

Нейронные сети и ГИС

СОДЕРЖАНИЕ

| | |
|--|-----------|
| 1. Нейронные сети: новое измерение в ГИС | 1 |
| Технологии искусственного интеллекта | 1 |
| Перспективы развития аналитических и прогностических свойств ГИС | 2 |
| Интеграция нейросетевых и геоинформационных технологий | 3 |
| 2. Нейронные сети в технике и биологии..... | 4 |
| Типы технических нейросетей | 6 |
| 3. Нейросетевые алгоритмы – математические аспекты | 7 |
| Алгоритм обратного распространения ошибки | 8 |
| Алгоритмы квантования данных и карты Кохонена..... | 14 |
| Прочие нейросетевые алгоритмы | 14 |
| 4. Связь ГИС с нейронными сетями..... | 15 |
| 5. Области применения нейросетевых ГИС | 17 |
| 6. Программное обеспечение | 19 |
| Программа ScanEx-NeRIS..... | 20 |
| Модуль Arc-SDM для ArcView | 21 |
| Литература..... | 21 |

1. Нейронные сети: новое измерение в ГИС

Технологии искусственного интеллекта

Одной из основных задач ГИС, которую ставят перед собой исследователи, является получение новых знаний, представлений о природе пространственных данных. В то же время пользователи часто недооценивают возможностей ГИС в области поддержки принятия решений, которые эти системы могут обеспечивать, уделяя основное внимание главным образом представлению данных. Ценность географической информации в системах поддержки принятия решений становится особенно значимой когда ГИС соединяется с программными средствами, базирующимися на технологиях и методах искусственного интеллекта (ИИ), получившими в последние годы значительное распространение в мире. Важность таких средств ИИ как экспертные системы и **нейронные сети** состоит в том, что они существенно расширяют круг практически значимых задач, которые можно решать на компьютере, и их решение приносит значительный экономический эффект.

Результатом успешного развития методов и технологий ИИ стало создание многочисленных приложений, ориентированных на конечных пользователей, включая специалистов в области ГИС. Интеграция систем ИИ с ГИС особенно эффективна в задачах оценки, контроля и принятия решения. В этом контексте развитие нейронных сетей, эволюционных вычислений (автономное и адаптивное поведение компьютерных приложений и робототехнических устройств), нечеткой логики, самоорганизующихся СУБД, обработки изображений, экспертных систем и ряда других технологий ИИ связано сегодня с расширением функциональных возможностей в части поддержки принятия решений. Есть все основания полагать, что в следующее поколение программного обеспечения ГИС будут встроены элементы ИИ.

Перспективы развития аналитических и прогностических свойств ГИС

ГИС технологии объединяют традиционные операции при работе с базами данных, такими как запрос и статистический анализ, с преимуществами визуализации и географического анализа, которые предоставляют цифровые карты. Используемые в ГИС алгоритмы, в принципе пригодны для анализа многомерных данных в абстрактных пространствах произвольной размерности, что делает перспективным их использование в задачах анализа и прогноза территориально-распределенных явлений и процессов. Это обеспечивает уникальные возможности ее применения в задачах, связанных с анализом и прогнозом территориально-распределенных процессов окружающего мира. Потребности комплексной оценки текущей ситуации, определения главных факторов и причин, прогноза возможных последствий, планирования стратегических решений и сценариев их реализации, выводят на круг задач лежащих за пределами возможностей традиционных ГИС технологий.

Это уже поняли ученые и недавно признали политики. Первый международный симпозиум по динамическим и многомерным ГИС состоялся 25 августа 1997 года в Гонконгском Политехническом Университете [5]. Необходимость развития этого направления отмечается и политиками самого высокого уровня. Так Ал Гор вице президентом США в своей речи посвященной развитию Цифрового будущего Планеты [6] отметил три ключевых проблемы:

- создание детального мульти-уровневого и мульти-масштабного представления Планеты;
- проблемы создания адекватного трехмерного описания Земли;
- проблемы обработки огромного количества разнородных геоданных об окружающем мире.

Для адекватного описания окружающего Мира используются разнородные по содержанию, структуре и представлению геоданные – о ландшафте, природных ресурсах, окружающей среде, погоде и климате, об импорте и экспорте, об инфраструктуре, демографии, преступности, и т.д. и т.п. Это многоуровневые данные с различной степенью детализации от сведений о земельных участках, городах и регионах, до описания стран, континентов, климата или геодинамики континентальных плит. В качестве одной из наиболее активно развивающихся информационных технологий анализа, структурирования и систематизации этих данных рассматриваются интеллектуальные технологии основанные на нейросетевых методах. Для решения прогностических задач важны сведения о событиях, изменении состояния объектов исследования, временная динамика процессов. Процессы, географические комплексы их описывающие, могут интерпретироваться как объекты в многомерном пространстве. В этом пространстве положение объекта характеризуется географическими координатами, временем, и количественными показателями характеристик вектора состояний. При исследовании процессов, развивающихся во времени, к трем пространственным осям добавляется четвертая ось – время, появляются возможности динамического моделирования. При этом множество других координатных осей многомерного пространства может быть использовано для определения меры близости характеристик различных объектов и процессов. Это важно при анализе причинно-следственных отношений событий, функциональных связей объектов, прогнозе развития процессов. Нейросетевые модели, построенные на статистическом анализе пространственных и временных рядов геоданных, пригодны для оценки будущего методами экстраполяции.

Геоинформационные системы, по естественным для человека причинам удобны для ввода, вывода территориально-ориентированной информации. Их математическое обеспечение может быть дополнено средствами обработки многомерных данных в абстрактных пространствах произвольной размерности. Геоданные могут храниться, как в самой ГИС, так и за ее пределами. То, что геоинформационные системы, по своей программно-

технологической природе, генетически связаны с базами данных, обеспечивает пользователю с их помощью естественный доступ к первичной исходной информации. Использование средств Интернет позволяет организовать доступ к информационным ресурсам находящимся практически в любой точке Планеты. Поскольку хранение, пересылка и поиск больших объемов информации на электронных носителях – задача технологически сложная, в ГИС обычно используют возможности работы с внешними СУБД. Как правило, информация, хранящаяся в БД требует специальной предобработки без которой пространственный анализ этих данных либо затруднен либо не возможен. Одним из наиболее эффективных при обработке больших массивов сырой разнородной информации являются нейросетевые алгоритмы. К тому же, параллельные по самой своей природе, они легко ложатся на вычислительные сети параллельной архитектуры.

ГИС имеет много общего с системами автоматизированного управления производственными объектами и технологическими процессами (САПР и АСУ). Если сложный объект управления представлен в виде структурно-функциональной схемы, то ГИС может стать удобным интерфейсом для доступа к первичной информации базы данных, электронным таблицам, датчикам. Геометрическая природа ГИС в принципе может обеспечить как построение, так и содержательный анализ структуры и топологии связей структурно-функциональных схем. Привлечение методов современной геометрии и статистики позволяет автоматизировать построение графов и сетей, выполнять количественный анализ функциональных связей, строить экстраполяционные, интерполяционные и адаптивные модели технологических процессов. Нейросетевые модели дают пример построения из нелинейных элементов многоуровневых схем практически любой разумной сложности. Они могут использоваться для автоматизации построения графов и технологических сетей, рассматриваемых как основа построения более сложных моделей включающих прямое математическое моделирование технологических процессов. При наличии значительного объема контрольно-измерительных данных на основе нейросетевых методов может быть построена и статистическая модель технологических процессов.

Интеграция нейросетевых и геоинформационных технологий

Хотя ГИС содержат геоданные, описывающие геометрические (топологические) свойства объектов (обычно точки, линии и полигоны), их функциональные возможности пространственного анализа относительно слабы. В математике средства многомерного пространственного анализа постоянно развиваются и хорошо обеспечены методами исследования геометрии, топологии и других свойств абстрактных объектов, их множеств и структур. Поэтому интеграция геоинформационных технологий, методов пространственного анализа данных, является естественной и многообещающей идеей.

