

Адаптивный метод сжатия и организации информации о рельефе местности

к.э.н. Боронников Д.А.¹, Пантюхин Д.В.², Данько С.В.²

¹ Университет машиностроения, ² Московский физико-технический институт
(495) 223-05-23, доб. 1510, unir@mami.ru

Аннотация. Разработан адаптивный нейросетевой метод сжатия и организации информации о рельефе местности, протестированный на примере конкретных данных. Разработана и исследована программная реализация метода в среде Матлаб. Экспериментальные исследования показали, что нейронная сеть успешно запомнила и обобщила входную информацию о рельефе местности (110149 пространственных точек), с ошибкой менее 0.5 метра. При этом было достигнуто сжатие исходной информации в 12 раз.

Ключевые слова: нейронные сети, пространственная организация данных, геоинформационные системы, недропользование

Работа выполнена при финансовой поддержке Министерства образования и науки РФ в рамках Госконтракта № 14.514.11.4096.

Введение

Геоинформационные системы (ГИС) являются основой для оперирования большими объемами пространственно-привязанных данных как в области недропользования, так и в других областях. Несмотря на различия в назначении геоинформационных систем, принципы их построения в целом совпадают.

Нейронные сети хорошо зарекомендовали себя как универсальные аппроксиматоры сложных многомерных функций для больших объемов данных. Поэтому целесообразно провести работу по исследованию возможностей применения нейронных сетей в ГИС.

Поскольку число видов различных данных, которые могут использоваться в ГИС, достаточно большое, то необходимо провести некоторую классификацию таких данных.

Пространственные объекты можно разделить на два типа:

- дискретные, когда объекты, измерением которых (например, фотографированием, лазерным сканированием и т.п.) получают данные, являются разнесенными в пространстве (например, отдельные деревья и т.п.);
- непрерывные – это различного рода площади, объемы и т.п. Примером может служить рельеф местности.

С другой стороны, данные об этих объектах практически всегда представляются в дискретном, цифровом виде. Так, например, рельеф местности, измеряемый лазерным сканером, представляется совокупностью точек - координат (x, y, z) , от которых отразился лазерный луч.

Налицо противоречие между реальностью и представлением о ней. По этой причине представление данных о непрерывных объектах также должно быть непрерывно. И если для дискретных представлений разработано множество различных алгоритмов и программ, то для непрерывных представлений это число довольно мало.

Основным способом представления непрерывных данных является технология нейронных сетей.

Детальный анализ работ позволяет провести сравнение различных методов представления пространственных данных и сделать следующие выводы о преимуществах и недостатках нейросетевого подхода в сравнении с классическими подходами к организации пространственных данных:

1. Данные представляются в непрерывном виде, закодированном в значениях весовых ко-

эффициентов нейронов. Возможно получение информации о данных, которые не были представлены в исходной выборке: автоматически выполняется обобщение (гладкая интерполяция) данных. Это свойство является одним из самых главных преимуществ применения нейросетевых технологий.

2. В нейронной сети одновременно происходят процессы запоминания и обобщения данных. С этим связана возможность дальнейшего извлечения знаний из данных. С одной стороны, это большое преимущество нейросетевого подхода, но с другой, возникает зависимость структуры нейронной сети от самих данных. Так, например, если данные полностью случайны, то нейронная сеть не сможет вывести общие закономерности (поскольку их нет) и тогда будет запоминать информацию. Однако, простое запоминание информации требует огромного количества вычислительных ресурсов, гораздо больше того, что потребовалось бы для запоминания того же объема данных прямым методом. Если же в данных есть закономерность, то нейронная сеть пытается вывести такую закономерность и тем самым снизить объемы запоминаемой информации.
3. Нейронная сеть обучается на данных. Наличие процесса обучения делает нейросетевой подход, с одной стороны, предпочтительным в случаях, когда данные необходимо не просто хранить, но и использовать для получения знаний из них, выявления связей. С другой стороны, в зависимости от данных, может случиться такой вариант, что нейронная сеть не обучится до заданного качества. И, хотя теория говорит, что нейронная сеть может аппроксимировать любую непрерывную функцию с заданной точностью, на практике для этого могут потребоваться нереальные вычислительные мощности. Поэтому нейросетевой подход следует использовать не сам по себе, а в сочетании с другими подходами, для тех данных, для которых преимущества превысят недостатки.
4. Нейросетевые алгоритмы обычно заключаются в двух этапах: обучение и непосредственное использование обученной нейронной сети (симуляция). Процесс симуляции – быстрый процесс. В силу изначальной параллельности нейронной сети этот процесс может быть относительно легко распараллелен на множество параллельных аппаратных ускорителей, например графические процессоры, что дает большой скачок в производительности системы, основанной на нейросетевой организации данных. Процесс же обучения гораздо более медленный. Часто он требует высоких вычислительных затрат, поэтому следует разносить эти процессы на разные вычислительные системы. Из этого следует, что нейросетевой подход в организации данных следует использовать для данных, которые часто используются, но редко обновляются.
5. Поскольку нейронная сеть выполняет обобщение данных, то она способна нивелировать возможные ошибки в данных. Так, наличие шумов или пропусков в данных учитывается автоматически, в процессе обучения. Повторим, что это свойство является как большим преимуществом нейросетевого подхода, в случаях, когда в исходных данных есть ошибки, так и недостатком в случае, когда требуется точное запоминание информации.

