

Московский Энергетический Институт
(технический университет)

Кафедра Электротехники и Интроскопии

Курсовая работа
Применение нейронных сетей для классификации и
параметризации дефектов труб парогенератора

Выполнил:
Студент группы А15-00
Мартынов Н. С.

Проверил:
Лунин В. П.

Москва 2004 год

Введение.....	3
Классификация.....	4
КОДИРОВАНИЕ ВЫХОДНЫХ ДАННЫХ.....	4
ИСПОЛЬЗУЕМЫЕ ПРИЗНАКИ.....	4
СТРУКТУРА НЕЙРОННОЙ СЕТИ.....	4
ОБУЧЕНИЕ.....	5
РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ.....	6
Параметризация.....	7
РАЗДЕЛЬНАЯ СХЕМА ПАРАМЕТРИЗАЦИИ.....	7
ОПРЕДЕЛЕНИЕ ГЛУБИНЫ.....	8
<i>Используемые признаки.....</i>	8
<i>Структура сети.....</i>	8
<i>Результаты обучения.....</i>	9
<i>Результаты тестирования.....</i>	9
ОПРЕДЕЛЕНИЕ ШИРИНЫ.....	10
<i>Используемые признаки.....</i>	10
<i>Структура сети.....</i>	10
<i>Результаты обучения.....</i>	11
<i>Результаты тестирования.....</i>	11
ОПРЕДЕЛЕНИЕ УГЛА.....	12
<i>Используемые признаки.....</i>	12
<i>Структура сети.....</i>	12
<i>Результаты обучения.....</i>	13
<i>Результаты тестирования.....</i>	13
Сравнение результатов определения класса и глубины реальных дефектов различными методами.....	13
Приложение.....	14
КОДИРОВАНИЕ ВХОДНЫХ ДАННЫХ.....	14
ФУНКЦИИ АКТИВАЦИИ.....	15
ВЫБОР КОЛИЧЕСТВА ОБУЧАЕМЫХ ПАРАМЕТРОВ СЕТИ.....	16
СТРАТЕГИЯ ОБУЧЕНИЯ.....	20
АЛГОРИТМ ОБУЧЕНИЯ.....	21
ОСТАНОВКА ОБУЧЕНИЯ.....	22
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	23

Введение

Парогенераторы являются ответственными изделиями так как являются оборудованием, применяемым на атомных электростанциях. Очевидно, что критически важно надежно выявлять дефекты трубок парогенератора. Более того, необходимо определять характер обнаруженного дефекта, его местоположение (внутренняя или внешняя поверхность трубки) и параметры. Важнейшими параметрами дефекта являются его глубина и ширина.

Контроль трубок парогенератора производится вихретоковым методом. Сигнал с датчика передается в компьютер, где подвергается обработке – фильтрации и кластеризации. Далее вычисляются признаки дефектов:

- Амплитуда
- Фаза
- Размах
- Энергия
- другие.

Полученные признаки подвергаются дальнейшей обработке с целью определения типа дефекта и его параметров:

- глубина
- ширина
- угол.

Для этого возможно применение таких методов как использование искусственных нейронных сетей и регрессионный анализ. В данной работе исследуется применение нейронной сети типа многослойный персептрон с обучением по алгоритму RPROP.

Классификация

Дефекты труб парогенератора необходимо классифицировать на находящиеся на внутренней и внешней стенках трубы.

Кодирование выходных данных

Для решения задачи классификации используется следующая схема кодирования выходных данных:

- если дефект принадлежит данному классу, то на соответствующем выходе устанавливается 1;
- если дефект не принадлежит данному классу, то на соответствующем выходе устанавливается 0;
- количество выходов соответствует количеству классов дефектов, то есть равно двум.

При использовании такого подхода, для каждого дефекта будут получены вероятности принадлежности его к каждому из классов.

Используемые признаки

Для классификации дефектов на внутренние и внешние были использованы следующие признаки:

Схема	Параметр	Частота	Коэффициент корреляции
Классификация	Фаза	60	0.69
	Фаза	130	0.75
	Фаза	280	0.74

Выбор признаков производился на основании коэффициентов корреляции. Вышеперечисленные признаки имеют максимальные коэффициенты корреляции с определяемым параметром (классом).

Структура нейронной сети

Нейронная сеть для классификации дефектов имеет следующую структуру:

Номер слоя	Роль слоя	Количество нейронов	Функция активации
1	Входной	3	Сигмоидальная
2	Скрытый	8	Сигмоидальная
3	Выходной	2	Сигмоидальная

Количество нейронов во входном слое обусловлено количеством признаков, в выходном – соответствует схеме кодирования выходных данных. Диапазон количества нейронов выбирается в соответствии с критерием приведенном в приложении, точное значение выбирается путем экспериментов.

