

Как и ожидалось, модель показала высокие результаты для обучающего множества и значительно более низкие для тестирующего. Особенно интересным в анализе результатов прогнозирования для компаний из тестирующего множества представляется факт высокой вероятности предсказания банкротства в первые 2 года. Требуется дальнейшее исследование для определения причин этого явления [1].

В данной работе было представлено, каким образом нейронные сети способны помочь людям в генерации знаний, которые основывались бы на всех первоначальных данных. Исследования в области нейронных сетей в основном достаточно наглядны. По сравнению с другими вычислительными методами в статистике и науке они имеют значительные преимущества. Но они никогда не смогут полностью заменить людей в процессе решения задачи. Нейронные сети должны использоваться для обобщения данных, а не для определения, атрибуты и критерии которого весьма важны при сборе данных. Нейронные сети адаптивны по своей природе, они могут подражать решению проблемы человеком, но они не сообщают, какой из критериев решения задачи должен быть принят во внимание перед сбором данных. Также были рассмотрены различные методики анализа и оценки рисков. Нейронные сети не являются панацеей, во многих случаях более эффективным будет использование традиционных методов статистики. Несмотря на это, во многих областях управления рисками нейронные сети зарекомендовали себя как более рациональное решение.

Литература

1. Короткий, С. Нейронные сети: алгоритм обратного распространения / С. Короткий. – 231 с.
2. Блум, Ф. Мозг, разум и поведение / Ф. Блум, А. Лейзерсон, Л. Хофстедтер. – М.: Мир, 1988. – 198 с.
3. Уоссермен, Ф. Нейрокомпьютерная техника: Теория и практика / Ф. Уоссермен. – Мир, 1992. – 154 с.

УДК 004.8.032.26

НЕЙРОСЕТЕВОЙ ПОДХОД К АРХИВАЦИИ ФАЙЛОВ

Шевеленков В. В.

УО «Брестский государственный технический университет», г. Брест

Введение

На сегодняшний день известно множество методов архивации файлов. Большинство персональных компьютеров используют в повседневной работе такие приложения, как WinRAR и WinZip, которые позволяют пользователям сжимать файлы в меньшие по размеру архивы. Особенно сжатые файлы удобно передавать и скачивать с удаленных ПК и распределенных систем.

Каждый известный компрессионный метод характеризуется такими параметрами, как скорость и максимально достижимый коэффициент сжатия. Целью написания данной работы является создание алгоритма сжатия файлов с коэффициентом, недостижимым распространенными на сегодняшний день приложениями.

Алгоритм, описанный ниже, основан на использовании алгоритма обратного распространения ошибки с использованием адаптивного шага обучения и некоторыми модификациями, для достижения необходимых результатов. Предложенный метод сжатия может быть применен к файлам любого типа и позволит достичь нового уровня информационной безопасности.

1. Общие принципы архивации

Первые работы, касающиеся сжатия информации были опубликованы Клодом Шенноном и Хартли. Они не были конструктивными, поскольку не содержали точного описания алгоритмов сжатия, а лишь описывали такую возможность. Практическое применение стало возможным лишь с появлением алгоритмов Шеннона-Фано (Shannon-Fano) и Хаффмана (Huffman), которые заключались в использования для каждого значения байта кода различной длины.

Наиболее известные архиваторы используют в процессе сжатия избыточность информации, арифметическое кодирование, энтропийное кодирование и др. Каждый из существующих методов имеет как сильные, так и слабые стороны. Например, методы, использующие в своей работе избыточность информации, показывают хорошую скорость работы, но при этом обеспечивают средний уровень сжатия (0,7-0,8 от исходного размера файла), а методы, использующие математическое кодирование, наоборот, работают медленнее, но и коэффициент сжатия у них выше.

Методика, предложенная в этой работе, основана на использовании алгоритма обратного распространения ошибки. Она показывает хорошие результаты по коэффициенту сжатия, но также имеет и весомые недостатки, выражающиеся в низкой скорости архивации. Причина написания данной работы – достижение коэффициента сжатия, равного 10 и выше.

