычислении градиента устанавливаются равными

$$\mu(x) = \frac{\partial H}{\partial Y(x)} = (Y(x) - Y^*(x)) \quad (3.9)$$

на вход при обратном функционировании поступают ошибки выходных сигналов. Такой метод вычисления градиента оценки называется методом обратного распространения ошибок.

Переменные обратного функционирования μ появляются как вспомогательные при вычислении производных сложной функции. Описать смысл двойственных переменных и легко представить двойственное функционирование сетей можно через метод неопределенных множителей Лагранжа. Это хорошо описано в [35, 85]. Покажем метод на примере слоистой ИНС с нелинейными Паде преобразователями. Уравнение функционирования Паде-нейрона:

$$x_i^{k+1} = \frac{\sum_{j=1}^{m_{k,i}} \alpha_{i,j}^k x_j^k}{c + \sum_{j=1}^{m_{k,i}} \beta_{i,j}^k x_j^k}; (i=0, \dots, n_k), (k=0, \dots, p), \quad (3.10)$$

где p – количество слоев нейросети, n_k – количество нейронов в k -м слое, $m_{k,i}$ – количество параметров у i -го нейрона k -го слоя. Другие обозначения: x^{p+1} – вектор выходных сигналов нейросети, \bar{x} – вектор ответов, $c > 0$.

Функционирование системы задается набором уравнений:

$$\varphi_i^k = x_i^{k+1} - \frac{\sum_{j=1}^{m_{k,i}} \alpha_{i,j}^k x_j^k}{c + \sum_{j=1}^{m_{k,i}} \beta_{i,j}^k x_j^k} = 0; \quad (i=0, \dots, n), (k=0, \dots, p), \quad (3.11)$$

Система (3.11) задает способ вычисления x . Пусть имеется функция $H(\alpha, \beta, x)$. Эта функция зависит от α, β и переменных функционирования x . Как правило, явно функция оценки зависит только от выходных переменных x^{p+1} . $H(\alpha, \beta, x) = H(x^{p+1})$.

Для задачи обучения требуется найти производные $\Delta_{\alpha_i}^k = \partial H(\alpha, \beta, x) / \partial \alpha_i^k$ и $\Delta_{\beta_i}^k = \partial H(\alpha, \beta, x) / \partial \beta_i^k$. для этого введем новые переменные μ (множители Лагранжа) и производящую функцию W :

$$W(\alpha, \beta, x, \mu) = H(x^{p+1}) + \sum_{k=1}^p \sum_{i=1}^n \mu_i^k \left(x_i^{k+1} - \frac{\sum_{j=1}^{m_{k,i}} \alpha_{i,j}^k x_j^k}{c + \sum_{j=1}^{m_{k,i}} \beta_{i,j}^k x_j^k} \right); \quad (3.12)$$

Уравнения (3.11) можно записать как

$$\frac{\partial W}{\partial \mu_i^k} = 0, \quad (i=1, \dots, n), (k=1, \dots, p). \quad (3.13)$$

Заметим, что для тех α, β, x , которые удовлетворяют уравнениям (3.11), при любых μ

$$W(\alpha, \beta, x, \mu) \equiv H(\alpha, \beta, x). \quad (3.14)$$

Это означает, что для истинных значений переменных функционирования x при данных α, β функция $W(\alpha, \beta, x, \mu)$ совпадает с исследуемой функцией H (функцией оценки).

Подбираем такую зависимость $\mu_i(\alpha, \beta)$, чтобы, используя (3.14), получить для $D_{\alpha_{i,j}^k} = \partial H(\alpha, \beta, x) / \partial \alpha_{i,j}^k$ и $D_{\beta_{i,j}^k} = \partial H(\alpha, \beta, x) / \partial \beta_{i,j}^k$ наиболее простые выражения. Выберем такие μ чтобы:

$$\frac{\partial H(\alpha, \beta, x)}{\partial x_j^k} + \sum_{i=1}^n \mu_i^k \frac{\partial \varphi_i^k(\alpha, \beta, x)}{\partial x_j^k} = 0. \quad (3.15)$$

Исходя из (3.15) находим μ_q^k ($k=1, \dots, p$), ($q=1, \dots, n$):

$$\mu_q^k = \sum_{i=1}^n \mu_i^{k+1} \frac{\alpha_{i,q}^{k+1} - \beta_{i,q}^{k+1} x_i^{k+2}}{c + \sum_j \beta_{i,j}^{k+1} x_j^{k+1}}, \quad \mu_q^p = -\frac{\partial H(x^{p+1})}{\partial x_q^{p+1}} = \bar{x}_q - x_q^{p+1}. \quad (3.16)$$

Если μ определены согласно (3.16), то:

$$D_{\alpha_{i,j}^k} = \partial H(\alpha, \beta, x) / \partial \alpha_{i,j}^k = - \sum_{i=1}^n \mu_i^k \left(\frac{x_j^k}{c + \sum_{j=1}^{m_{k,i}} \beta_{i,j}^k x_j^k} \right) \quad (3.17)$$

$$D_{\beta_{i,j}^k} = \partial H(\alpha, \beta, x) / \partial \beta_{i,j}^k = \sum_{i=1}^n \mu_i^k \left(\frac{x_j^k \cdot x_i^{k+1}}{c + \sum_{j=1}^{m_{k,i}} \beta_{i,j}^k x_j^k} \right) \quad (3.18)$$

эти выражения используются для вычисления производных сложной функции вычисляемой нейросетью.

Построение оценки

Если в качестве ответа нейронная сеть должна выдать число, то естественной оценкой является квадрат разности выданного сетью выходного сигнала и правильного ответа. Все остальные оценки для обучения сетей решению таких задач являются модификациями данной.

Приведем пример такой модификации. В [35, 36, 85, 96, 88] отмечается, что процесс обучения нейросети можно значительно ускорить, применяя более специализированную оценку. Пусть при составлении задачника величина $\bar{\alpha}$, являющаяся ответом, измерялась с некоторой точностью ε . Тогда нет смысла требовать от сети обучиться выдавать в качестве ответа именно величину $\bar{\alpha}$. Достаточно, если выданный сетью ответ попадет в интервал $[\bar{\alpha} - \varepsilon, \bar{\alpha} + \varepsilon]$. Оценка, удовлетворяющая этому требованию, имеет вид:

$$H = \begin{cases} 0, & \text{при } |\alpha - \bar{\alpha}| \leq \varepsilon, \\ (\alpha - \bar{\alpha} - \varepsilon)^2, & \text{при } \alpha > \bar{\alpha} + \varepsilon, \\ (\alpha - \bar{\alpha} + \varepsilon)^2, & \text{при } \alpha < \bar{\alpha} - \varepsilon. \end{cases} \quad (3.19)$$

Эту оценку будем называть оценкой числа с допуском ε .

Оценка вида (3.19) позволяет ускорить процесс обучения и получить более сглаженные нейросетевые функции. В этом случае требования к сложности нейросети становятся более мягкими. Сглаженная функция, в свою очередь, обладает более высокими экстраполяционными и интерполяционными способностями [103, 104].

Для задач классификации также можно пользоваться оценкой типа суммы квадратов отклонений выходных сигналов сети от требуемых ответов. Однако, эта оценка плоха тем, что во-первых, требования при обучении сети не совпадают с требованиями интерпретатора, во-вторых – такая оценка не позволяет оценить уровень уверенности сети в выданном

ответе. Для задач классификации целевая переменная является дискретной (номинальной). Ее кодирование проводится таким образом, чтобы сеть имела несколько выходных полей, каждому из которых соответствует определенный класс. Тогда ответом нейросети на предъявленный пример будет считаться номер класса, соответствующий номеру выходного параметра, на котором зафиксировано наибольшее значение сигнала (интерпретатор «победитель забирает все» [35, 36]). В этом случае в качестве оценки работы нейросети предпочтительнее использовать оценку вида «расстояние до множества правильных ответов», предложенную в [35, 84]:

$$H = \begin{cases} 0, & \forall j \neq k, \alpha_k - \alpha_j \geq \varepsilon \\ \rho, & \exists j \neq k, \alpha_k - \alpha_j < \varepsilon \end{cases} \quad (3.20)$$

где k – номер «истинного» класса, α^i – выходные сигналы сети, $i=1, \dots, P$, P – число выходных сигналов, ε – требуемый уровень отличия «истинного» сигнала от остальных, ρ – функция расстояния до множества правильных ответов

$$\rho = \sum_{j=0}^l \left(\frac{\sum_{i=0}^l \beta_i}{l+1} - \beta_j \right)^2, \quad (3.21)$$

где $\beta_0 = \alpha_k - \varepsilon$, $\beta_j = \alpha_i$ – текущие выходные сигналы (за исключением α_k), переобозначенные таким образом, что $\beta_j > \beta_{j+1}$, $j = 1, \dots, P-1$, P – число выходных сигналов, l – минимальный номер, такой, что верно неравенство

$$\frac{\sum_{i=0}^l \beta_i}{l+1} \geq \beta_{l+1} \text{ при } l < P-1, \text{ или равенство } l = P-1.$$

Модификация синаптической карты (обучение)

Кроме прямого и обратного функционирования, все элементы должны уметь выполнять еще одну операцию – модификацию параметров. Процедура модификации параметров состоит в добавлении к существующим параметрам вычисленных поправок (для сетей с непрерывно дифференцируемыми элементами вектор поправок является градиентом некоторой функции от выходных сигналов).

На основе формул (3.17) и (3.18) вычисляются производные по всем настраиваемым параметрам сети. Это позволяет использовать градиентные методы оптимизации, например метод наискорейшего спуска. Он заключается в итерационном изменении параметров в соответствии с правилами:

$$\alpha_{i,j}^k(t+1) = \alpha_{i,j}^k(t) - h \cdot \frac{\partial H(\alpha, \beta, x)}{\partial \alpha_{i,j}^k}, \beta_{i,j}^k(t+1) = \beta_{i,j}^k(t) - h \cdot \frac{\partial H(\alpha, \beta, x)}{\partial \beta_{i,j}^k}, \quad (3.22)$$

где t – номер итерации, h – шаг оптимизации. Величина шага вычисляется на каждой итерации путем одномерной оптимизации функции H при фиксированных α и β . В некоторых случаях бывает полезно использовать более сложную процедуру модификации карты.

Во многих работах отмечается, что при описанной выше процедуре модификации параметров происходит неограниченный рост величин параметров. Существует несколько различных методов решения этой проблемы. Наиболее простым является жесткое ограничение величин параметров некоторыми минимальным и максимальным значениями. Очевидно, что для Паде-нейронов нужно ограничивать параметр β в знаменателе. При использовании этого метода процедура модификации параметров β примет следующий вид:

$$\beta_{i,j}^k(t+1) = \begin{cases} 0, & \beta_{i,j}^k(t) + h \cdot D_{\beta_{i,j}^k} < 0, \\ \beta_{i,j}^k(t) + h \cdot D_{\beta_{i,j}^k}, & \beta_{i,j}^k(t) + h \cdot D_{\beta_{i,j}^k} > 0, \end{cases} \quad (3.23)$$

Алгоритмы обучения

Все алгоритмы обучения сетей методом обратного распространения ошибки опираются на способность сети вычислять градиент функции ошибки по обучающим параметрам. Таким образом, обучение состоит из вычисления градиента и модификации параметров сети. Обучение двойственных сетей с точки зрения используемого математического аппарата эквивалентно задаче многомерной оптимизации. Однако, существует множество не градиентных методов обучения, таких, как метод покоординатного спуска, метод случайного поиска и целое семейство методов Монте-Карло. Все эти методы могут использоваться при обучении нейронных сетей, хотя, как правило, они менее эффективны, чем градиентные методы.

Изучению градиентных методов обучения нейронных сетей посвящено множество работ [105-108]. Все градиентные методы объединены использованием градиента как основы для вычисления направления спуска.

Наиболее известным среди градиентных методов является **метод наискорейшего спуска**. Идея этого метода проста: поскольку вектор градиента указывает направление наискорейшего возрастания функции, то минимум следует искать в обратном направлении. Последовательность действий следующая:

1. Вычислить оценку H_2
2. $H_1 = H_2$
3. Вычислить градиент
4. Оптимизация шага

5. Модификация параметров
6. Вычислить оценку $H2$
7. Если $H1 - H2 < \text{Точность}$ то переход к шагу 2

Этот метод работает, как правило, на порядок быстрее методов случайного поиска. Чтобы ускорить обучение и исправить другие недостатки наискорейшего спуска, существует огромное число методов. Например, такие как итерационный и модифицированный партан-методы. Существует также большое семейство *квазиньютоновских методов*, позволяющих на каждом шаге проводить минимизацию в направлении минимума квадратичной формы. Идея этих методов состоит в том, что функция оценки приближается квадратичной формой. Зная квадратичную форму, можно вычислить ее минимум и проводить оптимизацию шага в направлении этого минимума. Наиболее часто используемые методы из семейства одношаговых квазиньютоновских методов – BFGS метод и метод сопряженных градиентов. Эти методы хорошо зарекомендовали себя при обучении нейронных сетей. Подробно ознакомиться с ними можно в работе [35].

Получение, интерпретация и отображение результатов

Результаты работы нейросети получаются после тестирования. Операция тестирования – это ни что иное, как прямое функционирование сети. При этом на вход сети подаются примеры из всего задачника, а не только обучающее множество. На выходе сети получаем ответы. По ответам можно определить несколько характеристик обученности сети. Если сеть обучена не полностью т.е. решает не все примеры из обучающего множества то можно говорить об ошибке обучения. Соответственно если она нулевая то сеть правильно решает все примеры.

Ошибка обобщения показывает насколько хорошо сеть решает примеры не из обучающей выборки. Иногда бывает полезно не стремиться к нулевой ошибке обучения. В результате можно уменьшить ошибку обобщения. Происходить это может из-за того, что обучающая выборка не репрезентативна т.е. не отражает все особенности данных.

Естественная операция – это интерпретация ответов сети. Этот вопрос подробно описан в [35, 36]. Отметим что для решения задачи предсказания числа обычно обезразмеренные выходные сигналы масштабируют в первоначальный интервал. Для классификации могут быть применены иные методы, например: знаковая, порядковая, нечеткая и другие интерпретации.

В предложенной технологии нейросетевого анализа данных ГИС отображение результата является одной из основных целей обработки данных. После тестирования нейросети данные сохраняются в таблицу атрибутов. Возможно сохранение в ту же таблицу, из которой брались данные для обучающего множества. Например, заполняются пробелы в тех столбцах таблицы, которые служили правильными ответами при обучении нейросети. Кроме этого можно записывать такие характеристики как

относительные значимости входов для получения правильного выхода. То есть насколько каждый входной параметр, относительно других, влияет на построение предсказываемого. Причем такая значимость может быть как общая – целиком для каждого параметра, так и отдельно для каждого примера (каждой точки сетки). Таким образом, при создании тематических слоев получаем поля значимости. Один из примеров того, что дают поля значимости – визуально определяются области для каждого входного слоя, наиболее влияющие на получение выходного.

Рассмотрим формально задачу определения значимости по входным параметрам. Пусть $F(\mathbf{x})$ – нейросетевая аппроксимация некоторой неизвестной функции $y(\mathbf{x})$, после окончания обучения параметры \mathbf{a} и \mathbf{b} фиксированы и не влияют на значение $F(\mathbf{a}, \mathbf{b}, \mathbf{x})$. Описанная нейросетевая парадигма обеспечивает построение непрерывной дифференцируемой функции от \mathbf{x} и позволяет эффективно вычислять $\frac{\partial F(\mathbf{x})}{\partial x_i}$ в любой точке X .

Значения производных могут быть интерпретированы как характеристика влияния входных параметров на значение выходного в данной точке. Если в некоторой точке \mathbf{x} выполняется условие $\frac{\partial F(\mathbf{x})}{\partial x_i} < 0$ то параметр x_i при увеличении будет уменьшать значение целевого параметра и наоборот.

$\left| \frac{\partial F(\mathbf{x})}{\partial x_i} \right|$ – характеризует скорость изменения $F(\mathbf{x})$ при изменении x_i . Оценка по всей выборке может производиться по-разному. Например, сумма модулей, максимум или минимум модуля.

3.1.3. Задачи для нейронных сетей

Приведем методику построения нейросети для решение слабоструктурированных задач анализа зависимостей:

1. Сбор данных, определение входных и целевых параметров, указание их типов (дискретные или непрерывные), диапазонов изменений.
2. Формирование обучающей и тестовой выборки.
3. Выбор вида, структуры, параметров нейросети.
4. Выбор метода оценки, интерпретатора ответов, метода оптимизации и определение их параметров.
5. Формирование нейросети.
6. Определение условий останова работы сети.
7. Обучение нейросети.
8. Определение числа повторений эксперимента с разными начальными условиями (инициализируемой картой синаптических весов нейросети) для сбора статистических данных.
9. Определение критериев выбора лучшей нейросети.
10. Сохранение значений параметров задачника, нейросети и процесса обучения.

11. Исследование модели на оптимальный выбор системы исходных показателей, дублирование исходных признаков, значимость исходных признаков для решения основной задачи. При необходимости – переход к п. 1.

12. Анализ полученных результатов, использование построенной нейросетевой модели.

Теперь коснемся некоторых перечисленных пунктов для конкретных задач в ГИС. Тривиальные и общие решения в некоторых пунктах не затрагиваются. Опишем только особенности связанные с задачами. Большинство модификаций относительно конкретной задачи затрагивает пункт 1. Правильно задав входы и выходы, построив оценку и интерпретатор ответов, облегчающих обучение, сводим решение специфических задач к решению стандартной.

Классификация процессов и явлений

Определение входных и целевых параметров. На вход нейросети подаются исходные данные. На выход, параметры, описывающие классификацию. Нейросеть учится классификации. После правильного решения всех примеров из обучающего множества сеть считается полностью обученной. При тестировании на тестовой выборке решается задача отнесения к определенному классу объектов, для которых он не был определен. Результаты записываются в таблицу. В ГИС создается тематический слой на основе выполненной классификации.

Районирование, типология

С точки зрения нейросетевой технологии районирование и типология не сильно различаются. Все различие в условии слитности участков территорий, однородных в смысле некоторого критерия или группы критериев для районирования и необязательности выполнения этого условия для типологии. Обе эти задачи, в сущности, сводятся к классификации территориальных комплексов. Однако можно разделить решение в зависимости от условий. Если заданы правила классификации то это обычная предыдущая задача. Если же правила не заданы или, как правило, заданы неявно, то решение такой задачи приобретает смысл решения задачи "классификации без учителя". Изменения затрагивают пункты с первого по седьмой. То есть используются специфические методы создания обучающей выборки и/или вида нейросети и соответственно алгоритмы ее обучения. Например, для решения подобных задач часто используются сеть Кохонена и подобные ей.

Выявление определяющих факторов

Выполняется в пункте 11. Описано в конце раздела 3.1.3.

