

МОСКОВСКИЙ ЭНЕРГЕТИЧЕСКИЙ ИНСТИТУТ (ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ)

Институт Автоматики и Вычислительной Техники

Кафедра Электротехники и Интроскопии

ВЫПУСКНАЯ РАБОТА

Тема:	определение типа и параметров дефектов
	труб нефте- и газопроводов с использованием
	нейронной сети типа многослойный персептрон

Студент	A15-00	Мартынов Н.С.
	группа	подпись фамилия, и., о.

Научный руководитель	зав. кафедрой	доцент	Лунин В. П.
	должность	звание	подпись фамилия, и., о.

“Работа допущена к защите”

Дата _____

МОСКВА

2004 г.

Аннотация

В работе исследуется возможность и особенности применения нейронной сети типа многослойный персептрон для определения типа дефектов труб нефте- и газопроводов по вектору признаков и определения их параметров. Исследованы особенности применения совмещенной (одна нейронная сеть) и раздельной (разные нейронные сети используются для определения класса и подкласса) схем классификации. В работе даны рекомендации по выбору количества скрытых слоев и нейронов в них. Исследована возможность обучения сети на зашумленных модельных данных и данных, полученных от тестовой трубы. Оценена возможность применения специальных методов ускорения процесса обучения и улучшения способности к обобщению, таких как алгоритмы RPROP и SARPROP, а также ранняя остановка процесса обучения.

Annotation

In this work it is investigated possibility and features of use of artificial neural networks of type multilayer perceptron for determination of type of defects of oil and gas pipe line and estimation of theirs parameters by features vector. Investigated features of using combined (one neural network) and separate (different neural networks used for determination of base class and subclass) classification schemes. Recommendations given for selecting count of hidden layers and count of neurons in them. Investigated possibility of training neural network using noised model data and data obtained from test pipe. Estimated possibility of use of advanced training techniques such as RPROP and SARPROP algorithms and early stopping for better performance and generalization abilities of neural network.

ВВЕДЕНИЕ

В нашей стране, богатой природными ресурсами, существует разветвленная сеть нефте- и газопроводов. Однако эксплуатируемые сети трубопроводов в значительной мере изношены, что может приводить к масштабным катастрофам, имеющим значительные экономические и экологические последствия. Для их предотвращения необходимо проводить ремонт трубопроводов. Целесообразной является выборочная замена участков, на которых располагаются дефекты, которые могут стать причиной аварии. Первым шагом в решении этой задачи является поиск дефектов труб и определение их параметров для оценки степени опасности.

В настоящее время существует несколько методов неразрушающего контроля. Основными из них являются:

- магнитный;
- вихретоковый;
- ультразвуковой.

В данной работе рассматривается магнитный способ неразрушающего контроля. Важнейшими задачами являются не только определение наличия или отсутствия дефекта, но определение его типа и параметров. Именно эта информация необходима для принятия решения об опасности дефекта и способе его устранения.

Для решения задач дефектометрии сигналы, полученные от преобразователей дефектоскопов, подвергаются обработке специальными методами, среди которых:

- регрессионный анализ;
- генетические алгоритмы;
- обработка с помощью искусственных нейронных сетей.

Искусственные нейронные сети - технология, появившаяся еще в 50-х годах прошлого столетия, однако только разработки 89-94 годов, позволили на практике использовать их потенциал, развитие этой области становится все более интенсивным.

Исследование возможности применения искусственной нейронной сети типа "многослойный персептрон" для решения задачи классификации дефектов и определения их параметров - цель данной работы. Необходимо разработать программную систему для определения типа дефекта и его параметров по вектору признаков, при этом ошибка определения глубины дефекта не должна превышать 20% от толщины стенки трубы (8 мм), ошибка определения длины и ширины дефекта не должна превышать 20 мм.

1. Схема обработки данных

1.1. Источник входных данных

Для контроля труб большого диаметра используется внутритрубный магнитный дефектоскоп (рис. 1).

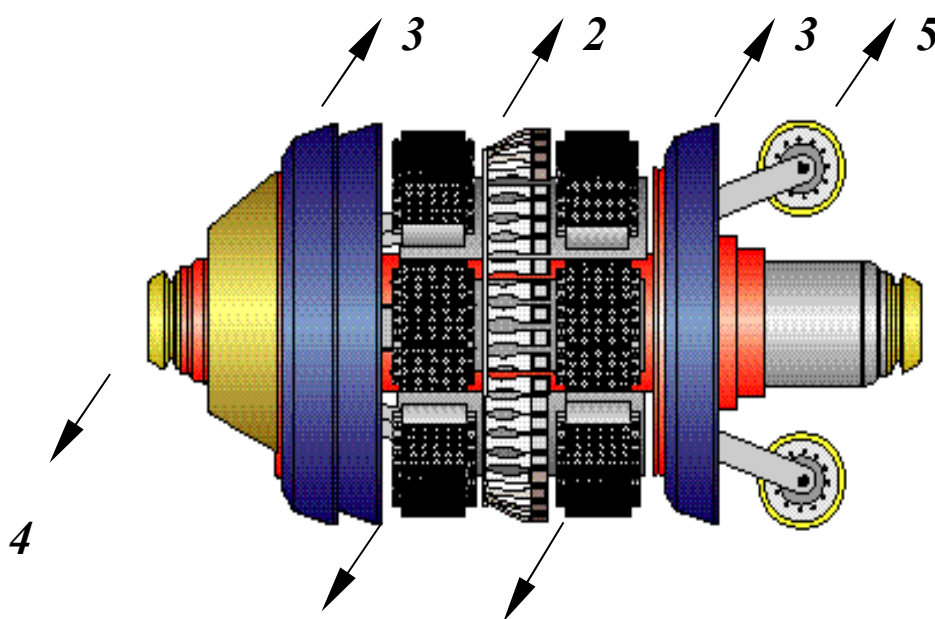


Рис. 1.1. Общий вид дефектоскопа

- 1 – щетки.
- 2 – датчики.
- 3 – центрирующие чаши.
- 4 – амортизатор.
- 5 – одометр

С помощью магнитной системы (см. рис. 1.2.), включающей постоянные магниты, магнитопровод и щетки создается магнитное поле.

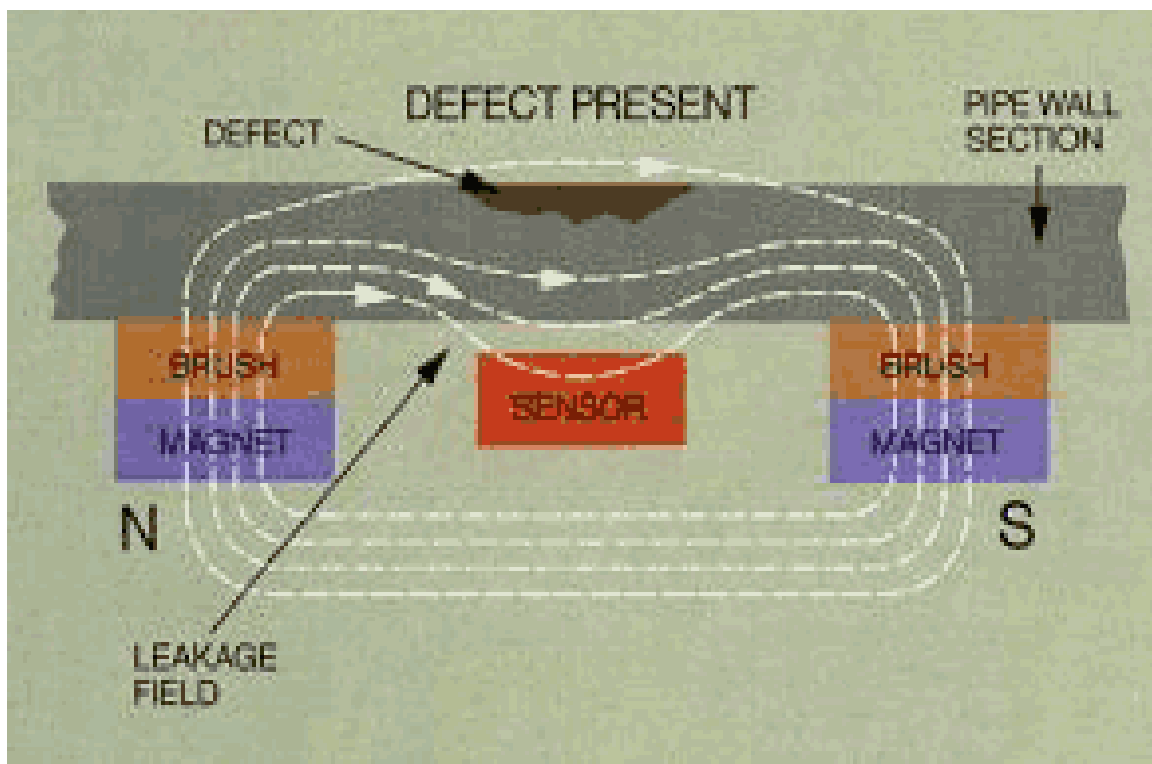


Рис. 1.2. Схема магнитной системы снаряда-дефектоскопа (магнитопровод не показан)

Поле рассеяния дефектов снимается с помощью измерительной системы, включающей датчики Холла и АЦП.

Сигнал от магнитного дефектоскопа поступает в виде двумерного массива значений для двух составляющих вектора магнитной индукции, одним "измерением" в котором, является номер датчика, а вторым - номер скана (см. рис. 1.3.). Расстояние между датчиками - 8.2 мм, расстояние между сканами 3.3 мм.

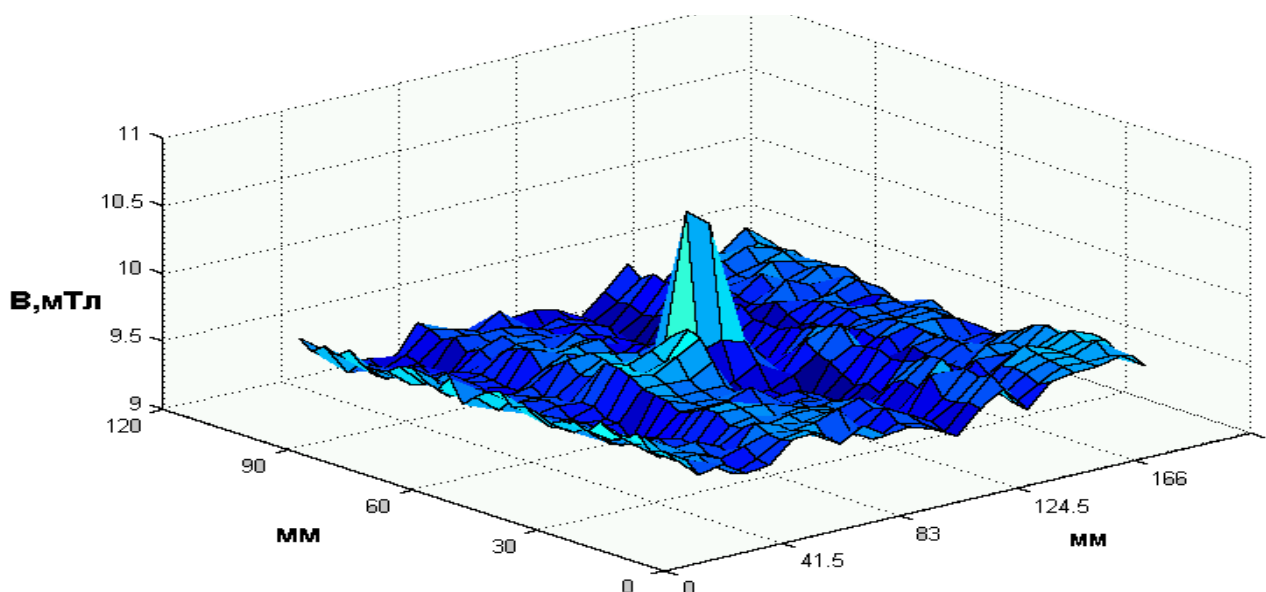


Рис. 1.3. Пример сигнала (внутренняя коррозия)

Прежде чем определять тип дефекта, необходимо сначала выявить так называемые области интереса (region of interest) или "рамки", содержащие сигнал от дефекта. Этот процесс называется кластеризацией.

Полученные "рамки" подвергаются обработке с целью вычисления так называемых признаков дефекта. В идеале значения (либо сочетания значений) признаков должны однозначно характеризовать дефект. Также их число должно быть минимально.

В данной работе для определения типа дефекта используются следующие признаки:

1. Разность максимального и минимального значения осевой составляющей поля.
2. Площадь сечения осевой составляющей поля по уровню 0,7 от максимума.
3. Количество датчиков, среди показаний которых есть значения превышающие уровень 0,7 от максимума.
4. Площадь сечения осевой составляющей поля по уровню 0,6 от максимума.
5. Количество датчиков, среди показаний которых есть значения превышающие уровень 0,6 от максимума.
6. Площадь сечения осевой составляющей поля по уровню 0,5 от максимума.
7. Количество датчиков, среди показаний которых есть значения превышающие уровень 0,5 от максимума.
8. Отношение второго к шестому элементов вектора признаков.

9. Отношение осей эллипса, полученного при сечении осевой составляющей поля по уровню 0,6 от максимума.
10. Значение максимальной гармоники при разложении показаний датчика, выдавшего максимальное показание, в ряд Фурье.
11. Номер максимальной гармоники.
12. Ширина спектра по уровню 0,5 от максимума.
13. Разность максимального и минимального значения угловой составляющей поля.

Вычисленные признаки затем обрабатываются с помощью нейронной сети, которая принимает решение о принадлежности дефекта к тому или иному типу. Затем с помощью нейронных сетей, обученных вычислению параметров дефектов данного типа, определяются геометрические размеры дефекта.

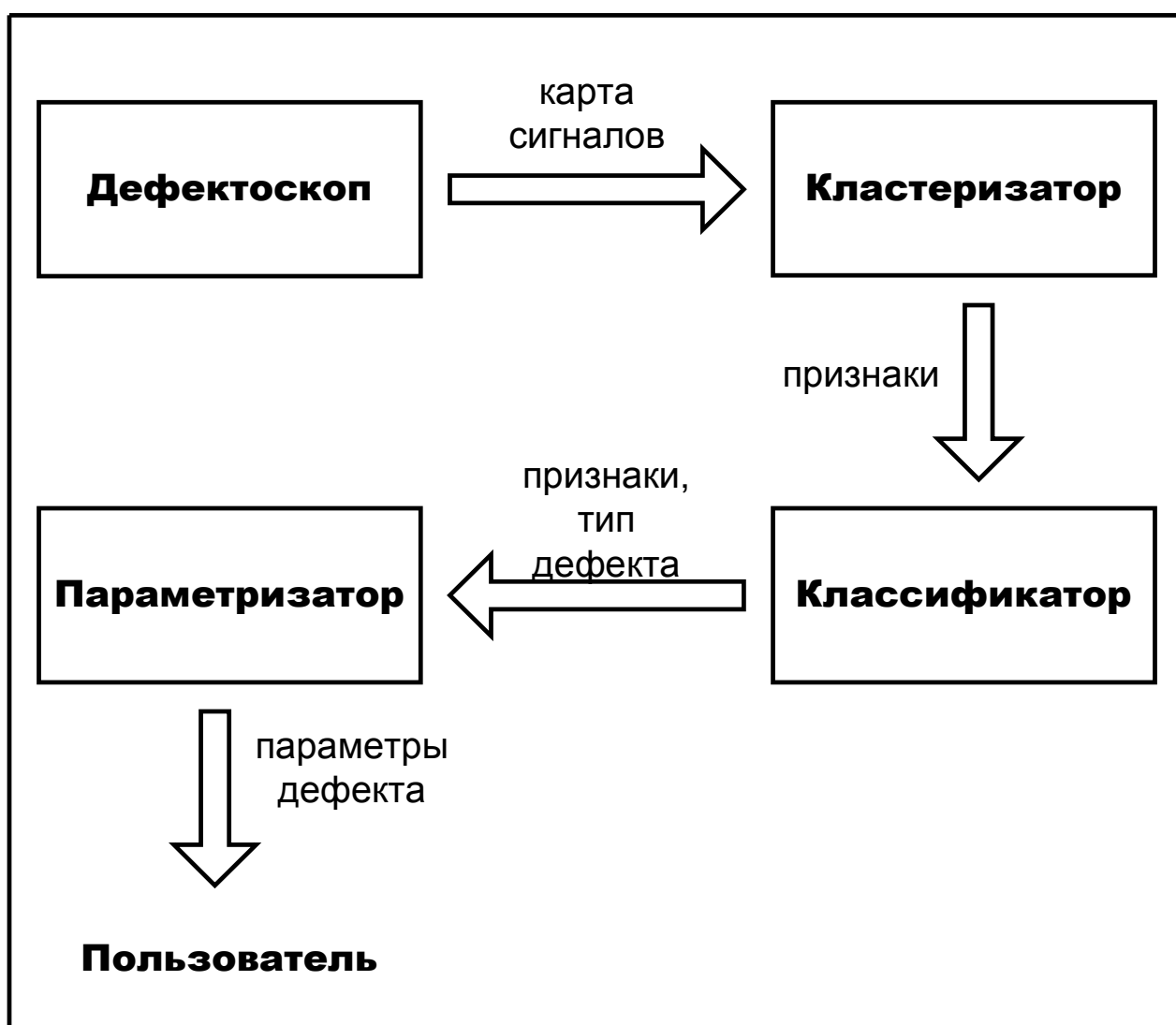


Рис. 1.4. Схема обработки данных

Нейронная сеть обучается с помощью данных полученных при математическом моделировании дефектов с различными параметрами с помощью метода конечных элементов (данные получены с помощью программного пакета Ansys). База модельных дефектов включает:

- 1550 дефектов типа внешняя коррозия (глубиной от 0.4 до 8 мм, длиной от 60 до 40 мм, шириной от 6 до 40 мм)
- 1225 дефектов типа внутренняя коррозия (глубиной от 0.4 до 8 мм, длиной от 60 до 40 мм, шириной от 6 до 40 мм)
- 3601 дефект типа внешняя трещина (глубиной от 0.4 до 8 мм, раскрытием от 0.5 до 2 мм, длиной от 40 до 120 мм)
- 3600 дефекта типа внешняя трещина (глубиной от 0.4 до 8 мм, раскрытием от 0.5 до 2 мм, длиной от 40 до 120 мм).