Логика этой интеграции такова:

1. расширение функциональной полноты традиционных методов, технологии и программных средств пространственного анализа в ГИС, за счет использования возможностей развитых математических методов анализа многомерных данных;
2. развитие новых методов, основанных на интеллектуальных вычислительных технологиях, как базы для создания следующего поколения удобных и более сильных инструментальных программных средств анализа геоданных в условиях все возрастающих объемов первичной информации;
3. создание новых моделей данных, информационных технологий и программных средств, специально предназначенных для многомерного анализа данных, моделирования и прогноза территориально распределенных процессов и обеспечивающих интеграцию с традиционными ГИС.

По первому пути ГИС развиваются практически с момента их появления. Второе и третье направление связано с фундаментальными исследованиями на стыке математики, информатики, и нейрофизиологии. За последние 10 лет исследователи хорошо разработа-

ли и выделяют целый класс статистических и адаптивных методов анализа многомерных данных, получивших название нейросетевых методов. Нейросетевые методы применяются не только для анализа данных, но и, что существенно, для построения моделей процессов, разворачивающихся в многомерных пространствах. Уже сегодня предлагаются интересные классы нелинейных моделей построенных на основе статистического анализа первичных данных. При этом средства информационных технологий используются для организации доступа и предобработки первичных данных хранящихся в ГИС и БД. Статистические и адаптивные методы анализа геоданных позволяют улучшить качество исходной информации и построить нейросетевую модель адекватную как назначению и качеству исходных данных, так и суждениям экспертов, и задачам исследователей.

Объем таких информационных моделей существенно определяется задачами пользователя, мало зависит от объемов первичных данных, а их аналитические и прогностические свойства существенно улучшаются по мере роста объема и полноты исходных геоданных. В мире стремительно растут объемы информационных ресурсов отражающих социально-экономические, демографические, экологические, климатические, геологические и другие процессы. Все острее стоит проблема «Что делать?» с этими «полезными ископаемыми». Интернет и клиент-серверные технологии обеспечивают распределенный доступ к исходным данным, нейросетевые методы анализа геоданных хорошо распараллеливаются, позволяют создавать прогностические модели над многомерными массивами разнородной информации, геоинформационные технологии дают прекрасный картографический интерфейс. Это в принципе позволяет решать колоссальные по вычислительной трудоемкости задачи создания моделей глобальных социально-экономических процессов и природных явлений на вычислительных кластерах. Значимость проблем и масштабность задач определяют устойчивый интерес развитию нейросетевых методов и перспективы интеллектуальных вычислительных технологий в создании виртуальных геоинформационных моделей окружающего мира на базе информационных и вычислительных ресурсов Интернет.

2. Нейронные сети в технике и биологии

Современные компьютеры устроены по так называемой схеме Фон-Неймана, реализующей быстрое последовательности большого числа бинарных операций. Видимо, такой подход отчасти был обусловлен структурой математики первой половины двадцатого века, когда высшие разделы математики опирались на арифметику, а та на бинарную логику. Вроде бы понятно, что если бы за основу было взято что-то иное, а не булева алгебра, могли бы получиться существенно иные компьютеры.

В качестве основной альтернативы подходу Фон-Неймана обсуждалась ориентация на воспроизведение принципов работы биологических нейронных сетей. Примерно в те же годы что и первый компьютер, была создана первая нейроподобная система – перцептрон Розенблатта. Некоторое время оба направления – Фон-Неймана и Розенблатта – развивались независимо, затем перцептронное направление пережило кризис, и возродилось уже в восьмидесятые годы под именем нейронных сетей, при этом на новом этапе бинарно-логический и бионический принципы стали сочетаться. Интересно, что кризис бионического направления во многом был обусловлен не техническими сложностями, и не отсутствием приложений, а содержательным математическим рассмотрением, проведенным Минским и Пейпертом – они показали, что не существует перцептрона, способного надежно определять топологические характеристики образа, такие как связность, и этого оказалось достаточно для резкого падения первоначального энтузиазма. Любопытно также, что и возрождение энтузиазма было связано не с опровержением выводов Минского и Пейперта, а просто с формированием обширной ориентированной на приложения сферы деятельности. Впрочем, в последние годы появились некоторые приложения нейроподобных алгоритмов и к задачам топологии.

Сопоставление машины Фон-Неймана и биологической нейронной сети приводит к следующей таблице:

| | Машина фон Неймана | Биологическая нейронная система |
|------------------------|---------------------------------|---------------------------------|
| Процессор | Сложный | Простой |
| | Высокоскоростной | Низкоскоростной |
| | Один или несколько | Большое количество |
| Память | Отделена от процессора | Интегрирована в процессор |
| | Локализована | Распределенная |
| | Адресация не по содержанию | Адресация по содержанию |
| Вычисления | Централизованные | Распределенные |
| | Последовательные | Параллельные |
| | Хранимые программы | Самообучение |
| Надежность | Высокая уязвимость | Живучесть |
| Специализация | Численные и символьные операции | Проблемы восприятия |
| Среда функционирования | Строго определенная | Плохо определенная |
| | Строго ограниченная | Без ограничений |

Более конкретные характеристики мозга человека: кора головного мозга образована нейронами поверхностью толщиной от 2 до 3 мм с площадью около 2.2 дм², содержит около 10¹¹ нейронов, каждый нейрон связан с 10³ - 10⁴ другими нейронами.

Нейроны взаимодействуют посредством короткой серии импульсов, как правило, продолжительностью несколько миллисекунд (мс). Сообщение передается посредством частотно-импульсной модуляции. Частота может изменяться от нескольких единиц до сотен герц, что в миллион раз медленнее, чем самые быстродействующие переключаемые электронные схемы, тем не менее, достаточно сложные решения по восприятию информации человек принимает за несколько сотен мс.

Сопоставим биологический нейрон с наиболее часто рассматриваемой схемой технического нейрона:

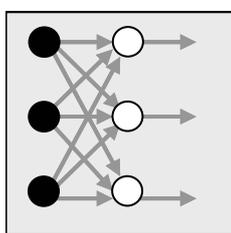
| | Технический нейрон | Биологический нейрон |
|-----------------|--|---|
| Схема | <p>Входной сумматор</p> <p>Нелинейный преобразователь</p> <p>Точка ветвления</p> | <p>Дендриты</p> <p>Синапс</p> <p>Тело клетки</p> <p>Ядро</p> <p>Аксон</p> |
| Функция отклика | <p>$\varphi(\alpha)$</p> <p>α</p> | <p>$\varphi(\alpha)$</p> <p>α</p> |

Оба типа нейронов реагируют на воздействие со стороны многих нейронов, в зависимости от величины связей с этими нейронами. В отличие от технических нейронов, реакция биологического нейрона всегда неотрицательная, причем, если воздействие на него не достигло критического уровня, реакции нет. Возможно, с этим связано одно из наиболее очевидных отличий биологических нейронных сетей от существующих сегодня нейропрограмм – один и тот же мозг, в зависимости от того, какие нейроны «молчат», может работать весьма различным образом, это выглядит так как будто мозг – «склад» процессоров, которые по разному соединяются при разных задачах. Hardware, реализующее технические нейроны, бывало самым разным – сначала релейные схемы, сейчас, например, операционные усилители, но чаще всего – эмуляция в обычном компьютере. Что до вычислительной мощности, то по экспертным оценкам современные ПК могут моделировать уровень нервной системы сложных червей, лучшие нейросетевые спецпроцессоры – уровень мухи.

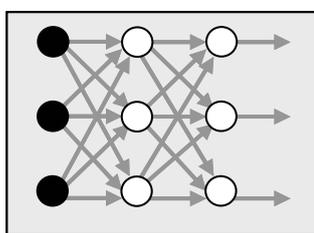
Если рассматривать биологические нейронные сети, то для них существует принципиальное внутреннее деление: нейронные сети могут реализовывать либо рефлекторное поведение, либо мышление. Нейрофизиологически рефлекторному поведению соответствует относительно короткий всплеск процессов в ответ на внешнее воздействие, с последующим возвращением в спокойное состояние, а мышлению – длительная работа сети, нередко с весьма умеренным, но постоянным, уровнем возбуждения мозга, при этом внешние воздействия скорее мешают. В технических системах воспроизводят в основном рефлекторное поведение, хотя возможно, что некоторые нейроалгоритмы, решающие «внутренне сложные» задачи, могут сопоставляться и с процессами мышления.