Таким образом, нейросетевой подход к организации данных имеет право на существование, но требует применения в ограниченных ситуациях. Поэтому в реальных ГИС нейросетевой алгоритм организации пространственных данных должен быть одним из многих, а выбор конкретного метода организации данных должен определяться свойствами данных и задачами их обработки.

Основными задачами в ГИС, для которых применяют нейросетевые методы решения, являются:

- оценка параметров (загрязненность вод, оптимальности положения и др.);
- предсказание (мест залежей, траекторий движения, и др.);
- кодирование информации (адресов и географических координат);
- обработка изображений (извлечение данных из карт);
- управление (например, трубопроводами).

Для решения этих задач, данные переводятся в нейросетевое представление.

Ниже проведен синтез нейросетевого алгоритма организации пространственных данных о рельефе местности для Озерного ГОК.

Нейросетевой алгоритм организации пространственных данных о рельефе местности

Физическая постановка задачи. Месторождение Озерного горно-обогатительного комбината (ГОК) расположено в Республике Бурятия в 160 км от Транссибирской железнодорожной магистрали.

Для реализации функций моделирования и проектирования необходимо разработать способ организации пространственных данных о рельефе местности.

Эти данные представляют собой координаты точек сканируемой лазерным методом поверхности с шагом 0.5 м. Данные с лазерного сканера поступают в среду Autodesk Revit, где проходят предварительную обработку. Затем данные выгружаются в текстовый файл, содержащий информацию о координатах точек поверхности рельефа местности. По выбору проектировщика данные могут быть представлены на регулярной или нерегулярной сетке, с шагом до 0.5 м. В данных возможны пропуски.

Рассматриваемый вариант реальных данных рельефа местности Озерного ГОК представлен на рисунке 1. Это площадка размером 50*20 м, с перепадом высот до 10 м. Есть пропуски измеренных данных, связанные с наличием дороги и съездов с нее.