Полигонные данные включают:

- 80 дефектов типа внешняя коррозия (глубиной от 1 до 7.94 мм, длиной от 5.98 до 48.49 мм, шириной от 5.98 до 100 мм)
- 32 дефекта типа внутренняя коррозия (глубиной от 1.1 до 4 мм, длиной от 15.68 до 31.49 мм, шириной от 15.68 до 31.49 мм)
- 15 дефектов типа внешняя трещина (глубиной от 1.7 до 4.36 мм, длиной от 28 до 170 мм, раскрытием от 0.5 до 2.2 мм)
- 2 дефекта типа внутренняя трещина (глубиной от 1.94 до 2.43 мм, длиной от 52 до 80 мм, раскрытием от 0.5 до 1 мм)

Толщина стенки трубы изменяется от 7.8 до 10 мм.

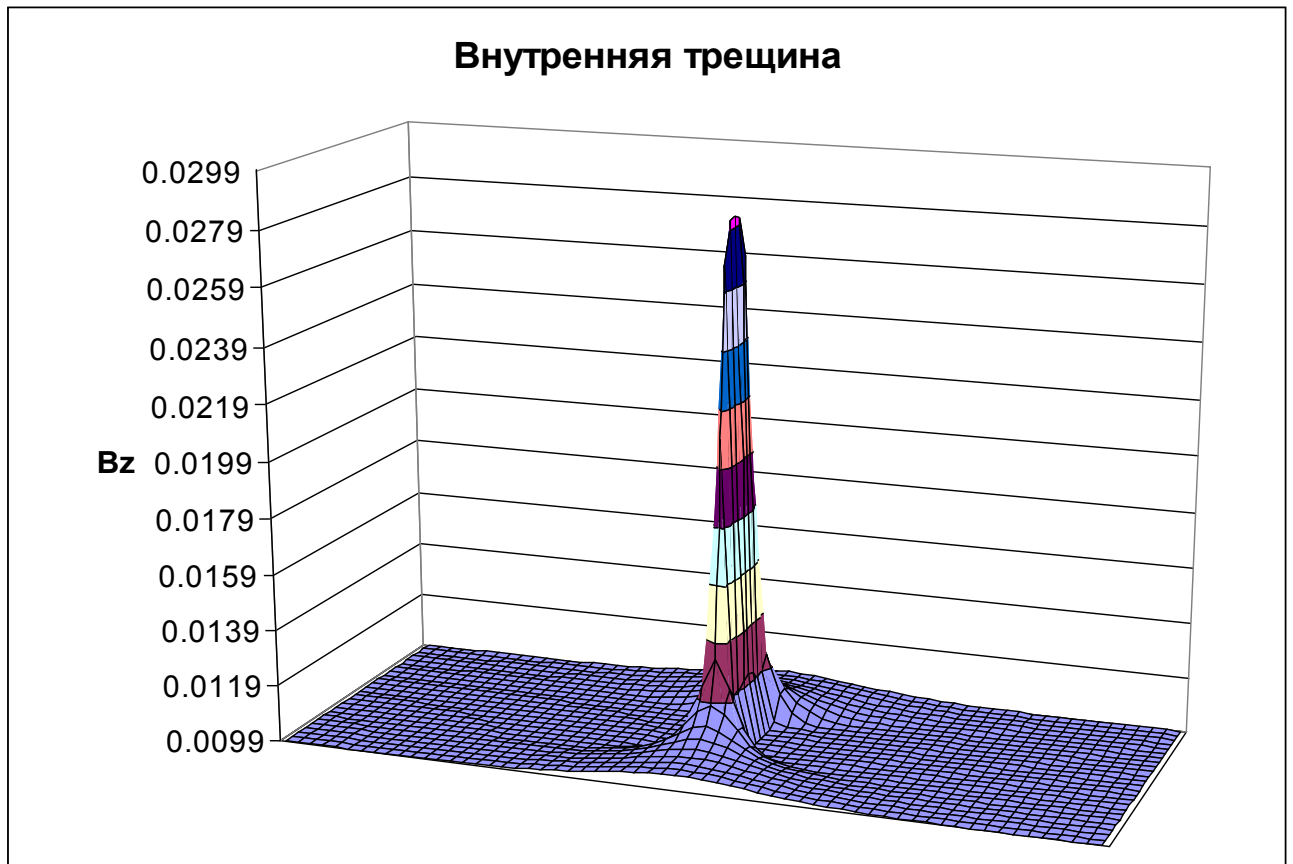


Рис. 1.5. Пример модельного сигнала для дефекта типа внутренняя трещина

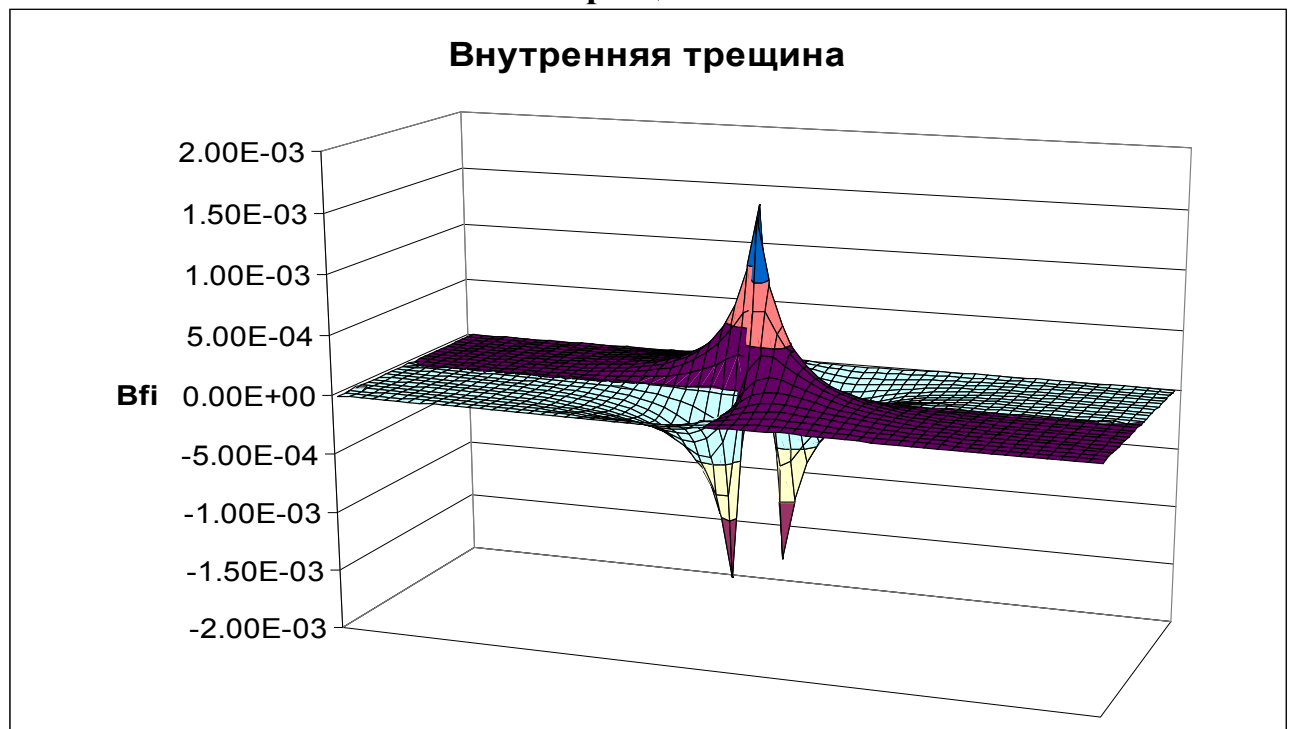


Рис. 1.6. Пример модельного сигнала для дефекта типа внутренняя трещина

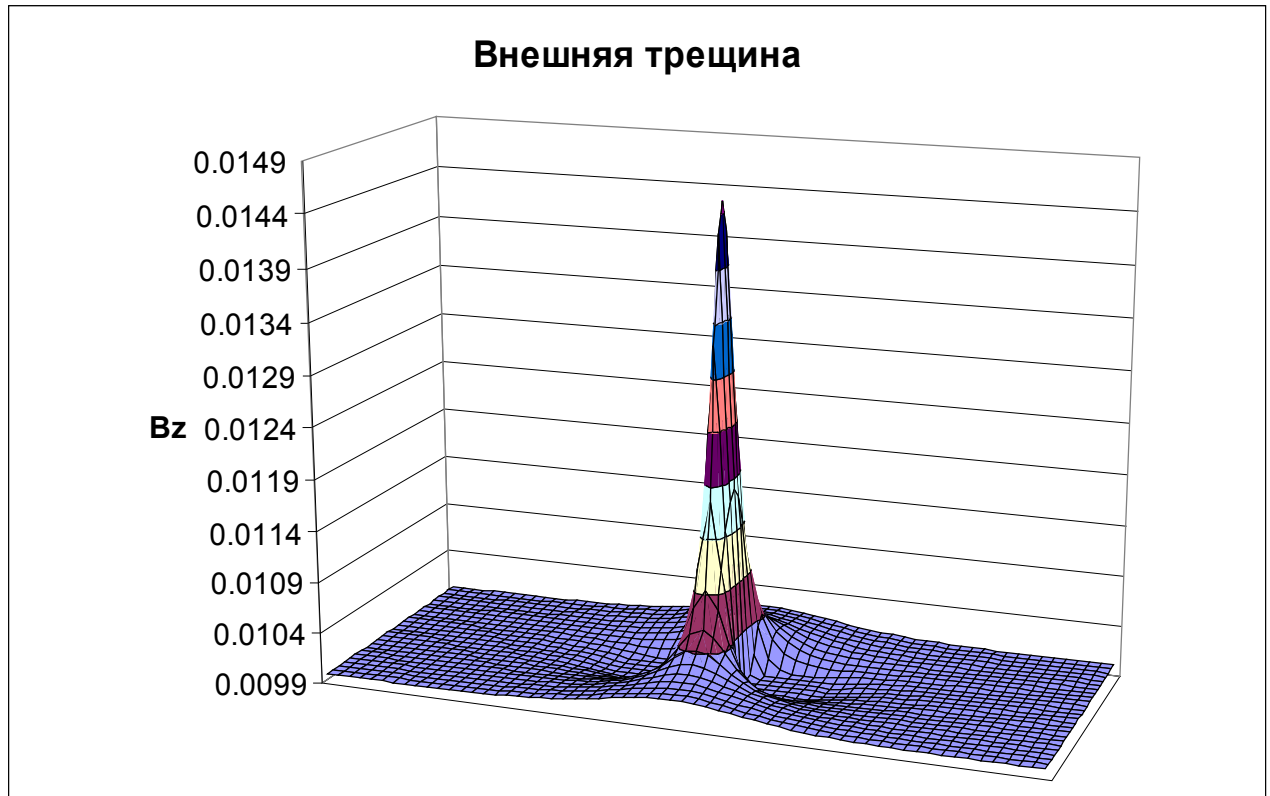


Рис. 1.7. Пример модельного сигнала для дефекта типа внешняя трещина

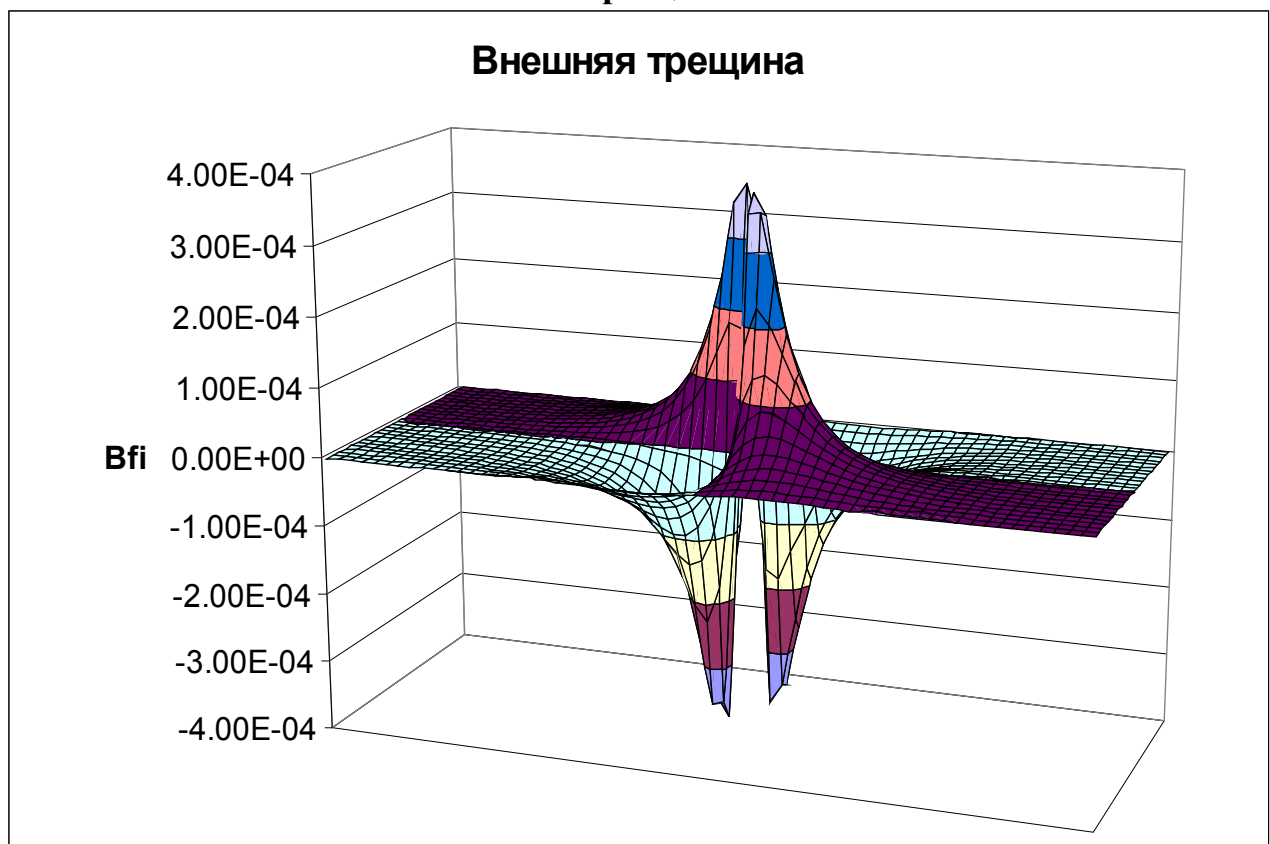


Рис. 1.8. Пример модельного сигнала для дефекта типа внешняя трещина

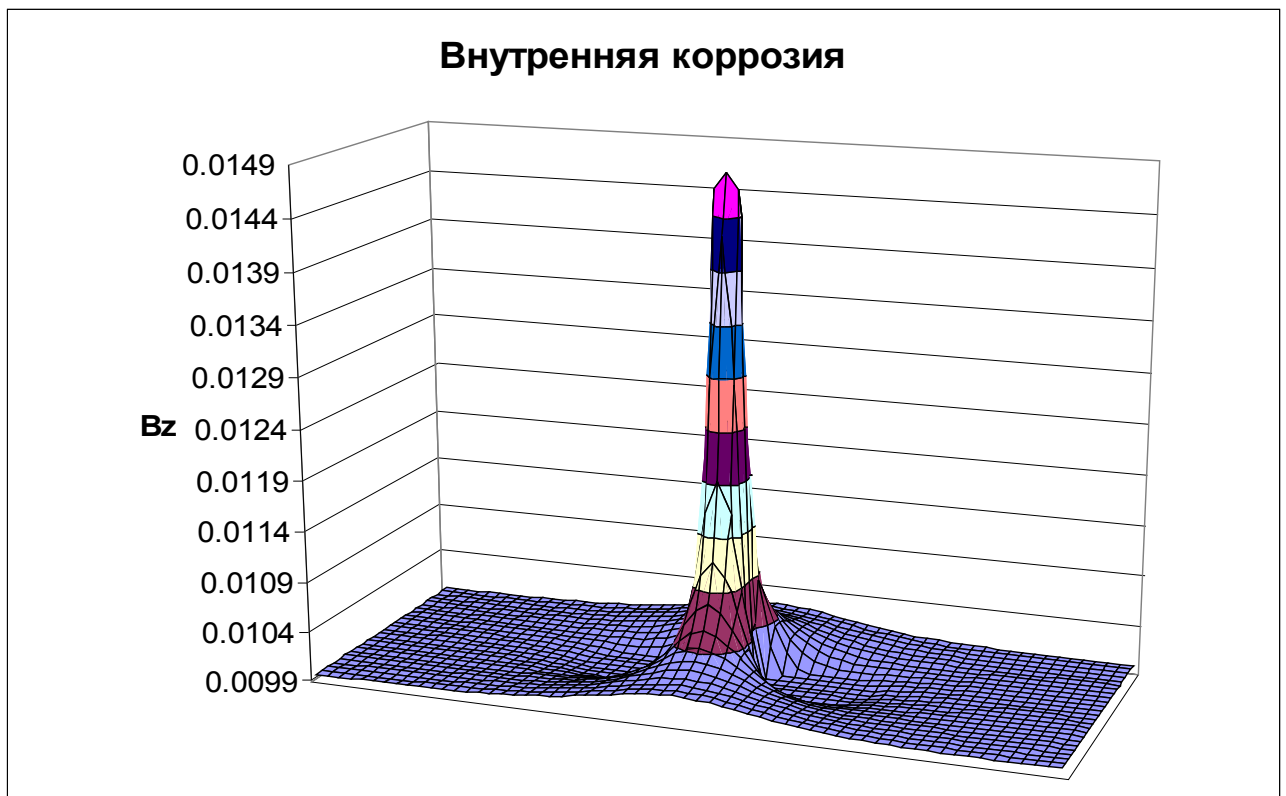


Рис. 1.9. Пример модельного сигнала для дефекта типа внутренняя коррозия

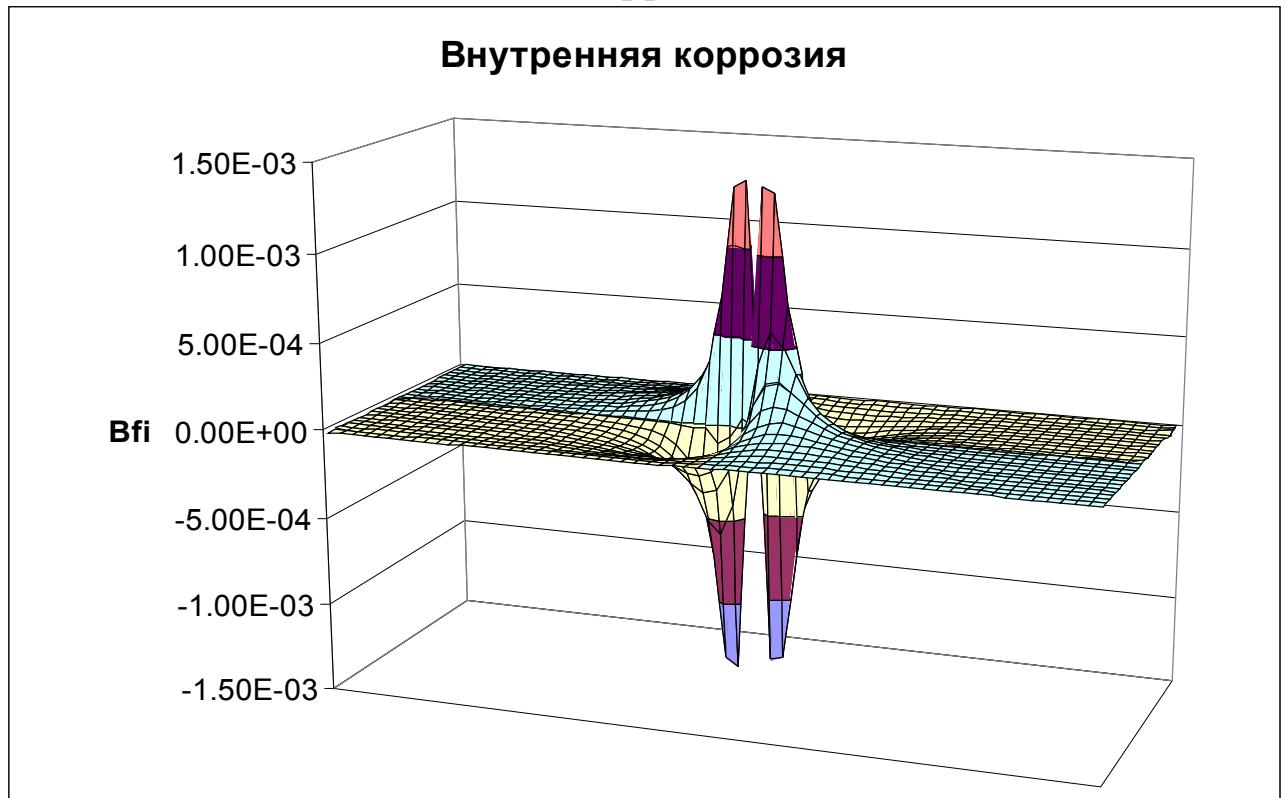


Рис. 1.10. Пример модельного сигнала для дефекта типа внутренняя коррозия