Временной анализ

Не углубляясь в анализ временных рядов, опишем построение задачника для решения таких задач. В сущности, решается первая задача, то есть классификация. Отличие состоит в формировании обучающей выборки. Если не изменять структуру задачника то просто надо добавить для каждого примера предысторию или вести обучение по скользящему блоку задачника. Например, в качестве примера N в момент времени t с предысторией глубиной два на вход нейросети подается $N(t-2)+N(t-1)+N(t)$. При обучении нужно требовать предсказание $N(t+1)$. Вопрос определение глубины предыстории в данной работе не рассматривается.

Создание моделей поверхности, интерполяция, анализ и прогнозное картирование пространственно распределенных данных

Создание моделей поверхности и интерполяция сводятся в данном случае к задаче построения функции по конечному набору значений. Нейросетевой метод позволяет строить сложную нелинейную нейросетевую функцию, описывающую данные. При этом на вход нейросети подаются пространственные и атрибутивные показатели, описывающие объект или явление. На выходе требуется получить целевой показатель, характеризующий желаемую поверхность.

Задача анализа и прогнозного картирования состоит из совокупности описанных задач. Например, классификации, предсказания и анализа временных рядов.

3.2. Программные средства и примеры использования

3.2.1. Реализация программного комплекса для нейросетевого анализа данных в ГИС

Особенности разработанной программной системы

Программную систему для нейросетевого анализа данных в ГИС функционально можно разделить на три подсистемы.

- Подсистема взаимодействия с ГИС-компонентой.
- Подсистема получения и подготовки данных.
- Подсистема нейросетевой обработки данных.

1. Подсистема взаимодействия с ГИС-компонентой обеспечивает интерфейс между динамически подсоединяемой библиотекой содержащей программный ГИС-объект и другими подсистемами. Транслирует запросы системы и вызывает внутренние методы ГИС-объекта, отслеживает события и обеспечивает получение и передачу данных. Позволяет выполнять набор следующих операций:

1) картографические операции:

- Открытие слоев GeoDraw для DOS и GeoDraw для Windows, косметических слоев ГеоГраф 1.5., растровых слоев. Все

сторонние форматы, поддерживаемые ГеоГраф 1.5 (SXF, DX90), могут использоваться только через импорт карт ГеоГраф 1.5 [109].

- Импорт картографических композиций созданных в ГеоГраф 1.1 и ГеоГраф 1.5.
- Функции векторного не топологического редактора для косметических слоев ГеоГраф 1.5.
- Масштабирование карты.
- Варьирование объектным составом карты.
- Создание тематических слоев.
- Оверлейные операции.
- Изменение способа отображения объектов (цвет, тип линии и т.п.), в том числе и определение символики через значения атрибутов, то есть синхронизация визуализации с изменениями в базах данных;

2) атрибутивные операции:

- Связь таблиц атрибутов со слоями карты.
- Работа с таблицами атрибутивных данных.
- Получение данных об объектах слоя.
- Возможность опрашивать через карту в режиме реального времени базы данных.

Формально состоит из блока работы с картой, блока работы с таблицами и блока "легенды". Физически ГИС-объект встроен в подсистему.

2. Подсистема получения и подготовки данных позволяет получать из карты данные в удобном для нейросетевой обработки виде и производить некоторую предобработку. Технология получения данных из ГИС описано в пункте 3.2.2.

3. Подсистема нейросетевой обработки данных (нейроимитатор) предназначена для построения моделей ИНС и решения с помощью ИНС задач ГИС. Задачи, решаемые нейроимитатором в рамках данной работы:

1) автоматизированное решение задач ГИС на основе нейросетевой парадигмы, описанной в пункте 3.1 третьей главы. Построение нейросетевых моделей, решающих задачи:

- классификации;
- предсказания;
- регрессии;
- прогнозирования значений временных рядов;

2) обеспечение функционирования построенных моделей в составе системы;

3) решение дополнительных информационных задач, в том числе получение дополнительной информации о характере зависимости между входными и выходными параметрами модели. Необходимость решения этой задачи связана с необходимостью получения дополнительной информации при анализе и оценке полученной модели, а также при

планировании принятия решений и исследовании моделируемого объекта или процесса. Это позволяет выявлять информацию о том, как влияет каждый компонент входного вектора нейросети на полученное решение в данной точке пространства входных параметров. На основе этого эксперт может сделать вывод о качестве полученной нейросетевой модели и получить новую информацию об исследуемом объекте при нейросетевой реализации информационной модели.

Структура и функции программной системы

Общая схема взаимодействия подсистем приведена в пункте 3.1.1. третьей главы (рис 3.2). Покажем общие схемы первых двух подсистем и остановимся подробно на нейроимитаторе.

Подсистема взаимодействия с ГИС-компонентой.



Рис. 3.8. Схема подсистемы взаимодействия с ГИС-компонентой

Подсистема получения и подготовки данных состоит из нескольких блоков (рис 3.9). Предназначена для:

- получения данных из ГИС и преобразования в вид удобный для нейросетевой обработки;
- предобработки полученных данных;
- работы с базами данных;
- обеспечения информационного взаимодействия между подсистемами;
- предварительной визуализации.

Подсистема нейросетевой обработки данных.

Потоки данных между блоками:

1. Таблица исходных данных (выборка).
2. Таблица обработанных данных.
3. Сохранение файлов параметров нейросети.
4. Чтение файлов параметров нейросети.
5. Заданные пользователем или прочитанные параметры процесса обучения.
6. Сохранение параметров обучения.
7. Параметры формирования структуры нейросети.
8. Результаты в терминах поставленной задачи.
9. Данные в формате пригодном для использования блоком обучения. После нормировки, центрирования, кодирования, разбиения на обучающую и тестовую выборки.

10. Созданная (загруженная) нейросеть.
11. Обученная нейросеть.
12. Параметры процесса обучения.



Рис. 3.9. Блоки подсистемы получения и подготовки данных

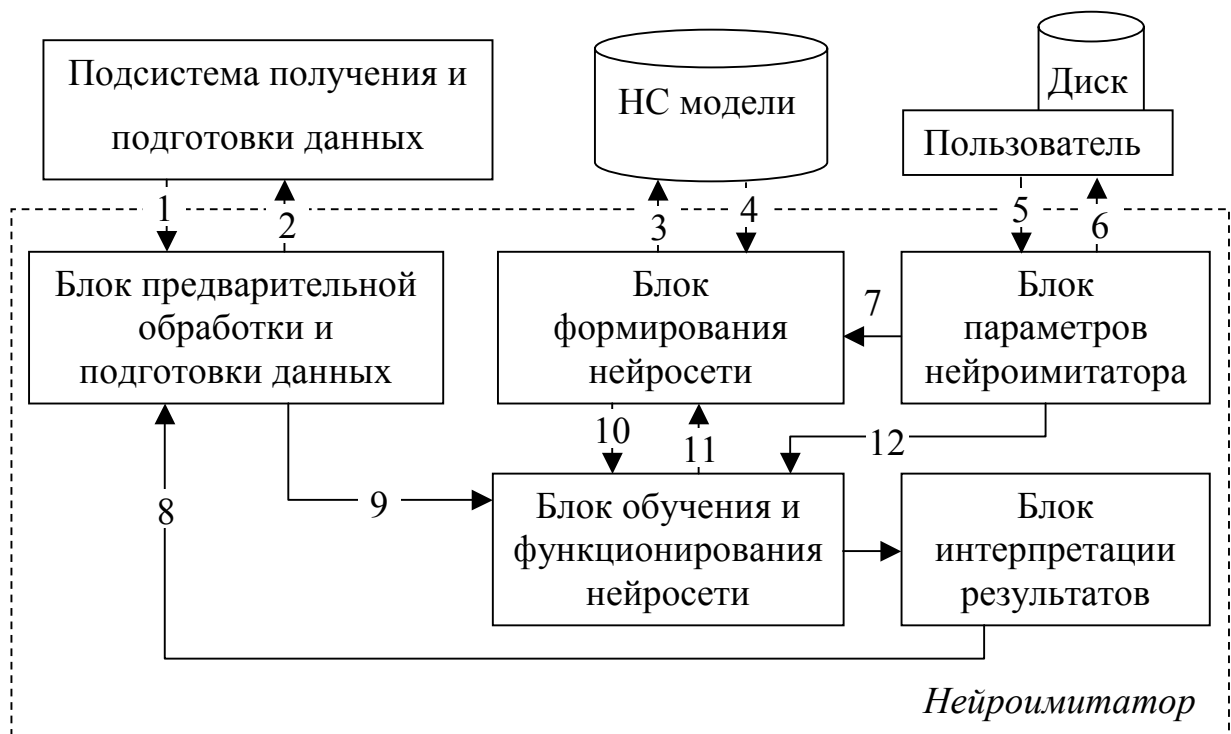


Рис. 3.10. Общая схема структуры нейроимитатора

Подсистема предназначена для непосредственных экспериментов с моделями ИНС и исследования полученных моделей в диалоге с пользователем. Реализованы стандартные функции нейроимитатора: предварительная обработка данных, возможность конструирования структуры многослойной сети, выбор метода оптимизации и вида оценки, обучение и тестирование нейросетевых моделей, определения минимального набора входных параметров, чтение и сохранение нейросетей на диске.

Файл параметров процесса обучения может быть сформирован для каждой задачи, и содержит значения основных параметров обучения, включая настройки формирования структуры нейросети. Файл уже обученной нейросети используется в режиме функционирования системы, а также для дообучения нейросети на новых данных.

Результатом функционирования модуля является файл результатов (таблица), в котором сохраняются значения выходных параметров нейросети, полученные на основе текущих данных. В файле нейросети программа сохраняет полученную нейросетевую модель для дальнейшего использования.

Программная реализация системы

Приведенные выше методы и функции реализованы автором в программной системе GISNNA, предназначенной для решения задач ГИС методами нейроинформатики.

Для реализации программы была выбрана среда программирования Borland Delphi, в состав которой входит компилятор объектно-ориентированного языка высокого уровня Object Pascal и интегрированная среда разработчика. Такой выбор связан со следующими достоинствами Delphi:

1. Среда Delphi предназначена для создания 32-х битных приложений для современной операционной системы Microsoft Windows 95/NT, которая в настоящее время является фактическим стандартом для персональных компьютеров.

2. Объектно-ориентированный язык программирования позволяет сделать процесс разработки ПО более эффективным.

3. Компилятор Delphi строит эффективный исполняемый код, что особенно важно для систем моделирования ИНС, связанного со значительными объемами вычислений.

4. Интегрированная среда разработчика позволяет создавать интерфейс программы с минимальными затратами времени и труда программиста, что повышает эффективность процесса разработки.

Разработанная программная система может быть использована в информационных системах, работающих под управлением ОС MS Windows 95/98/NT и допускающих реляционное представление данных для решения задач.

Основу программной системы составляют модули, описанные в п. 3.2.1.

Взаимодействие с полнофункциональной геоинформационной системой осуществляется на уровне обмена файлами. В программу может быть загружена карта созданная в ГИС Географ. Данные для анализа извлекаются из карты и атрибутивных таблиц. Сохранение результатов анализа производится в этих таблицах. Запись картографических объектов может производиться в файлы обменных форматов.

Интерфейс программ основан на стандартах GUI (Graphical User Interface) и использует стандартные элементы интерфейса Windows: окна, меню, кнопки, поля ввода и выбора и т.д.

3.2.2. Решение задачи восстановления пропусков

Технология получения данных из ГИС

Конечная цель любой обработки информации в ГИС – это получение картинки (тематического, композиционного, оверлейного и других слоев). Пространственные данные для обработки и для создания тематического слоя электронной карты могут быть получены несколькими способами. Например, по данным дистанционного зондирования или данным изысканий. В большинстве же случаев, пространственные данные вводятся при оцифровке существующих бумажных карт. В результате любого способа ввода пространственной информации образуется слой или слои электронной карты, на которые содержат идентифицированные пространственные объекты, связанные с базами атрибутивных данных.

При вводе пространственной информации как массива координат объектов или непосредственного ввода как точечного объекта электронного точечного слоя карты, как промежуточная задача (иногда как основная) возникает необходимость восстановить непрерывное поле признака по имеющимся дискретным значениям. Иногда наоборот появляется необходимость использовать пространственную информацию с оцифрованных карт изолиний. В явном виде, без аналитического представления данной поверхности, сделать это сложно. Поэтому, для решения задач обработки информации представленной в виде изолиний (в силу специфичности обработки), приходится решать обратную задачу восстановления дискретных значений и построения поверхности по полученным точкам.

Для нейросетевого анализа данных в ГИС удобно использовать данные в виде описанном в разделе 2.3 второй главы. То есть в виде данных на сетке или матричном виде. Это означает что в узлах сетки конкретным координатам этих узлов соответствуют некоторые количественные и качественные значения явлений (или различные показатели одного явления). Такая пространственная информация, выраженная в цифровой форме и взятая в одних и тех же точках со всех

исходных карт, позволяет строить пространственно – цифровые модели, проводить статистические и другие необходимые расчеты, проводить нейросетевую и не только обработку данных для поиска взаимосвязей, анализа рассматриваемых явлений, проведения численных экспериментов. В результате таких математических операций в узловых точках создаются "каркасы" из полученных данных, на основе которых можно создавать новые расчетные карты.

Если данные изначально представлены в сеточном (растровом) виде то можно перейти к сразу к задачам обработки и анализа. Если нет то получение данных требует описания. Задачу получения данных из ГИС в приемлемой для обработки форме может быть разбита на несколько подзадач.

- Создание сеточного слоя для растривания векторного.
- Получение матричной модели слоя по векторным данным в узлах регулярной сетки цифровой поверхности.
- Постобработка. Различного вида фильтрация (сглаживание, выделение контуров, контраст ...)
- Визуализация матричных моделей, ввод и тематическая доводка слоя в ГИС.

Опишем технологию поэтапно. Пусть есть карта представляющая собой композицию тематических векторных слоев описывающих разные аспекты (явлений, временные...) одной территории.

Сначала программно создается сеточный слой. Слой задается тремя типами объектов: полигонами – квадратики, линиями – границы квадратов сетки и точками – центры полигонов. Масштаб детализации задается размерами ячеек сетки, т. е. количеством узлов. Не имеет смысла строить поверхность с точностью превышающей точность исходных данных, другими словами, размер ячейки должен быть не меньше среднего шага цифрования (линейных отрезков, которыми выполнялась векторизация), зависящего от масштаба исходной карты. Затем слой добавляется в композицию тематических слоев средствами ГИС.

Второй этап состоит в получении данных в каждой точке сетки из каждого слоя. Чтобы найти значение в каком-либо слое в определенной точке, нужно решить задачу принадлежности точки полигону растрируемого слоя. Затем найти в таблице атрибутов по идентификатору полигона его значение и присвоить его точке сетки. В результате получается прямоугольная таблица. Столбцы – признаки (координаты X и Y, значения слоев), строки – точки. Из этой таблицы далее будет образован задачник для обучения нейросетей.

Выполнение третьего этапа не обязательно и продиктовано особенностями векторного представления полей признаков. При векторизации поля данных огрубляются путем определения линий уровня с некоторым шагом и создания полигонов на их основе. После растеризации полигональных слоев ситуация в точности полей

кардинально не меняется. Но иногда бывает полезно сгладить данные и убрать крупные линейные участки (рис. 3.11).

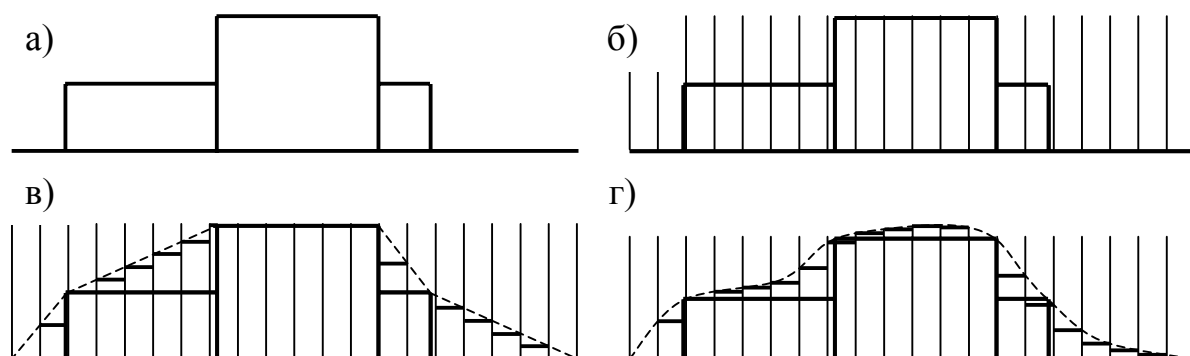
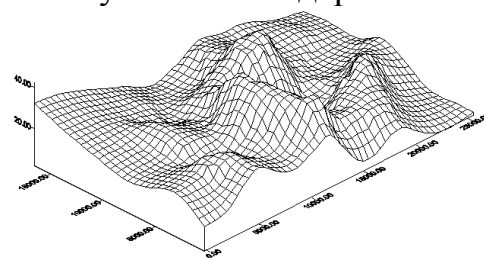
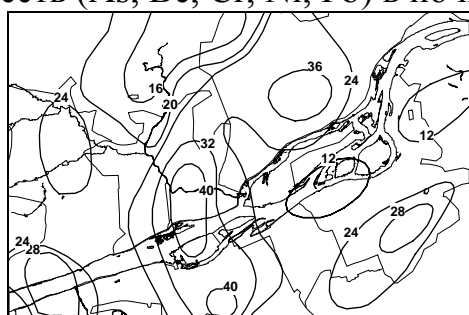


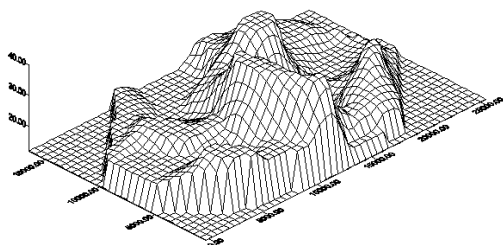
Рис. 3.11. Профильное сечение слоя: а) полигональный слой, б) растрованный по сетке, в) сглаженный кусочно-линейно, г) сглаженный фильтром

Существует большое семейство методов направленных на восстановление непрерывных полей признаков по дискретным данным и получения в дальнейшем значений в узлах регулярной сетке. Наиболее известные методы: минимальной кривизны, триангуляции, ближайшего соседа, кригинга, радиальных базисных функций, обратного расстояния в степени. Для этих же целей можно использовать и нейросетевые модели [17, 110].

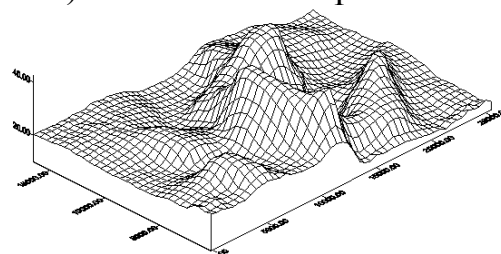
Для примера рассмотрим построение поверхности по данным изолинейного слоя "Суммарное содержание веществ с канцерогенными свойствами в почве" из электронного экологического атласа города Красноярск [111-114] (рис. 3.12). Суммарное содержание канцерогенных веществ (As, Be, Cr, Ni, Pb) в почве. Изолинии условных содержаний.