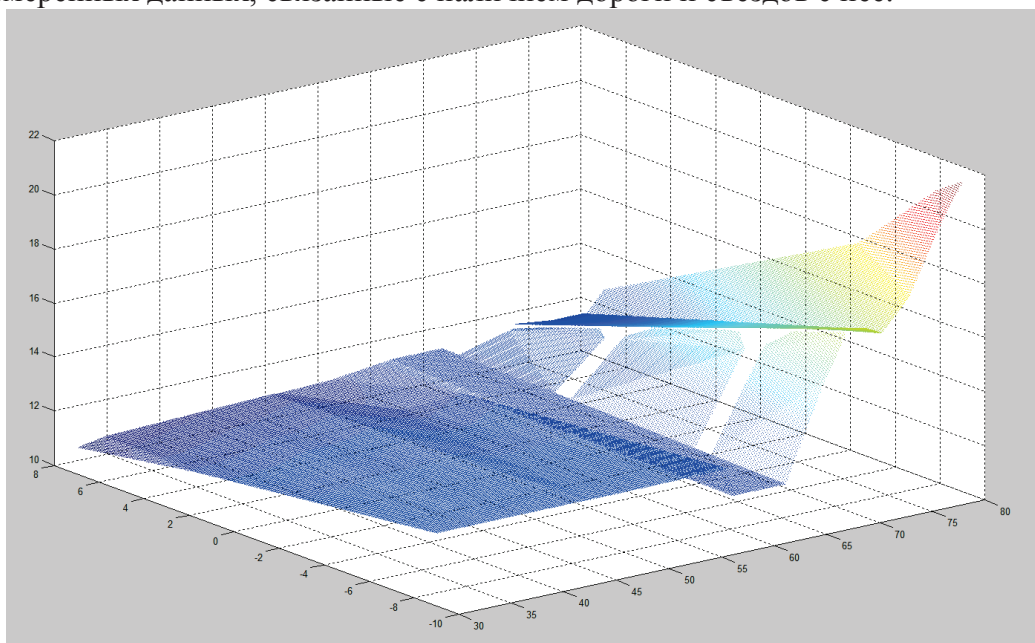


Рисунок 1. Рельеф местности

Математическая постановка задачи. Математически задача состоит в выполнении преобразования координат (x,y) сетки пространственного разбиения в координату z – высоты текущего участка сканируемой поверхности. Т.е. необходимо преобразовать участок пространства размерности два в участок пространства размерности один на основе имеющихся измеримых точек. При этом должна выполняться гладкая интерполяция результатов для точек, не представленных в исходных данных, но находящихся в пределах изучаемого участка.

Нейросетевая постановка задачи. В соответствии с физической и математической постановкой, исходными данными является набор координат (x,y) и соответствующая им высота z . Данные представлены в текстовом файле, содержащем тройки чисел (x,y,z) .

Входным сигналом нейронной сети будут текущие координаты (x, y) точки вычислительной сетки (2 входа).

Выходной сигнал нейронной сети будет интерпретироваться как координата z текущей точки рельефа (1 выход).

Желаемый выходной сигнал нейронной сети – истинное значение координаты z текущей точки рельефа.

Сигнал ошибки нейронной сети – разность между действительным (выходным) и желаемым выходным сигналом.

Функционал вторичной оптимизации нейронной сети – сумма квадратов ошибок на всех точках обучающего множества.

Метод поиска экстремума функционала вторичной оптимизации нейронной сети – метод скалированных сопряженных градиентов *trainscg* [1].

Тип нейронной сети – многослойный персептрон с двумя скрытыми слоями, число нейронов в слоях – варьируемый параметр.

Функции активации для скрытых слоев – гиперболический тангенс *tansig* [2], для выходного слоя – линейная.

Начальные условия при настройке нейронной сети – случайные числа из диапазона $(-1,1)$.

Начальные данные проходят предобработку – перенос и масштабирование к диапазону $(-100,100)$ для всех координат.

Выходные данные проходят постобработку и преобразуются в обратном порядке.

Тестирование проводится на примере реального рельефа местности Озерного ГОК из файла *Relief.txt*.

Математическое моделирование нейросетевого алгоритма организации пространственных данных в области недропользования проводится в соответствии с аналитическими выражениями, определяющими структуру синтезированной нейронной сети и выбранного метода обучения.

Формула преобразования, выполняемого одним нейроном:

$$g_i^k = f^k \left(\sum_{j=1}^{J^{k-1}} W_{ij}^k \cdot x_j^k + W_{i0}^k \right).$$

Здесь: $k=1, \dots, K$ – номер слоя нейронной сети;

K – число слоев нейронной сети;

g_i^k – выход i -ого нейрона k -ого слоя нейронной сети;

f^k – функция активации k -ого слоя нейронной сети;

W_{ij}^k – весовые коэффициенты связей i -ого нейрона k -ого слоя нейронной сети с j -ым нейроном $(k-1)$ -ого слоя;

x_j^k – j -ый вход в k -ый слой, для первого слоя нейронной сети это соответствующий вход в нейронную сеть, для остальных слоев это соответствующий выход предыдущего слоя нейронной сети $x_j^k = g_j^{k-1}$;

W_{i0}^k – свободный член i -ого нейрона k -ого слоя нейронной сети;

J^{k-1} – число нейронов в $(k-1)$ -ом слое нейронной сети, для $k-1=0$ это число входов в нейронную сеть;

$i, j=1, 2, \dots, J^k$ – номер нейрона в k -ом слое нейронной сети.