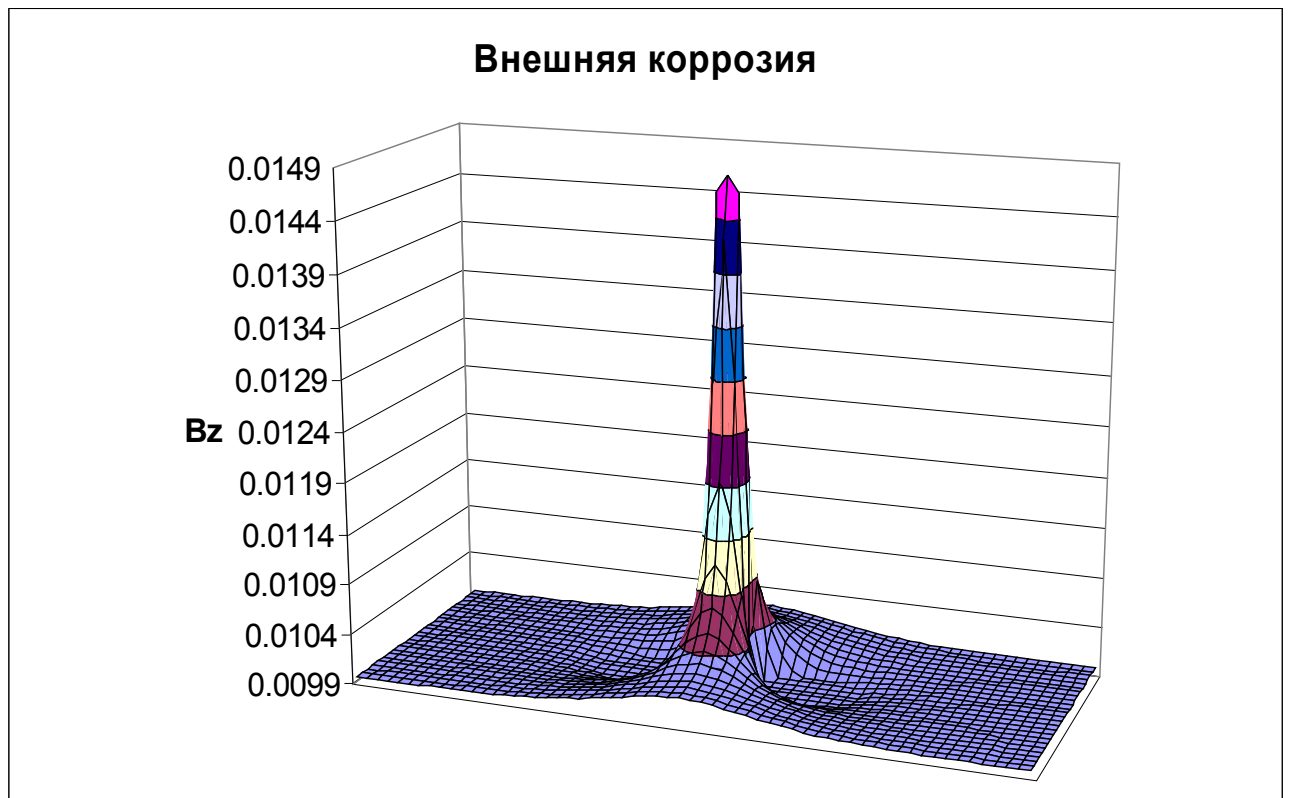


Рис. 1.11. Пример модельного сигнала для дефекта типа внешняя коррозия

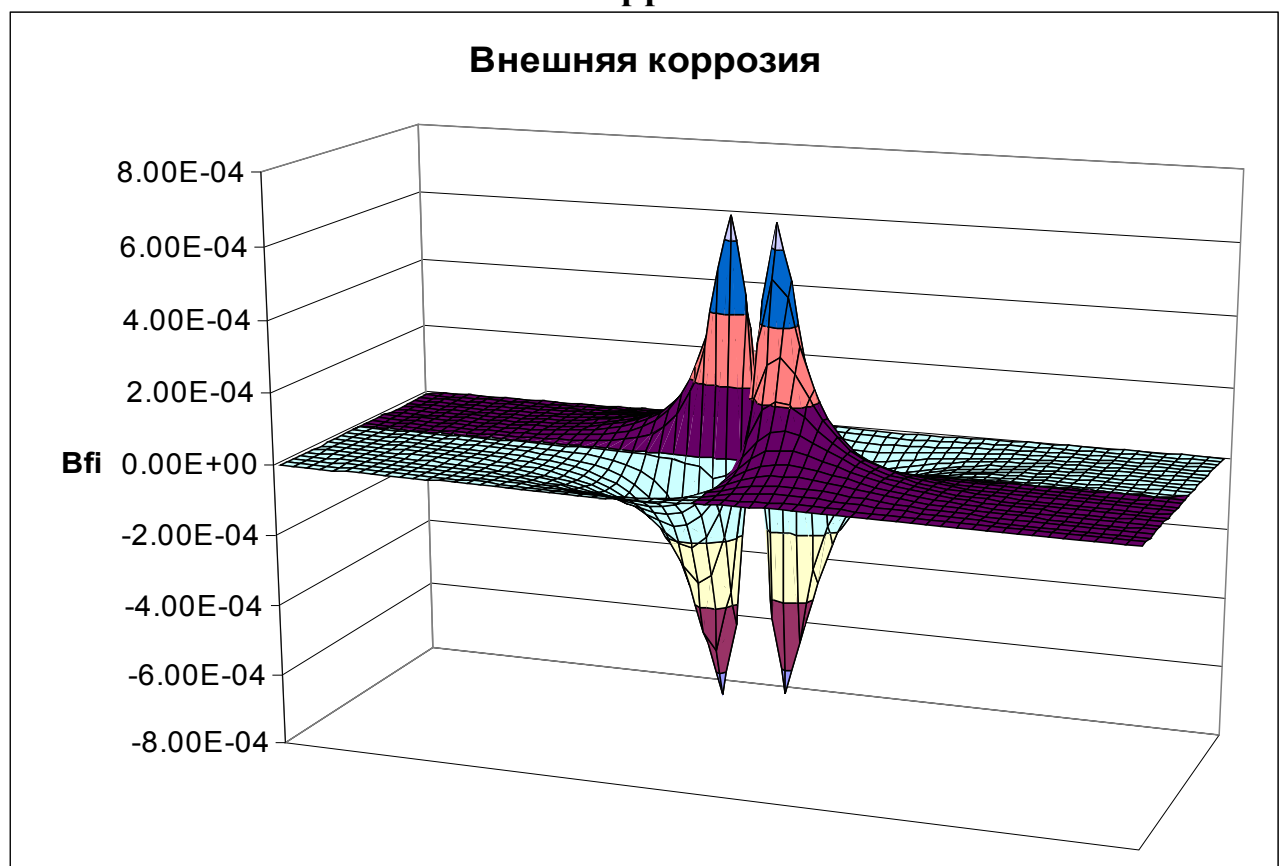


Рис. 1.12. Пример модельного сигнала для дефекта типа внешняя коррозия

Позднее вычисляются параметры обнаруженного дефекта, и производится оценка степени его опасности.

1.2. Программная система

Для исследования возможности применения различных алгоритмов и схем работы с нейронной сетью типа многослойный персептрон для решения задач классификации и параметризации была разработана программная система с архитектурой представленной на рис 1.13.

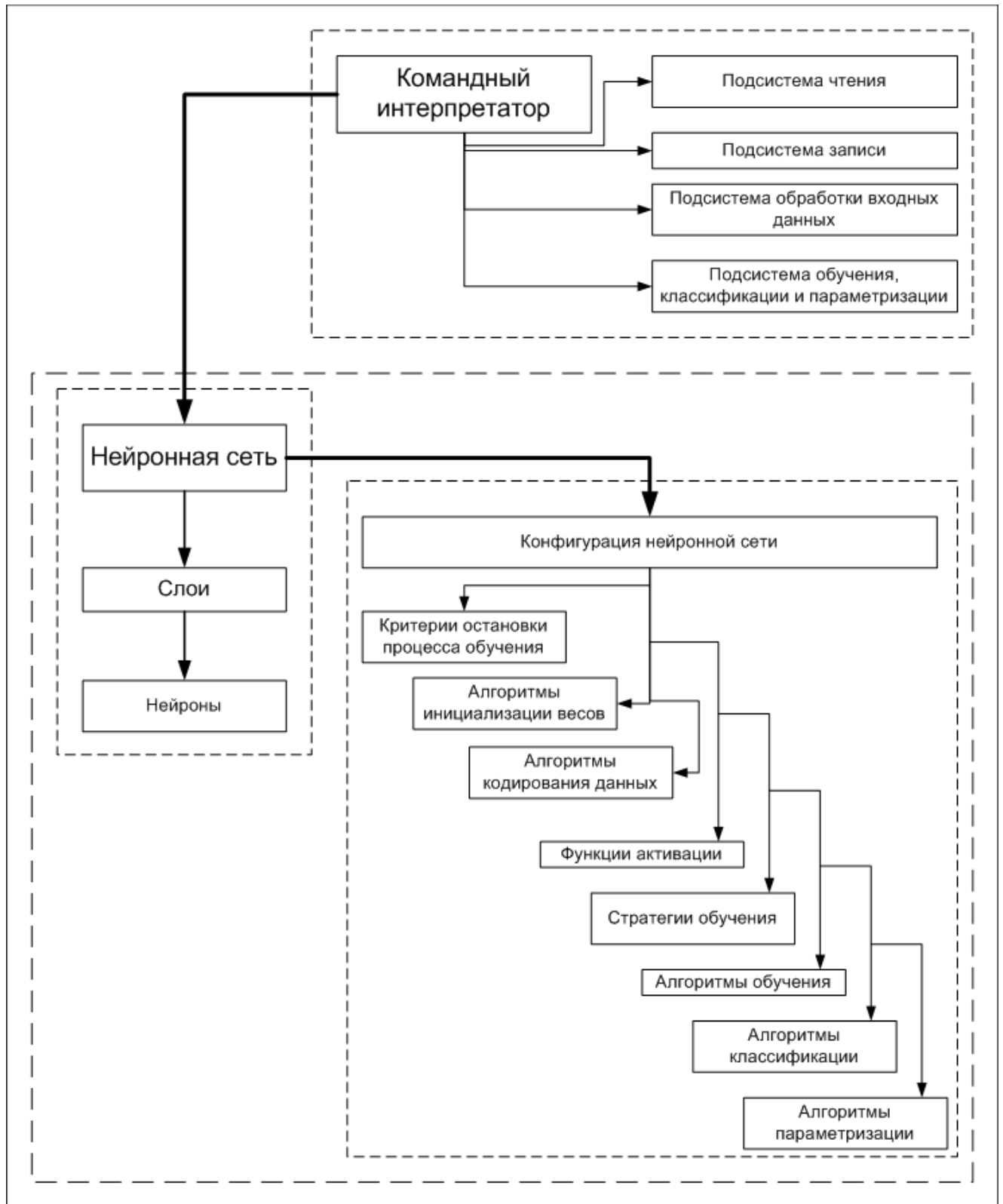


Рис. 1.13. Архитектура программной системы

Программа состоит из двух систем:

- командный интерпретатор
- система работы с нейронными сетями.

1.2.1. Командный интерпретатор

При исследовании большого количества схем работы с нейронной сетью необходимо выполнить значительное количество экспериментов. Для автоматизации этого процесса разработан командный язык для описания заданий и интерпретатор данного языка, позволяющий, подготовив инструкции по проведению одного и более экспериментов, выполнить их без интерактивного участия человека.