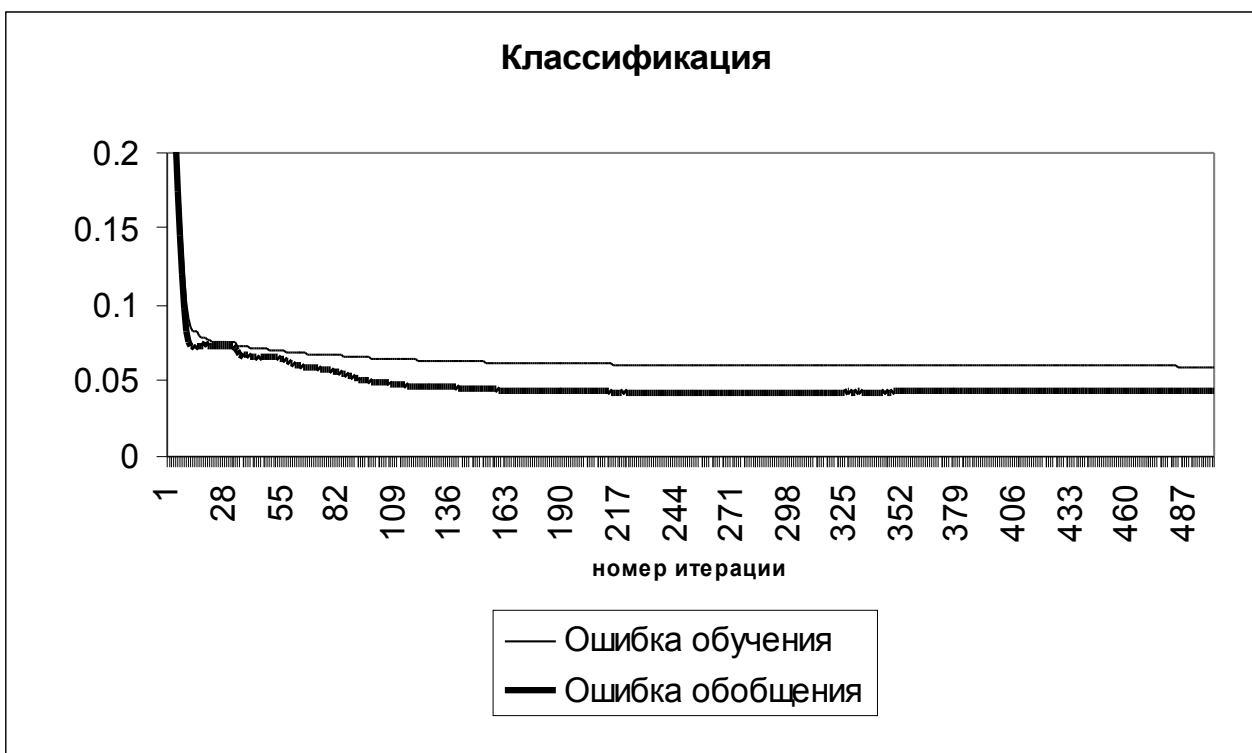
Обучение

При обучении нейронной сети для решения задачи параметризации база модельных дефектов была разделена на две части:

- обучающий набор;
- проверочный набор.

Вектора из первого набора, используются для обучения, вектора из второго для контролирования процесса обучения и оценки его качества.

Ошибка обучения характеризует способность нейронной сети обрабатывать данные, участвовавшие в ее обучении. Ошибка обобщения, вычисляемая для проверочного набора, характеризует способность нейронной сети к обработке данных, не участвовавших в ее обучении.



Как видно из приведенного графика в процессе обучения уменьшаются и ошибка обучения, и ошибка обобщения. Это свидетельствует о том, что улучшается способность нейронной сети обрабатывать как данные участвовавшие, так и данные не участвовавшие в обучении.

Результаты тестирования

Тестирование проводилось путем последовательной подачи на вход нейронной сети всех векторов из базы модельных дефектов: как участвовавших, так и не участвовавших в обучении. Результатом явилась не правильная классификация лишь 9.8% векторов. Неправильно классифицированные вектора соответствуют сквозным дефектам, которые могут быть отнесены как ко внешним так и ко внутренним.

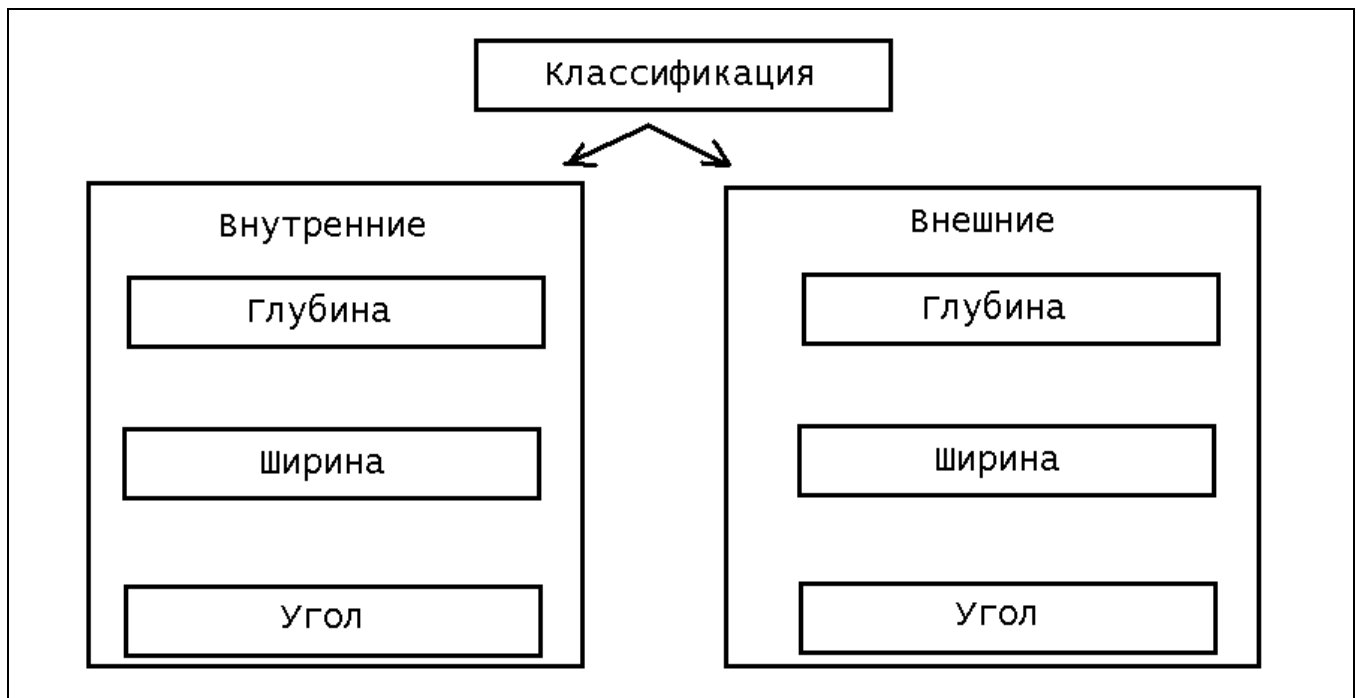
Параметризация

Для оценки степени опасности дефекта необходимо определить его глубину.

Раздельная схема параметризации

В данной работе используется раздельная схема параметризации:

- каждый из параметров определяется отдельной нейронной сетью
- для определения параметра дефектов каждого из классов используется своя нейронная сеть
- для каждой из нейронных сетей подбирается индивидуальный набор признаков.