Выражение для всей нейронной сети получается подстановкой данной формулы для каждого нейрона в нейронной сети в соответствии со связями нейронов.

В нашем случае, в соответствии с выбранными параметрами, имеем:

$$z = g^3 = \left(\sum_{j=1}^{100} W_j^3 \cdot f \cdot \left(\sum_{i=1}^{100} W_{ji}^2 \cdot f \cdot (W_{ix}^1 \cdot x + W_{iy}^1 \cdot y + W_0^1) + W_{i0}^2 \right) + W_{j0}^3 \right).$$

Здесь принято:

- количество слоев нейронной сети – три;

- один выход нейронной сети z , которой описывает координату z точки поверхности рельефа;
- число нейронов с двух скрытых слоев – 100 в каждом;
- число входов в нейронную сеть – два, координаты x и y сетки пространственного разбиения;
- функция активации последнего, третьего (выходного) слоя – линейная;
- функции активации $f()$ скрытых слоев одинаковые – *tansig*. Эта функция выполняет нелинейное преобразование входа x в выход $g(x)$ по формуле:

$$g(x) = \frac{2}{1 + e^{-2 \cdot x}} - 1.$$

Такая нейронная сеть может выполнять достаточно сложные нелинейные преобразования пространства входов размерностью два в пространство выходов размерностью один.

Обучение нейронной сети происходит в соответствии с методом шкалированных сопряженных градиентов *trainscg*, описание математического аппарата и вывод которого можно найти в [1]. Отметим, что это один из самых быстрых методов настройки нейронных сетей в пакете Матлаб.

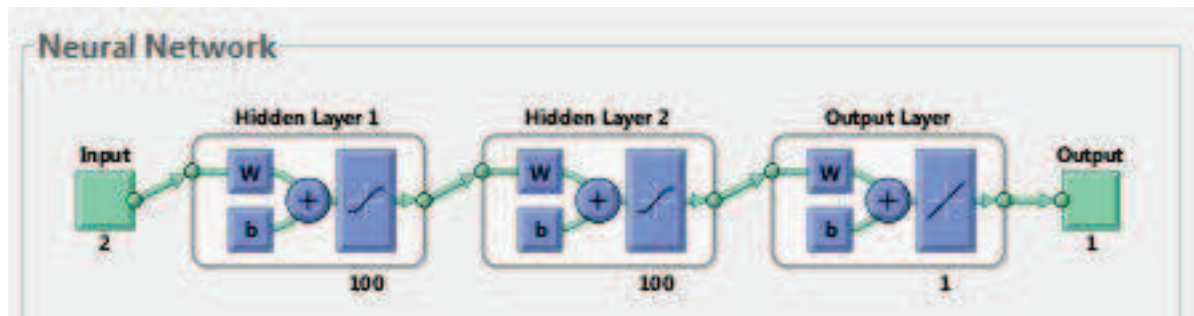


Рисунок 2. Структура трехслойной нейронной сети

Синтезированная структура нейронной сети представлена на рисунке 2.

Программная реализация

Для выполнения программной реализации был выбран пакет Матлаб фирмы Mathworks версии 2013a под управлением операционной системы Windows, который включает в себя Neural Network Toolbox – специализированный инструментарий для создания, моделирования, обучения нейронных сетей различного типа.

Программная реализация состоит из одного модуля, который позволяет:

1. Загружать пространственные данные в заданном выше формате с произвольным разрешением пространственной сетки. Загрузка данных происходит при помощи команд Матлаб открытия/закрытия и чтения текстовых файлов: *fopen()*, *fclose()*, *textscan()*, *fgetl()* и др.
2. Проводить обработку пространственных данных с произвольным пространственным разрешением. Применяемый способ вычислений позволяет узнать значение координаты z в произвольной точке внутри исследуемой области (x, y) , при этом точность вычислений зависит от шага пространственного разрешения исходных данных, а для промежуточных точек автоматически выполняется гладкая интерполяция результатов. Желаемая точность обучения нейронной сети подобрана таким образом, чтобы обеспечить максимальное отклонение обученной нейронной сети не больше 0.5 м.