Данная система включает следующие части:

- подсистема чтения, позволяющая загружать исходные данные из файлов различных форматов, а также конфигурацию и состояние нейронной сети
- подсистема записи, позволяющая записывать конфигурацию и состояние нейронной сети в любой момент времени, а также обработанные исходные данные и результаты классификации и параметризации
- подсистема обработки входных данных, с помощью которой можно комбинировать наборы данных (объединение наборов, усечение набора, перемешивание набора), формируя из них новые, а также производить преобразование этих данных (добавление шума)
- подсистема обучения, классификации и параметризации, использующая подсистему работы с нейронными сетями для использования сформированных данных для обучения, классификации или параметризации.

1.2.2. Подсистема работы с нейронными сетями

Данная подсистема включает в себя:

- собственно нейронную сеть, состоящую из набора слоев нейронов (входной, скрытые, выходной), необходимого для решения данной задачи; каждый из слоев характеризуется типом нейронов и их количеством
- конфигурацию нейронной сети, которая включает:
 - алгоритмы инициализации весов в начальный момент времени (установка в 0.5; установка в случайное значение из различных диапазонов, не зависящих от количества слоев и нейронов в них; инициализация на основе размера нейронной сети)
 - алгоритмы кодирования данных (без преобразования; приведение к заданному диапазону; устранение постоянной составляющей)
 - функции активации нейронных элементов (сигмоидальная униполярная; сигмоидальная биполярная; линейная)

- стратегии обучения (итеративная; групповая)
- алгоритмы обучения (классический алгоритм обратного распространения ошибки; его модификация с использованием демпфирования; модификация с использованием адаптивного шага обучения; RPROP; SARPROP)
- критерии остановки процесса обучения (превышение максимального количества итераций; достижение требуемой ошибки; достижение требуемого изменения ошибки; критерий потери обобщения)
- алгоритмы классификации (поиск нейрона с максимальных значением на выходе; учет порога достоверности; учет значений на выходе остальных нейронов; раздельная и совмещенная схемы)
- алгоритмы параметризации (раздельная и совмещенная схемы)

1.2.3. Использование объектной технологии

При создании данной модели использовался объектно-ориентированный подход к созданию программных систем. Это позволило разграничить область ответственности каждой из подсистем, что в свою очередь обусловило возможность замены алгоритмов при сохранении конфигурации остальной части системы. Таким образом, удалось добиться возможности изменения любых параметров схем работы с нейронной сетью с целью поиска конфигурации, дающей наилучший результат при решении данной задачи.

2. Общие вопросы использования нейронных сетей

В настоящее время разработано много различных типов ИНС (искусственных нейронных сетей). В данной работе исследована сеть типа "многослойный персептрон" (Multy Layer Perceptron).

Исследуемая сеть:

- статическая (ее структура не изменяется во времени)
- обучается с учителем
- имеет прямые связи (Feed Forward Network)
- полносвязна (Fully Connected Network), то есть все нейроны следующего слоя имеют связи со всеми нейронами предыдущего
- обучается по методу обратного распространения ошибки (Error BackPropagation)

Выбор такой сети обусловлен следующими факторами:

- разработан математический аппарат, доказывающий способность подобных сетей аппроксимировать любую функцию N переменных, сопоставив N входным значениям M выходных [3]
- при решении задач классификации такая сеть при обучении "с учителем" вырабатывает внутреннее представление для исходных данных - такое автоматическое выделение признаков является большим преимуществом, например, перед сетями с радиально-базисной функцией активации (RBF, Radial Basis Function Network), которые только могут симулировать схему классификации "ближайший сосед" [6].

2.1. Строение сети

Рассматриваемая сеть состоит из слоев следующих типов:

- входной
- скрытые
- выходной.

Входной слой один и он является первым слоем сети. Входные данные в виде признаков подаются именно на этот слой. Нейроны входного слоя имеют особую функцию активации, рассмотренную в разделе "Кодирование данных" (далее под функцией активации будет подразумеваться активационная функция для нейронов скрытых и выходного слоев). Этот слой не обучается. Количество нейронов во

входном слое равно количеству входных переменных, то есть количеству признаков.

Количество скрытых слоев и нейронов в них определяют емкость сети, то есть ее способность к обобщению.

Выходной слой один и является последним слоем сети. При обучении на него подаются входные данные в виде закодированных типов дефектов или их параметров, а при использовании с него снимаются выходные данные. Этот слой обучается наравне со скрытыми слоями. При решении задачи классификации нейроны выходного слоя имеют сигмоидальную функцию активации, при определении параметров для скрытого слоя необходимо использовать линейную функцию активации.

В дальнейшем, будут использоваться следующие обозначения:

- m - номер слоя ($0..N_m-1$), где $N_m > 2$ - число слоев в сети, включая входной
- n - номер нейрона в слое ($0..N_{n_m}-1$), где N_{n_m} - число нейронов в слое m

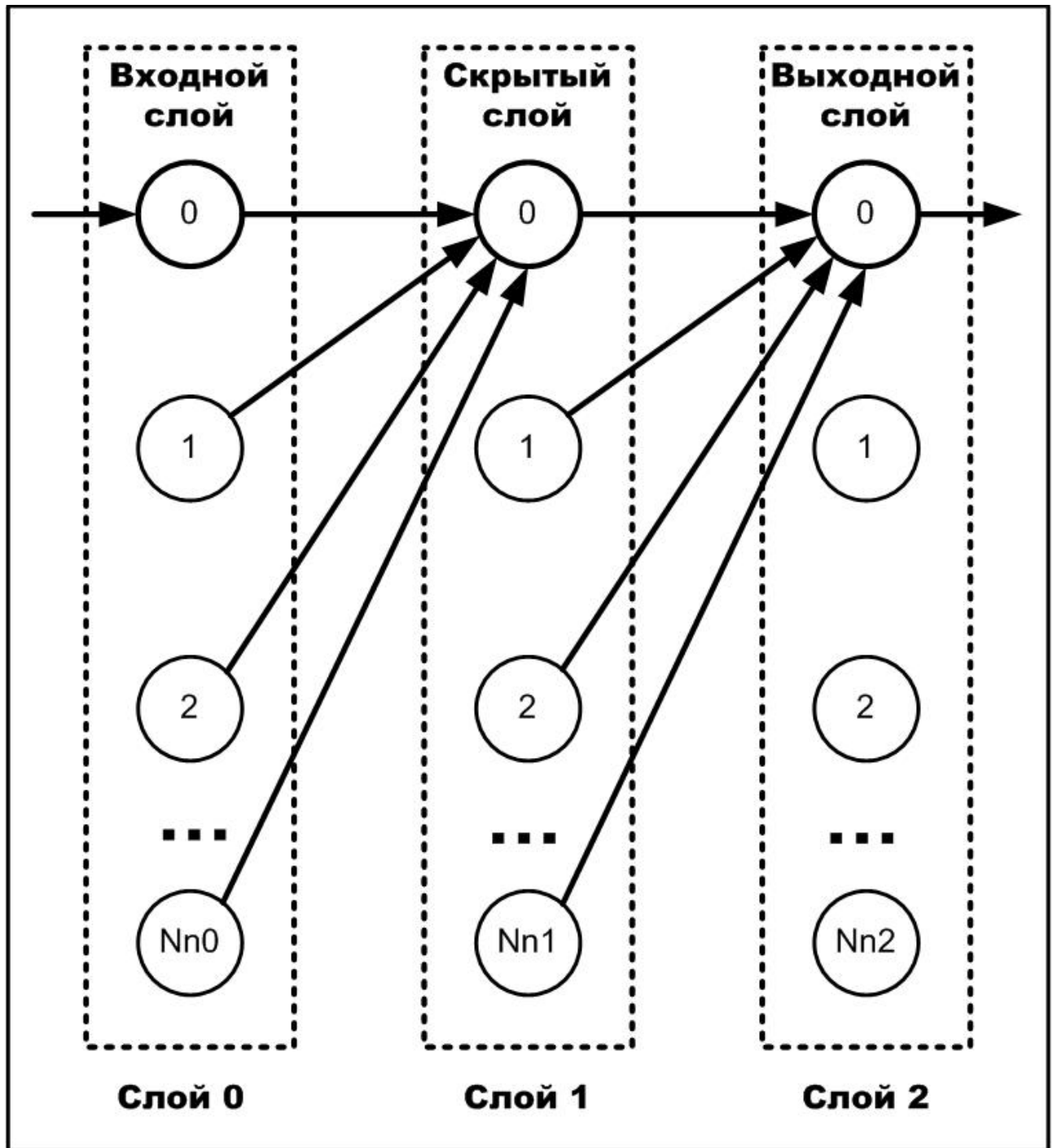


Рис. 2.1. Структура нейронной сети

2.2. Модель нейрона

Каждый нейрон рассматриваемой сети представляется набором следующих параметров:

- w - вектор входных весов нейрона, таким образом, w_{m,n,n_i} - вес, связывающий вход нейрона n в слое m с выходом нейрона n_i в слое $m-1$

- T - порог нейрона (threshold), представляющий вес связи с поляризационным нейроном (сами поляризационные нейроны в структуру сети не включены)
- y - выход нейрона, значение которого определяется как

$$y_{m,n} = F(S_{m,n}) \quad (1)$$

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

$$S_{m,n} = \sum_{ni} (y_{m-1,ni} * w_{m,n,ni}) - T_{m,n} \quad (2)$$

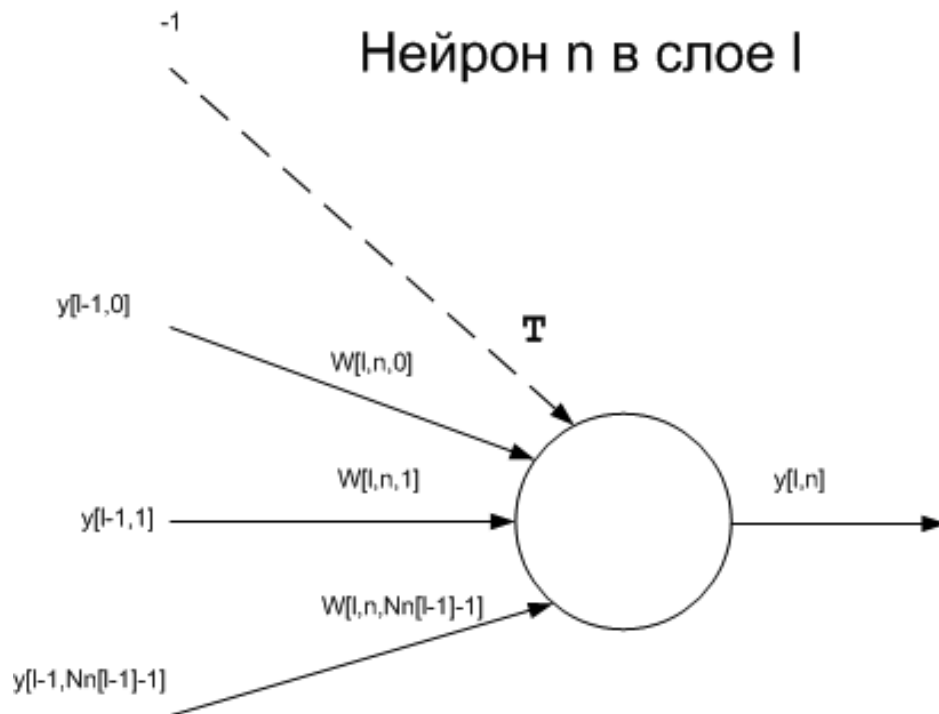


Рис. 2.2. Используемая модель нейрона

В приведенных далее алгоритмах нейроны также содержат следующие параметры:

- δ - произведение ошибки нейрона на значение производной функции активации
- $\partial E / \partial w$ - вектор, значений частной производной ошибки по весу
- $\partial E / \partial T$ - значение частной производной ошибки по порогу
- $\partial E / \partial w^{t-1}$ - вектор значений частной производной ошибки по весу в предыдущий момент времени
- $\partial E / \partial T^{t-1}$ - значение частной производной ошибки по порогу в предыдущий момент времени
- Δw - вектор приращений весов
- ΔT - приращение порога
- $\Delta \Delta w$ - вектор приращений приращений весов
- $\Delta \Delta T$ - приращение приращений порога

Для оптимизации и ускорения процесса вычислений добавлены следующие параметры:

- S - взвешенная сумма для нейрона
- FdS - значение производной функции активации для нейрона.

Назначение указанных выше параметров и способы их вычисления будут рассмотрены в соответствующих разделах.

2.3. Функция активации

С помощью функции активации вычисляется значение на выходе нейрона - степень его возбуждения (см. формулу (1)).

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

$$y = 1 / (1 + e^{-c \cdot x}) \quad (3)$$

где $c > 0$ - коэффициент, характеризующий ширину линейной части по оси абсцисс.

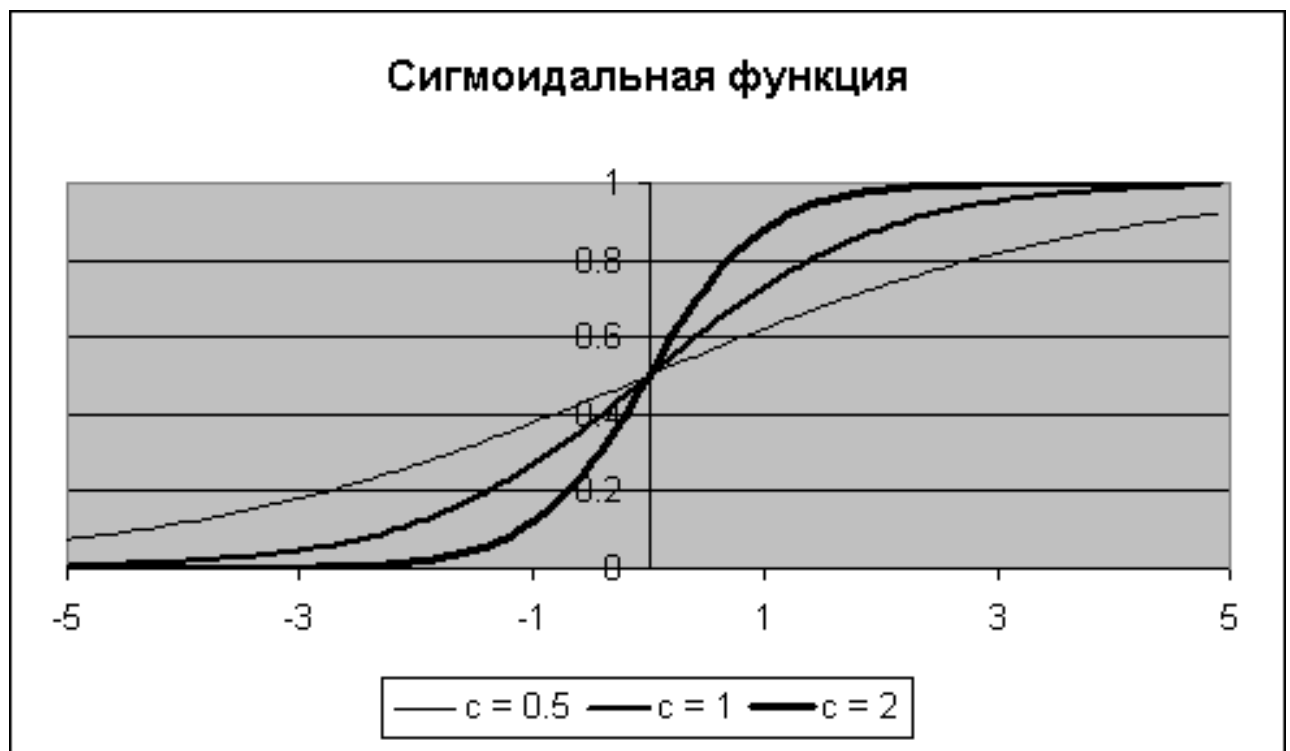


Рис. 2.3. Вид сигмоидальной функции

Как видно, сигмоидальная функция является монотонной и всюду дифференцируемой. Производную можно вычислить по следующей формуле:

$$y' = c * y * (1 - y) \quad (4)$$

Для вычисления параметров дефектов также используется линейная функция активации

$$y = k * x \quad (5)$$

ее производная

$$y' = k \quad (6)$$

2.4. Инициализация параметров нейронной сети

Выбор начальных значений весов и порогов влияет на скорость обучения сети.

Если начальные значения весов большие, то это вводит нейронные элементы в насыщение, что обуславливает малые значения градиентов изменения весов и, как результат, малую скорость обучения.

Если начальные значения весов слишком маленькие, это ведет к тому, что производная функции активации нейрона близка к нулю, что также замедляет скорость обучения [3].