а) минимальной кривизны



б) триангуляции



в) кригинга

Рис. 3.12. Поверхности построенные различными методами

Для визуализации данных, представленных в виде матрицы узловых значений сетки, необходимо подготовить полигональный слой, состоящий из полигональных объектов в виде квадратиков – пикселей. Размеры

пикселей берутся в соответствии с размерами используемой сетки. Таблица узловых значений сетки, которую необходимо визуализировать, сохраняется в качестве файла базы данных, читаемых ГИС – программой. Такая база данных загружается в качестве атрибутивной информации созданного полигонально-пиксельного слоя. После идентификации объектов слоя каждому пикселу будет присвоено значение определенного узла в качестве атрибутивной информации. В дальнейшем определяется шкала тематической раскраски и в готовом слое электронной карты с помощью цвета пикселей отображается пространственное изменение расчетного признака или явления по данной территории (рис. 3.13).

Технология восстановления пропущенных данных

Для формирования и настройки нейросетевых моделей использовалась технология и методы построения нейросетей, разработанные во второй главе данной работы.

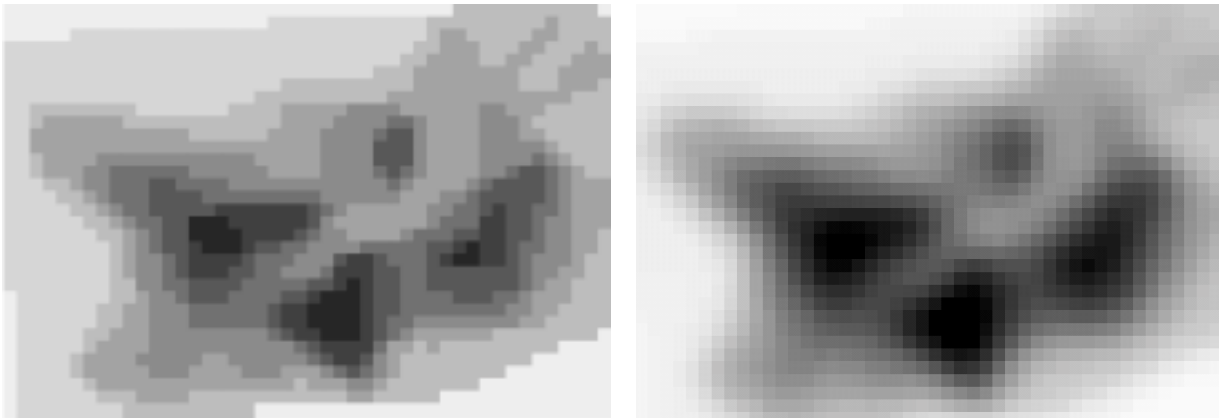


Рис. 3.13. Растрированный и затем сглаженный слой

Перейдем к основной части нейросетевой обработки данных в ГИС. Напомним, что основная задача, которая решается, это задача восстановления пропущенных данных то есть заполнения пробелов в таблице. Создание таблицы описано в разделе 3.2.1. Следующий шаг – это создание задачника для нейросети.

Задачник (обучающее множество) состоит из примеров вида "вход – правильный выход" [18, 35, 36, 80, 115]. Вход – вектор сигналов, предъявляемый нейросети, выход – вектор сигналов, который нейросеть должна выдать после функционирования. Затем производится нормировка. Например, непрерывные признаки нормируются согласно (3.1). Для получения ответов в реальном масштабе измерения целевой переменной производится обратное преобразование выходных сигналов сети.

По паре "требуемый выход – полученный выход" строится функция оценки (3.7).

Формируется нейросеть, вычисляющая функцию (3.2). Она состоит из нелинейных преобразователей сигнала, связанных между собой посредством адаптивных связей, которые линейно преобразующих

проходящий через них сигнал (рис. 3.5). Такие элементы называются нейронами. В нашем случае нейросеть состоит из Паде–нейронов (рис. 3.6).

Паде–нейрон состоит из двух адаптивных сумматоров и нелинейного элемента с двумя входами, вычисляющего частное от своих входных сигналов согласно (3.3) (рис. 3.7). Название происходит от аппроксимации Паде – метода рациональной аппроксимации функций.

С выбранным количеством слоев и нейронов в слое создается слоистая нейросеть. Начальные значения весовых коэффициентов адаптивных сумматоров нейронов инициализируются случайными величинами, распределенными равномерно в интервале $[-0.01, 0.01]$. После выбора метода обучения (см. гл.3. п. "Алгоритмы обучения") нейросеть запускается на обучение. Обучением называется процесс минимизации функции оценки. Он выполняется с привлечением градиентных методов оптимизации и состоит в поиске таких значений параметров, при которых нейросеть выдает правильный вектор выходных сигналов. (см. гл.3. п. "Оптимизационное обучение нейронных сетей")

В качестве метода оптимизации используется один из трех методов: модифицированный Partan, метод сопряженных градиентов, BFGS метод. В качестве функции ошибки используются модификации метода наименьших квадратов «МНК с люфтом» для задач регрессии и «расстояние до множества правильных ответов» для задач классификации. Останов процесса обучения происходит при выполнении одного из условий: 1) ошибка сети на обучающей выборке, полученная с помощью выбранного метода оценки, не превышает заданный уровень; 2) уменьшение ошибки сети не превышает заданного значения. Нейросеть считается обученной после достижения заданного (малого) значения функции оценки то есть при выполнении первого условия останова. Градиентные методы поиска приводят, в общем случае, к локальным экстремумам, которые часто не обеспечивают требуемого качества обучения. Применена следующая эвристическая схема борьбы с локальными минимумами. В случае прекращения уменьшения ошибки сети в процессе обучения производится добавление равномерно распределенной случайной величины к весам связей сети (процедура "удар") и обучение продолжается. Если в результате добавления не удастся существенно уменьшить оценку, увеличивается интервал значений случайной величины и "удар" повторяется. Ширина интервала значений последовательно расширяется от значения 0.02 до значения 0.3 с шагом 0.02. Если в результате применения наибольшего по величине сдвига и последующего обучения оценка сети не уменьшилась, процесс обучения сети данной структуры прекращается.

Встречаются ситуации, когда нейросеть плохо обучается из-за противоречивой обучающей выборки и увеличение размера нейронной сети мало помогает. Это бывает когда в обучающей выборке, присутствуют задачи с одинаковыми условиями, но разными ответами

(одинаковыми входными векторами данных, но разными выходными). Появление таких конфликтных примеров может, например, означать недостаточность набора входных признаков, поскольку при расширении признакового пространства конфликтным примерам могут соответствовать разные значения добавляемого признака и критическая ситуация будет исчерпана. В любом случае пользователь должен решить эту проблему, например, исключением конфликтных примеров из задачника.

После построения, нейросетевая модель решения сохраняется на диске, и далее в процессе функционирования ГИС-приложения загружается для решения задач. Адекватность и качество функционирования модели оценивается пользователем или экспертом. При необходимости, строится новая модель или производится настройка имеющейся на основе новых данных (дообучение).

В процессе использования программы, пользователь может дополнительно получить информацию о влиянии компонент входного вектора на значение целевого параметра в заданной точке.

В связи с тем, что нейронные сети являются вероятностными моделями, может существовать несколько различных нейросетей, решающих одну и ту же задачу. При этом результаты решения дополнительных информационных задач, полученные на основе этих нейросетей, могут быть различны. Поэтому для получения более точного представления о процессе получения значений целевого параметра может быть полезным использование нескольких нейросетевых моделей.

После обучения нейронной сети необходимо провести ее тестирование на тестовой выборке для определения точности решения не входивших в обучающую выборку задач. Точность правильного решения очень сильно зависит от репрезентативности обучающей выборки. Обычно при решении различных неформализованных задач в разных проблемных областях точность в 70-90% правильных ответов на тестовой выборке соответствует проценту правильных ответов при решении этих же задач специалистом-экспертом.

Для моделирования искусственных нейронных сетей использовался разработанный программный комплекс. Программа позволяет выполнять как элементарные операции с картой, так и традиционные операции над нейронными сетями (создание сети, ее обучение, тестирование), а также предобработку (растрирование, фильтрация) и визуализацию результатов. Вычислительные эксперименты проводились на персональных компьютерах.

Целью экспериментального использования разработанных методов была проверка работоспособности и экспериментальное изучение возможностей предложенных методов и программных средств.

В качестве примеров решения практических задач с использованием разработанных средств рассматривается задача в области анализа медико-экологических данных [116-119].

В работе использовалась электронная экологическая карта города Красноярска [112, 113]. Карта содержит данные о морфоструктуре рельефа, плотности и фитопатологическом состоянии растительного покрова в городе, загрязнениях некоторыми, в том числе и канцерогенными, веществами почв и снежного покрова. А также информацию о населении города. Это плотность жилищно-промышленной застройки, динамика показателя заболеваемости злокачественными новообразованиями жителей Красноярска.

Исходя из предположения о пространственной связи повышенных загрязнений атмосферы и почвы с ростом заболеваемости была поставлена задача заполнения пробелов в данных слоя заболеваемости на основе данных из других слоев [120-128].

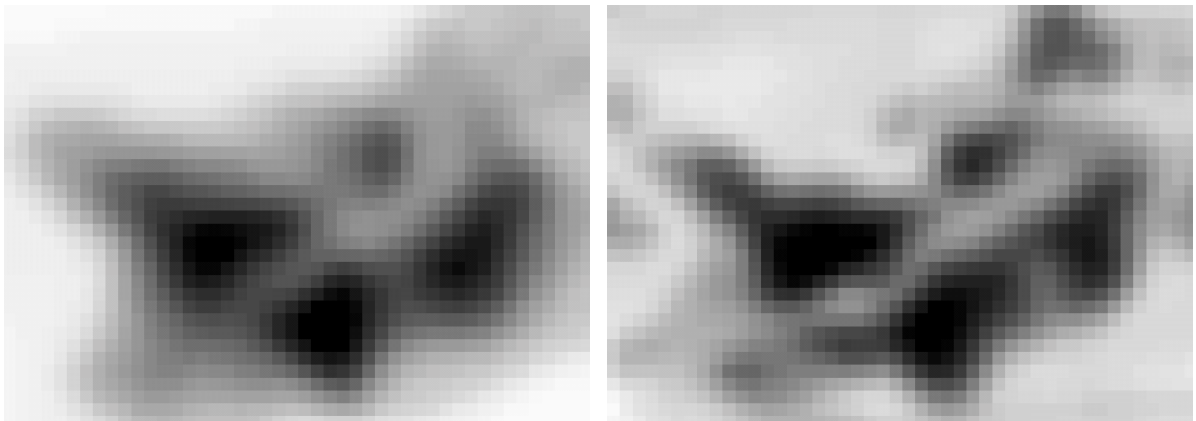


Рис. 3.14. Исходный слой и слой по результатам работы нейросети через несколько циклов обучения

Эксперимент проводился два раза. Каждый раз было обучено по десять нейросетей. Первый раз на вход подавались данные, характеризующие экологическое состояние города, на выходе нужно было получить количество заболевших. Во второй раз к исходным данным добавлялись координаты X, Y.

Обучающее множество создавалось по половине известной информации. Тестирование проводилось для всего слоя. Ошибка обучения была сведена к нулю. Ошибка обобщения составляла 10 – 20 %.

Замечено что нейросеть за несколько первых циклов обучения воссоздает поверхность качественном смысле и все остальное время обучения приближает ее до точных числовых значений (рис. 3.14).

Обученная нейросеть может распространять знания о зависимости между слоями на отсутствующие области карты. При тестировании нейросети по задачику и примерам, не входящим в задачник, результат записывается в таблицу, после чего, полигональный слой ("сетка") может быть раскрашен в соответствии с полученными при тестировании значениями.

Исследовался также вопрос значимости исходных показателей для решения задачи (рис. 3.15–3.16). Какие из входных сигналов являются доминирующими, (значимыми) при принятии нейросетью решения, а

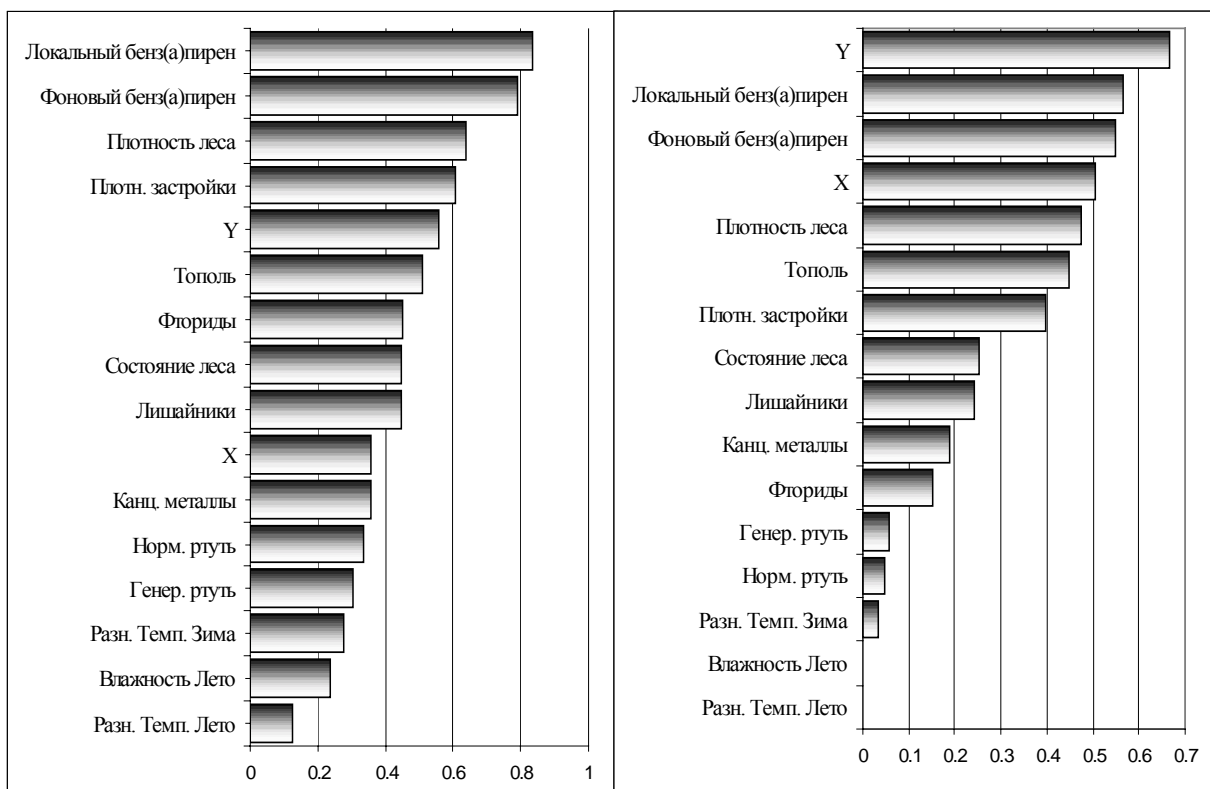


Рис. 3.15. Распределение относительной значимости по входным параметрам с участием координат. Левый график первоначальный. Правый после сокращения незначимых признаков.

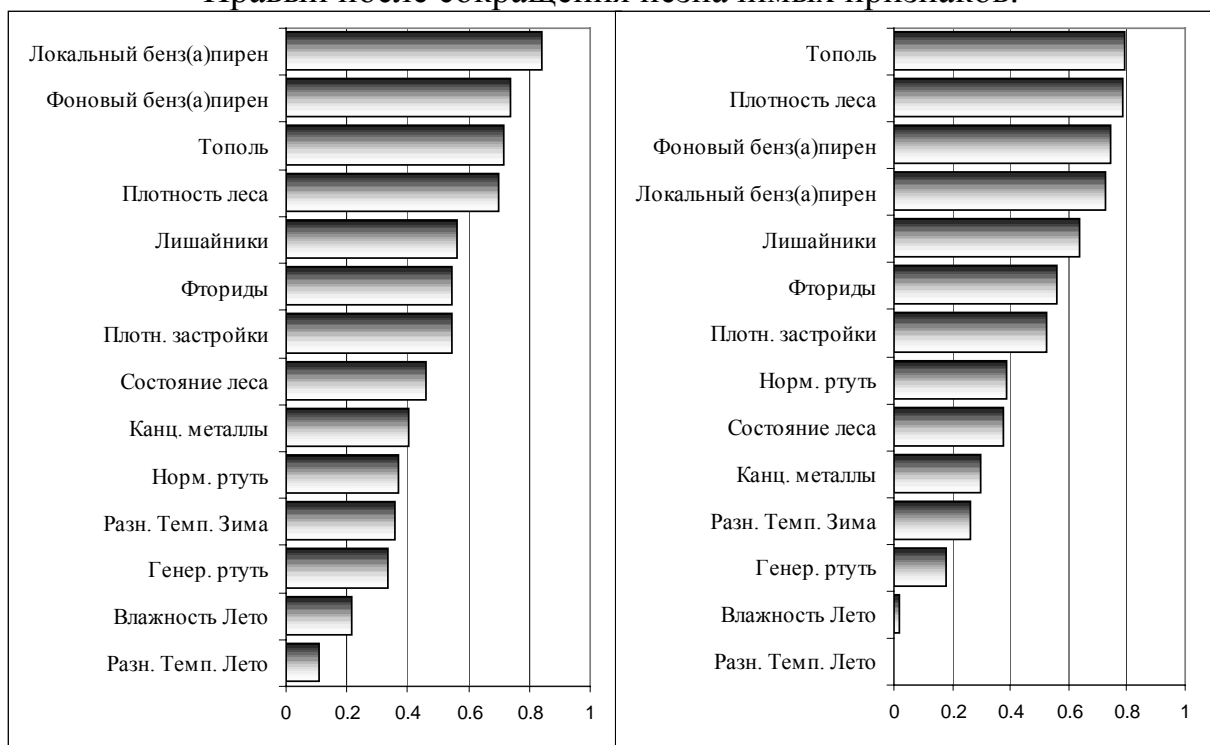


Рис. 3.16. Распределение относительной значимости по входным параметрам без участия координат. Левый график первоначальный. Правый после сокращения незначимых признаков.

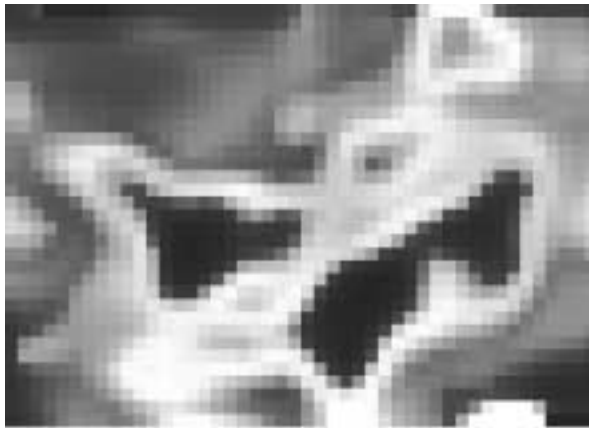
Поля значимости рассчитываются как частные производные нейросетевой функции от входных параметров (см. гл. 3.1. п. "Получение, интерпретация и отображение результатов"). Значения производных могут быть интерпретированы как характеристика влияния входных параметров на значение выходного в данной точке. Если в некоторой точке значение производной по параметру меньше нуля то при его увеличении будет уменьшаться значение целевого параметра и наоборот. На рис. 3.17 а), б) и в) светлые области соответствуют отрицательным значениям производных темные соответственно положительным. Серый цвет примерно соответствует нулевым изменениям. На рис. 3.17 г) все значения производных положительные. Белые области близки к нулевым значениям.