Обработка данных выполняется в несколько этапов:

На первом этапе загруженные данные проходят предварительную обработку таким образом, чтобы привести диапазон данных к диапазону $(-100, 100)$. Это достигается за счет операции масштабирования и параллельного переноса, это необходимо для обеспечения работоспособности модуля с реальными данными, представленными в абсолютных координатах.

Также это обеспечивает независимость модуля расчета от конкретных диапазонов данных.

На втором этапе создаются нейронные сети с заданными параметрами. Это выполняется командой *newff()*.

На третьем этапе созданная нейронная сеть обучается с помощью функции *train()* выбранным методом обучения. В процессе обучения выводится информация о текущем состоянии процесса.

На четвертом этапе нейронная сеть проверяется (симулируется) с помощью команды *sim()*. Полученная информация с выхода нейронной сети преобразуется к исходным диапазонам. Затем происходит вычисление полученного максимального отклонения обученной нейронной сети от истинных значений. Если это значение меньше заданного (0.5 м), то нейронная сеть считается успешно обученной. Она запоминается в файл (команда *save()*). Сохраняется также информация о масштабирующих коэффициентах.

Именно эта информация (структура и весовые коэффициенты обученной нейронной сети) является нейросетевым представлением исходных пространственных данных. Таким образом, исходные пространственные данные организованы в структуре нейронной сети, а конкретные данные кодируются значениями весовых коэффициентов нейронной сети.

Для извлечения данных необходимо загрузить сохраненную обученную нейронную сеть (команда *load()*), подать на ее вход значения координат интересующей точки вычислительной сетки и запустить симуляцию нейронной сети. Результат вычислений будет искомым значением третьей координаты точки рельефа.

3. Выгружать и отображать данные о ходе вычислений и результатах. В процессе вычислений в командном окне Матлаб выводится информация о ходе процесса (командами *fprintf()*). Результаты вычислений – структура обученной нейронной сети и масштабирующие коэффициенты – сохраняются в файл. Рассчитывается и отображается в командном окне максимальная ошибка обучения.

4. Измерять достигнутые характеристики процесса обработки:

- a. время обработки в секундах и циклах;
- b. точность обработки в процентах и максимальное отклонение;
- c. размер получаемого представления в байтах и числе весовых коэффициентов.

Эти данные выводятся в командное окно Матлаб.

Экспериментальные результаты

Для тестового варианта, содержащегося в файле *Relief.txt*, были показаны следующие характеристики процесса вычислений:

Количество обработанных точек: 110149.

Ошибка обучения нейронной сети: 0.91719 при заданной максимальной ошибке 0.91735.

Среднеквадратичное отклонение результатов составило 0.0025.

Максимальное отклонение обученных данных: 0.451 метра (2.109 процентов), что меньше заданного значения 0.5 метра.

Время обработки составило 6757.6 секунд (1482 циклов). Из этого времени 0.46 секунды проводилась загрузка исходных данных, 0.005 секунды проводилась предобработка данных (масштабирование), 6755.49 секунд длилось обучение нейронной сети, 1.719 секунды проводилась симуляция и постобработка результатов.

Процесс обучения показан на рисунке 3.

Размер получаемого представления обученной нейронной сети (объект *net2*): $b=2863876$ байт, что соответствует 10501 весовым коэффициентам. Из них $a=2643576$ байт соответствует исходным данным (поля *exampleInput* и *exampleOutput*), хранение которых обязательно, а служит только для возможности однозначного повторения процесса обучения. Таким образом, на запоминание весовых коэффициентов, структуру нейронной сети и прочую служебную информацию приходится $b-a=2863876-2643576=220300$ байт. Коэффициент

сжатия составил $a/(b-a)=12$ раз.