Начальные значения весов следует выбирать так, чтобы ожидаемая девиация взвешенной суммы лежала между линейной частью и насыщением функции активации [4]. Для этого начальные значения весов следует выбирать случайным образом из диапазона $[-w_{im_m} \dots w_{im_m}]$,

где w_{im_m} вычисляется для каждого слоя

$$w_{im_m} = 1 / \sqrt{Nn_{m-1}} \quad (7)$$

где Nn_m - число нейронов в слое m [1].

2.5. Кодирование данных

2.5.1. 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 [6].

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 [3].

Таким образом, одним из рациональных (но не оптимальным) является следующий способ нормализации входных значений [3]:

$$x_n' = (x_n - \text{mean}_n) / (\text{dev}_n) \quad (8)$$

где mean - среднее арифметическое, а dev - девиация.

В результате нормализации входных значений по (8) на входы нейронов первого скрытого слоя будут подаваться числа в диапазоне от -1 до 1, не имеющие сдвига.

$$\text{mean}_n = 1/N_{\text{mp}} * \sum_i x_n^i \quad (9)$$

$$\text{dev}_n = \max(|\text{mean}_n - \max x_n|, |\text{mean}_n - \min x_n|) \quad (10)$$

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

2.5.2. Кодирование выходных данных

Значения на выходе сети определяются функцией активации выходного слоя. Поэтому при решении задачи классификации в процессе обучения следует так кодировать номер класса, чтобы требуемое значение на выходе нейрона находилось в диапазоне значений функции активации.

Так как каждому типу дефекта соответствует свой нейрон, то его выход должен являться бинарным значением. Поэтому для кодирования выходных данных наиболее подходит следующая схема:

1. В выходных векторах обучающего набора элемент с номером, равным номеру класса, должен иметь значение, равное максимальному значению, которое может принимать функция активации.
2. В выходных векторах обучающего набора элемент с номером, не равным номеру класса, должен иметь значение, равное минимальному значению, которое может принимать функция активации.

При определении параметров дефектов для выходного слоя используется линейная функция активации, диапазон ее значений $(-\infty; +\infty)$. Поэтому специальных действий по кодированию выходных данных при решении задачи определения параметров дефектов не требуется.

2.6. 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 [3, 5]:

- 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

$$E = 1 / N_{mp} * \sum_k E_k \quad (11)$$

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

$$\partial E / \partial W = 1 / N_{mp} * \sum_k \partial E / \partial W_k \quad (12)$$

В качестве E_k может использоваться среднеквадратическая ошибка

$$E_k = 1 / N_{n_{Nm-1}} * \sum_n (y_{Nm-1,n} - out_{k,n}) \quad (13)$$

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:

- 1) низкая скорость обучения.

2.7. Алгоритм обучения

Если стратегия обучения определяет, когда будут изменяться параметры сети, то алгоритм обучения определяет само изменение.

Обучение по методу обратного распространения ошибки можно разделить на три фазы:

- 1) прямое распространение сигнала вдоль сети
- 2) обратное распространение ошибки
- 3) изменение параметров сети.

2.7.1. Прямое распространение сигнала

На фазе прямого распространения сигнала вычисляется выходная активность всех нейронных элементов сети:

- 1) подать сигнал, преобразованный по формуле (8), на входной слой
- 2) для каждого слоя, начиная с первого скрытого
 - а) для всех нейронов слоя
 - i) вычислить взвешенную сумму $S_{m,n}$ по формуле (2)
 - ii) вычислить выходную активность $y_{m,n}$ по формуле (1)

2.7.2. Обратное распространение ошибки

На фазе обратного распространения ошибки рекуррентно вычисляется градиент $\partial E / \partial W$ для каждого слоя через ошибку следующего слоя сети [5]:

- 1) для выходного слоя
 - а) для каждого нейрона выходного слоя вычислить:

$$\delta_{m,n} = (y_{m,n} - out_n) * F'(S_{m,n}) \quad (14)$$

$$\partial E / \partial T_{m,n} = -\delta_{m,n} \quad (15)$$
 - i) для каждого веса нейрона вычислить

$$\partial E / \partial w_{m,n,ni} = \delta_{m,n} * y_{m-1,ni} \quad (16)$$
- 2) последовательно для скрытых слоев, начиная с последнего
 - а) для каждого нейрона в слое вычислить:
 - i) $\delta_{m,n}$

$$\delta_{m,n} = F'(S_{m,n}) * \sum_{ni} \delta_{m+1,ni} * w_{m+1,ni,n} \quad (17)$$

здесь F' - производная функции активации следующего слоя.
 - ii) $\partial E / \partial T_{m,n}$ и $\partial E / \partial w_{m,n,ni}$ по (15) и (16)

2.7.3. Изменение весов

2.7.3.1. Алгоритм RPROP

Для изменения параметров сети в процессе обучения может применяться алгоритм RPROP.

Resilient error backpropagation - алгоритм, предложенный в 1992 году (последняя модификация в 1994) [8, 9]. 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:

$$\Delta_{ij}(t) = \begin{cases} \Delta_{ij}(t-1) * \eta^+, & \text{если } \partial E / \partial w_{ij}(t-1) * \partial E / \partial w_{ij}(t) > 0, \\ \Delta_{ij}(t-1) * \eta^-, & \text{если } \partial E / \partial w_{ij}(t-1) * \partial E / \partial w_{ij}(t) < 0, \\ \Delta_{ij}(t-1), & \text{иначе} \end{cases} \quad (18)$$

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

$$\Delta \Delta w_{m,n,ni} = \min(\Delta \Delta_+ * \Delta \Delta w_{m,n,ni}, \Delta \Delta_{\max}) \quad (19)$$

$$\Delta w_{m,n,ni} = -\text{SGN}(\partial E / \partial w_{m,n,ni}) * \Delta \Delta w_{m,n,ni} \quad (20)$$

$$w_{m,n,ni} = w_{m,n,ni} + \Delta w_{m,n,ni} \quad (21)$$

$$\partial E / \partial w_{m,n,ni}^{t-1} = \partial E / \partial w_{m,n,ni}^t \quad (22)$$

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

$$\Delta \Delta w_{m,n,ni} = \max(\Delta \Delta_- * \Delta \Delta w_{m,n,ni}, \Delta \Delta_{\min}) \quad (24)$$

$$\partial E / \partial w_{m,n,ni}^{t-1} = 0 \quad (25)$$

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

$$\Delta w_{m,n,ni} = -\text{SGN}(\partial E / \partial w_{m,n,ni}) * \Delta \Delta w_{m,n,ni} \quad (26)$$

$$w_{m,n,ni} = w_{m,n,ni} + \Delta w_{m,n,ni} \quad (27)$$

$$\partial E / \partial w_{m,n,ni}^{t-1} = \partial E / \partial w_{m,n,ni}^t \quad (28)$$

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:

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

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

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

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

Одним из немногочисленных недостатков алгоритма RPROP является то, что он, как и классический алгоритм обратного распространения ошибки может остановить процесс обучения в ближайшем минимуме целевой функции. Для преодоления этого недостатка разработан алгоритм SARPROP [13, 14].

2.7.3.2. Алгоритм SARPROP

Simulated Annealing (имитация отжига) RPROP - модификация алгоритма RPROP, предложенная в 1998 году, сочетает добавление в веса нейронов шума, а также их угасание.

Количество добавляемого шума ассоциируют с температурой системы, которая убывает по мере увеличения количества пройденных эпох обучения. Это позволяет на начальных этапах провести более широкий поиск минимума целевой функции, а затем по мере уменьшения температуры уточнить найденное решение.

Добавление шума происходит при выполнении двух условий:

- изменение знака градиента ошибки;
- величина изменения шага обучения меньше величины пропорциональной текущему значению среднеквадратической ошибки (RMS).

Величина добавляемого шума зависит как от значения среднеквадратической ошибки, так и от температуры.

Добавление шума только при изменении знака градиента ошибки необходимо для избежания нарушений в процессе адаптирования скорости изменения весов. При этом добавление шума происходит при малых значениях скорости изменения шага обучения. Это позволяет покидать локальные минимумы, в тоже время не нарушая процесс адаптации.

Второе изменение алгоритма RPROP - угасание весов (weights decay), основано на модификации целевой функции. Степень угасания весов также зависит от температуры, благодаря этому степень угасания уменьшается в процессе обучения. Новый градиент ошибки вычисляется как

$$\partial E / \partial w_{ij}^{\text{SARPROP}} = \partial E / \partial w_{ij} - k_1 * w_{ij} * 2^{-T * \text{epoch}} \quad (29)$$

Угасание весов вводится для того, чтобы сделать их значения по возможности меньшими на начальном этапе обучения. Это позволяет, увеличить эффективность поиска решения до того, как веса нейронов начнут расти (уточнение решения).

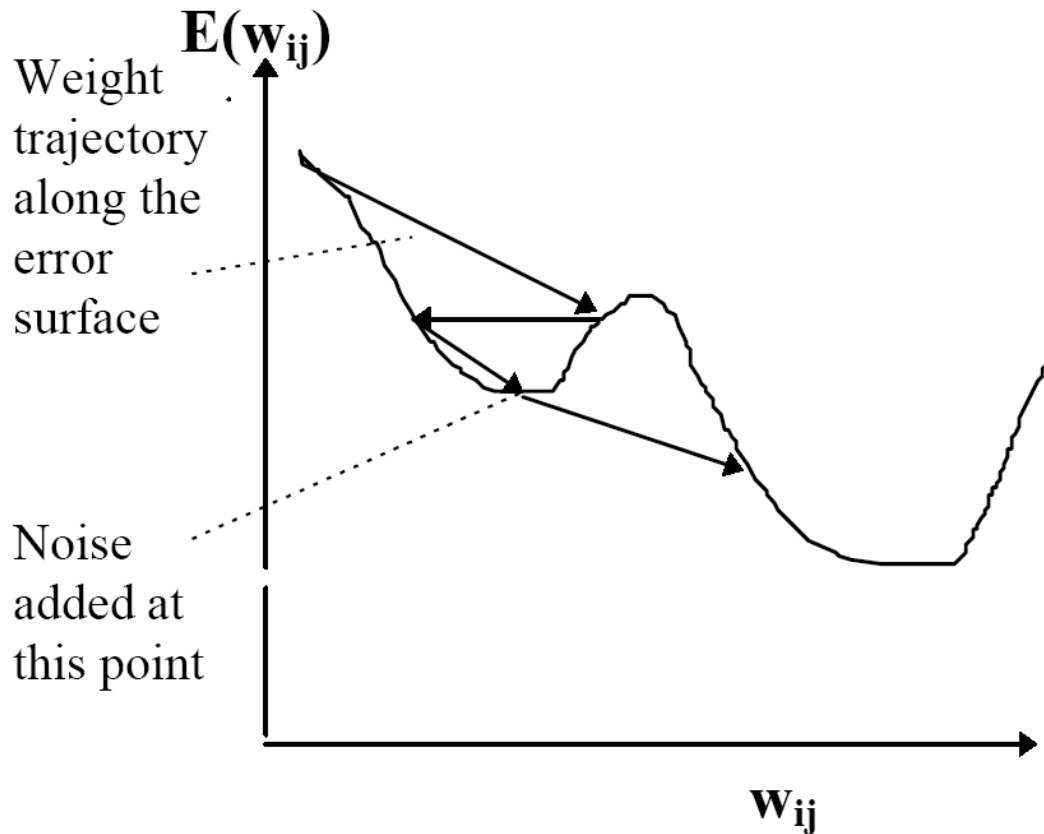


Рис. 2.4. Иллюстрация эффекта добавления шума в веса
Псевдокод алгоритма SARPROP:
ДЛЯ ВСЕХ ВЕСОВ И ПОРОГОВ

ЕСЛИ $\partial E / \partial w_{ij}(t-1) * \partial E / \partial w_{ij}(t) > 0$ ТОГДА

$$\Delta_{ij}(t) = \min(\Delta_{ij}(t-1) * \eta^+, \Delta_{\max})$$

$$\Delta w_{ij}(t) = w_{ij}(t) + \Delta w_{ij}(t)$$

$$\partial E / \partial w_{ij}(t-1) = \partial E / \partial w_{ij}(t)$$

ИНАЧЕ ЕСЛИ $\partial E / \partial w_{ij}(t-1) * \partial E / \partial w_{ij}(t) < 0$ ТОГДА

ЕСЛИ $\Delta_{ij}(t-1) < k_2 * \text{error}^2$ ТОГДА

$$\Delta_{ij}(t) = \Delta_{ij}(t-1) * \eta^- + k_3 * r * \text{error} * 2^{-T * \text{epoch}}$$

ИНАЧЕ

$$\Delta_{ij}(t) = \Delta_{ij}(t-1) * \eta^-$$

$$\Delta_{ij}(t) = \max(\Delta_{ij}(t-1) * \eta^-, \Delta_{\min})$$

$$\partial E / \partial w_{ij}(t-1) = 0$$

ИНАЧЕ

$$\Delta w_{ij}(t) = - \text{sign}(\partial E / \partial w_{ij}(t)) * \Delta_{ij}(t)$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$$

$$\partial E / \partial w_{ij}(t-1) = \partial E / \partial w_{ij}(t)$$

Здесь error - значение среднеквадратической ошибки, T - температура, k_1 , k_2 , k_3 - константы, r - случайное число из диапазона $[0; 1]$.

По результатам практических исследований установлено, что для большинства практических задач подходят следующие значения параметров алгоритма SARPROP [13]:

- $k_1 = 0.01$
- $k_2 = 0.1$
- $k_3 = 3$
- T выбирается из диапазона $[0.01; 0.07]$

2.7.4. Остановка процесса обучения

Процесс обучения нейронной сети носит итеративный характер. Рассмотрим критерии, по которым можно определить момент, когда обучение следует остановить.

Первый способ [1] заключается в том, что процесс обучения останавливают в тот момент, когда ошибка сети E станет меньше заданной:

$$E < E_{\text{stop}} \quad (27)$$

Такой подход обладает рядом недостатков:

- 1) если значение E_{stop} слишком маленькое, то сеть его может не достигнуть
- 2) если значение E_{stop} слишком большое, то возможности нейронной сети не будут использованы целиком

Второй подход заключается в остановке процесса обучения после некоторого заданного числа итераций:

$$I > I_{\text{stop}} \quad (28)$$

Такой способ также не лишен недостатков:

- 1) если значение I_{stop} слишком маленькое, то возможности сети не будут использованы полностью, так как возможно, что при большем числе итераций сеть сошлась бы к меньшему значению ошибки
- 2) если значение I_{stop} большое, то сеть может показать результаты хуже чем, если бы процесс обучения был остановлен раньше.

Последнее вызвано так называемым эффектом переобучения (overfitting).

2.7.4.1. Переобучение

Эффект переобучения заключается в том, что способность нейронной сети к обобщению в некоторый момент обучения начинает снижаться.

Чтобы обнаружить данный факт, необходимо использовать так называемый проверочный набор данных (validation set). Он формируется из части обучающей выборки (обычно 1/3). Нейронная сеть не обучается на выборках из проверочного набора, он используется только для обнаружения факта переобучения.

Графическую интерпретацию данного эффекта можно представить, используя кривые обучения и проверки (learning curve, validation curve).