2. Методика архивации

Процедура трансформации информации состоит из следующих шагов:

- нормализация информации;
- сжатие информации;
- сохранение информации.

В данной работе автор предлагает методику сжатия файлов любого типа, так же как и большинство используемых архивационных приложений.

2.1. Нормализация информации

Содержимое каждого файла представляется набором ASCII символов. Код каждого символа представляется числом в диапазоне от 0 до 255. Алгоритм обратного распространения использует при обучении и прогнозировании числа в диапазоне от 0 до 1, таким образом, необходимо нормализовать информацию, т.е. привести исходные данные к требуемому диапазону.

2.2. Сжатие информации

Для сжатия информации мы используем многослойную нейронную сеть, обучаемую при помощи алгоритма обратного распространения ошибки с адаптивным шагом обучения и варьированием количества нейронов в скрытых слоях. На рис.1 показана архитектура данной нейронной сети. [1]

Причем $X_1...X_N$ -входы нейронной сети, именно на эти входы будет подаваться нормализованная информация. W_1 , W_2 и W_3 – матрицы весовых коэффициентов. Y -выходной сигнал нейронной сети.

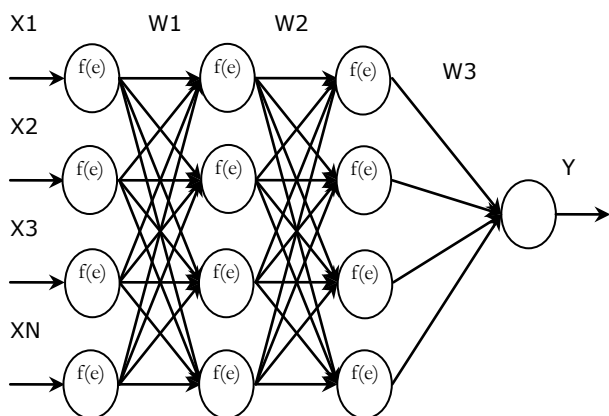


Рисунок 1 – Архитектура нейронной сети

Процедура обучения или тренировки нейронной сети – это итерационный процесс, который может быть разделен на следующие шаги:

- Каждый шаг обучения начинается с подачи входных образов из обучающей выборки. Для каждого образа вычисляются выходные значения для каждого нейрона скрытых слоев сети и выходного нейронного элемента.

- На следующем шаге алгоритма выходной сигнал сети сравнивается с желаемым (эталонным), который выбирается следующим, после поданного входного образа. Разница является ошибкой выходного нейрона, и основная идея методики заключается в прогнозировании требуемой информации, т.е. в получении значения выходного сигнала нейронной сети, равного желаемому последующему из обучающей выборки. Ошибки нейронов скрытых слоев невозможно вычислить, т.к. выходные значения этих нейронов неизвестны. Только в середине 80-х годов появился алгоритм обратного распространения ошибки, который заключался в распространении ошибки выходного нейрона на все нейроны скрытых слоев. Весовые коэффициенты W , которые используются для обратного распространения ошибки, равны весовым коэффициентам при прямом распространении информации.

- После этого необходимо скорректировать весовые элементы и пороговые значения для каждого нейронного элемента.

- Также мы изменяем шаг обучения, который влияет на скорость обучения сети.

- При достижении желаемой ошибки обучения мы можем использовать нашу сеть для прогнозирования.

Таким образом, обладая информацией о значениях весовых коэффициентов, первых 10 элементов, пороговых элементов и некоторых других параметрах, мы можем прогнозировать нашу обучающую выборку с высокой точностью, достигающей значений порядка 0,0001 и выше [2].

2.3. Сохранение информации

Обученная нейронная сеть представляет собой определенное число параметров, необходимых для сохранения: весовые коэффициенты (210 при архитектуре 10-10-1), стартовый набор эталонных значений (10) и т.д. Для восстановления архитектуры сети нужно сохранить эти данные в файле, требующем сжатия. Последовательность сохранения информации следующая:

- Параметры нейронной сети – количество слоев, количество элементов в каждом слое;
- Весовые коэффициенты;
- 10 первых элементов обучающей выборки;
- Количество прогнозируемых элементов.