а)



б)



в)



г)

Рис. 3.17. Пример распределения значимости по входным слоям

4. Технология картирования произвольных данных

4.1. Упругие карты

Описывается технология визуализации произвольных данных. Рассматривается представление данных и алгоритм построения упругой карты, моделирующей данные. Предложены различные способы проектирования многомерных данных на двумерную сетку. Продемонстрировано применение методов визуализации произвольных данных на примере картографирования экономических таблиц.

4.1.1. Введение

В самых разных областях человеческой деятельности (в медицине, биологии, экономике и т.д.) исследователи сталкиваются с необходимостью осмысления больших таблиц данных, собранных в результате наблюдения за свойствами объектов той или иной природы. Как правило, такие таблицы содержат информацию о состоянии нескольких сотен или тысяч объектов, по каждому из которых известны значения определенного набора интересующих исследователя свойств. Число таких свойств (*признаков*) может также достигать нескольких сотен. Естественно, среди исследуемых объектов могут найтись такие, некоторые свойства которых неизвестны или недоступны для измерения. В таких случаях данные называются неполными или *данными с пробелами*.

Традиционным приемом при анализе таблиц является их представление, когда каждому исследуемому объекту сопоставляется точка в некотором абстрактном *многомерном пространстве данных*. Размерность этого пространства равна числу свойств-признаков в наборе, который характеризует состояние каждого из объектов. Значение каждой из координат точки в пространстве данных равно значению соответствующего признака, так что близким (в той или иной метрике) точкам в пространстве данных соответствуют объекты со сходными свойствами.

Таким образом, таблица данных представлена как облако точек в многомерном пространстве. Следует заметить, что такая картина не совсем верна в случае неполных данных. Так, например, если для объекта неизвестно значение только одной из координат, то его правильнее представлять как прямую, параллельную соответствующей координатной оси. Если на значение отсутствующего признака наложены априорные ограничения, то прямая превращается в отрезок. В случае, когда число неизвестных признаков более одного, то объект представляется в виде *m*-мерной плоскости (*m* – число неизвестных признаков), параллельной *m* координатным осям, или, соответственно, многомерным прямоугольным параллелепипедом.

Необходимый элемент анализа (осмысления) набора данных – это их красивое, наглядное и компактное описание. Более того, для многих

случаев построение такого описания уже решает задачу анализа данных. Если же в ходе такого анализа удастся решать вопросы о восстановлении пропущенных данных, то, как правило, этого достаточно для большинства приложений. На сегодняшний день известно два основных способа такого описания. Первый из них заключается в том, что точки данных разбиваются на несколько больших классов (кластеров), затем в классах ищутся более мелкие подклассы и т.д. В конечном итоге таблица данных описывается как иерархическая система кластеров в облаке точек [35, 74, 129, 130].

Второй способ – сокращение размерности описания данных. Дело в том, что человеческий мозг неспособен к эффективному анализу объектов размерности более трех, поэтому возникает необходимость в построении различных проекций многомерных данных, причем эти проекции малой размерности, как правило, выбираются так, чтобы максимально полно сохранить информацию об имеющихся в наборе данных закономерностях. Такой подход, например, характерен для традиционного линейного факторного анализа, когда в качестве такой проекции выбирается ортогональная проекция на подпространство, натянутое на несколько главных собственных векторов корреляционной матрицы (*главных компонент*) [129-133].

Под *визуализацией данных* можно понимать такой способ описания данных, когда размерность их описания сокращается до *двух* измерений. В этом случае данные можно изображать, например, в виде точек на экране монитора компьютера. Если в процессе сокращения размерности будет сохранена существенная часть закономерностей, присущих данным (разбиение на классы, отношения соседства), то исследователь получает возможность наглядно представить себе исходный набор многомерных данных, сделать выводы об их распределении.

Для визуализации многомерных данных разработан метод, заключающийся в построении вложенного в многомерное пространство данных двумерного многообразия, называемого *картой*, которое определенным способом *моделирует* или аппроксимирует данные (то есть большая часть точек данных лежит в окрестности карты). После этого точки данных с помощью определенной процедуры переносятся или *проецируются* на карту (см. рис. 4.1.). После того, как найдены положения всех образов точек данных на карте, каждому объекту из набора данных можно поставить в соответствие пару координат, характеризующих положение образа на двумерной карте.

На левом рисунке (рис. 4.1.) показано облако точек данных в пятимерном пространстве (показана проекция на первые две координатные оси), и карта, моделирующая эти данные. Белый кружок с линией отображает процесс проецирования многомерных данных на двумерное многообразие. На правом рисунке (рис. 4.1.) двумерная карта с нанесенными на нее образами данных «развернута» и каждому объекту соответствует теперь только две координаты образа на карте.

Сама идеология построения моделирующих карт была предложена

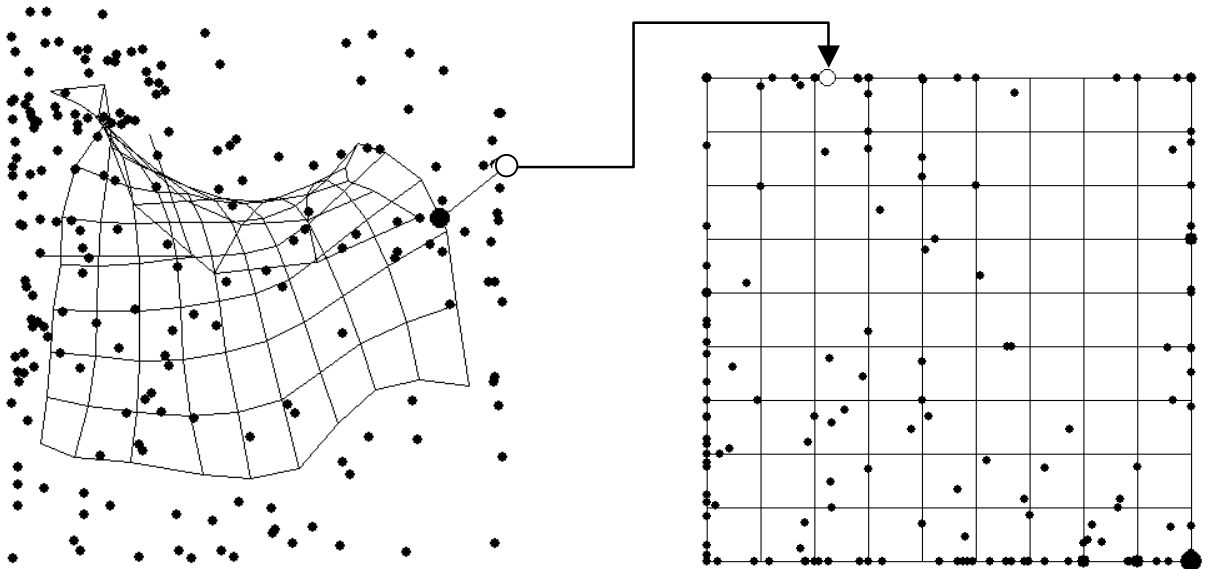


Рис. 4.1. Карта и ее развертка

Кохоненом [77]. Им же был разработан алгоритм построения самоорганизующихся карт (SOM – Self-Organizing-Maps), который с успехом применяется на практике. В дальнейшем алгоритм SOM совершенствовался в нескольких направлениях и на данный момент известно большое количество его модификаций. В данной работе предлагается иной алгоритм построения карт, названных *упругими* [68-71, 134, 135]. Кроме этого, описываются различные способы непрерывного проектирования данных на карту (а не только в узлы, как это было предложено Кохоненом), что позволяет посмотреть на принципы визуализации данных в ином ракурсе. Приведен пример картографирования экономических показателей двухсот крупнейших российских предприятий по данным журнала «Эксперт».

4.1.2. Постановка задачи

Проблема визуализации многомерных данных может быть разбита на несколько задач. Сформулируем эти задачи на математическом языке.

Рассматривается пространство R^n , в котором задано конечное множество точек X , которое интерпретируется как набор произвольных данных, полученных в результате наблюдений за состоянием исследуемой системы. Результат одного измерения (или набор признаков для одного объекта) изображаются точкой x^i ($i=1..N$, N – количество точек в X) в R^n . Расстояние между точками – обычное евклидово.

1) Построение карты

В целях исследования и визуализации множества X оно аппроксимируется двумерным многообразием. Можно указать два основных способа построения такого моделирующего многообразия. Во-первых, можно построить какое-либо стандартное линейное многообразие (например, плоскость двух первых главных компонент) и искать малые

отклонения от него. Такой подход можно назвать *квазилинейным*.

В работе предложен и используется принципиально нелинейный способ. В R^n размещается двумерная сетка G , расположение узлов которой r^{ij} , удовлетворяет определенным требованиям. Для упругой карты этими требованиями являются а) близость узлов сетки к данным; б) не слишком сильная «растянутость» сетки; в) не слишком сильная изогнутость сетки.

Будем считать, что на сетке изначально задана «внутренняя» ортонормированная система координат u, v так, что целым положительным значениям координат соответствуют вершины сетки (в этих координатах сетка является "прямоугольной"). Эти целые значения и используются в качестве значений индексов i, j для r^{ij} ($1 \leq i \leq n_1, 1 \leq j \leq n_2$, где n_1, n_2 – число узлов сетки по горизонтали и вертикали).

После того, как сетка G размещена в многомерном пространстве, она *доопределяется* до многообразия, которое обозначим через M . Это производится с помощью той или иной процедуры интерполяции между узлами. Самой простой такой процедурой является какой-либо вариант триангуляции, в результате чего получается кусочно-линейное многообразие M , составленное из треугольников.

2) Проектирование данных

Для того, чтобы можно было бы представлять X с помощью двумерного многообразия M , необходимо построить отображение $P: x \in X \rightarrow r \in M$, которое будем называть правилом проектирования или *проектором*. Для рассматриваемых задач желательно, чтобы проектор обладал следующими качествами:

а) проектор должен сохранять отношения соседства, то есть желательно, чтобы близким точкам в R^n соответствовали близкие точки на карте (по крайней мере, в некоторой окрестности сетки);

б) проектор, по крайней мере, в некоторой окрестности сетки должен быть однозначным;

в) проектор должен быть по возможности непрерывным, чтобы плавным изменениям состояния системы в X соответствовали непрерывные изменения положения образа в M .

г) проекция должна не слишком сильно отличаться от ближайшей вершины сетки G или, если доопределена карта, то от ближайшей точки карты.

Создателем SOM Кохоненом был применен самый очевидный и в некотором смысле естественный вариант проектирования – кусочно-постоянный. При этом каждой точке из X сопоставляется та вершина сетки, которая является ближайшей к этой точке:

$$P: x^k \rightarrow r^{ij}, x^k \in X, r^{ij} \in G, (x^k - r^{ij})^2 \rightarrow \min$$

Достоинством такого проектирования являются его логическая прозрачность и простота, очевидный его недостаток – разрывность, что не позволяет подробно изобразить картой структуру X . Тем не менее, такой вид проектирования успешно применяется на практике для раскраски карт

Кохонена и может изображать разбиение данных по кластерам, размеры кластеров и их взаимное расположение.

Для предложенной технологии построения упругих карт кусочно-постоянный способ проектирования малоприменим, поскольку, в отличие от SOM, каждый из узлов сетки G , вообще говоря, не располагается в центре локального сгущения точек данных. Напротив, упругая карта представляет собой более-менее равномерно натянутую на данные сетку, и поэтому существенная часть данных может быть расположена в промежутках между узлами. В этой работе рассмотрены два варианта кусочно-линейного непрерывного проектирования многомерных данных на двумерную карту.

4.1.3. Построение упругой карты

Основой для построения упругой карты является двумерная прямоугольная сетка G , вложенная в многомерное пространство, аппроксимирующая данные и обладающая регулируемыми свойствами упругости по отношению к растяжению и изгибу.

Расположение узлов сетки ищется в результате решения вариационной задачи на нахождение минимума следующего функционала:

$$D = \frac{D_1}{|X|} + \lambda \frac{D_2}{m} + \mu \frac{D_3}{m} \rightarrow \min, \quad (4.1)$$

где $|X|$ - число точек в X ; m - число узлов сетки (в случае прямоугольной сетки $m = n_1 \cdot n_2$); λ, μ - коэффициенты упругости, отвечающие за растяжение и изогнутость сетки соответственно; D_1, D_2, D_3 - слагаемые, отвечающие за свойства сетки, именно:

$$D_1 = \sum_{ij} \sum_{x \in K_{ij}} \|x - r^{ij}\|^2 \quad - \text{ является мерой близости расположения}$$

узлов сетки к данным. Здесь K_{ij} - подмножества точек из X , для которых узел сетки r^{ij} является ближайшим (*таксоны*):

$$x \rightarrow r^{ij}, \left\| r^{ij} - x \right\|^2 \rightarrow \min, K_{ij} = \left\{ x \in X \mid x \rightarrow r^{ij} \right\};$$

$$D_2 = \sum_{ij} \left\| r^{ij} - r^{i,j+1} \right\|^2 + \sum_{ij} \left\| r^{ij} - r^{i+1,j} \right\|^2 \quad - \text{ мера растянутости сетки};$$

$$D_3 = \sum_{ij} \left\| 2r^{ij} - r^{i,j-1} - r^{i,j+1} \right\|^2 + \sum_{ij} \left\| 2r^{ij} - r^{i-1,j} - r^{i+1,j} \right\|^2 \quad - \text{ мера}$$

изогнутости (кривизны) сетки.

Алгоритм решения вариационной задачи напоминает метод динамических ядер. На каждом шаге алгоритма:

1) Квадратичный функционал D минимизируется при заданном разбиении множества X на таксоны - подмножества точек K_{ij} , для которых данный узел сетки является ближайшим. При фиксированном разбиении

на таксоны функционал D квадратичен по переменным r^{ij} , поэтому задача нахождения его минимума сводится к решению системы линейных уравнений с матрицей размерами $m \times m$, причем коэффициенты самой матрицы системы зависят только от положения узлов r^{ij} , а сама матрица имеет очень простую – трехдиагональную структуру заполнения.

2) Производится новое разбиение множества X на таксоны, соответствующее новым положениям узлов сетки.

Общий критерий D уменьшается при каждой итерации алгоритма, он естественно ограничен снизу нулем, а число разбиений конечного множества X по таксонам конечно. Следовательно, алгоритм сходится.

При построении карты возникает общая для всех оптимизационных задач проблема нахождения глобального минимума функционала D , связанная с выбором начальной карты. В работе не предлагается ее абсолютного решения, однако, достаточно разумных экстремальных значений можно достичь следующим образом:

1) Первоначально карта располагается в плоскости первых двух главных компонент. Ее размеры выбираются равными порядка $1/2 - 1/3$ от длины двух главных полуосей эллипсоида рассеяния.

2) К положениям узлов добавляется малый случайный шум для того, чтобы вывести расположение узлов из одной плоскости и избежать возникновения различных вырожденных случаев.

3) При нахождении минимума функционала D параметры λ , μ меняются от больших значений (порядка 10) к малым (порядка 0,01). В результате карта в начале расчета очень «жесткая», ее кривизна невелика, в результате чего моделируются самые общие особенности распределения точек данных. При уменьшении коэффициентов упругости карта становится «мягкой» и ее узлы располагаются к точкам данных гораздо ближе. Такой метод можно назвать методом *отжига*.

В результате в пространстве данных располагается относительно равномерная сетка G , аппроксимирующая данные. После проектирования данных на сетку расстояния между образами и их скоплениями на карте отражают соответствующие особенности распределения данных в многомерном пространстве.

4.1.4. Проектирование многомерных данных на двумерную сетку

После построения сетки G , которая является точечной аппроксимацией множества X , она должна быть доопределена до многообразия. Самым простым способом является выбор какого-либо способа триангуляции, один из вариантов которой показан на рисунке 4.2.

В результате в многомерном пространстве данных располагается двумерное кусочно-линейное многообразие, на котором соседние вершины соединяются ребрами, а ребра ограничивают плоские грани сетки.

В качестве одного из самых очевидных способов проектирования многомерных данных на двумерную сетку можно предложить нахождение ближайшей точки карты (а не только ближайшей вершины, как в случае SOM). Задача состоит в нахождении ближайшей к данной точке данных x

точки кусочно-линейного многообразия M . Заметим, что ближайшая точка карты может оказаться на грани сетки, на ребре или в вершине сетки. Таким образом, пространство вокруг карты разбивается на области, где ближайшей точкой карты является точка грани, ребра или вершина.

Прежде всего, определим ближайшую к x вершину сетки и обозначим ее радиус-вектор как r^0 , а окружающие ее вершины перенумеруем изображенным на рисунке 2 образом.

Помимо ближайшей вершины r^0 , можно выбрать еще две наиболее близкие к x вершины r^k и r^l из ближайшего окружения r^0 , причем таким образом, чтобы вершины r^0 , r^k , r^l являлись вершинами грани карты при выбранном способе триангуляции. Таким образом, определяем *ближайшее ребро и ближайший треугольник* (грань) карты.