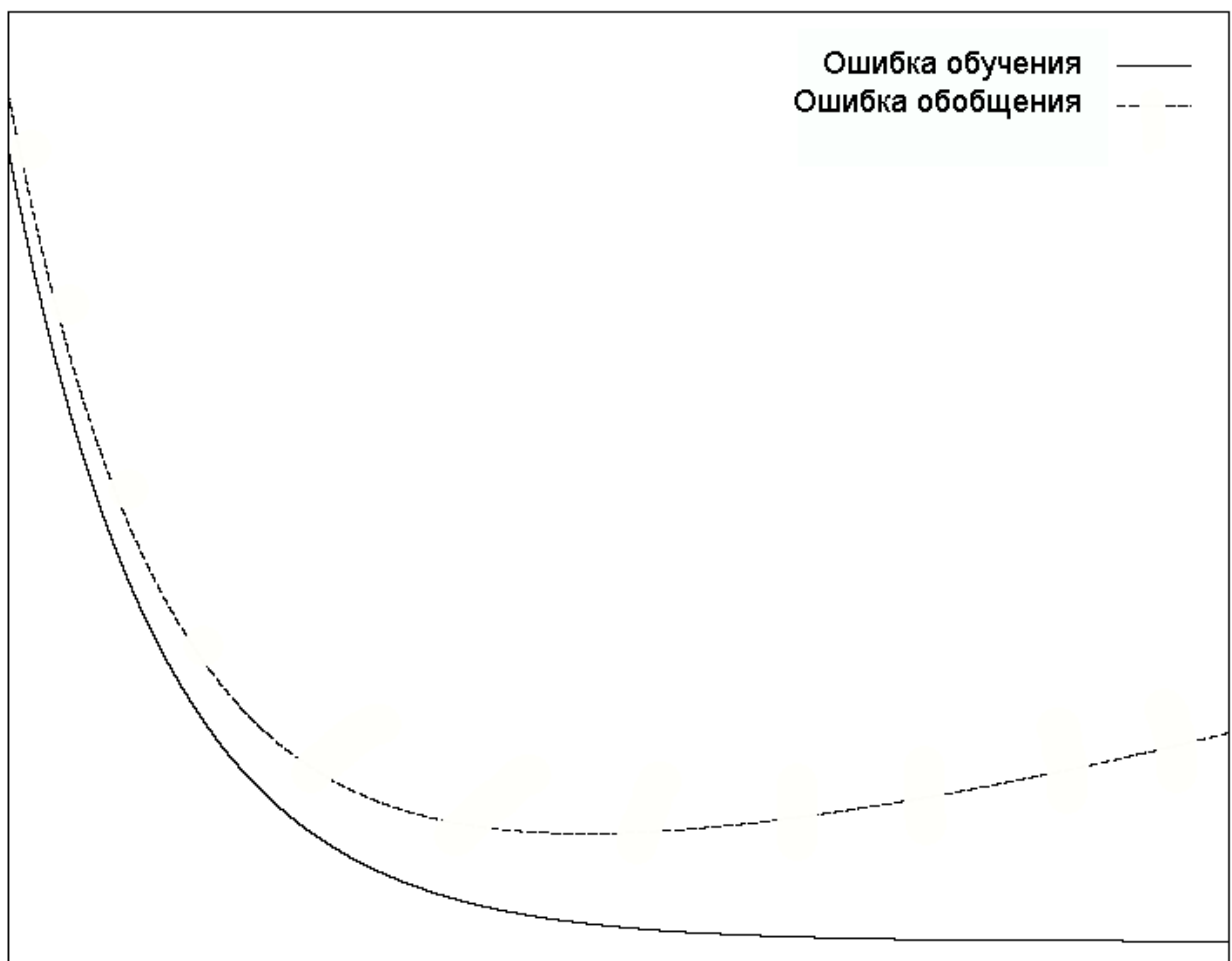


Рис. 2.5. Характерные кривые обучения и обобщения

Здесь нижняя кривая соответствует ошибке обучения, а верхняя ошибке обобщения. Так как сеть не обучалась на проверочном наборе данных, то можно ожидать, что результаты, полученные для него, будут соответствовать и другим входным данным при практическом использовании этой сети. Таким образом, при использовании сеть,

обучавшаяся после наступления эффекта переобучения, покажет худшие результаты чем, если бы обучение было остановлено раньше.

Для борьбы с этим эффектом разработано несколько способов:

- затухание весов (weights decay)
- ранняя остановка процесса обучения (early stopping)
- другие.

Рассмотрим второй способ подробно.

2.7.4.2. Ранняя остановка процесса обучения

Руководствуясь приведенным выше графиком (рис. 2.5.), можно сделать вывод, что обучение следует остановить после того, как ошибка обобщения (E_v) начала возрастать. Однако данный график представляет идеальный случай, реальная же картина подобна приведенной на рис. 2.6.

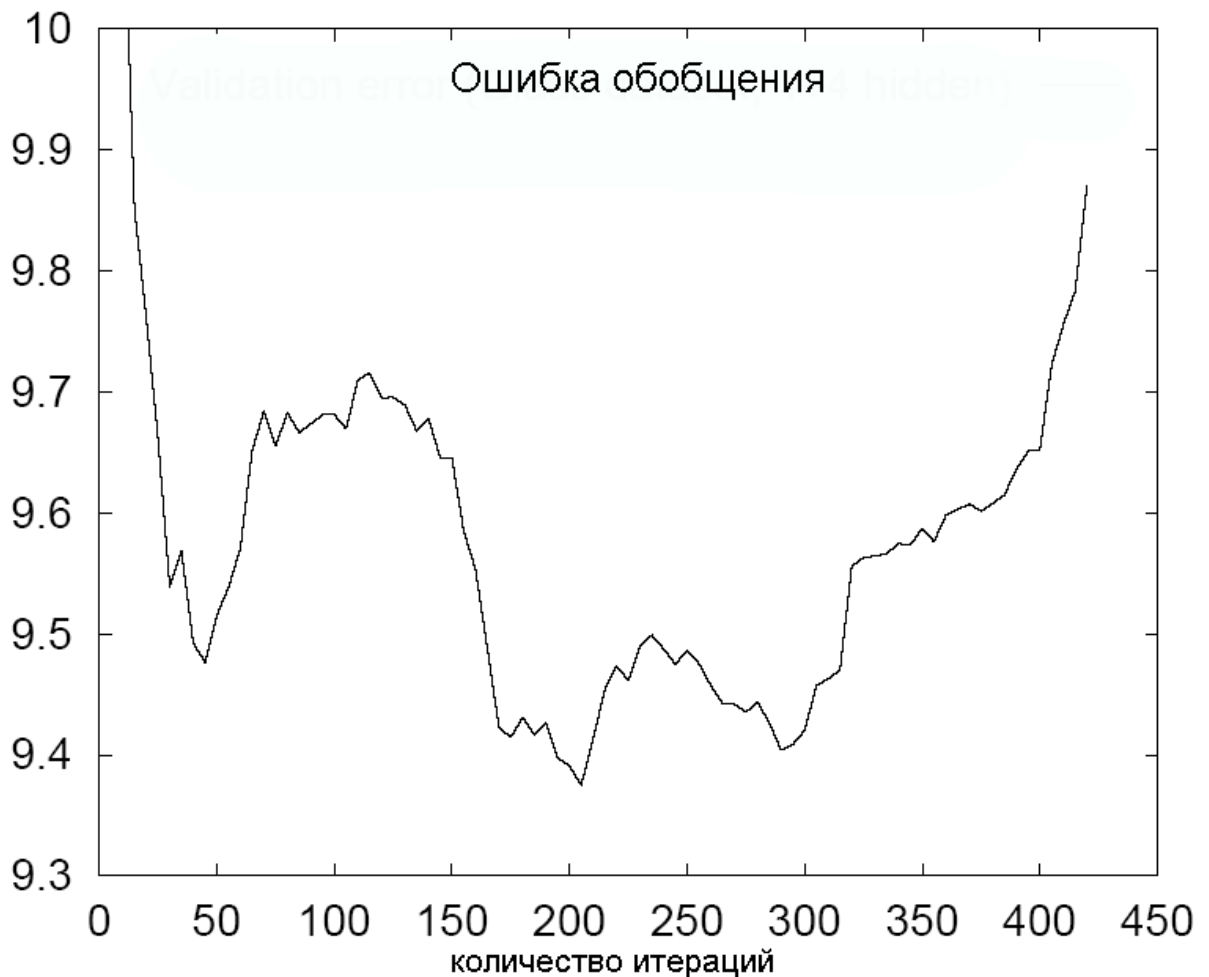


Рис. 2.6. Типичный вид кривой обобщения для практических задач

Как видно, ошибка обобщения имеет локальный минимум на 40-й итерации, в то время как глобальный минимум находится на 200-й.

Рассмотрим критерий остановки, основанный на потере обобщения (generalization loss) [7].

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

$$E_{opt}^t = \min_{t' < t} E_v^{t'} \quad (30)$$

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

Теперь введем функцию потери обобщения (generalization loss)

$$GL^t = 100 * (E_v^t / E_{opt}^t - 1) \quad (31)$$

Данная функция возвращает приращение ошибки обобщения относительно достигнутого минимума.

Критерием остановки процесса обучения является следующий:

$$GL_{\alpha}: GL^t > \alpha \quad (32)$$

Этот алгоритм можно представить следующим образом:

1) На каждой итерации вычислить потерю обобщения

$$GL = 100 * (E_v^t / E_{opt} - 1) \quad (33)$$

2) Вычислить E_{opt}

$$E_{opt} = \min(E_{opt}, E_v) \quad (34)$$

3) Принять решение об остановке процесса обучения

$$\text{остановить, если } GL > \alpha \quad (35)$$

Параметр α определяет, насколько быстро остановится процесс обучения после того, как началась потеря обобщения. Если значение этого параметра будет слишком малым, то обучение остановится в ближайшем локальном минимуме. Если его значение будет слишком большим, то процесс обучения может продолжаться неоправданно долго, при этом увеличение точности может быть малозначительным. На практике этот параметр берется из диапазона [5..10].

Следует заметить, что этот критерий не гарантирует остановку процесса обучения, поэтому его следует дополнить критерием ограничения числа итераций.

2.8. Выбор числа нейронов скрытого слоя

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. Разделение данных на обучающее и тестовое подмножества представлено на рис. 2.7.

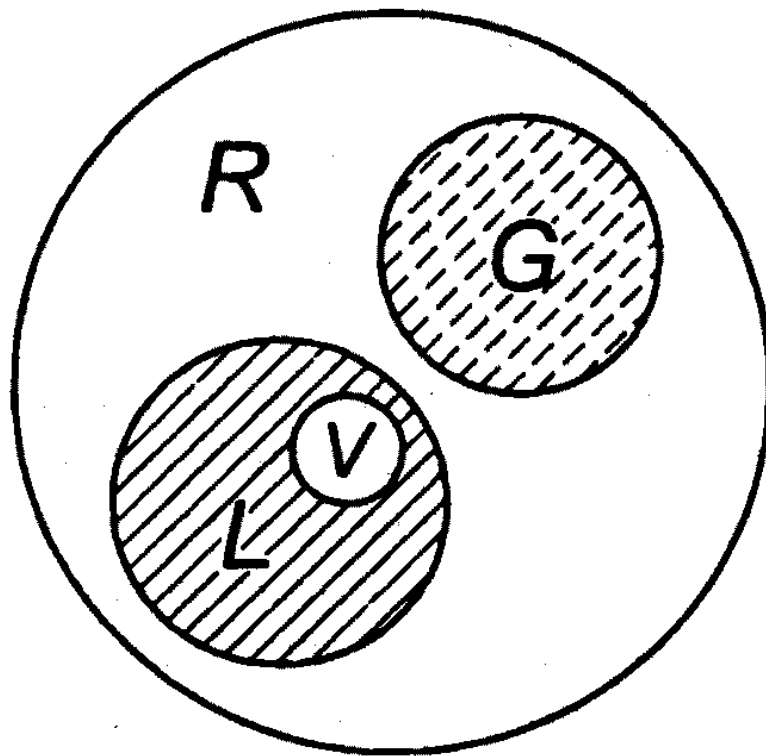


Рис. 2.7. Иллюстрация разделения данных, подчиняющихся правилу R , на обучающее подмножество L , тестовое подмножество G и контрольное подмножество V

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

$$E_G \Rightarrow E_L + \varepsilon(p/h, E_L) \quad (36)$$

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

$$2[K/2]N \leq VCdim \leq 2N_w(1 + \lg N_n) \quad (37)$$

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 выражения (36) следует, что нижняя граница диапазона приблизительно равна количеству весов, связывающих входной и скрытый слои, 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 [2].

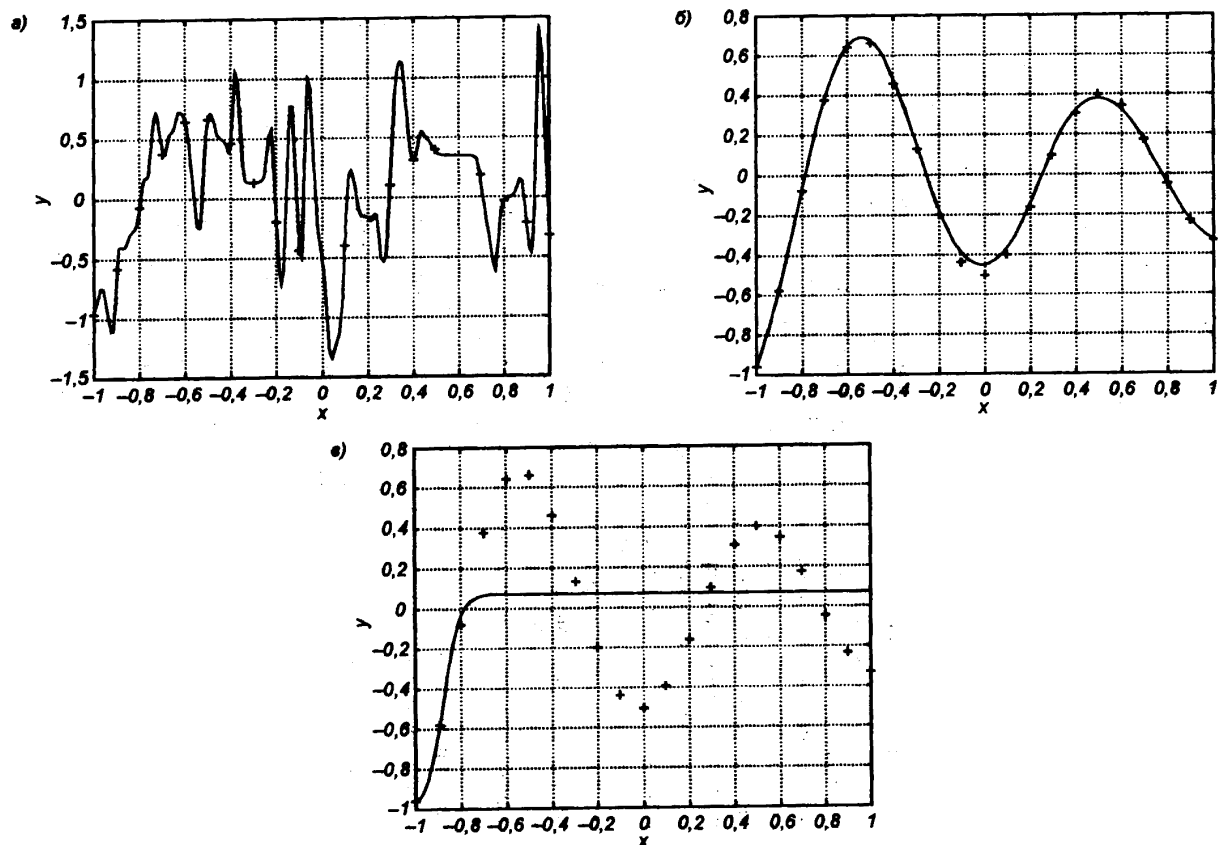


Рис. 2.8. Иллюстрация эффекта переобучения

На рис. 2.8. 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.

2.8. Выбор данных для обучения

2.8.1. Использование шума в обучающем наборе

Возможно улучшение способности нейронной сети к обобщению за счет специальной подготовки множества обучающих выборок. Для хорошо натренированной сети становится актуальной задача выработки у выходных сигналов нечувствительности к вариациям входных величин при условии, что эти вариации находятся в определенных допустимых границах, а сеть реализует монотонное отображение. Другими словами, аналогичные входные сигналы должны вызывать аналогичные реакции в случае, если они не входили в состав обучающего множества.

Для исследования данного явления нейронная сеть с 20 нейронами в скрытом слое обучалась на модельных данных и зашумленных модельных данных (5%, нормальное распределение) (набор из 500 векторов), в качестве проверочного набора использовались признаки, рассчитанные по сигналам, полученным от тестовой трубы, в качестве тестового набора также использовались тестовые данные.

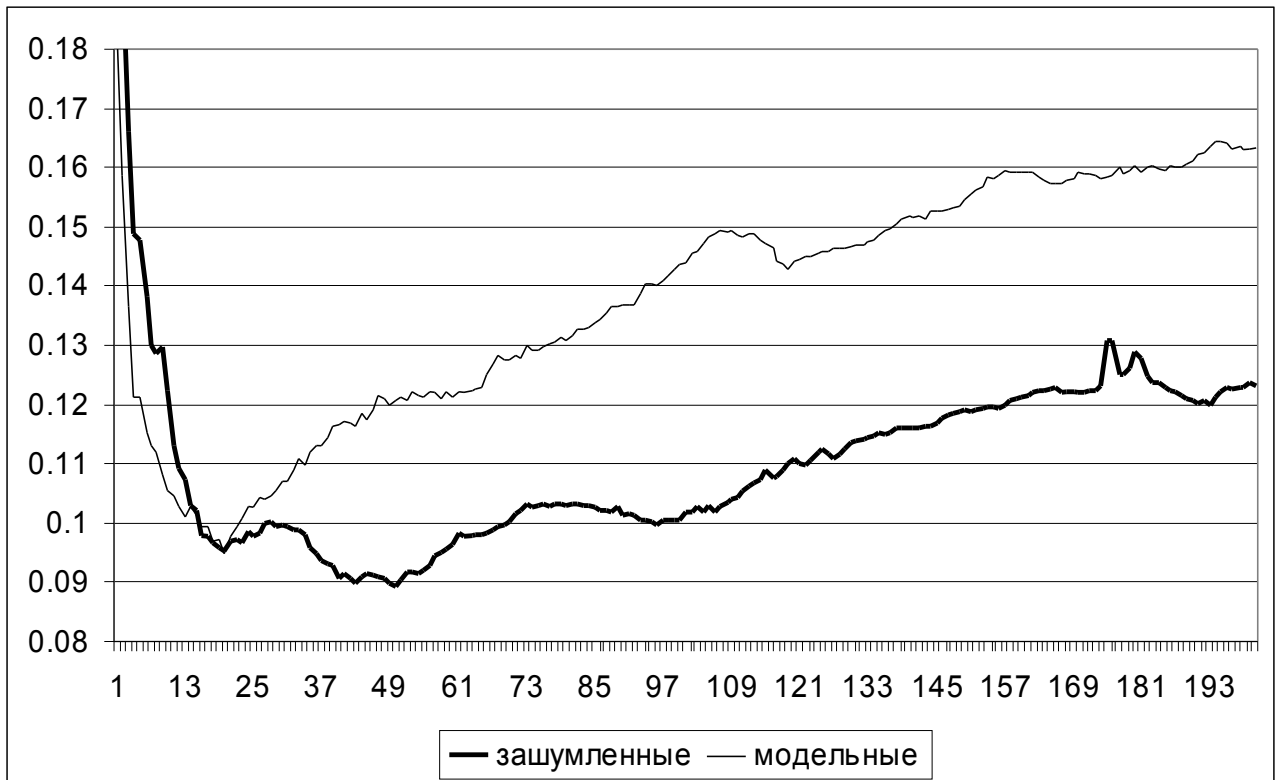


Рис. 2.9. Зависимость ошибки обобщения от числа итераций при обучении на различных наборах данных

Как видно из рис. 2.9. нейронная сеть, обучаемая на наборе модельных данных, который был возмущен, показывает большую способность к обобщению, чем нейронная сеть, обучаемая на наборе из модельных данных. Аналогичные результаты получены и для нейронных сетей с другим количеством нейронов в скрытом слое (5, 10, 15, 25, 30, 35, 40), а также при использовании в качестве проверочного набора вместо тестовых данных зашумленных модельных. При использовании сетей, полученных, как было сказано ранее, для классификации дефектов тестовой трубы получены следующие результаты:

- при обучении на модельных данных ошибка классификации 22.79%
- при обучении на зашумленных модельных данных ошибка классификации 21.88%

Таким образом, получено экспериментальное подтверждение того, что обучение на возмущенных данных способствует приобретению сетью помехоустойчивости и как следствие лучшей способностью обработки данных, полученных от измерительной системы дефектоскопа.