После сохранения данных числовых вещественных значений процесс сжатия можно считать законченным.

3. Модификация нейронной сети

Изложенный выше алгоритм успешно архивирует большинство файлов. На некоторых из них нейронная сеть обучается довольно быстро, т.е. скорость высока, однако некоторые файлы не поддаются сжатию, т.е. нейронная сеть на обучающей выборке, полученной из этих файлов, не может в процессе обучения достичь требуемой среднеквадратичной ошибки. Для решения данной проблемы автором была разработана методика сжатия почти каждого информационно массива (файла). Если нейронная сеть не может в процессе обучения достичь желаемой ошибки, тогда алгоритм изменяет параметры нейронной сети, такие как количество нейронов в скрытых слоях и количество слоев сети. Если эти модификации не помогают достичь минимальной ошибки, нам остается дообучать нейронную сеть в процессе прогнозирования, включением некоторого числа эталонных элементов в спрогнозированный ряд.

Таким образом, варьируя параметры нейронной сети и дообучая сеть в процессе прогнозирования, мы можем достичь высокой точности тренировки сети.

4. Экспериментальные результаты

На рисунке 2 мы видим архивационную программу, созданную автором по алгоритму, предложенному выше. Она позволяет выбирать файлы, обучать сеть, сохранять сжатые файлы. Также представляется информация о текущем состоянии нейронной сети, количестве нейронов в слоях и текущая ошибка. Если программа не может достичь требуемой ошибки, она автоматически производит модификацию сети.

Программа тестировалась на файлах различных типов и результаты сравнились со стандартным приложением ПК WinRAR. При сжатии некоторых типов файлов (*.jpg, *.mp3, *.Rar) коэффициент сжатия достигал 5-9, причем приложение WinRAR эти файлы сжать не могло, и их размер после применения только увеличивался. Недостатком данного алгоритма является низкая скорость сжатия, равная 10 минутам при файле размером 10 Мб. Однако основная цель работы – достижение высоких коэффициентов сжатия без учета скорости. Простейшее решение проблемы избыточности времени – распараллеливание алгоритма и использование другой архитектуры нейронной сети.

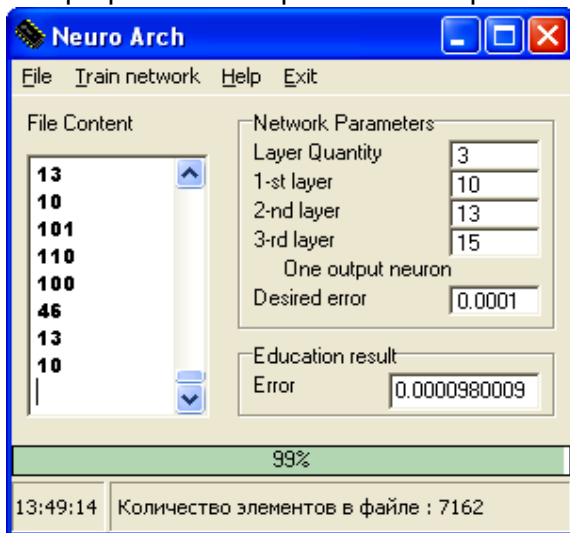


Рисунок 2 – Архиватор

Заключение

В данной работе представлена новая методика архивации файлов. Результаты экспериментов показывают положительный результат применения данного алгоритма к файлам многих типов. Более того, данный алгоритм имеет множество возможностей для усовершенствования. Конечно, алгоритм не лишен недостатков, но в следующих работах автор попытается от них избавиться.

Литература

1. Rumelhart D., Hinton G., Williams R. Learning representation by backpropagation errors// Nature.-1986.-N323.-P.533-536.
2. Golovko V., Savitsky J., Gladischuk. Predicting Neural Net// Proceedings Int. Conf. CM NDT-D5.-Berlin:DGZfP.-1995.-P348-353.