Типы технических нейросетей

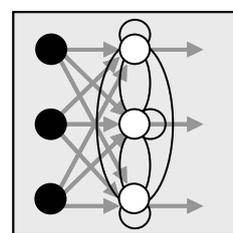
В литературе заметное внимание уделяется вопросам архитектуры технических нейронных сетей, приведем вариант соответствующей классификации схем:



однослойные



многослойные



полносвязные

Более принципиальным является разбиение нейроалгоритмов на два класса – Supervised (обучающиеся по образцу, с Учителем) и Unsupervised (обучающиеся без образца, без Учителя). В первом случае обучение организовано как воспроизведение набора правильных образцов (обучающей выборки), после чего сеть может адекватно реагировать и на примеры, которых не было в обучающей выборке, во втором случае образцы правильной реакции исходно отсутствуют. В части русской литературы утвердились термины обучение с учителем и обучение без учителя, что не является точным переводом с английского, и не вполне точно по нормам русского языка. Видимо, нейросети, обучающиеся по образцу, неплохо воспроизводят рефлекторное поведение. Нейросети, обучающиеся без образца, быть может, иногда моделируют более интересную вещь – мышление, однако, делают это несравненно менее успешно.

Нейросети, обучающиеся по образцу, произошли от перцептронов, и в современной трактовке могут рассматриваться как варианты и модификации сетей с обратным распространением ошибки (иногда как результат примитивизации такого рода сетей, с целью упрощения реализации). К этому классу можно отнести, например, однослойный и многослойный перцептрон, машину Больцмана, сети, обучающиеся по правилу Хебба, рекур-

рентные слоистые и полносвязные сети обратного распространения ошибки, сети, использующие радиальные базисные функции. Различия между указанными системами порой достаточно велики, но всегда есть немало общего, а детали классификации различаются у разных авторов.

Нейросети, обучающиеся без образца, пожалуй, более разнообразны, хотя стоящая за ними теория математически порой более примитивна – это карты Кохонена, системы с множественными локально устойчивыми состояниями, такие как сеть Хопфилда, сети, настраиваемые на основе адаптивного резонанса. Прямые аналогии между данными классами не просматриваются, хотя часто исходно имеются необработанные данные, а в итоге возникают их образы, построенные в ходе работы нейросети, либо сама сеть, меняя свою структуру, моделирует образы данных. В живой природе есть аналогии и этому – строились карты возбуждения участков коры мозга в зависимости от возбуждения участков тела, получившие названия «гомункулусов», оттого, что на этих картах формируется узнаваемый образ человека, только ладони, например, получаются увеличенными, а спина – уменьшенной.

Наконец, разрабатываются, хотя далеко не стали широкоупотребительными, комбинированные подходы. Идеология такого комбинирования порой заставляет вспомнить лозунг «человека создал труд». За основу берется алгоритм обучения по образцу, произвольно устанавливающий некое первоначальное, можно сказать «абстракционистское», соответствие между «сырыми данными» и «обработанными данными». Затем «обработанные данные» меняются в ходе внешней «трудовой деятельности» с учетом «свойств материала» так что соответствие улучшается. Устанавливается новое соответствие между исходными и обработанными данными, вновь меняются обработанные данные, и так далее. В итоге нейросеть порождает с одной стороны «художественный образ» ситуации, а с другой стороны – собственный навык быстрого, как бы рефлекторного, соотнесения реальных данных и их образов. Например, так можно проверять связность образа данных – если «природа материала» не позволяет ему изменять связность, а выше обозначенный подход работает, то значит и у образа данных связность та же. Данный пример мы привели, чтобы напомнить о кризисе нейросетевого подхода, имевшим место до восьмидесятых годов, как раз в связи с пессимизмом по части возможностей использования нейроалгоритмов в задачах топологии.

В заключение укажем списки задач, которые принято относить к чаще всего рассматриваемым в рамках нейросетевого подхода:

- обучение по образцу – классификация образов, аппроксимация функций, предсказание, управление, анализ данных, категоризация внутри класса, сжатие данных;
- обучение без образца – категоризация, категоризация внутри класса, анализ данных, сжатие данных, ассоциативная память.

3. Нейросетевые алгоритмы – математические аспекты

Под именем нейросетевых алгоритмов сегодня объединяется несколько подходов к обработке данных, которые их авторы, не согласовывая друг с другом, сочли напоминающими принципы организации биологических нейронных сетей. Видимо, сыграла роль привлекательность названия, вместе с тем обстоятельством, что по настоящему принципы работы таких сложных биологических систем, как мозг человека, никому не известны, и в этом смысле все равны и свободны. Это несколько нарушает существующие в математике традиции логически обоснованной классификации алгоритмов, но поскольку некоторые нейроалгоритмы достаточно эффективны, приходится считаться с установившейся практикой. Здесь мы обсудим два типа нейроалгоритмов, наиболее часто используемых в приложениях – алгоритмы обратного распространения ошибки (back error propagation algorithms; BackProp; в российских публикациях восьмидесятых годов использовалось мате-

матически более культурное название: алгоритмы двойственного функционирования; АДФ) и карты Кохонена (самоорганизующиеся карты, self-organization maps, SOM).

Алгоритм обратного распространения ошибки

Нейроалгоритмы обратного распространения ошибки исходно создавались во многом для решения классической задачи математической статистики – задачи регрессии табличных данных. Хорошо известны простейшие задачи регрессии – такие как задача проведения прямой, наилучшим образом приближающей облако точек (например, прямой, сумма квадратов расстояний точек до которой минимальна). В этом примере речь идет о линейной регрессии, алгоритм которой должен определить параметры положения прямой – ее углы наклона к осям координат и координаты одной из точек прямой.

Если речь идет о нелинейной регрессии, в качестве аппроксимирующего облако точек многообразия выступает не прямая, не плоскость или гиперплоскость, а некоторая гладкая кривая, поверхность или гиперповерхность. Чтобы описать такие нелинейные многообразия, требуется увеличить число параметров по сравнению со случаем линейной регрессии: возникает задача многопараметрической нелинейной регрессии.

Задолго до возникновения алгоритма обратного распространения ошибки были известны сложности проведения нелинейной многопараметрической регрессии данных:

- при большом числе параметров регрессии увеличивается время счета;
- параметры регрессии определяются неоднозначно (т.н. плохая обусловленность задач многопараметрической регрессии);
- неясно, как выбрать лучший из множества вариантов нелинейной регрессии.

Алгоритм обратного распространения ошибки успешно преодолел первую из указанных сложностей, после чего был накоплен практический опыт в отношении двух оставшихся сложностей, так что сегодня эти сложности уже не смущают, хотя полной математической теории по их поводу так и не было создано. Можно говорить о том, что нейроалгоритм обратного распространения ошибки эффективно решает задачи многопараметрической нелинейной регрессии: имеются программы, осуществляющие «хорошую» нелинейную регрессию для десятков и сотен тысяч точек, с определением десятков тысяч регрессионных параметров, за приемлемое время (до нескольких суток счета).

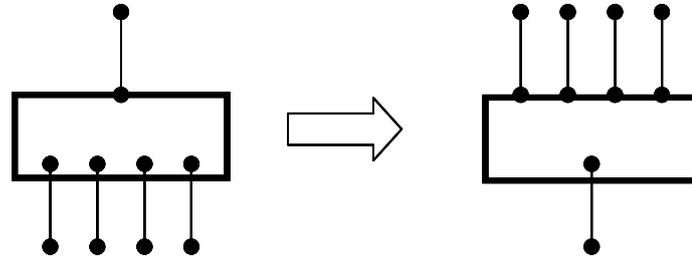
Общая постановка задачи многопараметрической нелинейной регрессии состоит в следующем:

Имеется метод вычисления величины отклонения регрессионного многообразия от точек в пространстве данных, в зависимости от формы этого многообразия. Тем самым, можно говорить об оценке величины отклонения H , как о сложной, часто неявно заданной, функции от множества регрессионных параметров β , задающих многообразие. Требуется найти значения параметров β , при которых $H(\beta)$ минимальна.