Определение глубины

Используемые признаки

Схема	Класс дефекта	Параметр	Частота	Коэффициент корреляции
Определение глубины	Внутренний	Фаза	130	0.68
		Фаза	280	0.61
		W07	130	0.39
	Внешний	Фаза	60	0.82
		Фаза	130	0.95
		Фаза	280	0.96

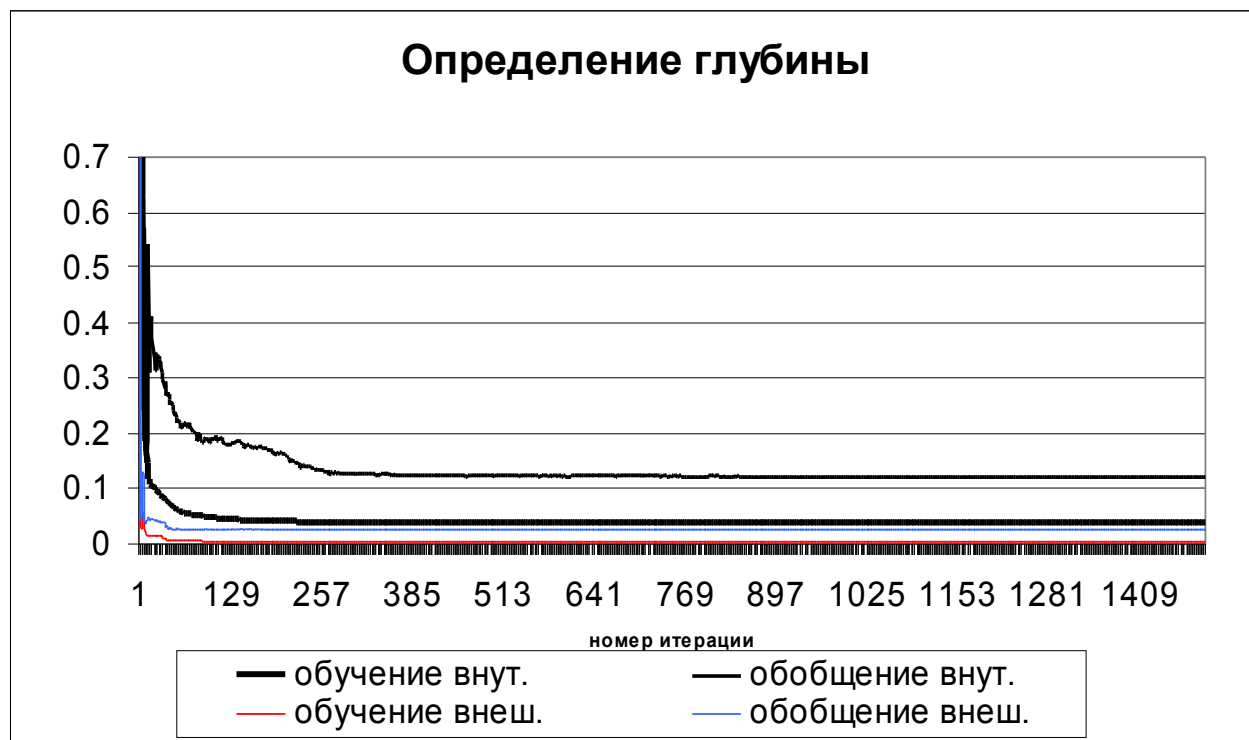
Структура сети

Номер слоя	Роль слоя	Количество нейронов	Функция активации
1	Входной	3	Линейная
2	Скрытый	3	Сигмоидальная
3	Выходной	1	Линейная

Для входного и выходного слоев используется линейная функция активации. Это позволяет на выходе нейронной сети получить непосредственное значение параметра (глубины).

Результаты обучения

Рассмотрим кривые ошибок обучения и обобщения при определении глубины дефекта.



Как видно из графика, способность нейронной сети определять глубину внутренних дефектов значительно ниже чем для внешних.

Результаты тестирования

Для оценки качества обучения нейронной сетью было произведено определение глубины модельных дефектов. Относительная погрешность составила 9.96% для внешних дефектов и 22.99% для внутренних. Это подтверждает, что полученная нейронная сеть пригодна для вычисления глубины реальных дефектов.