Алгоритм нахождения ближайшей точки карты заключается в следующем:

1. Выполняется ортогональное проектирование x на плоскость, натянутую на вектора $m=r^k-r^0$ и $n=r^l-r^0$. Образом x окажется точка $r^* = r^0 + um + vn$, где

$$u = \frac{(x-r^0, m) \cdot n^2 - (n, m)(x-r^0, m)}{m^2 n^2 - (m, n)^2}, \quad v = \frac{(x-r^0, n) \cdot m^2 - (n, m)(x-r^0, n)}{m^2 n^2 - (m, n)^2}.$$

В случае, если $u \geq 0, v \geq 0, u + v \leq 1$, то ближайшей точкой карты является точка r^* . Иначе:

2. Выполняется ортогональное проектирование на ближайшее ребро, натянутое на вектор $m=r^k-r^0$. При этом

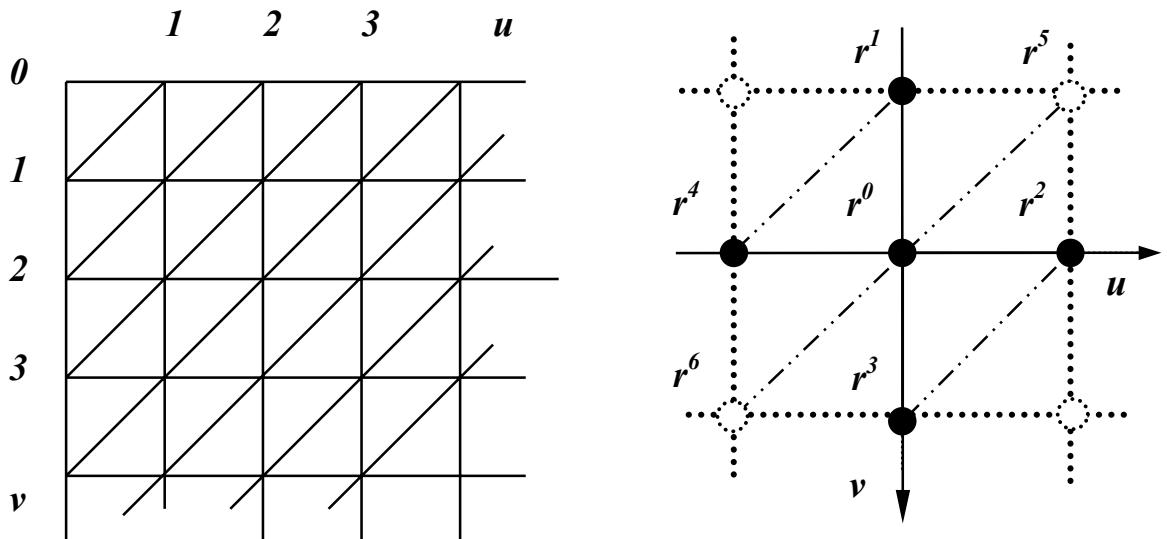


Рис. 4.2. На левом рисунке показан один из вариантов триангуляции сетки. На правом – выбранный способ нумерации узлов, окружающих ближайший к данной точке данных узел.

$$x \rightarrow r^{**} = r^0 + um, \quad u = \frac{(x - r^0, m)}{m^2},$$

и если $0 \leq u \leq 1$ - ближайшей точкой карты является r^{**} . Иначе ближайшей точкой карты является вершина r^0 .

Другой вариант кусочно-линейного проектора. Для случая моделирования конечного множества точек данных с помощью самоорганизующейся кривой (SOC) [131-133] был предложен вариант центральной проекции на одномерную сетку. Точки данных сначала ортогонально проектируются в плоскость трех вершин на сетке – ближайшей к заданной точке данных и двух ее соседей, после этого на плоскости выполняется центральное проектирование, в качестве центра проекции O выбирается точка пересечения серединных перпендикуляров к ребрам ломанной, непосредственно прилегающих к ближайшей к точке данных x вершине (см. рис. 4.3).

Можно перенести этот вариант проектирования на случай двумерной карты, если считать, что проектирование поочередно выполняется сначала для одной, а потом для другой координатной линии двумерной сетки. Таким образом, можно определить значения внутренних координат u, v для точки проекции.

Следует заметить, что для описанных выше алгоритмов (как при

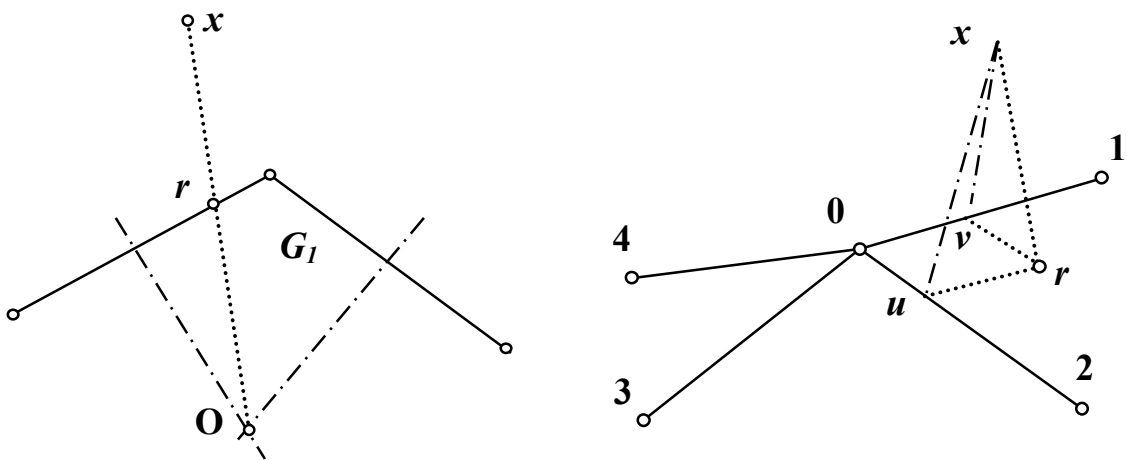


Рис. 4.3. На правом рисунке показан способ нахождения центральной проекции точки x на ломаную, которая является одной из координатных линий u, v двумерного многообразия M . На левом рисунке изображен способ нахождения координат на карте образа точки x .

построении карты, так и при проектировании) не является принципиальным наличие в данных пробелов. В случае, если по какой-либо точке данных известен не полный набор признаков, то вычисление всех необходимых расстояний от точки данных до карты выполняется в пространстве меньшей размерности, где для выбранной точки данных известны все координаты. Таким образом, на карте находят свое

положение также и неполные данные, что позволяет использовать изложенные алгоритмы в качестве средства восстановления пробелов в данных. Хотя практика показывает, что точность восстановления данных при таком способе невысока, тем не менее, пользователь может получить представление о месте данных с пробелами среди всей совокупности данных. Кроме того, ниже рассматривается способ существенно увеличить точность восстановления данных.

4.1.5. Использование карты для анализа распределения данных

После того как карта построена, ее можно «развернуть» (см. рис. 4.1.) и наносить разного рода слои информации [68-71, 134, 135]. В результате карта приобретает разные раскраски, дающие представление об исходном распределении данных. Такие раскраски могут быть построены с использованием богатого арсенала средств и методов ГИС. Таким образом, открывается новое широкое поле деятельности для использования ГИС-технологий для картирования информации самого разного происхождения.

Во-первых, на карте можно изобразить сами данные. При этом можно отображать различные разбиения на подмножества данных, в соответствии со значением того или иного признака.

Во-вторых, на карте можно изобразить произвольные функции координат, поскольку каждой точке с координатами u, v на двумерной карте соответствует точка в многомерном пространстве данных. Самыми простыми раскрасками являются раскраски, отражающие значение той или иной координаты в точках размещения карты. Кроме этого, на карте можно отображать такие координатные функции, как плотность распределения данных в пространстве или плотность того или иного подмножества данных. Сама плотность может быть рассчитана с помощью какой-либо непараметрической оценки. Кроме собственно плотностей подмножеств интерес могут представлять раскраски, отвечающие значению относительных плотностей подмножеств на фоне общего распределения.

В-третьих, на карте можно изображать различные свойства самого моделирующего многообразия. Например, можно получить раскраски, отражающие значения метрических коэффициентов карты или ее кривизны, свидетельствующие о мере ее растянутости и изогнутости на разных участках. Также интерес представляет такая раскраска карты, на которой цветом отражены расстояния от точки карты до ближайшей точки данных в многомерном пространстве. Такая раскраска позволяет оценить степень близости моделирующей карты к точкам данных в разных областях пространства.

Наконец, весьма плодотворной оказалась идея составления по множеству X нескольких карт, которые можно по аналогии с начертательной геометрией образно назвать «главными проекциями множества». Первая из таких карт визуализирует сами данные, вторая «главная проекция» визуализирует погрешности аппроксимации данных (то есть разности между положениями точек данных и их образами на

карте) и позволяет оценить точность описания данных первой картой, и так далее. Последовательность таких «главных проекций» позволяет моделировать данные с высокой точностью, что в случае неполных данных позволяет правдоподобно восстанавливать пропущенные или ремонтировать недостоверные данные.

4.2. Применение методов визуализации произвольных данных к картографированию экономических таблиц

В качестве примера применения описанной технологии метод визуализации произвольных данных был применен для картографирования таблицы крупнейших российских предприятий, взятой из журнала «Эксперт-200» [136]. Файлы исходных данных были получены с официального сайта журнала <http://www.expert.com>.

Исходная таблица содержала информацию об экономическом положении двухсот крупнейших российских предприятий, ранжированную в порядке убывания валового объема производства продукции. Изначально таблица содержала следующие поля-признаки (часть из них является независимыми признаками, часть рассчитывается по явным формулам):

- Название предприятия;
- Регион местонахождения предприятия;
- Отрасль, к которой относится предприятие;
- Валовой объем производства продукции в 1998 году;
- Валовой объем производства продукции в 1997 году;
- Темпы роста предприятия
- Валовой объем производства в 1998 году, выраженный в долларовом эквиваленте;
- Балансовая прибыль предприятия;
- Прибыль предприятия после налогообложения;
- Прибыльность предприятия;
- Число работающих на предприятии;
- Производительность труда.

В работе [137] уже была предпринята попытка визуализации таблицы предприятий, взятой из журнала «Эксперт» за 1997 год. В этой работе были использованы традиционные самоорганизующиеся карты Кохонена и диаграммы Хинтона. Там же было предложено использовать в качестве координат пространства данных отношения некоторых независимых признаков из таблицы. Было предложено четыре таких координаты.

В данной работе было решено расширить пространство исходных данных еще одним измерением, в результате чего был получен набор независимых признаков.

Таблица 1

Описание полей обрабатываемой таблицы

Обозначение признака	Значение	Условное Название
----------------------	----------	-------------------

LG_VO1998	Логарифм валового объема производства продукции в 1998 году	Размер
TEMP	Валовой объем производства продукции в 1998 году / Валовой объем производства продукции в 1997 году	Темп роста
PROFIT_BAL	Балансовая прибыль предприятия / Валовой объем производства продукции в 1998 году	Балансовая прибыль
PROFIT_NAL	Прибыль предприятия после налогообложения / Валовой объем производства продукции в 1998 году	Чистая Прибыль
PRODUCTIV	Прибыль предприятия после налогообложения / Число работающих на предприятии	Производительность

В результате была составлена таблица из двухсот записей с пятью полями. Часть записей содержала неполную информацию (по отдельным признакам информация отсутствовала).

Данные были предварительно нормированы по формуле $\tilde{x}_i = th\left(\frac{x_i - M}{\sqrt{D}}\right)$, где \tilde{x}_i, x_i, M, D – новое, старое значения признака, среднее значение и дисперсия признака соответственно, th – гиперболический тангенс.

Карта, с помощью которой осуществлялась визуализация множества данных, была построена по описанному выше алгоритму построения упругих карт. Первоначальная сетка содержала 10 узлов по вертикали и 10 по горизонтали. Для нахождения локального минимума функционала применялся описанный выше метод отжига. Параметры μ и λ медленно (так чтобы при каждом изменении карта успевала перейти в близлежащий локальный минимум) менялись от значений $\mu = 5, \lambda = 5$ до $\mu = 0.01, \lambda = 0.01$.

После построения упругой карты, данные из пространства признаков, были спроектированы на карту, с помощью описанной выше процедуры нахождения ближайшей точки карты в случае кусочно-линейной интерполяции между узлами.

В качестве иллюстрации анализа экономических данных ниже приведены раскраски полученной карты по координатным полям, а также слой рассчитанной плотности данных в точках карты. На раскрасках большими точками с номерами выделена группа предприятий, принадлежащих нефтегазовой промышленности. Такое выделение позволяет проанализировать место той или иной отрасли промышленности среди других предприятий.

4.2.1. Раскраска по признакам

На рис.4.4а изображено значение признака LG_VO1998 в точках карты. При этом более светлым участкам соответствуют более высокие показатели признака. Самый яркий цвет соответствует первым 10%

предприятий с самым большим валовым объемом производства. Для примера кружками с цифрами выделены предприятия нефтегазовой промышленности. Цифрам соответствуют следующие названия предприятий:

1 – ОАО «Газпром»; 2 – Нефтяная компания «ЛУКОЙЛ»; 3 – Башкирская топливная компания; 4 – Нефтяная компания «Сургутнефтегаз»; 5 – Тюменская нефтяная компания; 6 – «Татнефть»; 7 – Нефтяная компания «Славнефть»; 8 – Нефтяная компания «Роснефть»; 9 – Оренбургская нефтяная компания «Онако»; 10 – Центральная топливная компания; 11 – Нефтяная компания «КомиТЭК».

Рис.4.4б изображает раскраску по показателю ТЕМР. Как видно из рис.4.4б, область крупнейших предприятий не пересекается с областью наиболее высоких темпов роста. В правом нижнем углу карты, например, располагаются предприятия пищевой промышленности, цветной металлургии и другие, быстро развивающиеся отрасли.

На рис. 4.4в, 4.4г, 4.4д показаны раскраски по признакам PROFIT_BAL, PROFIT_NAL, PRODUCTIV. Раскраски этих признаков схожи, что указывает на корреляцию последних трех признаков. Вместе с этим различия в раскраске позволяют выделить предприятия, которые

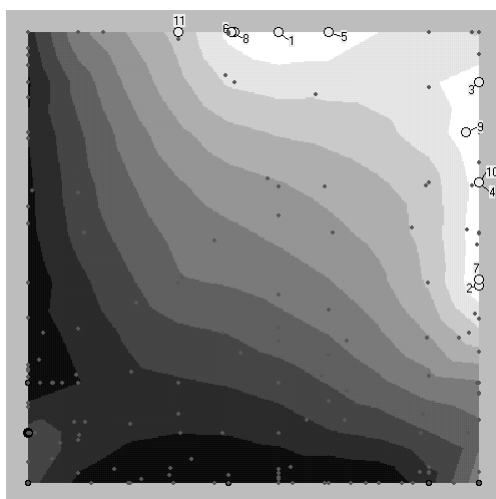


Рис. 4.4а

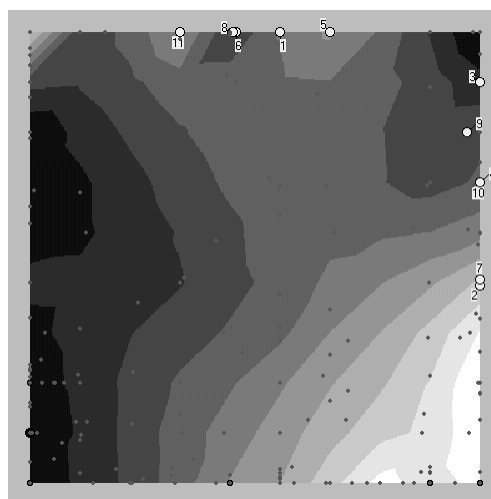


Рис. 4.4б

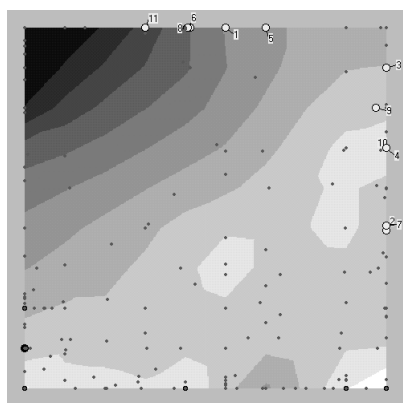


Рис. 4.4в

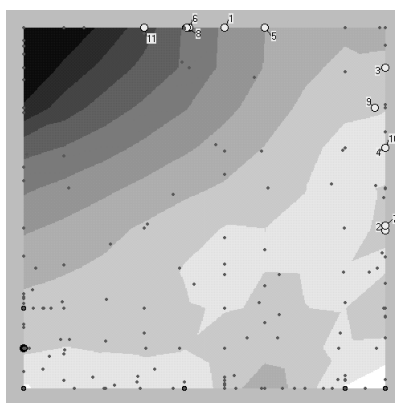


Рис. 4.4г

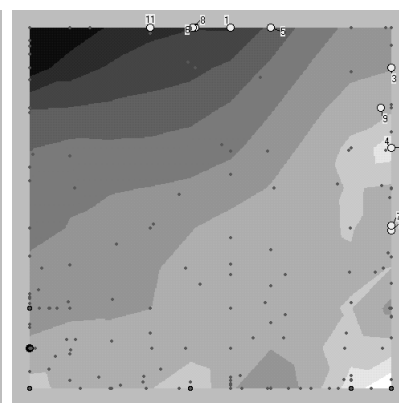


Рис. 4.4д

выпадают из корреляционной зависимости.

4.2.2. Раскраска по плотности данных

На рис. 4.5а, 4.5б, 4.5в показана раскраска карты по плотности

данных, оцененной с помощью какой-либо непараметрической оценки. Существует два способа оценить плотность данных. Во-первых, можно рассматривать двумерное распределение точек на карте. Во-вторых, можно рассчитать плотность точек в исходном n -мерном пространстве, и изображать на карте значения этой плотности в точках расположения карты. На рисунках изображено применение первого способа. Более темным участкам соответствуют более высокие значения плотности.

Рис.4.5а изображает двумерное распределение общей плотности данных. На рис.4.5б – распределение плотности предприятий нефтегазовой промышленности. Рис.4.5в отражает удобную для оценок относительную плотность предприятий нефтегазовой промышленности (то есть отношение первых двух плотностей).

На рис.4.6 отражено расстояние от каждой из точек карты до ближайшей точки данных. Более темным участкам соответствуют большие

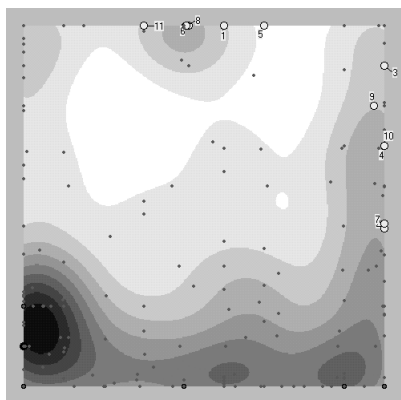


Рис. 4.5а

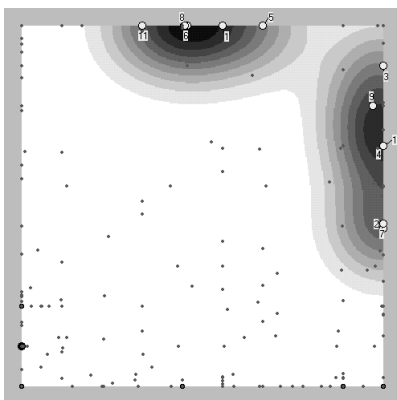


Рис. 4.5б

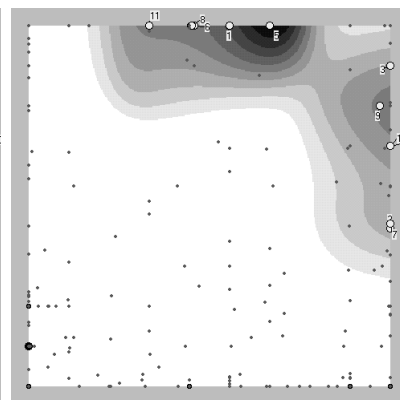


Рис. 4.5в

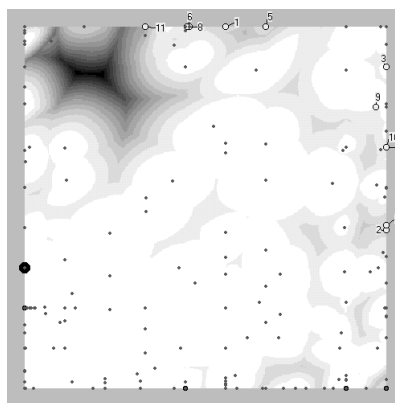


Рис. 4.6

расстояния. Видно, что в целом данные достаточно плотно прилегают к карте, за исключением участка в левом верхнем углу (впрочем, точки данных там отсутствуют и темный цвет указывает на то, что точки в левом верхнем углу карты расположены в многомерном пространстве достаточно далеко от основного массива данных).