Несложно привести пример алгоритма, как-то решающего данную задачу: пусть множество параметров регрессии имеет N элементов, возьмем первый элемент и немного изменим его. Если H увеличится, изменим первый элемент в другую сторону, и тогда она, как правило, уменьшится. Затем также изменим второй регрессионный параметр, и через N шагов очередь снова дойдет до первого параметра. Недостаток данного подхода в том, что время счета будет пропорционально $N * T$, где T - время расчета H . Достижение метода обратного распространения ошибки состоит в том, что это время сокращается в N раз. Поскольку в интересных случаях N порядка от сотен до десятков тысяч, это очень существенное сокращение.

Приведем простую аналогию, иллюстрирующую основную идею метода. Пусть у нас имеется система из резервуара и N входных трубок разной пропускной способности, сливающихся в одну выходную трубку. Требуется измерить пропускные способности всех

входных трубок. Медленный способ состоит в том, что по очереди наливаем жидкость в каждую трубку, и смотрим, как она вытекает. В N раз более быстрый способ состоит в том, что конструкция переворачивается, жидкость наливается в выходную трубку, и смотрится, как жидкость потечет из каждой из входных трубок. Аналогичное «переворачивание» имеет место и в методе обратного распространения ошибки, чем и обусловлено его название.



Математически все выглядит следующим образом: имеется неявная функция $H(\beta)$, например задаваемая соотношениями

$$H = H(x_i), \psi_j(x_i, \beta_k) = 0, k = 1, \dots, N; i, j = 1, \dots, M; \quad (1)$$

то есть имеется M величин x и столько же уравнений на них, причем в эти уравнения входят параметры β , а функция H явно зависит от x . Если мы научимся решать общую задачу о нахождении минимума функции $H(\beta)$, то и задачу многопараметрической нелинейной регрессии решим, поскольку, как мы увидим впоследствии на примерах, эта задача как правило может быть представлена в форме (1).

Введем множители Лагранжа μ_j и производящую функцию

$$W(\alpha, \beta, \mu) = H(\alpha) + \sum_{j=1}^M \mu_j \cdot \psi_j(\alpha, \beta) k \quad (2)$$

С помощью (2) уравнения на x , входящие в (1), можно записать в форме

$$\partial W / \partial \mu_j = 0, j = 1, \dots, M. \quad (3)$$

Если (как будет предполагаться ниже), условия (1) и (3) выполнены, то $W = H$, следовательно

$$\begin{aligned} dH / d\beta_i &= dW / d\beta_i = \partial W / \partial \beta_i + \\ &+ \sum_{j=1}^M (\partial W / \partial \alpha_j \cdot \partial \alpha_j / \partial \beta_i + \partial W / \partial \mu_j \cdot \partial \mu_j / \partial \beta). \end{aligned} \quad (4)$$

Чтобы найти экстремумы функции $H(\beta)$, во многих алгоритмах требуется иметь эффективный способ вычисления полных производных $H(\beta)$ по параметрам β . Выражение (4) явилось результатом тождественных преобразований, с учетом (1) и (3). Если максимально упростить это выражение, его можно будет использовать для нахождения $dH / d\beta_i$. Заметим, что в силу (3) последние члены в (4) можно не учитывать – они равны нулю. До сих пор множители Лагранжа μ были произвольными введенными чисто формально параметрами. За счет этого произвола, выбрав в качестве дополнительных условий на множители Лагранжа уравнения

$$\partial W / \partial \alpha_j = 0, j = 1, \dots, M, \quad (5)$$

можно обратить в ноль и предпоследние члены в (4), сведя (4) к простому виду

$$dH / d\beta_i = \partial W / \partial \beta_i. \quad (6)$$

(3), (5) и (6) и дают искомый простой способ вычисления $dH/d\beta_i$. Если сопоставлять приведенные соотношения с простой аналогией, с которой мы начинали, то можно сказать что появившиеся дополнительные уравнения (5) описывают «поток в искусственно перевернутой системе», если считать что (3) описывают «поток в исходной системе».

Специалист по прикладной математике узнает в приведенных рассуждениях стандартную схему, восходящую еще к Лагранжу, и широко используемую во многих приложениях – например, в теории оптимального управления. Естественно, что использование такого подхода в задачах неявно заданной многопараметрической регрессии оказалось эффективным.

Что касается алгоритма обратного распространения ошибки, то достаточно прийти в каком-либо конкретном случае к формулировке задачи регрессии в виде (1), и формулы (3), (5) и (6) как раз дадут нам этот алгоритм.

Приведем конкретный пример. Пусть имеются T записей $(y^t, x_1^t, \dots, x_Q^t), t = 1, \dots, T$, и мы хотим найти такую функцию, $y(x_1, \dots, x_Q)$, для которой минимальна сумма квадратов расстояний

$$H = \sum_{t=1}^T (y(x_1^t, \dots, x_Q^t) - y^t)^2. \quad (7)$$

Если бы функция y была очень простой, можно было бы написать алгоритм ее поиска и без формул (3), (5) и (6). Однако, рассмотрим вариант более сложного задания функции. Пусть y вычисляется в итоге работы следующего алгоритма:

$$\alpha_i^g = f\left(\sum_{j=1}^S \beta_{ij} \cdot \alpha_i^{g-1}\right); g = 1, \dots, G; y = \alpha_1^G, \quad (8)$$

причем зависимость $y(x_1, \dots, x_Q)$ обусловлена тем, что на начальной итерации

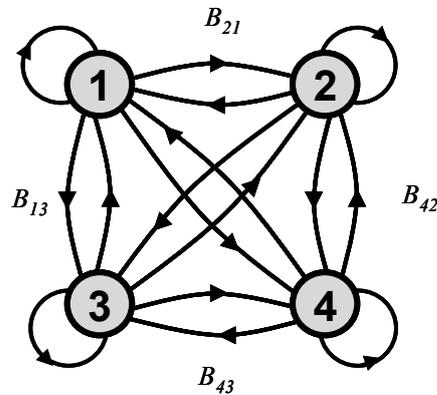
$$\alpha_i^0 = x_i; i = 1, \dots, Q;$$

что естественно предполагает $S \geq Q$. В (8) $f(z)$ - некоторая нелинейная функция, например $f = z/(1+|z|)$ или $f = \arctg(z)$, или $f = \sin(z)$. Алгоритм (8) определяет функцию $y(x_1, \dots, x_Q)$ за G итераций, как результат сложного сочетания нелинейных функций f и линейных операций суммирования. Даже при фиксированном выборе вида f , изменяя входящие в (8) параметры β_{ij} , можно получать существенно различающиеся виды функций $y(x_1, \dots, x_Q)$. Количество регрессионных параметров β_{ij} равно S^2 , и быстро растет с S - при $S = 40$, например, будет 1600 регрессионных параметров.

Прежде чем вернуться к вопросу о быстром алгоритме поиска параметров β_{ij} , при которых минимизируется величина (7), скажем несколько слов о том, почему мы задаем функцию $y(x_1, \dots, x_Q)$ посредством алгоритма (8), а не, как сказал бы Козьма Прутков, «каким либо другим, более похожим на функцию, выражением». Секрет здесь в том, что все время подразумевается, что функция – сложная, требующая при реализации в компьютере большого времени. Поэтому интересны такие способы представления функций, которые можно очень хорошо реализовать на параллельно работающих спецпроцессорах. Допустим, у нас есть спецпроцессор, позволяющий параллельно вычислять 40 нелинейных функций \arctg , за время τ_{\arctg} , и за одну операцию осуществлять умножение квадратной матрицы на вектор из 40 элементов, за время τ_{matr} . На таком спецпроцессоре итерация алгоритма (8) будет осуществляться за время $\tau_{\arctg} + \tau_{matr}$, и это может быть очень маленькое время. Если у нас есть локальная сеть, либо аналоговое устройство преобразования ин-

формации, алгоритм (8) также нетрудно реализовать. Наконец, (8) у нас иллюстрирует общие принципы работы со «сложно заданными» функциями, поэтому если окажется что для какого-либо варианта hardware оптимально использовать другое задание функции, так и надо сделать.