Определение ширины

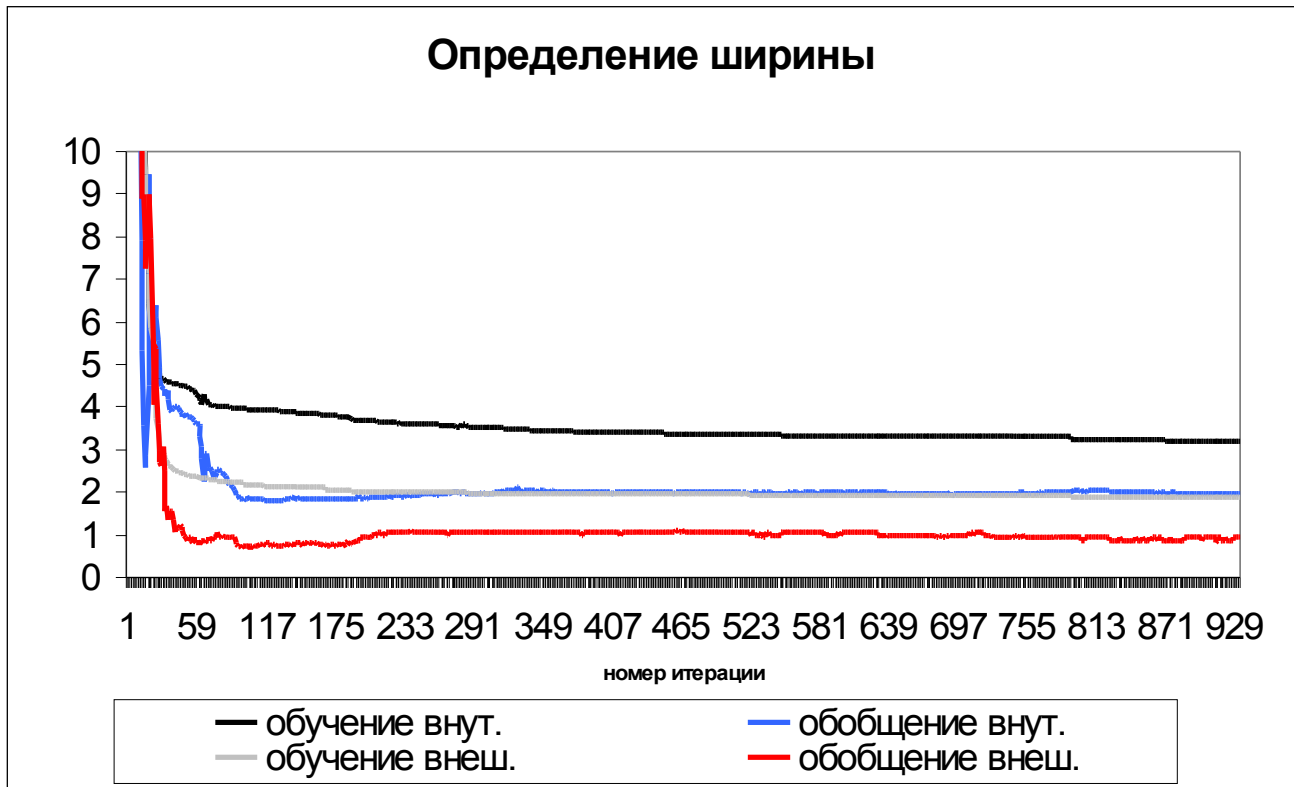
Используемые признаки

Схема	Класс дефекта	Параметр	Частота	Коэффициент корреляции
Определение ширины	внутренний	DX	280	0.93
		W07	280	0.72
		W05	280	0.74
	внешний	W05	280	0.74
		DX	130	0.93
		W07	130	0.79
		W05	130	0.80
		W05	130	0.81

Структура сети

Номер слоя	Роль слоя	Количество нейронов	Функция активации
1	Входной	4	Линейная
2	Скрытый	4	Сигмоидальная
3	Выходной	1	Линейная

Результаты обучения



Как видно из графика, способность нейронной сети определять ширину внутренних дефектов ниже чем для внешних.

Результаты тестирования

Для оценки качества обучения нейронной сетью было произведено определение ширины модельных дефектов. Относительная погрешность составила 59.55% для внешних дефектов и 92.35% для внутренних. Полученный результат показывает, что при использовании данной базы модельных дефектов и данного набора признаков определение ширины реальных дефектов может производиться только приближенно.

Определение угла

Используемые признаки

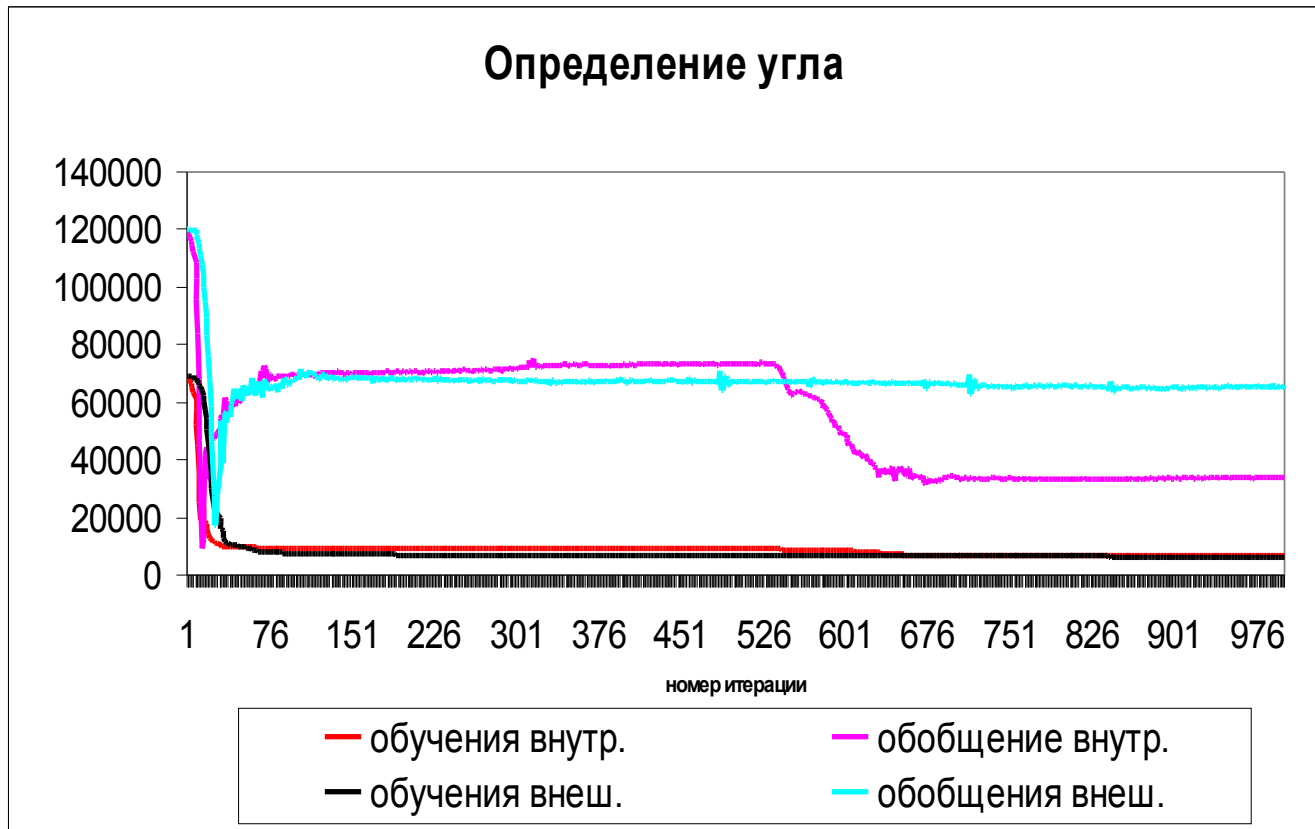
Схема	Класс дефекта	Параметр	Частота	Коэффициент корреляции
Определение ширины	Внешний	Амплитуда	60	0.51
		Реальная компонента фазы	60	0.65
		Реальная компонента фазы	130	0.58
	Внутренний	Энергия	60	0.40
		DX	130	0.29
		Амплитуда	280	0.55

Следует заметить, что коэффициенты корреляции между определяемым параметром (углом) и признаками невелики. Это может привести к низкой способности нейронной сети определять угол дефектов.