Беглый взгляд на рисунки позволяет сделать, например, такие выводы. Предприятия нефтегазовой промышленности являются лидерами по объему валового производства, но темпы роста этой области промышленности невелики по сравнению, например, с пищевой промышленностью. Предприятия нефтегазовой промышленности распадаются на две группы, которые существенно отличаются по прибыльности производства. В целом, набор таких рисунков могут служить удобным средством анализа для специалистов в макроэкономике.

На приведенных рисунках карты изображены во внутренних координатах сетки, в которых сетка является прямоугольной. Можно также изображать карту в проекции на координатные плоскости: данные проецируются на карту, а карта, в свою очередь, на одну из координатных

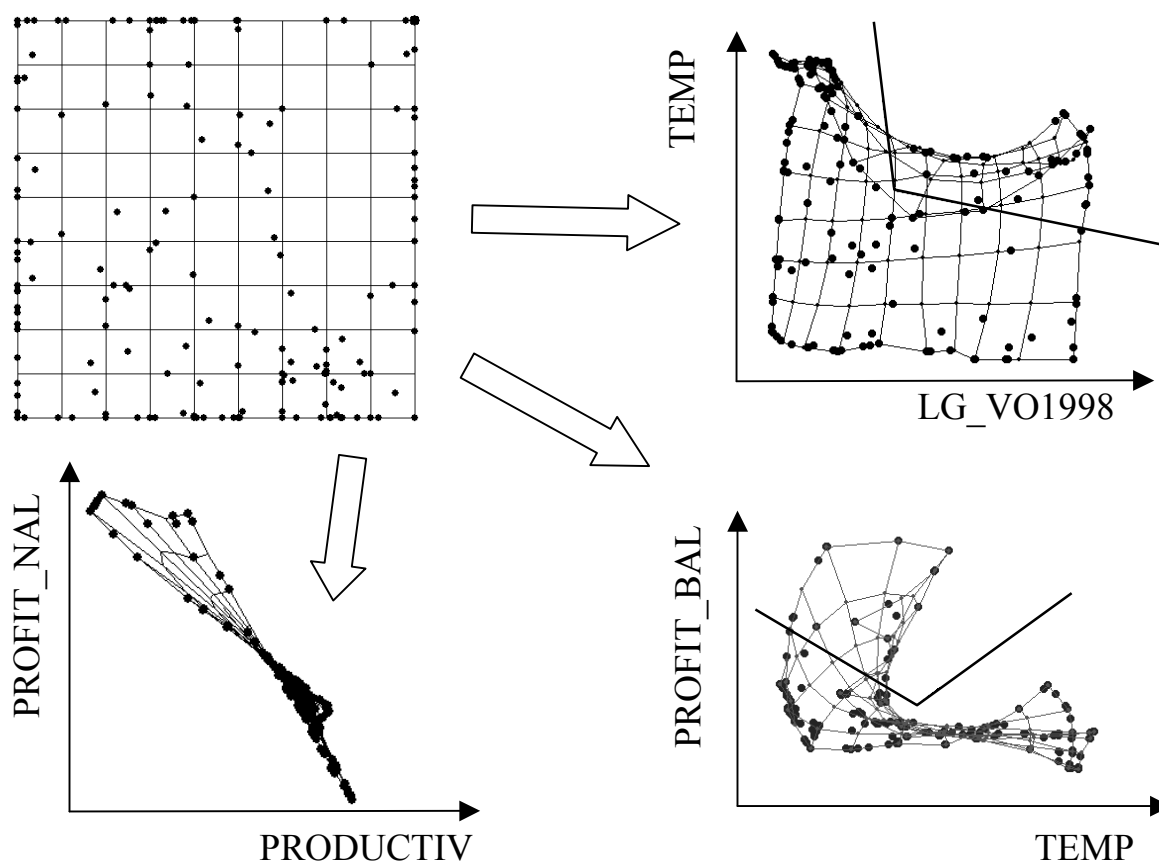


Рис. 4.7. Слева сверху изображена развернутая карта во внутренних координатах. Справа и ниже изображены три проекции карты на три разные пары координатных плоскостей. Две линии, расположенные вне карты, показывают направления первых двух главных компонент.

плоскостей (см. рис.4.7). Еще один из многих вариантов проекций – проекции на плоскости главных компонент.

4.2.3. Восстановление данных, регрессия и прогноз

Как уже упоминалось, данные, предоставляемые для визуализации, могут содержать недостающие значения. Тем не менее, наличие пробелов

в данных нисколько не препятствует построению упругих карт. Все описанные алгоритмы претерпевают минимальную модификацию, которая заключается в следующем: все расстояния между точками при наличии пробелов рассчитываются в пространстве меньшей размерности, где значения признаков известны полностью. На практике это означает, что в

формуле расчета расстояния $r(x, x^0) = \sqrt{\sum_{i=1}^N (x_i - x_i^0)^2}$, где x – точка, чьи

координаты известны не полностью, в сумме исключаются те слагаемые, для которых значение координаты x_i неизвестно. Это замечание относится к вычислению расстояния от точки до узла сетки и к вычислению сумм в функционале D_l .

Таким образом, по неполным данным можно построить упругую карту, положения узлов которой в пространстве известны *полностью*, что позволяет использовать такую карту в качестве средства для восстановления данных с пробелами.

Задачу восстановления данных можно сформулировать следующим образом:

Пусть дано конечное множество точек в многомерном пространстве, причем значения отдельных координат некоторых точек неизвестны. Требуется построить математическую модель множества, с помощью которой можно было бы правдоподобно восстановить недостающие значения координат. К сожалению, не так просто точно определить понятия *правдоподобия* [35]. Поскольку при построении модели данных не используется никаких гипотез о подчинении распределения данных какому-либо статистическому закону (не существует никакой генеральной совокупности), применение статистических оценок не является уместным. Единственным способом оценить правдоподобность процедуры восстановления значений координат является эмпирическое тестирование модели данных, когда на время делается предположение о том, что значение той или иной координаты неизвестно, затем координата восстанавливается с помощью модели и сравнивается с ее известным «истинным» значением.

С практической точки зрения задача восстановления данных может быть сформулирована как задача *регрессии*, то есть построения эмпирических зависимостей одних данных от других. С другой стороны, методы заполнения пробелов в данных могут быть использованы с целью *прогноза*, то есть предоставления правдоподобной оценки каких-либо параметров изучаемой системы при определенных предположениях о других параметрах.

Естественно, что значения восстановленных координат зависят от метода построения моделирующего многообразия и от способа проектирования данных на него. Так при использовании в качестве моделирующего многообразия SOM точка данных, отдельные координаты которой подлежат восстановлению, сопоставляется с ближайшим *узлом*

сетки, что означает замену точки на центр локального сгущения точек данных в пространстве. При использовании упругих карт также возможно кусочно-постоянное проектирование, однако более естественным являются кусочно-линейные способы проектирования (например, алгоритм нахождения ближайшей *точки карты*), описанные выше. Можно ожидать, что в случае кусочно-линейных проекторов точность восстановления будет выше.

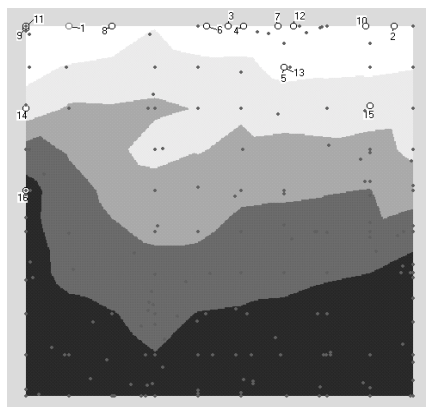
Точность восстановления пропущенных данных можно увеличить, делая упругую карту все более «мягкой». Однако возможности такого подхода ограничены, прежде всего тем, что количество узлов карты ограничено. Кроме того, совершенно «мягкая» карта фактически начинает моделировать случайный шум. Более перспективным является уже упомянутое «мультикартирование», когда после построения основной карты данных для каждой точки вычисляются «остатки» – разности между положением точки и ее проекцией на карту. Полученные остатки также картографируются, в результате чего возникает карта погрешностей. Процедура может быть продолжена до получения необходимой точности.

Рассмотрим процедуру восстановления данных с помощью нескольких карт более формально. Обозначим через M первую карту (то есть карту, построенную по самим данным), а через $P(x)$ – проектор на M , то есть функцию, сопоставляющую произвольной точке данных $x \in X$ точку $r \in M$. Обозначим через X^n , M_n , P_n – множество точек $\{x^n\}$, представляющих собой n -ые остатки (погрешности), карту построенную на основе этих остатков и процедуру проектирования на M_n соответственно.

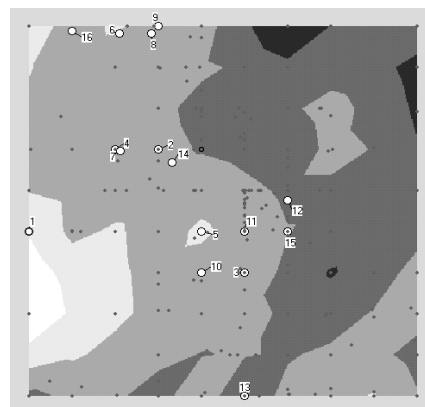
Остатки рассчитываются по формулам $x^1 = x - P(x)$, $x^2 = x^1 - P_1(x^1)$... $x^n = x^{n-1} - P_{n-1}(x^{n-1})$. Заменяя в тождестве $x = P(x) + P_1(x^1) + \dots + P_{n-1}(x^{n-1}) + x^n$ значение x^n на его смоделированное значение $P_n(x^n)$, получаем формулу для восстановления пропущенных значений в данных при помощи картирования n остатков:

$$\tilde{x} = P(x) + \sum_{i=1}^n P_i(x^i), \text{ где } \tilde{x} - \text{восстановленный радиус-вектор (в правой}$$

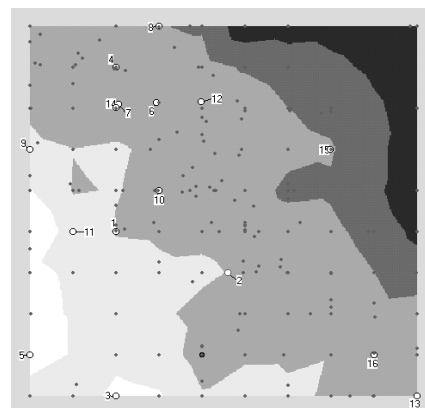
части выражения все координаты известны, несмотря на то, что координаты точки x и остатков x^i могут содержать пропущенные значения). Проиллюстрируем применение изложенных идей на примере картографирования экономических таблиц, приведенном в предыдущем разделе. На рис. 4.8 показаны первые три карты (карта данных, карты первых и вторых остатков). Все три карты приведены с нанесенной раскраской по признаку объема производства. При раскраске применена пятицветная раскраска, средний полутон отвечает значениям координат, близким к нулю. Это означает, что на карте первых остатков точки, попавшие в область среднего оттенка, наилучшим образом описываются моделью данных.



Карта данных



Карта первых остатков



Карта вторых остатков

Рис. 4.8. Карты данных и погрешностей модели.

Раскраска по признаку LG VO1998.

4.2.4. Проблема экстраполяции

Опыт визуализации данных показывает, что в случае не слишком «мягкой» карты проекции на карте имеют тенденцию к группировке вблизи краев карты, поскольку крайние точки карты оказываются ближайшими для большей части точек данных. Вследствие этого возникает идея экстраполяции карты, по крайней мере, на некоторую ее окрестность для того, чтобы не возникало подобных эффектов. В качестве условий, налагаемых на процедуру экстраполяции естественно взять требование того, чтобы вдали от данных карта становилась бы плоской (например, совпадала бы с плоскостью первых двух главных компонент).

При аппроксимации данных одномерными моделями [131-133] приемлемые результаты дает применение формул Карлемана [138].

4.3. Применение технологии для визуализации и анализа таблицы экологических измерений

Для практического применения описанной технологии была создана база данных по результатам измерений экологических показателей (примесей в слое приземного воздуха) за период с 1994-го по 1997 годы на территории города Красноярск.

Данные любезно предоставлены сотрудниками ООО "Экология" А.А. Дубковым, В.Н. Шевниным и А.В. Кучеренко.

По таблице 2 было построено два набора данных:

Набор №1 (Полный). Полный набор данных, содержащий все числовые поля и все измерения.

Набор №2 (Средние значения). Набор данных, в котором были усреднены значения условий и примесей по всем измерениям для каждой из точек, в которой проводились измерения. В результате каждая точка отображает средние многолетние значения измеряемых величин.

Оба набора данных состояли из одинакового набора количественных признаков.

Таблица 2

Описание полей исходной таблицы данных

Названия групп полей	Название поля	Описание
Географическое положение точек на карте	X Y ZONE REGION POINT	географические координаты на карте города Красноярска условные зоны города части зон условные названия точек
Даты и времена замеров	DATE TIME	день, месяц, год доля суток от полуночи
Метеорологические характеристики условия замеров	WS WD TEMP RH BP SR	скорость ветра, м/с направление ветра, градусы температура воздуха, °C влажность, % атмосферное давление, мм.рт.ст. солнечная радиация, МДж/м ²
Концентрации примесей в приземном слое атмосферы, в %	CO NO NO2 CO2 H2S SO2 O3 HCL	окись углерода окись азота диоксид азота диоксид углерода сероводород диоксид серы озон хлористый водород

Таблица 3

Способы нормировки признаков

Признак	Название	Нормировка
XX	Координата X	На среднеквадратичное отклонение
YY	Координата Y	На среднеквадратичное отклонение
CO	Концентрация CO	На среднеквадратичное отклонение
NO	Концентрация NO	На среднеквадратичное отклонение
NO2	Концентрация NO ₂	На среднеквадратичное отклонение
CO2	Концентрация CO ₂	На среднеквадратичное отклонение
H2S	Концентрация H ₂ S	Логарифм. преобразование и на ср.кв.отк.
SO2	Концентрация SO ₂	Логарифм. преобразование и на ср.кв.отк.
O3	Концентрация O ₃	Логарифм. преобразование и на ср.кв.отк.
HCL	Концентрация HCl	Логарифм. преобразование и на ср.кв.отк.
WS	Скорость ветра	Логарифм. преобразование и на ср.кв.отк.
WD	Направление ветра	На среднеквадратичное отклонение
TEMP	Температура	На среднеквадратичное отклонение
RH	Влажность	На среднеквадратичное отклонение
BP	Давление	На среднеквадратичное отклонение
SR	Солнечная радиация	Логарифм. преобразование и на ср.кв.отк.

Как видно из таблицы 3, к некоторым признакам предварительно

применялось логарифмическое преобразование в связи с тем, что их значения отличались на порядки.

По *Набору №1* была построена информационная модель полных данных, на основе которой был получен *Атлас №1* информационных раскрасок, представленный в стандартном ГИС-формате ГеоГраф 1.5. Атлас описан в разделе в разделе "*Атлас №1, №2 информационных раскрасок*". Отдельные слои из этого Атласа приведены в приложении.

По *Набору №2* была построена информационная модель усредненных данных, на основе которой был получен *Атлас №2* информационных раскрасок в формате ГеоГраф 1.5. Отдельные слои из *Атласа №2* приведены в приложении.

Проведенные эксперименты показали высокую правдоподобность данных в таблице. В частности, проводились эксперименты по предсказанию содержания примесей в атмосфере в зависимости от места, времени и атмосферных условий. При этом было выявлено, что более чем в 50% случаях относительная ошибка предсказания не превышает 10%. На долю же сильных ошибок (более 100% ошибки) выпало максимум 10% всех примеров, что может говорить о том, что в данных примерах имеют место ошибки в сборе и представлении информации. Причем по разным показателям эта доля варьируется – к примеру, содержание примесей серы предсказывается с ошибкой 5-10% более чем в 80% примерах и на долю больших ошибок приходится не более 1-2% всех примеров. Такая же картина и по содержанию озона и хлористого водорода. Несколько другая картина с соединениями углерода и азота – здесь процент ошибок выше, но все равно позволяет сделать вывод о правдоподобности этих данных.

При использовании моделей для предсказания содержания определенных примесей в атмосфере по всем остальным параметрам и соединениям ошибка достаточно мала и лишь в небольшом проценте примеров есть серьезные отклонения предсказанных значений от исходных. Что говорит о высокой правдоподобности данных и наличии небольшого числа, возможно, ошибочных значений.

Следует заметить, что наличие искажений не обязательно влечет за собой вывод об ошибочной их природе. Искажения могут быть вызваны влиянием каких-нибудь неучтенных фактора, поэтому большое наличие неправдоподобных значений может потребовать пересмотра всей модели данных и используемых при этом признаков.

Атлас №1, №2 информационных раскрасок

Основа визуализации данных – *карта данных*. Точки, близкие в исходном многомерном пространстве, оказываются близкими на карте данных. В нашем случае точка данных представляет собой проведенное измерение экологической обстановки. Это означает, что близкие на карте данных близкие точки изображают измерения, показавшие «схожесть» экологической обстановки по совокупности измеряемых величин. На рис. 1-16 приложения показана карта данных с нанесенными на нее точками данных.

Можно сказать, что карта данных подобна географической карте с той разницей, что на географической карте рядом оказываются объекты с близкими географическими координатами, а на карте данных – *объекты с близкими свойствами*. На географической карте рядом могут оказаться объекты с совершенно различными характеристиками, кроме того, некоторые объекты просто не имеют никакой привязки к географии. На карте данных учитывается сходство объектов «по совокупности свойств» в смысле расстояния в многомерном пространстве.

Простое изображение данных дает для исследователя не слишком много информации. Для того, чтобы представить себе какие особенности измерений соответствуют различным областям карты, исследователь получает представленный на рис. 1–16 (см. приложение) ГИС-атлас раскрасок. Сравнивая различные раскраски карты, исследователь получает представление о наборе данных. Схожесть раскрасок указывает на скореллированность соответствующих признаков.

Важной особенностью предлагаемой технологии является представление конечного продукта – атласа, в формате ГИС ГеоГраф 1.5. С помощью стандартной ГИС исследователь получает возможность генерации собственных раскрасок, подчеркивая те или иные интересующие его особенности набора данных, видоизменяя представление точек данных, накладывая различные слои друг на друга и находя их пересечения.

Атлас №2 имеет структуру, аналогичную *Атласу №1*. Разница состоит лишь в том, что он построен по *Набору данных №2* (усредненные значения). Соответственно, если на первом атласе показано 560 точек измерений, то на втором – 113 точек, соответствующих разным точкам города, в которых проводились измерения.