Можно нарисовать графическую схему информационных потоков в (8), или, если угодно, схему устройства соответствующего спецпроцессора (для случая $S = 4$):



На схеме показано кружками четыре нелинейных преобразования, соответствующих функции f , а линиями со стрелками – 16 регрессионных параметров, при ненулевых значениях которых результат нелинейной операции на предыдущей итерации влияет на следующую итерацию. На нулевой итерации стартовые значения величин $(\alpha_1^0, \alpha_2^0, \alpha_3^0, \alpha_4^0) = (x_1, x_2, x_3, x_4)$, а на последней итерации снимается значение $y = \alpha_1^G$, которое в итоге оказывается функцией от стартового состояния x . Приведенная схема очень характерна для нейроинформатики. Кружочки при этом называют нейронами, а стрелки синапсами, тогда схема информационных потоков начинает напоминать принятые в нейрофизиологии схемы взаимодействия нейронов через свои разветвленные окончания - синапсы. Потому-то (8) и представляет собой «нейроалгоритм». Не исключено, что разветвленность биологических межнейронных связей также отчасти связана с проблемой быстрого действия – биологический нейрон работает медленно, примерно за миллисекунды, поэтому при последовательной схеме соединения большие группы нейронов работали бы крайне неэффективно. Впрочем, попытки найти полное соответствие между алгоритмами, которые мы здесь обсуждаем, и биологическими нейронными сетями, пока успехом не увенчались.

А теперь вернемся к алгоритму обратного распространения ошибки. Сначала демонстрационно-учебная

Задача 0: построить алгоритм минимизации величины (7), при условии (8), для нелинейного преобразования $f(z) = 1/(1 + |z|)$.

Решение:

1). Строим производящую функцию - аналог (2). Поскольку в (7) функция y фигурирует T раз, и, соответственно, алгоритм (8) будет использоваться T раз, введем дополнительный индекс $t = 1, \dots, T$, нумерующий обращения к алгоритму (8). В качестве ψ возьмем соотношения (8), в которых правая часть равенства перенесена налево ($\psi = 0$):

$$\psi_i^{g,t} = \alpha_i^{g,t} - f\left(\sum_{j=1}^S \beta_{ij} \cdot \alpha_i^{g-1,t}\right);$$

тогда
$$W(\alpha, \beta, \mu) = H(\alpha) + \sum_{i=1, g=1, t=1}^{S, G, T} \mu_i^{g,t} \cdot \psi_i^{g,t}(\alpha, \beta).$$

Комментарий 1: различие с (2) в том, что вместо индексов j фигурируют тройки индексов i, g, t .

2). Находим уравнения – аналог (5).

$$\partial W / \partial \alpha_k^{r,\tau} = 0; \Rightarrow \partial H(\alpha) / \partial \alpha_k^{r,\tau} + \sum_{i=1, g=1, t=1}^{S, G, T} \mu_i^{g,t} \cdot \partial \psi_i^{g,t}(\alpha, \beta) / \partial \alpha_k^{r,\tau} = 0.$$

В приведенной сумме многие частные производные обратятся в ноль, поскольку лишь в некоторых $\psi_i^{g,t}$ имеются $\alpha_k^{r,\tau}$, а именно, останутся только члены с $g = r$ и $i = k$, и с $g = r + 1$; и только при $t = \tau$. Дифференцируя ψ , получим

$$\partial H(\alpha) / \partial \alpha_k^{r,\tau} + \sum_{i=1}^S \mu_i^{r+1,\tau} \cdot \partial \psi_i^{r+1,\tau}(\alpha, \beta) / \partial \alpha_k^{r,\tau} = 0, \Rightarrow$$

$$\partial H(\alpha) / \partial \alpha_k^{r,\tau} + \mu_k^{r,\tau} \cdot \partial \psi_k^{r,\tau}(\alpha, \beta) / \partial \alpha_k^{r,\tau} + \sum_{i=1}^S \mu_i^{r+1,\tau} \cdot \partial \psi_i^{r+1,\tau}(\alpha, \beta) / \partial \alpha_k^{r,\tau} = 0, \Rightarrow$$

$$\partial H(\alpha) / \partial \alpha_k^{r,\tau} + \mu_k^{r,\tau} - \sum_{i=1}^S \mu_i^{r+1,\tau} \cdot f'(\sum_{j=1}^S \beta_{ij} \cdot \alpha_i^{r,\tau}) \cdot \beta_{ik} = 0.$$

Последнее выражение является уравнением на множители Лагранжа μ .

Комментарий 2: $f'(z) = (z / (1 + |z|))' = 1 / (1 + |z|)^2$, эта производная может быть выражена через саму функцию f .

Комментарий 3: Решить уравнение для μ можно итерациями, если в уравнении для α_i^g верхний индекс, соответствующий номеру итерации, пробегает значения от 0 до G , то в уравнении для μ_k^r верхний индекс пробегает значения от $G + 1$ до 1, причем все μ_j^{G+1} должны полагаться равными нулю (это следует из того, что таких μ в выражении для производящей функции нет).

Комментарий 4: Частная производная оценки $\partial H(\alpha) / \partial \alpha_k^{r,\tau}$ в нашем случае отлична от нуля только на последней итерации, при $r = G$, и только для $k = 1$, и равна $2 \cdot (\alpha_1^{G,t} - y^t)$, а при прочих $k > 1$ и $r < G$ равна нулю.

3). Аналогично (5), находим производные оценки по регрессионным параметрам:

$$dH / d\beta_{ij} = \partial W / \partial \beta_{ij} = - \sum_{g=1, t=1}^{G, T} \mu_i^{g,t} \cdot f'(\sum_{p=1}^S \beta_{ip} \cdot \alpha_p^{g-1,t}) \cdot \alpha_j^{g-1,t}.$$

Общий комментарий: Сложность вычисления множителей Лагранжа, и сама структура уравнений на них, соответствует сложности вычислительной реализации алгоритма (8). Если известны множители Лагранжа μ и параметры функционирования α , все производные по параметрам регрессии находятся по простой формуле, требующей немного вычислительных ресурсов. Тем самым, задача быстрого нахождения «направления наискорейшего спуска» в пространстве регрессионных параметров решена.

Задача 1: Чтобы проследить самостоятельно все выше сделанные выкладки, возьмите количество записей $T = 5$, число итераций $G = 6$, количество нелинейных элементов – нейронов $S = 4$, и реализуйте вышеприведенную схему в более явном виде.

Задача 2: Реализуйте ту же схему для других случаев нелинейных функций: $f = \arctg(z)$ и $f = \sin(z)$.

Задача 3: Возьмите модификацию (8), при которой на каждой итерации работают различные регрессионные параметры:

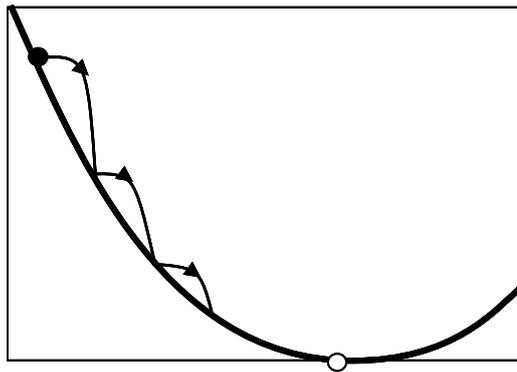
$$\alpha_i^g = f\left(\sum_{j=1}^S \beta_{ij}^S \cdot \alpha_i^{g-1}\right); g = 1, \dots, G; y = \alpha_1^G,$$

реализуйте общую схему построения алгоритма нахождения производных оценки в этом случае, и подумайте, почему этот вариант называется «многослойной нейросетью».