Структура сети

Номер слоя	Роль слоя	Количество нейронов	Функция активации
1	Входной	3	Линейная
2	Скрытый	3	Сигмоидальная
3	Выходной	1	Линейная

Результаты обучения



Как видно из графиков для ошибки обобщения нейронная сеть не обладает способностью определять угол для дефектов.

Результаты тестирования

Для оценки качества обучения нейронной сетью было произведено определение угла модельных дефектов. Относительная погрешность составила 400% для внешних дефектов и 500% для внутренних. Это подтверждает, что использование данной нейронной сети для определения угла дефектов невозможно.

Сравнение результатов определения класса и глубины реальных дефектов различными методами

Для решения задачи классификации реальных дефектов и определения их параметров было разработано три различных системы:

- нейронная сеть с комбинированным обучением по методу сопряженных градиентов и обратного распространения ошибки;
- нейронная сеть с обучением по методу RPROP;
- система регрессионного анализа.

Во всех случаях использовались различные наборы и сочетания признаков.

Сравнительная таблица классификации и определения глубины реальных дефектов

N°	Глубина (NN1)		Положение (NN1)	Глубина (NN2)		Положение (NN2)	Глубина (REG)		Положение (REG)
	%	mm		%	mm		%	Mm	
1	26	0.39	Внутренний	20.5	0.30693	Внутренний	18.8	0.282	Внутренний
2	17.7	0.266	Внутренний	16.6	0.2493	Внутренний	22.7	0.34	Внутренний
3	62.8	0.941	Внутренний	75.0	1.12513	Внутренний	47.3	0.709	Внутренний
4	27.8	0.418	Внешний	30.5	0.45718	Внешний	0	0	Внешний
5	98.8	1.482	Внешний	94.7	1.42124	Внутренний	100	1.5	Внутренний
6	25.3	0.379	Внутренний	17.7	0.26522	Внутренний	38	0.57	Внутренний
7	61	0.915	Внутренний	20.3	0.3039	Внутренний	59.6	0.893	Внутренний
8	52.1	0.782	Внутренний	63.6	0.95333	Внутренний	41.5	0.622	Внутренний

Как следует из таблицы все три системы показывают схожие результаты классификации дефектов и определения их глубины.

Приложение

Кодирование входных данных

Входными данными являются значения признаков дефектов. Это числа, в общем случае, как целые, так и действительные.

Теоретически на скрытые слои можно подавать непосредственно необработанные значения признаков. Однако на практике это делать нежелательно.

Согласно обобщенному дельта правилу (generalized delta rule), изменения весов скрытых слоев прямо пропорциональны выходам нейронов входного слоя.

Если выход входного нейрона равен нулю или близок к нулю, то вес, связывающий его с нейроном в следующем слое, не изменяется или меняется слабо. Это явление (low input prime factor) блокирует процесс обучения.

Если входные значения сдвинуты (абсолютное значение больше девиации), то веса нейронов первого скрытого слоя могут только вместе убывать или вместе возрастать, так как градиенты будут иметь один знак. Это означает, что значения весов будут изменяться "зигзагами". Это сильно замедлит процесс обучения, так как сдвиги в значениях на входе нейронов обуславливают преимущественные градиенты изменения весов.

Таким образом, одним из рациональных (но не оптимальным) является следующий способ нормализации входных значений:

$$x_n' = (x_n - \text{mean}_n) / (\text{dev}_n)$$

где mean - среднее арифметическое, а dev - девиация.

В результате нормализации входных значений на входы нейронов первого скрытого слоя будут подаваться числа в диапазоне от -1 до 1, не имеющие сдвига.

$$\text{mean}_n = 1/N_{mp} * \sum_i x_n^i$$

$$\text{dev}_n = \max(|\text{mean}_n - \max x_n|, |\text{mean}_n - \min x_n|)$$

Среднее арифметическое и девиация рассчитываются для каждого признака в отдельности для всего обучающего набора, а их значения хранятся в векторе весов нейронов входного слоя.

Функции активации

Значение на выходе нейрона определяется как

$$y_{m,n} = F(S_{m,n})$$

где F - функция активации нейрона, $S_{m,n}$ - взвешенная сумма для нейрона n в слое m и определяется как

$$S_{m,n} = \sum_{ni} (y_{m-1,ni} * w_{m,n,ni}) - T_{m,n}$$

С помощью функции активации вычисляется значение на выходе нейрона - степень его возбуждения.

Для классификации используется сигмоидальная функция активации, принимающая значения в диапазоне [0; 1]. Значение сигмоидальной функции вычисляется по формуле:

$$y = 1 / (1 + e^{-c*x})$$

где $c > 0$ - коэффициент, характеризующий ширину линейной части по оси абсцисс. Производную можно вычислить по следующей формуле:

$$y' = c * y * (1 - y)$$

Для вычисления параметров дефектов также используется линейная функция активации

$$y = k * x$$

ее производная

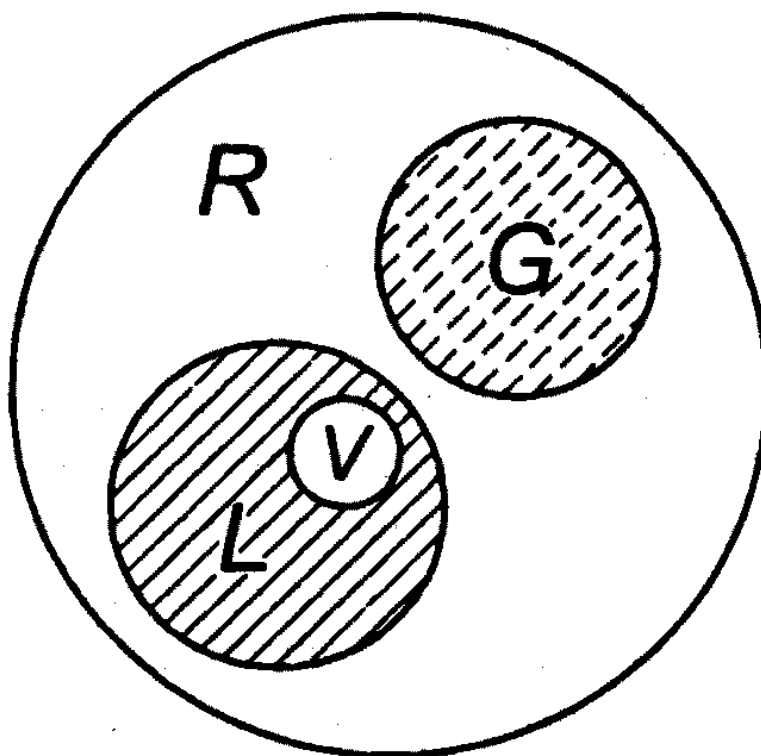
$$y' = k$$

Выбор количества обучаемых параметров сети

Для решения какой-либо задачи с применением искусственной нейронной сети следует, прежде всего, спроектировать структуру сети, адекватную поставленной задаче. Это предполагает выбор количества слоев сети и нейронов в каждом слое, а также определение необходимых связей между слоями.