2.8.2 Использование в обучающем наборе данных с тестовых участков труб

Иным способом обучения нейронной может быть использование в качестве обучающего набора не модельных данных, а полученных с прогонов дефектоскопа по тестовым трубам, для которых заранее известны параметры расположенных на них дефектах.

В этом случае количество доступных векторов для использования в обучающей выборке будет меньше, чем при использовании базы модельных сигналов. Однако сеть, обученная на данных от тестовой трубы, будет обладать лучшей способностью к классификации реальных же данных при промышленном применении, т.к. она будет лучше адаптирована к возможным искажениям сигнала и вызванным этими искажениями вариациям значений признаков.

Для проверки данной гипотезы нейронная сеть с различным количеством нейронов в скрытом слое обучалась на 2/3 части всех доступных векторов признаков для дефектов тестовой трубы, для проверки использовалась оставшаяся треть векторов, а для тестирования способности к классификации использовался весь набор из 329 векторов.

2.9. Результаты тестирования

2.9.1. Алгоритм обучения

Условия тестирования:

1. топология сети 13-25-4
2. сигмоидальная функция активации, $s=1$
3. кодирование входных данных по формуле (8)
4. кодирование выходных данных как 0 и 1
5. инициализация весов по формуле (7)
6. групповая стратегия обучения
7. применение алгоритма RPROP
8. использование критерия ранней остановки GL_{10}
9. 4 типа дефектов
 1. размер обучающей выборки - 200 векторов (по 50 для каждого класса дефекта)
 2. размер проверочной выборки - 100 векторов (по 25 для каждого класса дефекта)
 3. размер тестовой выборки - 100 векторов (по 25 для каждого класса дефекта)

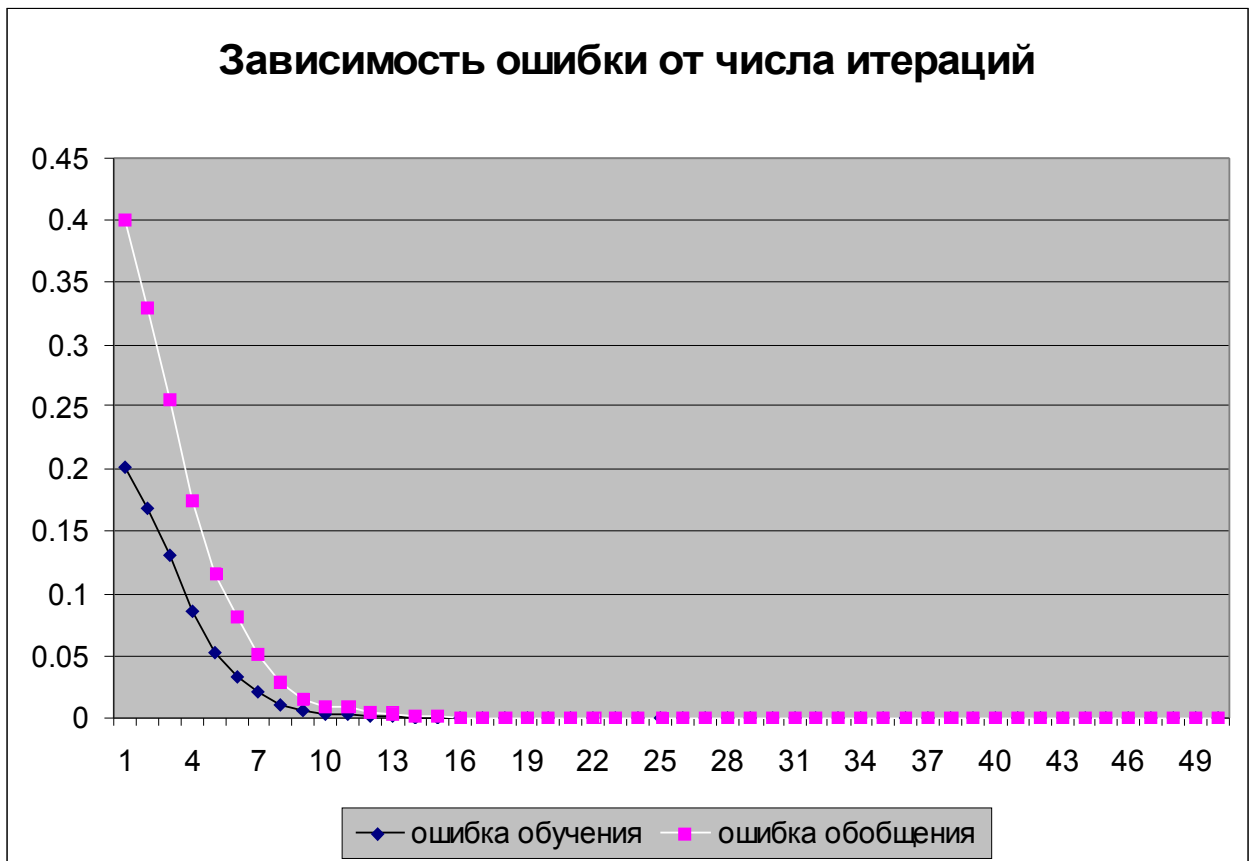


Рис. 2.9. Зависимость ошибки обучения и обобщения от числа итераций для исходной схемы

За результат классификации принимался номер нейрона с максимальной выходной активностью.

Результат классификации:

1. внешняя коррозия, ошибка 8.0%
2. внутренняя коррозия, ошибка 12.0%
3. внешняя трещина, ошибка 0.0%
4. внутренняя трещина, ошибка 0.0%

Средняя ошибка классификации: 5.0%.

Нейронная сеть обучалась и тестировалась на модельных данных.

2.9.2. Выбор числа нейронов скрытого слоя

Исследуемая нейронная сеть содержит три слоя, таким образом, количество весов сети может быть вычислено следующим образом

$$(37) N_w = (N_1 + N_3)N_2$$

где N_w - количество весов в сети, N_m - количество нейронов в слое m .

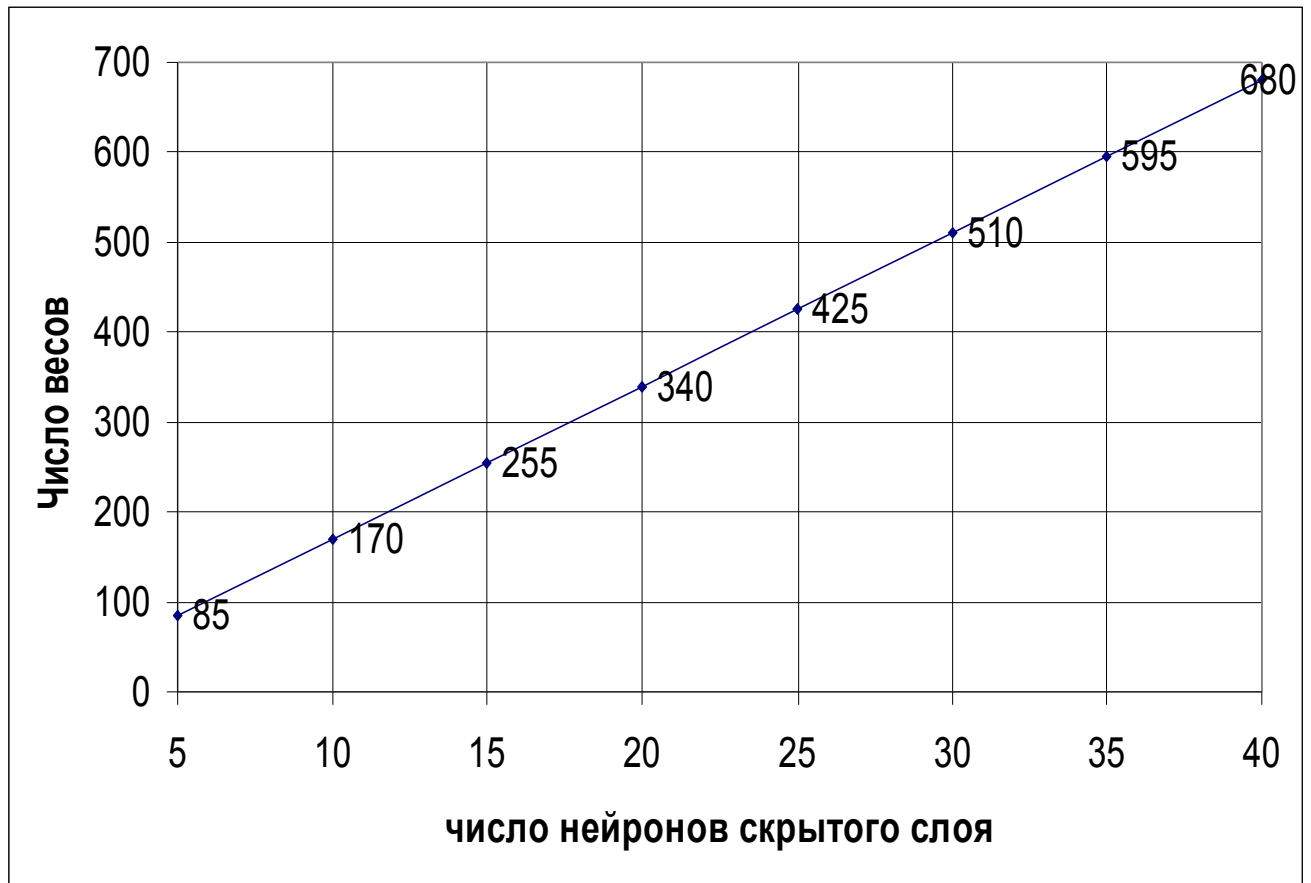


Рис. 2.10. Зависимость количества весов в сети от числа нейронов в скрытом слое

Для исследования влияния отношения количества векторов в обучающем наборе к мере VCdim нейронная с различным числом нейронов в скрытом слое обучалась на наборе модельных данных из 500 векторов. Теоретически, наилучшей способностью к обобщению должна обладать сеть, имеющая в скрытом слое около 20 нейронов (см. рис. 2.10.).

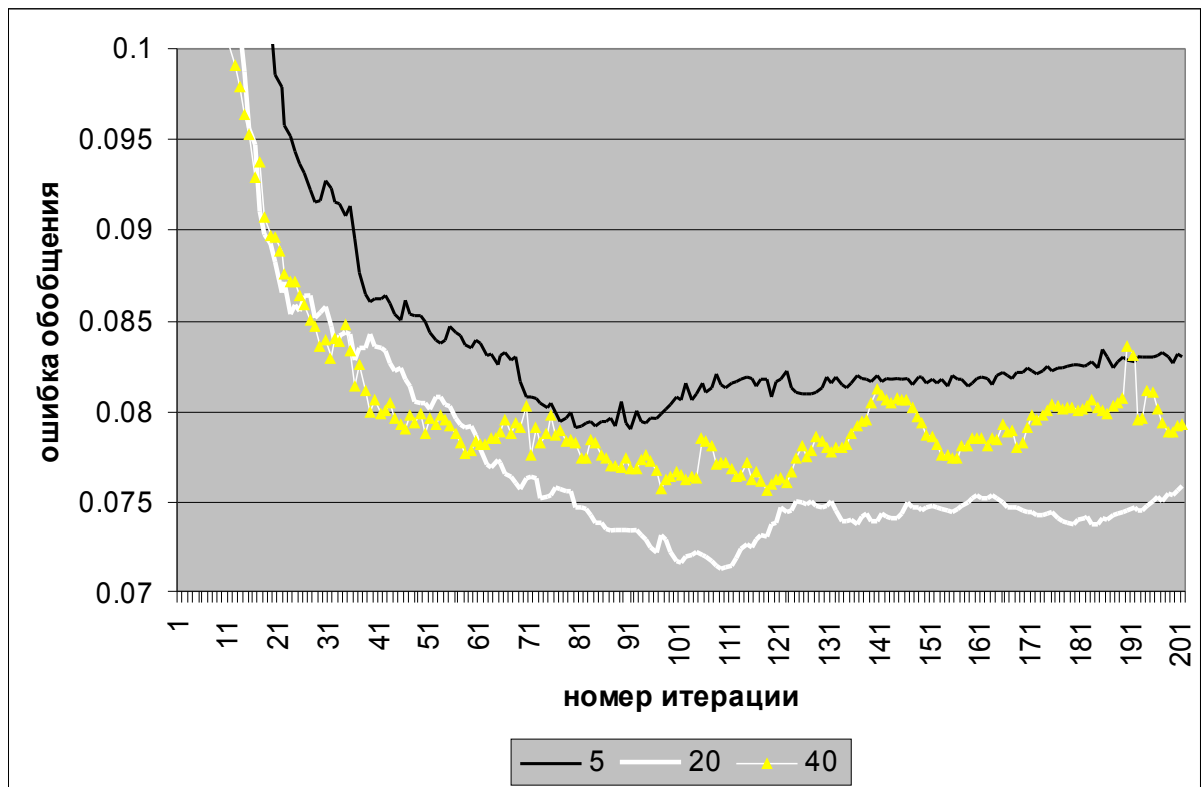


Рис. 2.11. Зависимость ошибки обобщения от числа нейронов в скрытом слое

Из рис. 2.11. видно, что нейронная сеть, имеющая в скрытом слое 5 нейронов и 85 весов, обладает низкой способностью к обобщению, т.к. не способна обеспечить низкое значение ошибки обучения (см. формулу (35)). Сеть с числом нейронов равным 40 (680 весов) также не обеспечивает минимальной ошибки обобщения т.к. мало отношение $p/h=0.78$, а число "лишних" весов велико. Минимальную же ошибку обобщения показывает нейронная сеть с 20 нейронами в скрытом слое (340 весов, $p/h=1.47$), что соответствует теоретическим предположениям.

Таким образом, теоретическая рекомендация по выбору размера скрытого слоя на основе соотношения p/h подтверждается результатами экспериментов (аналогичные результаты получены и для других наборов обучающих и проверочных данных, а также схемы классификации с отдельным определением базового класса и подкласса дефекта). Для достижения наименьшей ошибки обобщения следует выбирать такое количество нейронов в скрытом слое, чтобы для соотношения p/h выполнялось следующее условие

$$5 \leq p/h \leq 1$$

(38)

2.9.3. Использование тестовых данных при обучении

Проводилось обучение нейронной сети на модельных и тестовых данных, полученная сеть использовалась для классификации дефектов тестовой трубы.