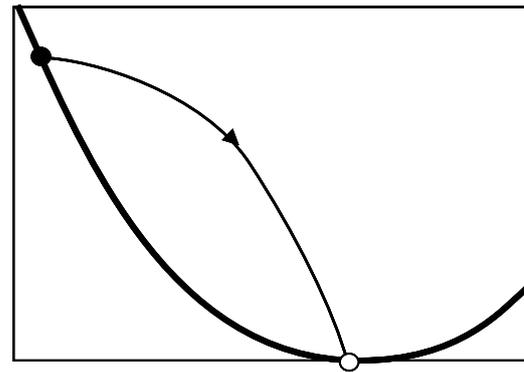
Когда определено направление сдвига регрессионных параметров, алгоритм уменьшения ошибки регрессии строится стандартными методами теории оптимизации. Простейший вариант - алгоритм спуска с постоянным шагом, когда новые вариационные параметры β^{new} вычисляются по формуле

$$\beta_{ij}^{new} = -\lambda \cdot dH / d\beta_{ij},$$

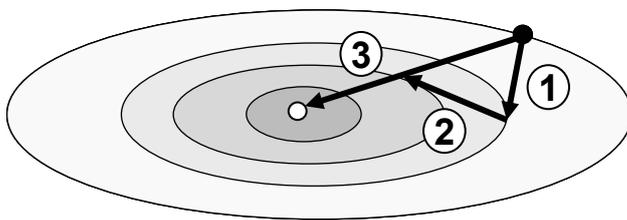
где λ – параметр метода; если этот параметр мал, то изменения идут слишком медленно, а если велик, то есть риск уйти из области, в которой можно полагаться на определенное нами направление спуска. Имеются варианты с определением оптимального значения λ (квазиньютоновский метод наискорейшего спуска), и с определением шага с учетом «опыта» предыдущих шагов (метод сопряженных градиентов). Эти три алгоритма иллюстрируют следующие схемы «в пространстве регрессионных параметров», на которых в качестве «рельефа» указана функция оценки регрессии:



Наискорейший спуск с постоянным шагом (одномерная проекция)



Квазиньютоновский наискорейший спуск (одномерная проекция)



Метод сопряженных градиентов (двумерная проекция)

Задача 4: Взяв за основу метод наискорейшего спуска с постоянным шагом, напишите для алгоритма из [Задачи 0](#) программу осуществляющую регрессию данных (например, в среде MathLab – это язык высокого уровня, программа будет короткой, и отладить ее будет нетрудно; кроме того, в этой среде имеется довольно типичное на сегодняшний день нейросетевое дополнение, в Help можно познакомиться с тем как оно устроено. Ваша программа видимо будет работать хуже, требовать больше итераций – поскольку пакет в MathLab использует более совершенные алгоритмы (включая и такие интересные, как алгоритм Левенберга-Макварта), чем спуск с постоянным шагом – но лиха беда начало, если

постараться и почитать учебники по оптимизации, и продумать, как улучшить скорость, матлабовский пакет можно очень сильно «перегнать».)

На этом мы закончим ознакомление с наиболее сложной частью «кухни» метода обратного распространения ошибки. Акцент на рассмотренных вопросах сделан и потому, что аналогичные приемы используются в теории оптимального управления, в других разделах прикладной математики – то есть, силы, потраченные на понимание «быстрых» алгоритмов регрессии, будут потрачены не напрасно.

Алгоритмы квантования данных и карты Кохонена

Карты Кохонена – это вариант алгоритмов квантования данных, то есть представления N точек данных с помощью меньшего числа точек-образцов. Изложим здесь один из вариантов – Batch SOM.

1. Выбирается регулярная сетка M узлов, расположенная так чтобы примерно соответствовать наиболее важной части пространства данных, обычно M существенно меньше N .
2. Каждая точка данных «приписывается» к ближайшему для нее узлу.
3. Определяется среднее арифметическое положение векторов «приписанных» групп, пусть для узла i это будет \bar{r}_i .
4. Определяется среднее арифметическое положение векторов «приписанных» к первым соседям групп, пусть для узла i это будет \bar{p}_i .
5. Новое положение узлов задается вектором $\bar{p}_i + \lambda \cdot \bar{r}_i$, где λ - параметр метода порядка десятых единицы.
6. Шаги 2-5 повторяются несколько раз.

В итоге получается сетка, не лишенная черт регулярности, но сгущающаяся там, где густы исходные данные. Такая сетка может рассматриваться как компактизованная модель исходного множества данных, либо как средство классификации новых данных – этим данным можно приписывать тот же класс, который ранее был приписан ближайшему к вектору данных узлу карты Кохонена.

Прочие нейросетевые алгоритмы

Комбинируя два изложенных подхода, можно получать новые варианты алгоритмов обработки данных. Например, если данных много, то можно сначала по ним построить компактную карту Кохонена (точнее, ее многомерную версию), и уже к этой карте применять методы нелинейной регрессии. Однако не все нейроалгоритмы сводятся к такого рода комбинациям – с ними следует знакомиться по многочисленным специализированным изданиям. Мы же представили вам два в каком-то смысле предельных варианта нейросетевых алгоритмов – алгоритм обратного распространения ошибки, относительно сложный математически, и связанный со многими традиционными разделами математики, и относительно простой алгоритм Кохонена, тем не менее решающий ряд практически значимых задач.

4. Связь ГИС с нейронными сетями

(Задачи математической картографии и геоинформационного моделирования, выводящие на использование нейросетевых алгоритмов для их решения)

ГИС являются хорошей средой и средством для внедрения методов искусственного интеллекта и экспертных систем. Математико-картографическое моделирование и геоинформационное картографирование – одна из основ ГИС-технологий, применяемых при принятии решений, управлении, проведении экспертиз, составлении прогнозов и т.п. Основой для математической обработки является выделение однородных в каком либо смысле и отграниченных друг от друга объектов. Эти задачи решают процедуры классификации, районирования, таксации, бонитировки, квантирования признаков.

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. Собственно математическое моделирование предполагает более углубленный анализ, нежели просто вычисление количественных показателей. 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.

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. Большинство аналитических задач такого рода можно сформулировать как задачу классификации географических комплексов. При этом предварительно осуществляется классификация показателей и факторов, описывающих эти комплексы. В этих исследованиях нейросетевые алгоритмы могут быть использованы для восстановления функции по конечному набору значений и для разбиения конечного множества объектов на классы.

Для содержательной интерпретации требуется анализ близости выделенных классов между собой и выявление их связи с различными сторонами исследуемого явления. Для этого используются различные алгоритмы факторного или корреляционного анализа. Использование этих методов неэффективно при больших объемах, а в особенности при нелинейном характере взаимосвязей в первичных данных. В этом случае целесообразно использование нейросетевых алгоритмов, которые можно интерпретировать как обобщение методов линейной статистики на нелинейный или локально-линейный случай.

Очень важным основанием географических классификаций являются временные, генетические (наследуемые) характеристики изучаемых комплексов. В географических исследованиях комплексы обычно рассматриваются 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.) результаты которых отображаются на географических картах. Используемая при этом многомерная классификация и критерии достоверности алгоритмов анализа многомерных данных могут быть неприменимы к реальным географическим задачам и представлению соответствующих данных в ГИС. Относительно гибкие нейросетевые подходы могут оказаться более эффективными.

Одной из существенных проблем анализа геоданных является выбор оптимальной системы базовых показателей в соответствии с сущностью изучаемых явлений. Естественное

желание всестороннего рассмотрения явления в максимально полной системе показателей может привести к избыточности включенных в анализ первичных данных. Данные, поставляемые из различных источников и представляемые как независимые, могут многократно дублироваться или вычисляться одни на основе других (производные данные). Это может исказить значимость признаков и привести к ошибкам в результатах анализа. Самую существенную помощь здесь может оказать глубокое проникновение в сущность исследуемого комплекса территориально-распределенных процессов, что позволяет найти критерий значимости показателей и сформировать их систему адекватную исследовательской задаче. В автоматизированном режиме возможно использование методов он-лайн коррекции моделей и техники вычислительного эксперимента, обеспечивающих исправление ошибок в данных, автоматизированное выявление причинно-следственных связей, снижение размерности задач многомерной классификации и анализа геоданных.