Подбор количества нейронов во входном слое обусловлен размерностью входного вектора x . Подобная ситуация и с выходным слоем, в котором количество нейронов принимается равным размерности ожидаемого вектора d . Серьезной проблемой остается подбор количества скрытых (внутренних) слоев и числа нейронов в каждом из них. Теоретическое решение этой задачи в смысле условия достаточности было предложено математиками, занимающимися аппроксимацией функции нескольких переменных. Следует отметить, что ИНС выступает в роли универсального аппроксиматора обучающих данных (x, d) . В процессе обучения подбираются его функциональные коэффициенты (векторы весов отдельных нейронов). На этапе функционирования при зафиксированных значениях весов производится простой расчет значения аппроксимирующей функции при заданном входном векторе.

Одно из важнейших свойств нейронной сети - это способность к обобщению полученных знаний. Сеть, натренированная на некотором множестве обучающих выборок, генерирует ожидаемые результаты при подаче на ее вход данных, относящихся к тому же множеству, но не участвовавших непосредственно в процессе обучения



Множество данных, на котором считается истинным некоторое правило R , разбито на подмножества L и G , при этом в составе L , в свою очередь, можно выделить определенное подмножество контрольных данных V , используемых для верификации степени обучения сети. Обучение проводится на данных, составляющих подмножество L . Способность отображения сетью элементов L может считаться показателем степени накопления обучающих данных, тогда как способность распознавания данных, входящих во множество G и не использованных для обучения, характеризует ее возможности обобщения (генерализации) знаний. Данные, входящие и в L , и в G , должны быть типичными элементами множества R . В обучающем подмножестве не должно быть уникальных данных, свойства которых отличаются от ожидаемых (типичных) значений.

Феномен обобщения возникает вследствие большого количества комбинаций входных данных, которые могут кодироваться в сети с N входами. Если в качестве простого примера рассмотреть однослойную сеть с одним выходным нейроном, то для нее может быть составлено 2^N входных выборок. Каждой выборке может соответствовать единичное или нулевое состояние выходного нейрона. Таким образом, общее количество различаемых сигналов составит 2^N . Если для обучения сети используются p из общего числа 2^N входных выборок, то оставшиеся незадействованными $(2^N - p)$ допустимых комбинаций характеризуют потенциально возможный уровень обобщения знаний.

Подбор весов сети в процессе обучения имеет целью найти такую комбинацию

их значений, которая наилучшим образом воспроизводила бы последовательность ожидаемых обучающих пар (x, d) . При этом наблюдается тесная связь между количеством весов сети (числом степеней свободы) и количеством обучающих выборок. Если бы целью обучения было только запоминание обучающих выборок, их количество могло быть равным числу весов. В таком случае каждый вес соответствовал бы единственной обучающей паре. К сожалению, такая сеть не будет обладать свойством обобщения и сможет только восстанавливать данные. Для обретения способности обобщать информацию сеть должна тренироваться на избыточном множестве данных, поскольку тогда веса будут адаптироваться не к уникальным выборкам, а к их статистически усредненным совокупностям. Следовательно, для усиления способности к обобщению необходимо не только оптимизировать структуру сети в направлении ее минимизации, но и оперировать достаточно большим объемом обучающих данных.

Со статистической точки зрения погрешность обобщения зависит от уровня погрешности обучения E_L и от доверительного интервала ϵ . Она характеризуется отношением

$$E_G \Rightarrow E_L + \epsilon(p/h, E_L)$$

Значение ϵ функционально зависит от уровня погрешности обучения E_L и от отношения количества обучающих выборок p к фактическому значению h параметра, называемого мерой Вапника-Червоненкиса и обозначаемого $VCdim$. Мера $VCdim$ отражает уровень сложности нейронной сети и тесно связана с количеством содержащихся в ней весов. Значение ϵ уменьшается по мере возрастания отношения количества обучающих выборок к уровню сложности сети.

По этой причине обязательным условием выработки хороших способностей к обобщению считается грамотное определение меры Вапника-Червоненкиса для сети заданной структуры. Метод точного определения этой меры не известен, о нем можно лишь сказать, что ее значение функционально зависит от количества синоптических весов, связывающих нейроны между собой. Чем больше количество различных весов, тем больше сложность сети и соответственно значение меры $VCdim$. Часто верхнюю и нижнюю границы этой меры определяют в виде

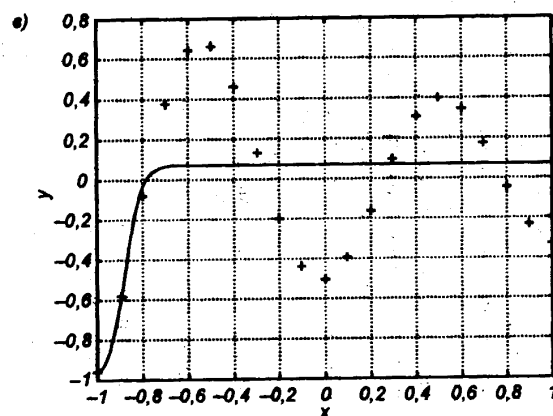
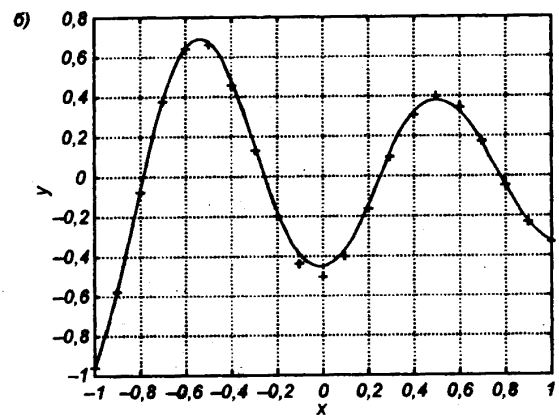
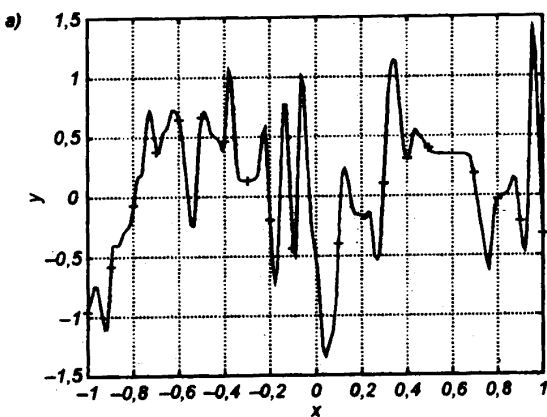
$$2[K/2]N \leq VCdim \leq 2N_w(1 + \lg N_n)$$