Итак, разработана и опробована на практике технология визуализации или картографирования многомерных данных (в которых могут содержаться пробелы) с помощью вложенных в пространство данных двумерных многообразий, названных упругими картами. И алгоритм построения этих многообразий, и общая идеология визуализации данных с их помощью существенно отличается от общепринятой на сегодняшний день технологии SOM. Особенностью технологии также является возможность непрерывного проектирования данных на карту, что существенно повышает точность представления данных.

Еще раз стоит отметить, что описанная технология открывает перспективы для использования всего арсенала методов и средств, накопленных в ГИС-технологиях для картирования данных самой различной природы, без привязки к географическим координатам. Можно сказать, что вместо географической карты в описанной технологии используется подложка, образованная структурой самих данных.

Атлас информационных раскрасок №1

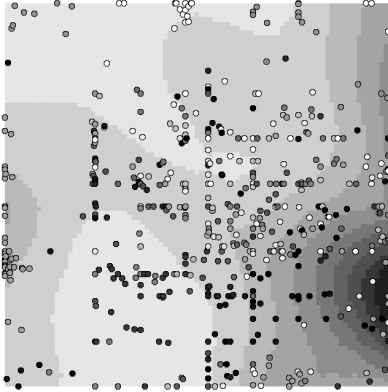


Рис. 1. Концентрация CO

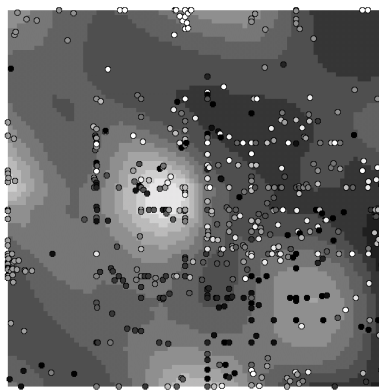


Рис. 2. Концентрация CO2

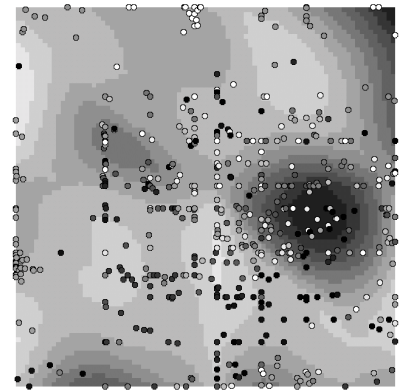


Рис. 3. Концентрация NO

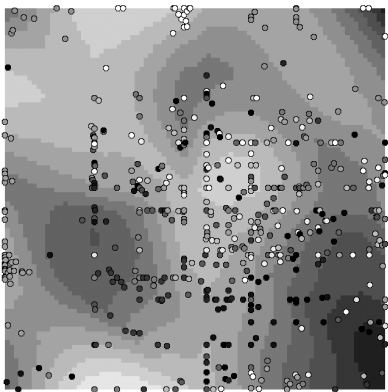


Рис. 4. Концентрация NO2

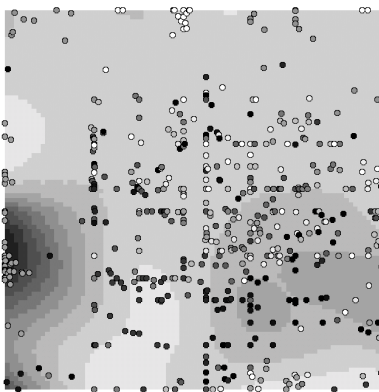


Рис. 5. Концентрация H2S

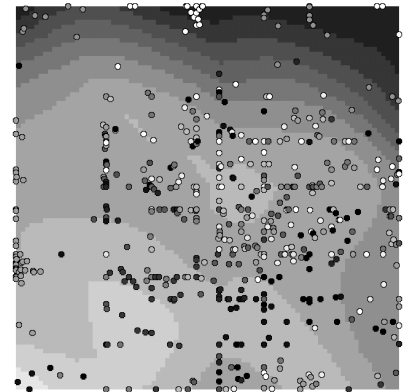


Рис. 6. Концентрация SO2

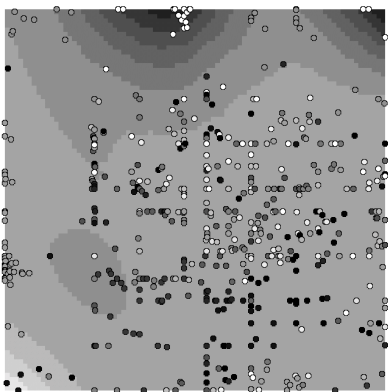


Рис. 7. Концентрация O3

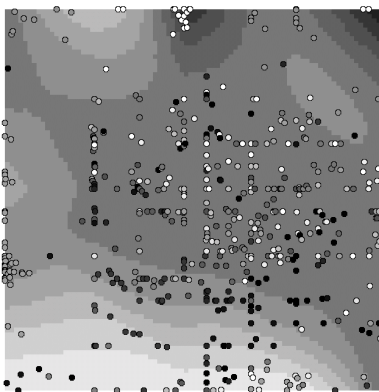
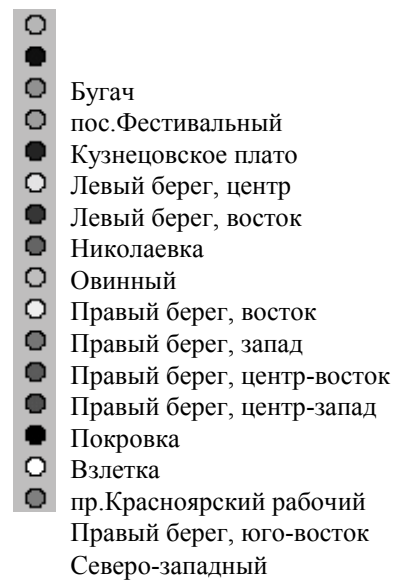


Рис. 8. Концентрация HCL



Атлас информационных раскрасок №2

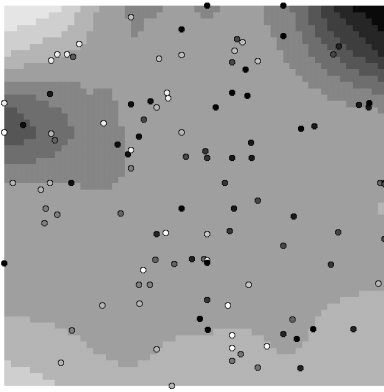


Рис. 9. Концентрация CO

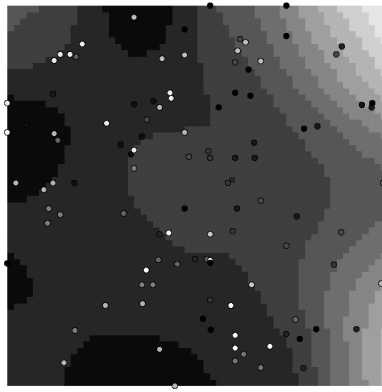


Рис. 10. Концентрация CO2

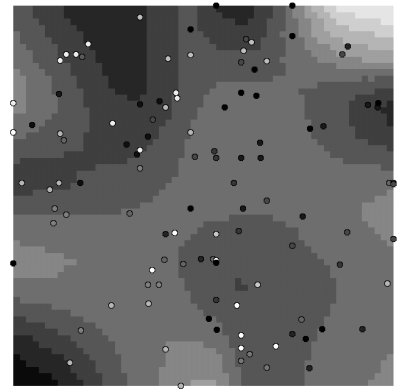


Рис. 11. Концентрация NO

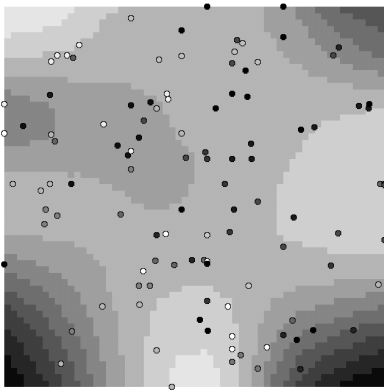


Рис. 12. Концентрация NO2

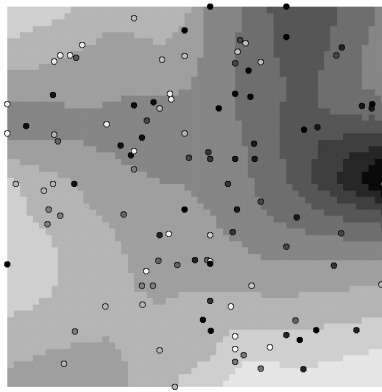


Рис. 13. Концентрация H2S

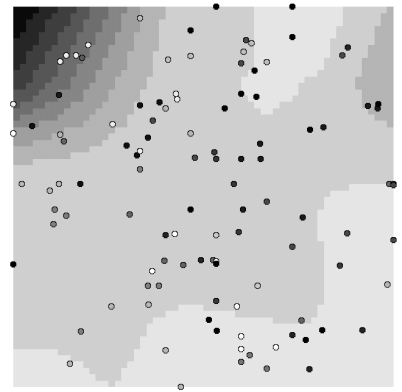


Рис. 14. Концентрация SO2

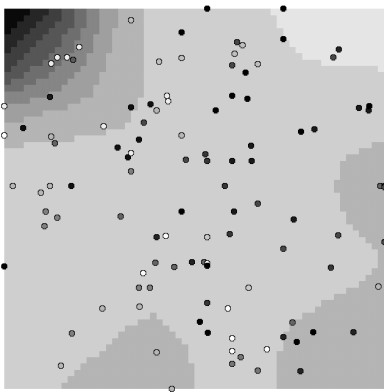


Рис. 15. Концентрация O3

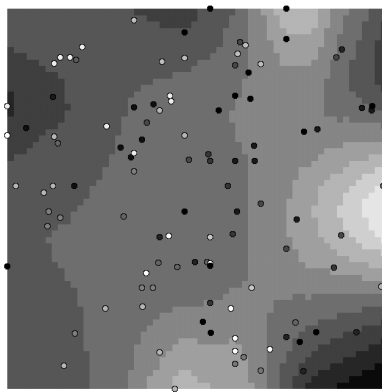
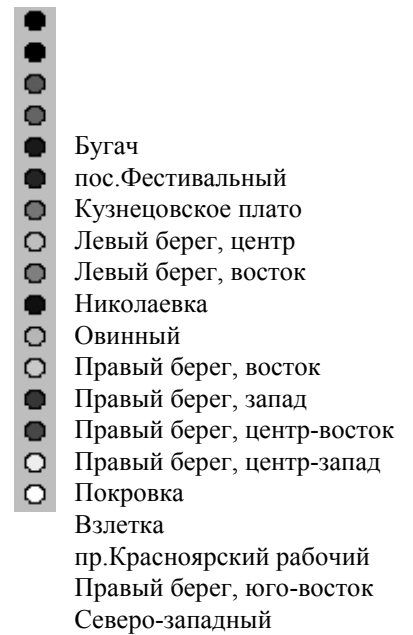


Рис. 16. Концентрация HCL



Заключение

В работе разработаны и реализованы технология и методы нейросетевого анализа в геоинформационных системах (ГИС) направленные на решение широкого спектра 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. В том числе получены следующие результаты:

1. Проведен анализ задач в ГИС и их решения в нейросетевом базисе, возможностей и методов применения нейросетевых моделей в составе прикладных геоинформационных систем.

2. Разработана и реализована технология применения искусственных нейронных сетей в геоинформационных системах. Технология включает в себя методы интеграции ИНС и ГИС и решение задач с помощью нейронных сетей.

3. Разработана технология решения задач в ГИС, включающая технологию получения данных из ГИС, методы создания и обучения ИНС с нелинейными Паде-преобразователями и технологию решения аналитических задач в ГИС. Впервые реализованы сети Паде-нейронов.

4. Разработана технология визуализации, картографирования и анализа многомерных данных (в которых могут содержаться пробелы) с помощью вложенных в пространство данных двумерных многообразий, названных упругими картами. Особенностью технологии также является возможность непрерывного проектирования данных на карту, что существенно повышает точность представления данных.

5. Разработана и реализована программная система для решения задач ГИС с помощью нейросетевых моделей. Система включает: подсистему взаимодействия с ГИС-компонентой, подсистему получения и подготовки данных, подсистему нейросетевой обработки данных. Программа позволяет выполнять как стандартные операции ГИС, так и традиционные операции полнофункционального нейроимитатора общего назначения, а также предобработку и визуализацию результатов.

6. Проведено экспериментальное исследование разработанных технологий с использованием реальных данных, которое подтвердило защищаемые положения. Метод нейросетевого анализа данных в ГИС апробирован на данных электронного экологического атласа города Красноярска.

Литература

1. Коновалова Н.В., Капралов Е.Г. Введение в ГИС. – М. – 1997. – 160с.
2. Хаксхолд В. Введение в городские географические информационные системы. New York: Oxford. – 1991. – 321 с.
3. Bouille F. Towards 2000: The actual main trends in future GIS// Eur. Transit.: Context of GIS: Conf. Proc., Brno, Aug. 28th-31st, 1994. – Brno. – 1994. – P.13-27.
4. Кошкарев А.В., Тикунов В.С., Геоинформатика. – М., Картгеоцентр-Геодезиздат. – 1993. – 348 с.
5. Берлянт А.М. Геоэконика. – М. – 1996. – 208 с.
6. Мартыненко А.И. Картографическое моделирование и геоинформационные системы // Геод. и картогр. –1994. – N 9. –С. 43-45.
7. Тикунов В.С. Моделирование в картографии: Учебник. М.: Изд-во МГУ. – 1997. – 405 с.
8. Берлянт А.М. Теоретические проблемы картографии. М.: Изд-во МГУ. – 1993. – 116 с.
9. Лучко С.А. Анализ данных с помощью ГИС. Региональная информационная система. // Труды международной конференции «ГИС для оптимизации природопользования в целях устойчивого развития территорий» (ИНТЕРКАРТО – 4). –Барнаул: Изд-во Алт. гос. ун-та. – 1998. – С. 91-93.
- 10.Питенко А.А. Новый подход к решению аналитических задач в ГИС // Материалы конференции молодых ученых Института вычислительного моделирования СО РАН, март 1999г. – Красноярск: ИВМ СО РАН. – 1999. – С.89–90.
11. Питенко А.А. Использование нейросетевых технологий при решении аналитических задач в ГИС // Методы нейроинформатики: сборник научных трудов / Под ред. А.Н. Горбаня; Отв. за вып. М.Г.Доррер. – Красноярск: КГТУ. – 1998. – С.152–163.
- 12.Fischer M.M. From conventional to knowledge-based geographic information systems // Comput., Environ, and Urban Syst. – 1994. – 18, N 4. – P. 233-242.
- 13.Zhang J., Zhang J. 90'S GIS software system desing consideration// Cehui Xuebao Acta geo-daet. et cartogr.sin. – 1994. – 23.N2. – P.127-134.
- 14.Марков Н.Г., Горяев Р.И., Захарова А.А., Ковин Р.В., Черноусов М.В. Математический аппарат для построения тематических карт при изучении и использовании недр // Трансферные технологии в информатике. – Томск: Изд-во ТПУ. – 1999. – Вып 1. – С. 53-61.
- 15.Виноградов Б.В., Сорокин А.Д., Федотов П.Б., Фролов Д.Е., Картографирование долговременной динамики сложных экосистем с помощью повторных аэрокосических съемок и динамических ГИС технологий // Труды международной конференции «ГИС для

- оптимизации природопользования в целях устойчивого развития территорий» (ИНТЕРКАРТО – 4). – Барнаул: Изд-во Алт. гос. ун-та. – 1998. – С. 26-37.
16. Захарова А.А., Ковин Р.В., Марков Н.Г., Напрюшкин А.А. ГИС – технология для проведения ландшафтно – экологических исследований // Труды международной конференции «ГИС для оптимизации природопользования в целях устойчивого развития территорий» (ИНТЕРКАРТО – 5). – Якутск: Изд-во Якутского ун-та. – 1999. 3ч. – С. 36-45.
17. Кравецкий А.С., Демьянов В.В., Каневский М.Ф., Савельева Е.А., Тимонин В.А., Чернов С.Ю. Картирование пространственных данных при помощи многослойного перцептрона и геостатистики. Препринт : № ИВРАЕ-99-03. Москва: Институт проблем безопасного развития атомной энергетики РАН. – 1999. – 41 с.
18. Питенко А.А. Нейросетевая аналитическая обработка и оценка значимости параметров для ГИС // Труды международной конференции «ГИС для оптимизации природопользования в целях устойчивого развития территорий» (ИНТЕРКАРТО – 4). – Барнаул: Изд-во Алт. гос. ун-та. – 1998. – С. 161-167.
19. Питенко А.А. Картографирование всех и всяческих данных // Труды международной конференции «ГИС для оптимизации природопользования в целях устойчивого развития территорий» (ИНТЕРКАРТО – 5). – Якутск: Изд-во Якутского ун-та. – 1999. 1ч. – С. 71-79.
20. Александрова Т.Д. Статистические методы изучения природных комплексов. М.: Наука. – 1975. – 96 с.
21. Крауклис А.А. Географический прогноз и результаты изучения динамики геосистем // Модели и методы оценки антропогенных изменений геосистем. – Новосибирск: Наука. – 1986. – С. 12-22.
22. Гуссейн-Заде С.М., Тикунов В.С. Проблемы использования методов автоматической классификации в географии. Вестн. Моск. ун-та, сер. геогр. – 1988. – С.80-86.
23. Гуссейн-Заде С.М., Тикунов В.С. Состояние, проблемы и перспективы классификации в географии. – В кн.: Классификация в современной науке. Новосибирск, Наука, Сибирское отд. – 1989. – С.119-129.
24. Тикунов В.С. Классификации в географии: ренессанс или увядание? (опыт формальных классификаций). – Москва-Смоленск: Изд-во СГУ. – 1997. – 367с.
25. Трофимов А.М., Заботин Я.И., Панасюк М.В., Рубцов В.А. Количественные методы районирования и классификации. – Казань: изд-во Казанск. ун-та. – 1985. – 120с.
26. Griffith D. Toward a theory of spatial statistics. – Geographical Analysis. – 1980. – P.325-339.
27. Scott L.M. Identification of GIS attribute error using exploratory data analysis// Prof. Geogr. – 1994. – 46, N 3. – P. 378- 386.