Показатели лежащие в основе большинства классификационных задач аналитической географии имеют различную природу и могут носить как количественный так качественный характер. Поэтому алгоритмы анализа геоданных должны уметь работать с характеристиками как числовой так и нечисловой природы. Это накладывает определенные ограничения на возможности использования всего многообразия методов математического анализа. В ГИС атрибутивное описание объектов анализа уже формализовано и представлено в форме электронных таблиц. Это облегчает использование нейросетевых технологий и позволяет преодолеть трудности анализа, связанные с большими объемами первичной информации, пробелами в данных, разнородностью количественных и качественных характеристик объектов. Нейросетевой подход позволяет проводить компактизацию и параметризацию данных, создавать на основе данных простые и наглядные модели данных.

Перечень приведенных проблем анализа территориально-распределенных процессов, представленных в виде географических комплексов фактически выводит на необходимость использования методов нейросетевого анализа для структуризации многомерных данных. Выявление объектов или их комплексов (они сами по себе тоже объекты со своей сложной структурой), установление между ними различных пространственных отношений и мер близости. Комплексование объектов и агрегирование данных существенно зависит от постановки исследовательской задачи и позволяет значительно сократить число объектов анализа. При переходе от аналитических задач к задачам оценки и прогноза становится определяющим построение моделей исследуемых процессов. При этом:

- выявленные в результате анализа объекты становятся базовыми элементами модели,
- показатели географического комплекса переходят в числовые показатели свойств объекта и его связей,
- анализ географических комплексов и их показателей выводит на определение функциональных связей и возможности структурирования объектов в соответствии с их функциональными связями.

Для построения моделей важен также анализ причинно-следственных отношений событий (изменение состояний объектов вовлеченных в тот или иной процесс), структура функциональных связей выделенных объектов, показатели обменных процессов реализуемых на этих связях, определение характеристик полевого типа и т.д.

Весьма существенно и то, что часто трудоемкость решения задачи традиционными методами и на основе нейроалгоритмов резко различаются. В ситуации, когда необходимо вникнуть в структуру данных, затем написать программу обработки данных, затем отладить эту программу, бывает достаточно просто загрузить «сырые» данные в нейропрограмму, подождать некоторое время (если база данных велика, например, подождать ночь), и наутро получить готовые результаты.

Одним из эффективных средств решения слабо формализованных задач на основе примеров являются нейронные сети. 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.
3. Устойчивость при работе с зашумленными и недостоверными данными.
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.

Нужно также отметить, что нейросети в совокупности с геоинформационными технологиями могут использоваться для решения задач, выходящих за рамки традиционных приложений ГИС. Например, они находят применение при визуализации и картографировании многомерных данных (в которых могут содержаться пробелы) с помощью вложенных в пространство данных двумерных многообразий. Особенностью технологии является возможность непрерывного проектирования данных на карту, что существенно повышает точность представления данных. Основой для построения такой карты является двумерная прямоугольная сетка, вложенная в многомерное пространство, аппроксимирующая данные и обладающая регулируемыми свойствами упругости по отношению к растяжению и изгибу – поэтому ее называют упругой картой. Расположение узлов сетки ищется в результате решения вариационной задачи на нахождение минимума функционала, вид и параметры которого зависят от исследовательской задачи и определяют упругие свойства карты.

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. В качестве достоинств нейронных сетей 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.

5. Области применения нейросетевых ГИС

Интегрированные с геоинформационными системами нейронные сети – мощный инструмент для решения широкого класса задач, обеспечивающий эффективную поддержку принятия решений. В качестве входных и выходных данных нейронная сеть может использовать пространственные данные ГИС. Программы, созданные на основе нейросетевых алгоритмов, будут динамически модифицировать слои электронной карты, изменять характеристики существующих объектов, создавать новые объекты. В результате обработки массива имеющихся данных могут также возникать новые слои карты, в то время как существующие слои будут приобретать динамические свойства.

Сегодня уже можно привести много примеров, демонстрирующих эффективность нейросетевых ГИС, их круг потребителей чрезвычайно широк. Наиболее выигрышно они проявляют себя в ситуациях, когда приходится иметь дело с большими массивами инфор-

мации, хранящимися в крупных организациях, на основе которых принимаются решения. В них нуждаются специалисты, оценивающие и прогнозирующие состояние какой-либо области человеческой деятельности, например, рынков сбыта продукции, реальной стоимости недвижимости, загрязнения территории. Планирование очередности действий при развитии территорий и их инвестиционной привлекательности, выявление зон с наиболее напряженной экологической, социальной или экономической ситуацией, анализ характеристик геологических объектов – эти и многие другие задачи уже невозможно решать на современном уровне без привлечения интеллектуальных геоинформационных систем.

Приведем несколько простых примеров.

В сельском хозяйстве одни слои ГИС могут содержать сведения о посеве зерновых культур, а другие – о достигнутой урожайности. Нейросеть в этом случае будет обобщать практический опыт методов и технологий выращивания зерновой культуры, с учетом конкретных климатических, почвенных и прочих характеристик выбранной территории.

В решении задачи лесоустройства с помощью нейросети можно анализировать динамику приростов деревьев по высоте, диаметру и объему. Обработка тематических слоев ГИС с такой информацией поможет спланировать лесоустроительные работы, например, при выращивании насаждений сосны в лесопарковых хозяйствах и в зеленых зонах.

Прогностические нейросетевые модели могут использоваться **в демографии и организации здравоохранения**, опираясь на пространственные данные по плотности населения, медицинской статистике, загрязнению окружающей среды, представленные в виде слоев ГИС. Экспертная система будет определять, например, вероятностную продолжительность жизни, взаимосвязь различных категорий заболеваемости от экологического состояния территории, прогнозировать вспышки эпидемий.

Обработка ДЗ – одна из традиционных задач, решаемых сегодня геоинформационными системами. Анализ изображений с точки зрения математики опирается на теорию распознавания образов, когда по входным данным необходимо отнести объект к тому или иному классу. В данной области нейронные сети, наряду с методами нечеткой логики, нашли наиболее широкое применение. Такие методы оказались тем адекватным языком, на котором можно описать правила классификации, не прибегая к точным математическим значениям (используя понятные человеку термины типа «небольшой», «значительный» и т.д.). И, наоборот, извлекать из обученных нейронных сетей правила классификации по исходным данным, представляя их на обычном языке. В отличие от традиционных статистических методов, основанных на вычислениях в рамках той или иной математической формализации, классификаторы, основанные на нейронных сетях, используют адаптацию в процессе обучения, не требующую предварительного обоснования модели. В то же время доказано, что результаты классификации и в том и в другом случае могут совпадать, т.е. нейронная сеть способна сама построить соответствующую математическую формализацию.

Для классификации используются различные нейросетевые алгоритмы. В нейронной сети с обучением без образца проводится анализ цветных или черно-белых пикселей снимка без привязки к другим слоям карты, с целью выделения однородных фрагментов изображения (объектов) по тону, структуре, оттенку. Обучение по образцу полагается на доступные пространственные данные для выбранного участка территории. Если известно априори, что выбранному фрагменту снимка соответствует, например, лесной массив с известной степенью антропогенной нарушенности, то эта информация может быть использована нейронной сетью для классификации изображения.

Нейронные сети все чаще используются в задаче выявления пространственно-однородных участков изображения. Эта задача является очень актуальной при разработке теоретических и методологических основ новых альтернативных систем земледелия, принципов экологически безопасного землепользования и проектов землеустройства на

ландшафтной основе. Отдельные элементы ландшафта могут быть выявлены на основе анализа фрагментов изображения, их формы, цвета, взаимосвязей, неоднородности. Нейронная сеть также позволит оценить взаимосвязи отдельных элементов ландшафта.