где $[]$ обозначена целая часть числа, N - размерность входного вектора, K - количество нейронов скрытого слоя, N_w - общее количество весов сети, а N_n общее количество нейронов сети.

Из вышеприведенного выражения следует, что нижняя граница диапазона приблизительно равна количеству весов, связывающих входной и скрытый слой, тогда как верхняя граница превышает двукратное суммарное количество всех весов сети. В связи с невозможностью точного определения меры $VCdim$ в качестве ее приближенного значения используется общее количество весов

нейронной сети.

Таким образом, на погрешность обобщения оказывает влияние отношение количества обучающих выборок к количеству весов сети. Небольшой объем обучающего подмножества при фиксированном количестве весов вызывает хорошую адаптацию сети к его элементам, однако не усиливает способности к обобщению, так как в процессе обучения наблюдается относительное превышение числа подбираемых параметров (весов) над количеством пар фактических и ожидаемых выходных сигналов сети. Эти параметры адаптируются с чрезмерной (а вследствие превышения числа параметров над объемом обучающего множества - и неконтролируемой) точностью к значениям конкретных выборок, а не к диапазонам, которые эти выборки должны представлять. Фактически задача аппроксимации подменяется в этом случае задачей приближенной интерполяции. В результате всякого рода нерегулярности обучающих данных и измерительные шумы могут восприниматься как существенные свойства процесса. Функция, воспроизводимая в точках обучения, будет хорошо восстанавливаться только при соответствующих этим точкам значениях. Даже минимальное отклонение от этих точек вызовет значительное увеличение погрешности, что будет восприниматься как ошибочное обобщение. По результатам разнообразных численных экспериментов установлено, что высокие показатели обобщения достигаются в случае, когда количество обучающих выборок в несколько раз превышает меру $VCdim$.



На рисунке представлена графическая иллюстрация эффекта гиперразмерности сети (слишком большого количества нейронов и весов). Аппроксимирующая сеть, скрытый слой которой состоит из 80 нейронов, на основе интерполяции в 21-й точке адаптировала свои выходные сигналы с нулевой погрешностью обучения. Минимизация этой погрешности на слишком малом (относительно количества весов) количестве обучающих выборок спровоцировала случайный характер значений многих весов, что при переходе от обучающих выборок к тестовым стало причиной значительных отклонений фактических значений y от ожидаемых значений d . Уменьшение количества скрытых нейронов до 5 при неизменном объеме обучающего множества позволило обеспечить и малую погрешность обучения, и высокий уровень обобщения. Дальнейшее уменьшение количества скрытых нейронов может привести к потере сетью способности восстанавливать обучающие данные, т.е. к слишком большой погрешности обучения E_L . Подобная ситуация иллюстрируется на нижнем графике рис. 11, где задействован только один скрытый нейрон. Сеть оказалась не в состоянии корректно воспроизвести обучающие данные, поскольку количество ее степеней свободы слишком мало по сравнению с необходимым для такого воспроизведения. Очевидно, что в этом случае невозможно достичь требуемого уровня обобщения, поскольку он явно зависит от погрешности обучения E_L . На практике подбор количества скрытых нейронов (и связанный с ним подбор количества весов) может, в частности, выполняться путем тренинга нескольких сетей с последующим выбором той из них, которая содержит наименьшее количество скрытых нейронов при допустимой погрешности обучения.

Решение по выбору окончательной схемы сети может быть принято только после полноценного обучения (с уменьшением погрешности до уровня, признаваемого удовлетворительным) различных вариантов ее структуры. Однако нет никакой уверенности в том, что этот выбор будет оптимальным, поскольку тренируемые сети могут отличаться различной чувствительностью к подбору начальных значений весов и параметров обучения.

Стратегия обучения

Стратегия обучения определяет общий способ обучения нейронной сети. В рассматриваемой схеме используется групповое обучение.

При групповом обучении (batch training) параметры нейронной сети изменяются после того, как на нее будет подан весь набор векторов обучающей выборки. Это можно представить следующим алгоритмом:

- 1) Для всего набора векторов обучающей выборки
 - а) Подать на сеть один вектор из обучающей выборки и вычислить градиент $\partial E / \partial W$

б) Аккумулировать $\partial E / \partial W$

2) Изменить параметры сети в соответствии с градиентом $\partial E / \partial W$

Если минимизируется ошибка сети

$$E = 1 / N_{mp} * \sum_k E_k$$

где N_{mp} - число векторов в обучающей выборке, E_k - ошибка сети для k -го вектора, то аккумулировать $\partial E / \partial W$ следует следующим образом

$$\partial E / \partial W = 1 / N_{mp} * \sum_k \partial E / \partial W_k$$

В качестве E_k может использоваться среднеквадратическая ошибка

$$E_k = 1 / N_{n_{Nm-1}} * \sum_n (y_{Nm-1,n} - out_{k,n})$$

где $N_{n_{Nm-1}}$ - число нейронов в выходном слое, $y_{Nm-1,n}$ - выход n -го нейрона выходного слоя, $out_{k,n}$ - ожидаемое значение n -го выходного элемента для k -го вектора обучающей выборки.

Достоинства используемой стратегии:

- 1) гарантированная сходимость к локальному минимуму
- 2) большое количество приемов улучшения сходимости

Недостатки:

низкая скорость обучения.