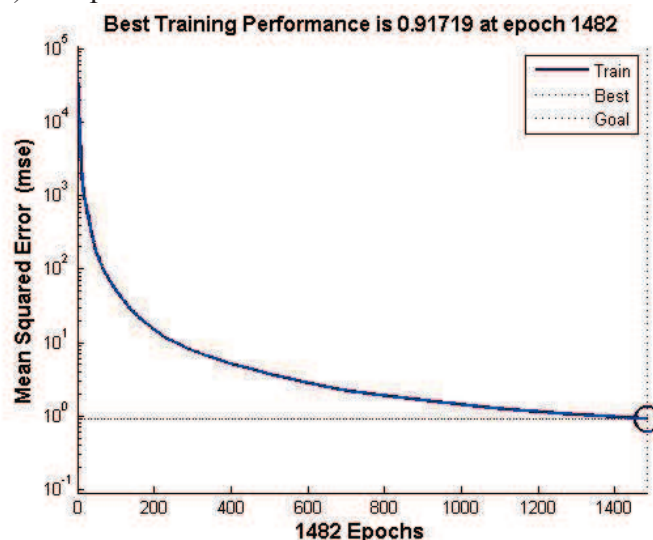


Рисунок 3 - График процесса обучения нейронной сети

Заключение

В процессе экспериментов было обнаружено, что используемые размерности нейронных сетей (два скрытых слоя по 100 нейронов в каждом) требуют достаточно заметных вычислительных затрат. Поэтому дальнейшее увеличение числа обрабатываемых пространственных точек делает необходимым переход на параллельные аппаратные вычислительные средства и разработку соответствующих программ для них.

На сегодняшний день наиболее эффективными аппаратными средствами являются ЭВМ с графическими процессорами. Этим объясняется наш выбор этого типа аппаратных ускорителей при разработке планов дальнейших исследований и разработок.

Дополнительно было изучено, что сжатие файла исходных данных с помощью упаковщика 7zip алгоритмом LZMA с максимальной степенью сжатия приводит к представлению объемом 823412 байт, т.е. в 3.74 раза хуже, чем достигнутое нами значение (220300 байт). При этом нейросетевое представление сразу готово к вычислениям, без необходимости распаковки упакованных данных.

Процесс прямого расчета (симуляции) уже обученной нейронной сети является довольно быстрым (1.719 секунды).

Результаты экспериментальных исследований подтверждают, что разработанный нейросетевой алгоритм организации пространственных имеет большие перспективы для реализации в составе ГИС.

Реализация таких алгоритмов позволит:

- уменьшить объемы запоминаемой информации;
- увеличить скорость обработки информации;
- унифицировать различные операции по обработке пространственных данных в нейросетевом представлении, например такие как, аппроксимация данных, классификация данных, экстраполяция данных, кодирование данных и др.

Литература

1. M.F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," Neural Networks, Vol. 6, 1993, pp. 525–533.
2. T.P. Vogl, J.K. Mangis, A.K. Rigler, W.T. Zink, and D.L. Alkon, "Accelerating the convergence of the backpropagation method," Biological Cybernetics, Vol. 59, 1988, pp. 257–263.

VI Школа-семинар «Современные проблемы термовязкопластичности в прикладных задачах анализа конструкций и технологий высоких параметров»

д.ф.-м.н. проф. Бондарь В.С.
Университет машиностроения
д.ф.-м.н. проф. Васин Р.А.
Институт механики МГУ
д.ф.-м.н. проф. Кийко И.А.
МГУ им. М.В. Ломоносова

В апреле 2013г. в Московском государственном машиностроительном университете (МАМИ) состоялась очередная школа-семинар, посвященная проблемам термовязкопластичности. В работе школы приняли участие свыше 50 человек из Москвы, Санкт-Петербурга, Нижнего Новгорода, Тулы, Твери, Королёва, Дмитровграда – сотрудники университетов (МГУ, СПбГУ, ТулГУ, Университета машиностроения, ТвГТУ, ГТУ МАДИ) и ведущих специализированных научно-исследовательских учреждений (ФГУП ЦНИИКМ «Прометей», ФГУП ЦНИИМАШ, ФГУП ЦИАМ, ГНЦ РФ НИИАР). Как всегда, программа школы содержала только заказные доклады. Авторы 12 заслушанных докладов представляли как университетскую науку, так и научно-исследовательские институты.



По тематике доклады можно условно разделить на следующие группы: информация о научных школах по механике деформируемых твердых тел (МДТТ); общие проблемы прочности и ресурса изделий; проблемы построения термомеханических моделей в МДТТ; конкретные задачи МДТТ, имеющие важное прикладное значение. Как правило, в каждом докладе содержалось обстоятельное обсуждение работ по основной теме сообщения; некоторые доклады по тематике можно было отнести к нескольким из перечисленных выше групп.

В докладе большого коллектива авторов (НИИ механики Нижегородского государственного университета), от имени которого выступал проф. А.В. Кочетков, была представ-