В транспортной отрасли нейронная сеть может стать эффективным дополнением к ГИС мониторинга автомобильных дорог. Здесь речь идет об анализе транспортной нагрузки и состояния транспортного полотна, выборе оптимальных коридоров для строительства новых трасс и определении приоритетов в строительстве, анализе различных стратегий проведения ремонтных работ и соответственном распределении финансовых вложений. Задачей для нейронной сети может стать оперативное принятие решения по оптимизации распределения транспортной нагрузки на автомобильные дороги в случае транспортного происшествия в определенном месте, повлекшего за собой скопление автомашин (пробку). Нейронная сеть будет использовать в качестве входных данных слой ГИС с автодорогами, местоположением аварии, текущими погодными условиями и прочими характеристиками, которые влияют на скорость движения. Все параметры системы, имеющие отношение к указанному происшествию, будут скорректированы нейронной сетью, и их откат в исходное состояние станет возможен после стабилизации ситуации. Следующим шагом в развитии данного направления должна стать возможность онлайн-вого доступа водителей из автомобиля с помощью карманных ПК (КПК, PDA) и других мобильных устройств к картографическому серверу, содержащему оперативную информацию об обстановке на дорогах и подготовленные с помощью нейросети рекомендации по выбору оптимального маршрута.

Задача идентификации личности стала особенно актуальной в связи с недавними террористическими событиями. В аэропортах и других местах большого скопления людей принимаются усиленные меры по обеспечению безопасности. ГИС может эффективно использоваться для мониторинга подобных объектов, обеспечивая необходимой оперативной информацией службу безопасности. Нейронная сеть может использоваться совместно с ГИС для локализации подозрительных ситуаций на основе оперативной обработки данных, попадающих в объектив видеокамеры наблюдения.

6. Программное обеспечение

На рынке программного обеспечения в настоящее время имеется множество самых разнообразных программ для моделирования нейронных сетей. Поиск в Интернете дает сотни ссылок на зарубежные и российские сайты. Можно выделить несколько основных функций, которые реализованы во всех этих программах:

- формирование, конструирование нейронной сети;
- обучение нейронной сети;
- имитация функционирования (тестирование) обученной нейронной сети.

С точки зрения компьютерной технологии и программных интерфейсов они опираются на современные стандарты – от простых программ, ориентированные на платформу Unix с текстовым интерфейсом, до сложных модульных продуктов, базирующихся на последних технологических решениях от Microsoft.

Интегрированные решения на основе ГИС и нейронных сетей пока представлены слабо, несмотря на то что повышение функциональной мощности геоинформационных пакетов за счет интеграции специальных модулей расширения или ГИС-приложений – одна из важнейших черт современных геоинформационных систем. Проблема интеграции нейронных сетей и ГИС может быть решена по крайней мере тремя способами:

- интеграция (встраивание) нейросетевых моделей в ГИС с использованием специализированных средств геоинформационной системы (программирование на встроенных языках типа Avenue, MapBasic и т.п.);
- развитие интерфейса между отдельными приложениями нейросетевого анализа и ГИС, как самостоятельными системами (например, через обмен данными по DDE);

- создание прикладного программного обеспечения нейросетевых систем с элементами ГИС (например, с использованием библиотек классов типа MapObjects, GeoConstructor, MapX и проч.).

Выбор конкретного варианта связан с требованиями и постановкой задачи, имеющимися ресурсами и опытом работы. Особенности реализации определяются возможностями используемого ПО – как со стороны нейронной сети, так и со стороны ГИС, например, встроенным языком программирования, средствами OLE и DDE, функциональными возможностями DLL и ActiveX. На основе выбранной технологии создается соответствующее нейросетевое ГИС-приложение.

Ниже приводятся два конкретных примера – программных продукта, которые уже созданы на основе нейросетей и ГИС.

Программа ScanEx-NeRIS

Программа NeRIS предназначена для тематической интерпретации пространственных данных, в первую очередь данных дистанционного зондирования Земли. Основным инструментом, реализованным в программе, являются нейронные сети Кохонена, используемые для ординации, классификации и тематической интерпретации. Являясь одним из методов классификации многомерных данных, нейронные сети Кохонена обладают важными дополнительными свойствами, на которых основана значительная часть используемых в программе алгоритмов.

Возможности пакета тематической обработки растровых изображений в программе ScanEx-NeRIS:

- оценка количества классов, требуемых для описания тематики и составления тематической карты;
- оценка внутренней дробности, неоднородности тематических объектов (контуров);
- оценка распределения свойств экспертных объектов в признаковом поле дистанционной модели;
- оценка вероятностей присутствия тематических объектов, заданных экспертом в поле признаков снимка (выделение на изображении областей с различным уровнем оценки: оптимистическим, реалистическим, пессимистическим);
- построение иерархических классификаций с оценкой близости классов между собой (удобный для специалистов по тематической картографии интерфейс обработки «карты расстояний» («деревя расстояний»)) с целью уточнения и корректировки строения легенды и географического содержания тематической карты;
- создание тематически ориентированных нейронных сетей для последующей обработки растра с целью выявления тематических объектов;
- автотрассировка (векторизация) результатов покласовой обработки;
- поддержка системы координат наиболее распространенных отечественных и зарубежных картографических проекций;
- экспорт растровых покрытий и векторных слоев в обменных форматах наиболее распространенных географических информационных систем;
- представление результаты классификации для всех видов нейронных сетей как присвоением индекса класса каждому классифицированному пикселу, так и созданием растровых слоев "вероятности" (possibility) принадлежности пиксела одному конкретному классу (создание нескольких таких слоев с последующей их визуализацией позволяет наглядно представлять результаты классификации, например, выявлять "белые пятна" (неклассифицированные области пространства) и представлять данные для окончательной классификации традиционными методами).

Модуль Arc-SDM для ArcView

Модуль Arc-SDM – одно из свободно доступных расширений ArcView для моделирования в ГИС на основе алгоритмов нечеткой логики и нейронных сетей. С точки зрения пользователя ГИС процесс пространственного моделирования с использованием этого модуля состоит в построении нового тематического слоя на основе нескольких уже существующих. Arc-SDM использует два нейросетевых алгоритма, которые вынесены в самостоятельный программный модуль DataXplore. Первый построен на основе нейросети, использующей радиальные базисные функции, второй на основе кластеризации в нечеткой логике. Нейросеть, использующая радиальные базисные функции должна пройти этап обучения, в результате которого будет сгенерирован набор параметров, определяющий взаимосвязь между входными слоями данных и выходным (результатирующим) слоем. После этого для классификации данных можно использовать обученную нейросеть.

Например, в геологической задаче исследования полезных ископаемых, результирующий слой содержит сведения о наличии или отсутствии месторождений. Входные данные, используемые в процессе обучения, можно разделить на два типа – местоположения известных месторождений и участки территории, про которые известно, что там полезных ископаемых нет. На основе исходных векторных тематических слоев геоинформационной системы создается grid-тема, далее подготовленный набор данных передается в программный модуль DataXplore. Результат вычислений отображается в виде нового тематического слоя.

Литература

1. Горбань А.Н. Обучение нейронных сетей. М.: СП ПараГраф, 1990. 160 с.
2. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996. 276 с.
3. Питенко А.А. Нейросетевой анализ в ГИС. // Автореф. дисс. канд. техн. наук 05.13.16, Красноярск, 2000.
4. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применения в экономике и бизнесе. М 1998.
5. Proceedings of International Workshop on Dynamic and Multi-Dimensional GIS. 25-26 August 1997, The Hong Kong Polytechnic University, 252 pp. // 1997.
6. Al Gore. The Digital Earth: Understanding Our Planet in the 21st Century, 1998. http://www.regis.berkeley.edu/rhome/whatsnew/gore_digearth.html.
7. Д. Добрынин, А. Савельев. Нейронные технологии для тематического дешифрирования ДДЗ. // Гис-Обзор, №1, 1999, стр.12-14. <http://www.scanex.ru/>.
8. Kemp, L.D., Bonham-Carter, G.F., Raines, G.L. and Looney, C.G., 2001, Arc-SDM: Arcview extension for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural network analysis. <http://ntserv.gis.nrcan.gc.ca/sdm/>.
9. Миркес Е.М. Нейрокомпьютер. Проект стандарта: Новосибирск, 1998, 319 с.
10. Зиновьев А.Ю. Визуализация многомерных данных // Красноярск, 2000, 168 с.