Алгоритм обучения

Для изменения параметров сети в процессе обучения применяется алгоритм RPROP.

Resilient error backpropagation - алгоритм, предложенный в 1992 году (последняя модификация в 1994). Это прямой адаптивный алгоритм (он адаптирует изменение весов на основе градиента ошибки), не зависящий от величины производной функции активации. Таким образом, в этом алгоритме вычисляется адаптивное изменение для каждого параметра:

$$\Delta_{ij}(t) = \begin{cases} \Delta_{ij}(t-1) * \eta^+, & \text{если } \partial E / \partial w_{ij}(t-1) * \partial E / \partial w_{ij}(t) > 0, \\ \Delta_{ij}(t-1) * \eta^-, & \text{если } \partial E / \partial w_{ij}(t-1) * \partial E / \partial w_{ij}(t) < 0, \\ \Delta_{ij}(t-1), & \text{иначе} \end{cases}$$

где $0 < \eta^- < 1 < \eta^+$.

Алгоритм RPROP можно описать следующим образом:

1) для каждого слоя сети, начиная с первого скрытого

а) для каждого нейрона слоя

i) Если $\partial E / \partial w_{m,n,ni}^t * \partial E / \partial w_{m,n,ni}^{t-1} > 0$

$$\Delta \Delta w_{m,n,ni} = \min(\Delta \Delta_+ * \Delta \Delta w_{m,n,ni}, \Delta \Delta_{\max})$$

$$\Delta w_{m,n,ni} = -\text{SGN}(\partial E / \partial w_{m,n,ni}) * \Delta \Delta w_{m,n,ni}$$

$$w_{m,n,ni} = w_{m,n,ni} + \Delta w_{m,n,ni}$$

$$\partial E / \partial w_{m,n,ni}^{t-1} = \partial E / \partial w_{m,n,ni}^t$$

ii) Если $\partial E / \partial w_{m,n,ni}^t * \partial E / \partial w_{m,n,ni}^{t-1} < 0$

$$\Delta\Delta w_{m,n,ni} = \max(\Delta\Delta_{-} * \Delta\Delta w_{m,n,ni}, \Delta\Delta_{\min})$$

$$\partial E / \partial w_{m,n,ni}^{t-1} = 0$$

iii) Если $\partial E / \partial w_{m,n,ni}^t * \partial E / \partial w_{m,n,ni}^{t-1} = 0$

$$\Delta w_{m,n,ni} = -\text{SGN}(\partial E / \partial w_{m,n,ni}) * \Delta\Delta w_{m,n,ni}$$

$$w_{m,n,ni} = w_{m,n,ni} + \Delta w_{m,n,ni}$$

$$\partial E / \partial w_{m,n,ni}^{t-1} = \partial E / \partial w_{m,n,ni}^t$$

iv) Аналогично для порога

Таким образом, если алгоритм проскакивает локальный минимум, то изменение параметра не происходит, а шаг изменения уменьшается, если знак производной ошибки не изменяется, то происходит увеличение шага изменения параметра.

$\Delta\Delta_{\max}$ задает максимальное значение изменения параметров.

$\Delta\Delta_{\min}$ задает минимальное значение изменения параметров.

$\Delta\Delta_{+}$ задает скорость увеличения изменения параметров.

$\Delta\Delta_{-}$ задает скорость уменьшения изменения параметров.

Авторы алгоритма рекомендуют следующие значения для приведенных выше величин:

$$\Delta\Delta_{\max} = 50$$

$$\Delta\Delta_{\min} = 10^{-6}$$

$$\Delta\Delta_{+} = 1,2$$

$$\Delta\Delta_{-} = 0,5$$

Остановка обучения

В работе используется критерий остановки, основанный на потере обобщения (generalization loss).

Введем функцию минимальной ошибки в проверочном наборе (lowest validation set error)

$$E_{opt}^t = \min_{t' < t} E v^{t'}$$

здесь t - номер итерации обучения. Эта функция возвращает минимальную ошибку обобщения для первых $t-1$ итераций.

Теперь введем функцию потери обобщения (generalization loss)

$$GL^t = 100 * (E_v^t / E_{opt}^t - 1)$$

Данная функция возвращает приращение ошибки обобщения относительно достигнутого минимума.

Критерием остановки процесса обучения является следующий:

$$GL_{\alpha}: GL^t > \alpha$$

Этот алгоритм можно представить следующим образом:

1) На каждой итерации вычислить потерю обобщения

$$GL = 100 * (E_v^t / E_{opt} - 1)$$

2) Вычислить E_{opt}

$$E_{opt} = \min(E_{opt}, E_v)$$

3) Принять решение об остановке процесса обучения
остановить, если $GL > \alpha$

Библиографический список

1. Yann Le Cun, Efficient BackProp. - Holmdel, 1996, 75 с.
2. Donald R. Tvetter, Backpropagator's Review, 1996
3. Martin Riedmiller, Advanced Supervised Learning in Multi-layer Perceptrons - From Backpropagation to Adaptive Learning Algorithms - Karlsruhe, 1994, 10 с.
4. Merten Joost, Speeding up backpropagation algorithms by using cross-entropy combined with pattern normalization, 1997, 20 с.
5. Lutz Prechelt, Automatic early stopping using cross validation: quantifying the criteria. - Karlsruhe, 1997, 12 с.
6. Martin Riedmiller, RPROP - A fast adaptive learning algorithm. - Karlsruhe, 1992, 12 с.
7. Martin Riedmiller, RPROP - Description and implementation details. - Karlsruhe, 1994, 16 с.
8. W. Shiffman, Comparison of optimized backpropagation algorithms. - Koblenz, 1993, 8 с.
9. Stewe Lawrence, What size neural network gives optimal generalization? Convergence properties of backpropagation. - Oxford, 1996, 37 с.
10. B. D. Ripley, Statistical ideas for selecting network architectures. - Princeton, 1995, 8 с.
11. N.K. Treadgold, T.D. Gedeon, THE SARPROP ALGORITHM: A SIMULATED ANNEALING ENHANCEMENT TO RESILIENT BACK PROPAGATION. - Sydney, 1998, 7 с.
12. P. A. Castilo, SA-Prop: Optimization of multilayer perceptron parameters using simulated annealing. - Granada, 1998, 10 с.