28. Shen Q. An application of GIS to the measurement of spatial autocorrelation//Comput., Environ. and Urban Syst. – 1994. – 18, N 3. – P. 167-191.
29. Тикунов В.С. Математизация тематической картографии. – Владивосток. – 1986. – 24с.
30. Тикунов В.С. Метод классификации географических комплексов для создания оценочных карт. – Вестник Моск. ун-та, сер. геогр. – 1985. – С.28-36.
31. Wang F. The use of artificial neural networks in a geographical information system for agricultural land-suitability assessment// Environ, and Plann. A. – 1994. – 26, N2. – P.265-284.
32. Цветков В.Я. Геоинформационные системы и технологии. – М.: Финансы и статистика. – 1998. – 288 с.
33. Питенко А.А. Нейросетевая парадигма решения аналитических задач в ГИС // "Студент и научно-технический прогресс": Информационные технологии. Материалы XXXVII международной научной студенческой конференции.–Новосибирск: НГУ.– 1999. – С.34.
34. Питенко А.А. Нейросети для геоинформационных систем. // Всероссийская научно-техническая конференция “Нейроинформатика-99”. Сборник научных трудов. Ч.1. М.: МИФИ.– 1999. С.65–68.
35. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. – Новосибирск: Наука. – 1996. – 276с.
36. Миркес Е.М. Нейрокомпьютер. Проект стандарта. Новосибирск.: Наука. – 1998. – 188с.
37. Little R.J., Schlushter M.D. Maximum likelihood estimation for mixed continuous and categorical data with missing values. – Biometrika. – 1985. – Vol. 72. – P. 497-512.
38. Beale E.M., Little R.J. Missing values in multivariate analysis. – J. Roy. Statist. Soc. B. – 1975. – Vol. 37. – P. 129-145.
39. Buck S.F. A method of estimation of missing values in multivariate data. – J. Roy. Statist. Soc. B. – 1960. – Vol. 22. – P. 202-206.
40. Afifi A.A., Elashoff R.M. Missing observations in multivariate statistics. – J. Amer. Statist. Assoc. – 1966. – Vol. 61. – P. 595-604.
41. Gilev S.E., Gorban A.N., Mirkes E.M. Several Methods for Accelerating the Training Process of Neural Networks in Pattern Recognition // Adv. Modelling & Analysis, A. AMSE Press. – 1992. – Vol. 12, № 4 – P. 29–53.
42. Gleason T.C., Staelin R. A proposal for handling missing data. – Psychometrika. – 1975. – Vol. 40. – P. 229-252.
43. Hocking R.R., Marx D.L. Estimation with incomplete data: an improved computational method and the analysis of nested data. – Commun Statist. A. – 1979. – Vol. 8. – P. 1151-1181.
44. Little R.J., Rubin D.B. Statistical analysis with missing data. – New York, Wiley. – 1987. – 430 p.
45. Загоруйко Н.Г., Ёлкина В.Н., Лбов Г.С. Алгоритмы обнаружения эмпирических закономерностей. — Новосибирск: Наука, 1985. — 110с.

46. Загоруйко Н.Г., Ёлкина В.Н., Тимеркаев В.С. Алгоритм заполнения пропусков в эмпирических таблицах (алгоритм "ZET") // Вычислительные системы. – Новосибирск. – 1975. – Вып. 61. Эмпирическое предсказание и распознавание образов. – С. 3-27.
47. Горбань А.Н., Миркес Е.М., Свитин А.П. Полуэмпирический метод классификации атомов и интерполяции их свойств // Математическое моделирование в биологии и химии. Новые подходы. — Новосибирск: Наука. Сиб. отделение. – 1992. – С.204-220.
48. Горбань А.Н., Новоходько А.Ю. Нейронные сети в задаче транспонированной регрессии, Второй Сибирский Конгресс по Прикладной и Индустриальной Математике, Тезисы докладов. Новосибирск. – 1996. – С.160-161.
49. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука. – 1979. – 448с.
50. Вапник В.Н., Червоненкис А.Ф. Теория распознавания образов. - М.: Наука. – 1974.
51. Исследование зависимостей. -М.: Финансы и статистика. – 1985.
52. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. — М.: Наука. – 1976. – 736 с.
53. Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука. – 1973. – 900 с.
54. Корнейчук Н.П. Сплаины в теории приближения, -М: Наука. – 1984. – 352 с.
55. Крылов В.И., Бабков В.В., Монастырный П.И. Интерполирование и интегрирование - Минск: Наука и техника. – 1983. – 287 с.
56. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. - М.: Финансы и статистика. – 1989.
57. Коваленко И. Н., Филиппова А.А. Теория вероятностей и математическая статистика. М.: Высш. школа. – 1982. – 256 с.
58. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. – М.: Мир. – 1982. – 488с.
59. Ahmad S., Tresp V. Classification with missing and uncertain inputs, Proc. of the 1993 IEEE ICNN. – P. 1949–1954.
60. Рао С.Р. Линейные статистические методы. — М.: Наука. – 1968. –548 с.
61. Юл Дж. Э., Кендэл М. Дж. Теория статистики. М.: Госстатиздат. – 1960. – 376 с.
62. Червяков В.А. О картографо-статистическом методе // Математические методы в географии. М.: Изд-во МГУ. – 1968. – С.140-142.
63. Дуда Р., Харт П. Распознавание образов и анализ сцен. - М.: Мир. – 1976. – 512с.
64. Загоруйко Н.Г. Методы распознавания и их применение. М.: Сов. радио. – 1972. – 206 с.
65. Фор А. Восприятие и распознавание образов.- М.: Машиностроение. – 1989. – 272с.

66. Киселев В.Г., Яковлев А.Ф. О некоторых математических задачах, возникающих при реализации на ЭВМ географических информационных систем // Кибернетика и вычислительная техника. Вып. 3. – М.: Наука. – 1987. – С. 277-296.
67. Питенко А.А. Визуализация и моделирование различных данных. // Нейроинформатика и ее приложения. Тезисы докладов VII Всеросс. семинара.– Красноярск: КГТУ. – 1999. – С.114-115.
68. Зиновьев А.Ю., Питенко А.А. Визуализация произвольных данных. // I Всесибирский конгресс женщин математиков. Тезисы докладов. ИВМ СО РАН: Красноярск. – 2000. – С.76.
69. Зиновьев А.Ю., Питенко А.А. Визуализация произвольных данных методом упругих карт // Материалы конференции молодых ученых Красноярского научного центра СО РАН, апрель 2000г. – Красноярск: КНЦ СО РАН. – 2000. – С.18–20.
70. Зиновьев А.Ю., Питенко А.А., Россиев А.А. Проектирование многомерных данных на двумерную сетку. // 2-я Всероссийская научно-техническая конференция “Нейроинформатика-2000”. Сборник научных трудов. Ч.1. М.: МИФИ. – 2000. – С.80-88.
71. Зиновьев А.Ю., Питенко А.А. Картографирование произвольных данных. // "Студент и научно-технический прогресс": Информационные технологии. Материалы XXXVIII международной научной студенческой конференции.–Новосибирск: НГУ.– 2000. – С.38.
72. Нейроинформатика / А.Н. Горбань, В.Л. Дунин-Барковский, А.Н.Кирдин, Е.М.Миркес, А.Ю.Новоходько, Д.А.Россиев, С.А.Терехов, М.Ю.Сенашова, В.Г.Царегородцев.-Новосибирск: Наука, Сибирская издательская фирма РАН. – 1998. – 296 с.
73. Gorban A.N., Novokhodko A.Yu. Neural Networks In Transposed Regression Problem, Proc. INNS WCNN '96.
74. Демиденко Е.З. Линейная и нелинейная регрессия. - М.: Финансы и статистика. – 1981. – 302 с.
75. Россиев А.А. Моделирование данных при помощи кривых для восстановления пробелов в таблицах. // Методы нейроинформатики: Сб. Научных трудов. Красноярск: Изд-во КГТУ. – 1998. – 204 с.
76. Горбань А.Н., Новоходько А.Ю., Царегородцев В.Г. Нейросетевая реализация транспонированной задачи линейной регрессии, Нейроинформатика и ее приложения: Тезисы докладов IV Всероссийского семинара. Красноярск. – 1996. – С.37–39.
77. Kohonen T. Self-Organizing Maps. Springer: Berlin – Heidelberg, 1997.
78. Нейроинформатика / А.Н.Горбань, В.Л.Дунин-Барковский, А.Н.Кирдин, Е.М.Миркес, А.Ю.Новоходько, Д.А.Россиев, С.А.Терехов, М.Ю.Сенашова, В.Г.Царегородцев.-Новосибирск: Наука, Сибирская издательская фирма РАН. – 1998. – 296 с.
79. Bishop C.M. Theoretical foundation of neural networks. Aston Univ., UK Tech.Rep.NCRG-96-024, Neural computing research group. – 1996. – 8p.

80. Методы нейроинформатики. Сб. научн. трудов. / Под ред. А.Н. Горбаня. Красноярск: КГТУ. – 1998. – 204 с.
81. Fiesler E. Neural network classification and formalization. - Computer Standarts and interfaces. V.I 6, Elseiver Science publishers, Amsterdam. – 1994. – 13p.
82. Kernsley, D.H., T. R. Martinez. A Survey of Neural Network Research and Fielded Applications // International Journal of Neural Networks: Research and Applications. – Vol. 2, No. 2 / 3 / 4. – 1992. – P.123-133
83. Гилева Л.В., Гилев С.Е., Горбань А.Н. Нейросетевой бинарный классификатор "CLAB" (описание пакета программ). Красноярск: Ин-т биофизики СО РАН. – 1992. – 25 с. Препринт № 194 Б.
84. Гилева Л.В., Гилев С.Е., Горбань А.Н., Гордиенко П.В., Еремин Д.И., Коченов Д.А., Миркес Е.М., Россиев Д.А., Умнов Н.А. Нейропрограммы. Учебное пособие: В 2 ч. // Красноярск, Красноярский государственный технический университет. – 1994. – 260 с.
85. Горбань А.Н. Обучение нейронных сетей. М.: изд. СССР-США СП «ParaGraph». – 1990. – 160 с.
86. Горбань А.Н. Нейрокомпьютер, или Аналоговый ренессанс. Мир ПК. – 1994. – № 10. – С. 126-130.
87. Armitage W.D., Lo J.-C. Enhancing the robustness of a feedforward neural network in the presence of missing data, Proc. of the IEEE ICNN, Orlando, FL, USA. – 1994 June. – Vol.2. – P.836–839.
88. Гилев С.Е., Коченов Д.А., Миркес Е.М., Россиев Д.А. Контрастирование, оценка значимости параметров, оптимизация их значений и их интерпретация в нейронных сетях. // Нейроинформатика и ее приложения: Материалы III Всероссийского семинара; КГТУ. Красноярск. – 1995. – 229 с.
89. Искусственный интеллект; В 3 кн. Кн.2 Модели и методы; Справочник / Под ред. Д.А. Поспелова. - М.: Радио и связь. – 1990. – 304 с.
90. Искусственный интеллект; В 3 кн. Кн. 1. Системы общения и экспертные системы; Справочник / Под ред. Э.В. Попова. – М.: Радио и связь. – 1990. – 464с.
91. Нильсон Н. Искусственный интеллект. – М.: Мир. – 1973. – 270 с.
92. Маккалок Дж., Питтс У. Логические исчисления идей, относящихся к нервной деятельности. // Автоматы. М.: ИЛ. – 1956.
93. Минский М., Пайперт С. Перцептроны. М.: Мир. – 1971.
94. Розенблатт Ф. Принципы нейродинамики. М.: Мир. – 1965.
95. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применения в экономике и бизнесе. – М.: МИФИ. – 1998. – 224 с.
96. Горбань А.Н. Быстрое дифференцирование, двойственность и обратное распространение ошибки / Нейроинформатика Новосибирск: Наука, Сибирская издательская фирма РАН. – 1998. – С. 73-100.
97. Дунин-Барковский В.Л. Информационные процессы в нейронных структурах. – М.: Наука. – 1978.

98. Соколов Е.Н., Вайткявичус Г.Г. Нейроинтеллект: от нейрона к нейрокомпьютеру. М.: Наука. – 1989. – 238 с.
99. Rumelhart D.E., Hinton G.E., Williams R.J. Learning internal representations by error propagation. – Parallel Distributed Processing: Exploration in the Microstructure of Cognition, D.E.Rumelhart and J.L.McClelland (Eds.), vol. I, Cambridge. – MA: MIT Press. – 1986. – P. 318 - 362.
100. Poggio T. and F. Girosi A Theory of Networks for Approximation and Learning. MIT AI memo 1140. – 1989. – 87 p.
101. Hecht-Nielsen R. Theory of the backpropagation neural network. - Neural Networks for Human and Mashine Perception. H.Wechsler (Ed.). Vol. 2. Boston, MA: Academic Press. – 1992. – P. 65 - 93.
102. Olmsted D. D. History and Principles of Neural Networks <http://www.neurocomputing.org/history.htm>
103. Bishop C.M. Regularization and complexity control in feed-forward networks. Aston University, Tech.Rep.NCRG-95-022, Neural computing research group. – 1995. – 8p.
104. Sarle W. Stopped training and other remedies for overfitting. In Proc. of the 27th Symposium on the Interface. – 1995. – 10 p.
105. Полак Э. Численные методы оптимизации. Единый подход. М., Мир. – 1974. – 376 с.
106. Самарский А.А., Гулин А.В. Численные методы: Учеб. пособие для вузов. М.: Наука. – 1989. – 432 с.
107. Хемминг Р.В. Численные методы. М.: Наука. – 1972. – 400 с.
108. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. – М.: Мир. – 1985. – 509 с.
109. ГеоГраф, GeoDraw – профессиональные отечественные ГИС для широкого круга пользователей // ГИС – обозрение. Зима. – 1994. – С.24 – 25.
110. Червяков В.А., Черванев И.Г., Кренке А.И. и др. Модели полей в географии. – Новосибирск: Наука. – 1989. – 143 с.
111. Витязь В.И., Витязь О.В., Дьякова Ю.Д., Дыхно Л.И., Дыхно Ю.А., Пузанов А.А., Хлебопрос Р.Г. Экологические структуры Красноярска. - Препринт 119Б, Красноярск. – 1990. – 68с.
112. Иванова Ю.Д., Питенко А.А. Электронный экологический атлас города Красноярска. – Препринт – "Информационный материал и руководство пользователя", Красноярск: ИВМ СО РАН. – 1999. – 16с.
113. Иванова Ю.Д., Питенко А.А. Электронный экологический атлас города Красноярска // Материалы конференции молодых ученых Института вычислительного моделирования СО РАН, март 1999г. – Красноярск: ИВМ СО РАН. – 1999. – С. 38–42.
114. Воробьев В.В., Батуев А.Р., Белов А.В., Богоявленский Б.А. Комплексный экологический атлас как фундаментальная база знаний и данных геоинформационной системы по управлению устойчивым региональным развитием // Труды международной конференции «ГИС для оптимизации природопользования в целях устойчивого

- развития территорий» (ИНТЕРКАРТО – 5). – Якутск: Изд-во Якутского ун-та, 1999. 1ч. – С. 54-63.
115. Питенко А.А., Иванова Ю.Д. Нейросетевая аналитическая обработка экологической информации в ГИС (пример)// ГИС для оптимизации природопользования в целях устойчивого развития территорий: Материалы международной конференции, 1-4 июля 1998г. – Барнаул: Издательство Алтайского государственного университета. – 1998. – С.167–168.
116. Келлер А.А. Медико – географический подход к изучению здоровья населения регионов // Медико-географические аспекты оценки уровня здоровья населения и состояния окружающей среды. Санкт-Петербург: Изд-во НИИ ГПМ РФ. – 1992. – С.37-45.
117. Крутько В.Н. // Методологические проблемы экологии человека. Новосибирск: Наука, сиб. отд. – 1988. – С. 85-94.
118. Рященко С.В. Медико–географический прогноз в территориальных комплексных системах охраны природы // Модели и методы оценки антропогенных изменений геосистем. – Новосибирск: Наука. – 1986. – С. 22 - 32.
119. Хлебович И.А., Ротанова И.Н. Комплексное картографирование проблемных медико – экологических ситуаций. – География и природные ресурсы. – 1997. – № 4. – С. 43-49.
120. Худолей В.В., Дятченко О.Т., Мерабишвили В.М., Шабашова Н.Я. Экологическая обстановка, демография и злокачественные новообразования в Санкт-Петербурге // Вопросы онкологии. – 1998. – №3. – С. 270-282.
121. Воздействие на организм человека опасных и вредных экологических факторов. Метрологические аспекты. Под ред. Исаева Л. К. Том 1. – М.: ПАИМС. – 1997. – 512 с.
122. Киреев Г.В., Татарский В.П., Задолинная С.Д., Рязанова Е.В. Зависимость онкологической заболеваемости от загрязнения атмосферного воздуха // Гигиена и санитария. М.: Медицина. – 1997. – № 2. – С 38-45.
123. Худолей В.В., Мизгирёв И.В., Экологически опасные факторы, Санкт – Петербург, Изд-во “Банк Петровский”. – 1996. – 186с.
124. Шубик В.М. Проблемы экологической иммунологии. Л.: Лениздат. – 1976. – 215 с.
125. Бочков Н.П., Чеботарев А.Н. Наследственность человека и мутагены внешней среды. – М.: Медицина. – 1989. – 272 с.
126. Быкорез А. Н., Рубенчик Б.Л., Слепнян Э.И. Экология и рак. Киев: Наукова думка. – 1985. – 180 с.
127. Янышева Н.Я. Экспериментально-гигиенические основы установления предельно допустимых концентраций бенз(а)пирена в атмосферном воздухе // Эпидемиология рака легкого. Ростов-на-Дону: Изд-во Ростовского университета. – 1990. – С. 198-208.

128. Шабад Л.М., О циркуляции канцерогенов в окружающей среде, Москва, Медицина. – 1973. – 367с.
129. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика. – 1974. – 240 с.
130. Дорофеюк А.А. Алгоритмы автоматической классификации: Обзор // Автоматика и телемеханика. – 1971. – № 12. – С. 78-113.
131. Горбань А.Н., Россиев А.А. Итерационный метод главных кривых для данных с пробелами // Проблемы нейрокибернетики: Труды 12 Международной конференции по нейрокибернетике. Ростов-на-Дону: Издательство СКНЦ ВШ. – 1999. – С. 198-201.
132. Россиев А.А. Моделирование данных при помощи кривых для восстановления пробелов в таблицах // Методы нейроинформатики / Под ред. А.Н.Горбаня. Красноярск: Изд-во КГТУ. – 1998. – С. 6-22.
133. Gorban A.N., Rossiev A.A. Wunch II D.C. Neural Network Modelling of Data with Gaps: Method of Principal Curves, Carleman's Formula and Other// Радіоелектроніка. Інформатика. Управління, Запоріжжє. – 2000. – № 1. – С. 47-55
134. Зиновьев А.Ю., Питенко А.А. Система визуализации произвольных данных. // 2-я Всероссийская научно-техническая конференция “Нейроинформатика-2000”. Ч.1. М.: МИФИ.– 2000. С.75-80.
135. Зиновьев А.Ю., Питенко А.А. Визуализация данных методом упругих карт //Радіоелектроніка. Інформатика. Управління, Запоріжжє. – 2000. – № 1. – С.76-85.
136. "Expert" magazine. – №36. – 1999.
137. Shumsky S.A., Kochkin A.N. Self-Organizing maps of financial indicators of the 200 biggest Russian companies. Proc. All-Russia science conference "NeuroInformatics-99". Moscow. – 1999. – Part 3. – P. 122-127.
138. Айзенберг Л.А. Формулы Карлемана в комплексном анализе. Первые приложения. – Новосибирск: Наука. – 1990.