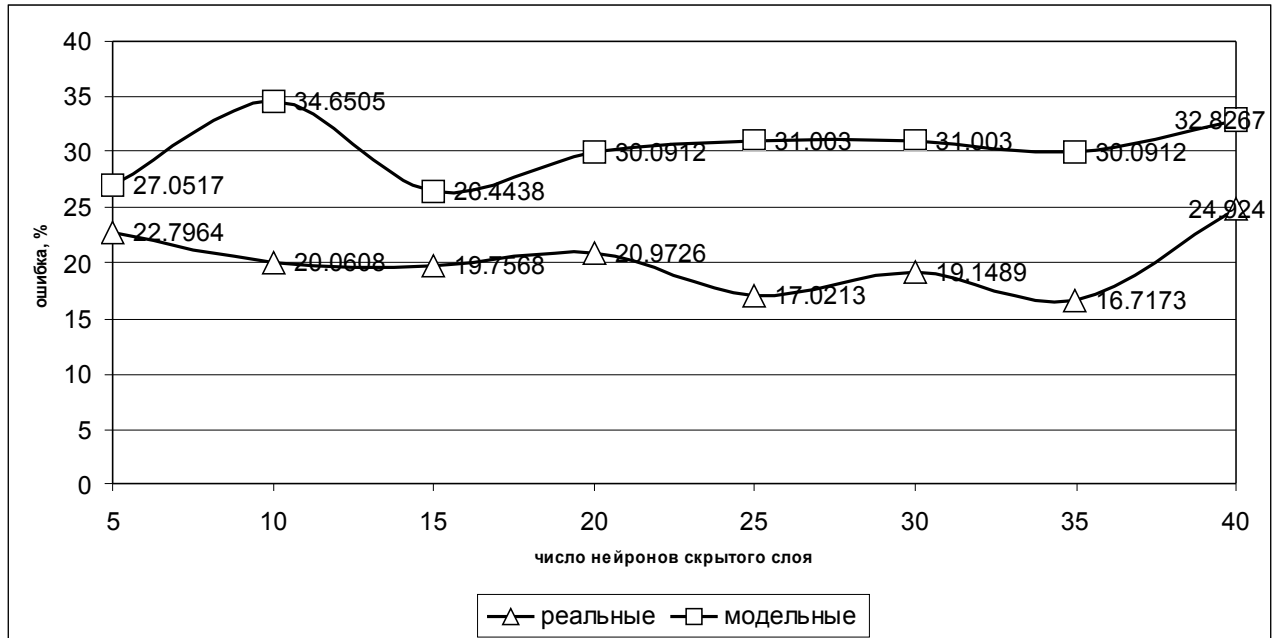


Рис. 2.12. Зависимость ошибки классификации дефектов тестовой трубы от числа нейронов при обучении на тестовых и модельных данных

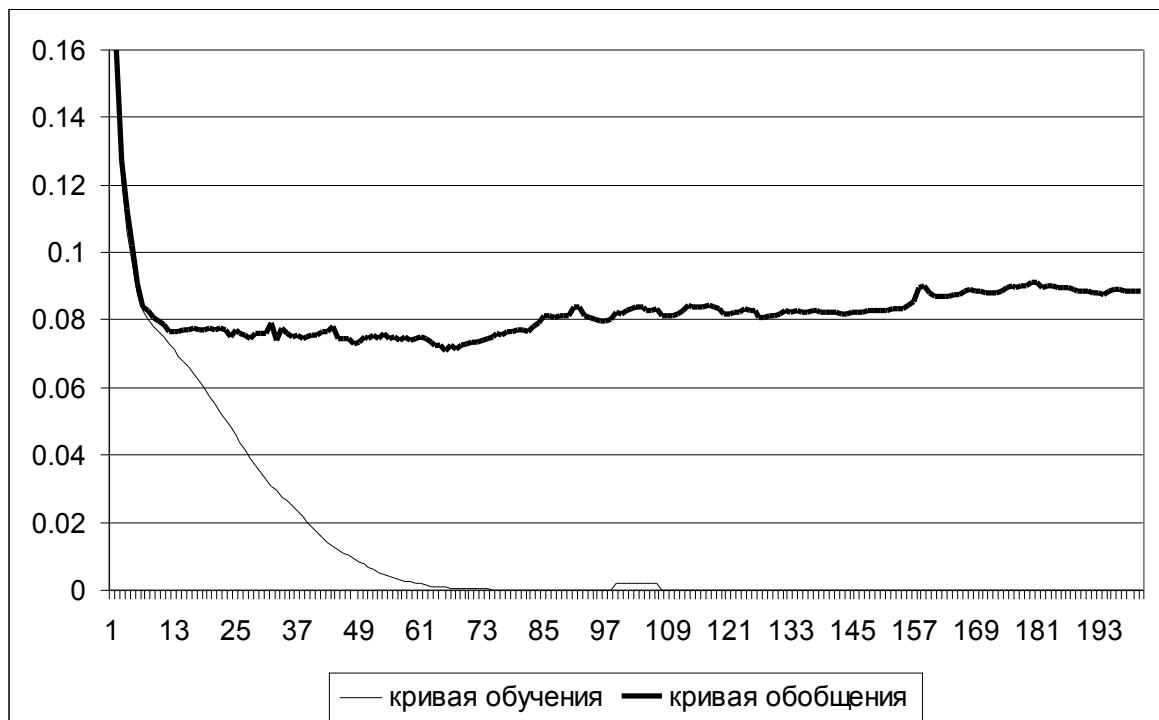


Рис. 2.13. Зависимость ошибки обучения и обобщения для нейронной сети с 35 нейронами в скрытом слое

Из рис. 2.12. видно, что нейронная сеть, обученная таким образом, обеспечивает меньшую ошибку классификации, чем при обучении на модельных или возмущенных модельных данных. Однако, как видно из рис. 2.13, результаты, полученные для большого числа нейронов (относительно количества векторов в обучающем наборе равном 220), вызваны не повышенной способностью сети к обобщению, а тем, что сеть запомнила использованные в процессе обучения вектора признаков, в то время как не использовавшиеся при обучении вектора подобны использовавшимся. Тем не менее, сеть с 10 нейронами в скрытом слое, для которой $p/h=1.29$, обеспечивает меньшую ошибку классификации, чем сети, обученные на модельных или возмущенных модельных данных (однако, следует учесть, что при обучении на модельных и возмущенных модельных данных использовалось только 500 векторов в обучающем наборе из 12000 доступных).

Таким образом, при наличии достаточного количества тестовых данных возможно обучение нейронной сети именно на них, либо на совместном наборе из модельных и реальных данных.

Данные от тестовой трубы также можно использовать в качестве проверочного набора для вычисления ошибки обобщения с целью оценки степени адекватности базы модельных дефектов, сигналам полученным с измерительного снаряда.

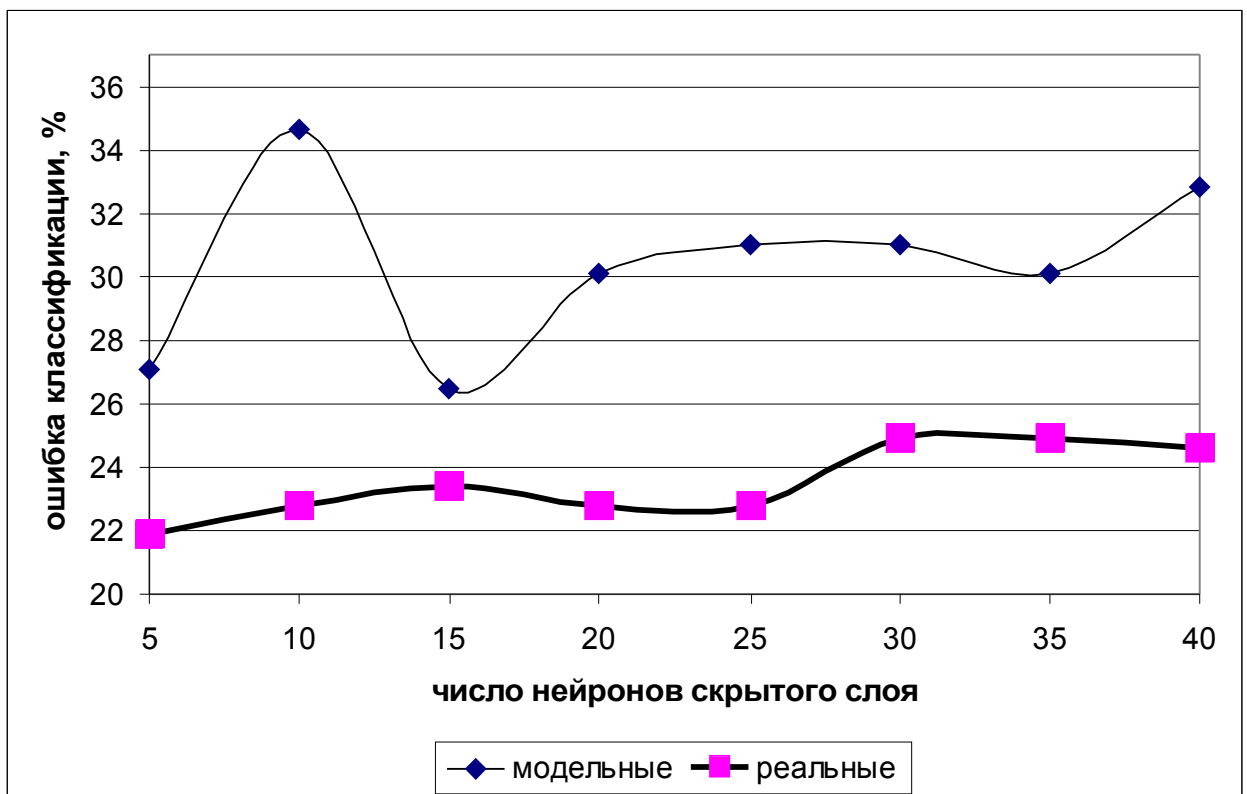


Рис. 2.14. Зависимость ошибки классификации от числа нейронов скрытого слоя при использовании в качестве проверочного набора модельных и тестовых данных

На рис. 2.14. приведены зависимости ошибки классификации дефектов тестовой трубы сетями, обученными при использовании в обучающем наборе модельных сигналов, а в проверочном наборе модельных и данных от тестовой трубы. Разность заключается в том, что в результате того, что при использовании в проверочном наборе данных от тестовой трубы ошибка обобщения больше и процесс обучения заканчивается гораздо раньше (число итераций уменьшается от 5 до 10 раз). Из-за этого варьируется способность сети к классификации модельных и данных от тестовой трубы. Видно, что при использовании в проверочном наборе модельных данных ошибка классификации значительно выше. Это может быть вызвано тем, признаки, рассчитанные для сигналов с измерительного прибора, значительно отличаются от признаков для модельных дефектов. Для преодоления этого недостатка следует использовать те признаки, которые наименее чувствительны к искажению сигнала, либо включать в базу модельных дефектов искажения характерные для сигналов, полученных с дефектоскопа.

3. Особенности классификации

3.1. Определение типа дефекта

При обучении нейронной сети для решения задачи классификации на ее выход подаются кодированные значения, обозначающие тип дефекта. В данной работе используется схема, когда количество распознаваемых классов совпадает с числом выходов нейронной сети. Так как при классификации для выходного слоя используется сигмоидальная функция активации, то во время обучения при принадлежности дефекта к данному классу на соответствующем выходе нейронной сети устанавливается 1, на остальных выходах 0. При использовании, обученной таким образом, нейронной сети определение типа дефекта производится путем поиска нейрона с максимальным значением на выходе. Номер этого нейрона будет обозначать тип дефекта, предсказанный нейронной сетью.

3.2. Оптимизация схемы распознавания

Возможны различные схемы использования нейронных сетей для определения типа дефекта:

- для классификации используется единственная сеть обученная распознавать все типы дефектов одновременно

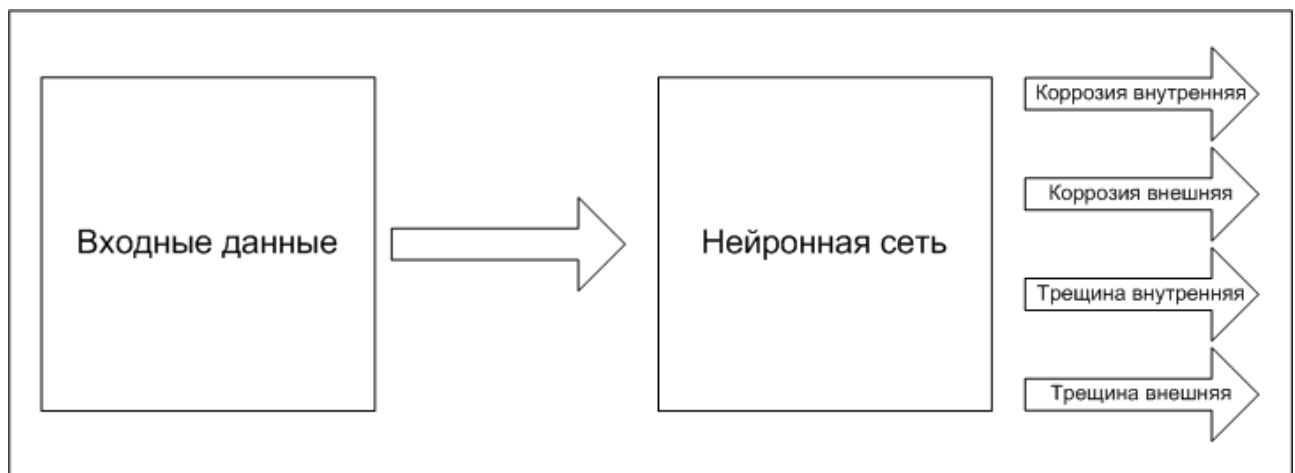


Рис. 3.1. Совмещенная схема классификации

- для определения базового класса (трещина/коррозия) используется отдельная нейронная сеть, после того как определен базовый класс, определяется подкласс дефекта (внутренний/внешний)

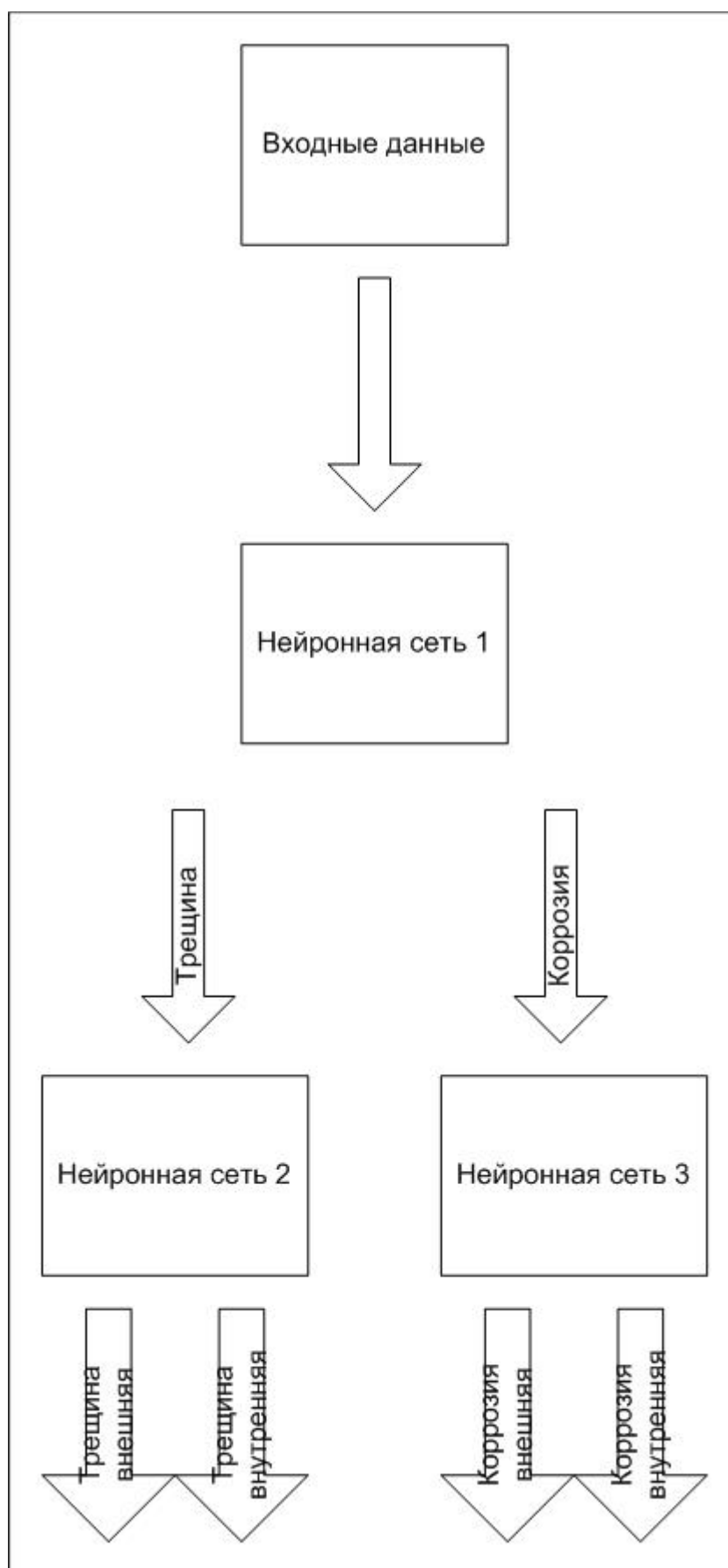


Рис. 3.2. Раздельная схема классификации

Теоретически раздельная схема классификации должна показать более точные результаты, т.к. функции, которые аппроксимируют нейронные сети в этом случае, имеют меньший порядок.

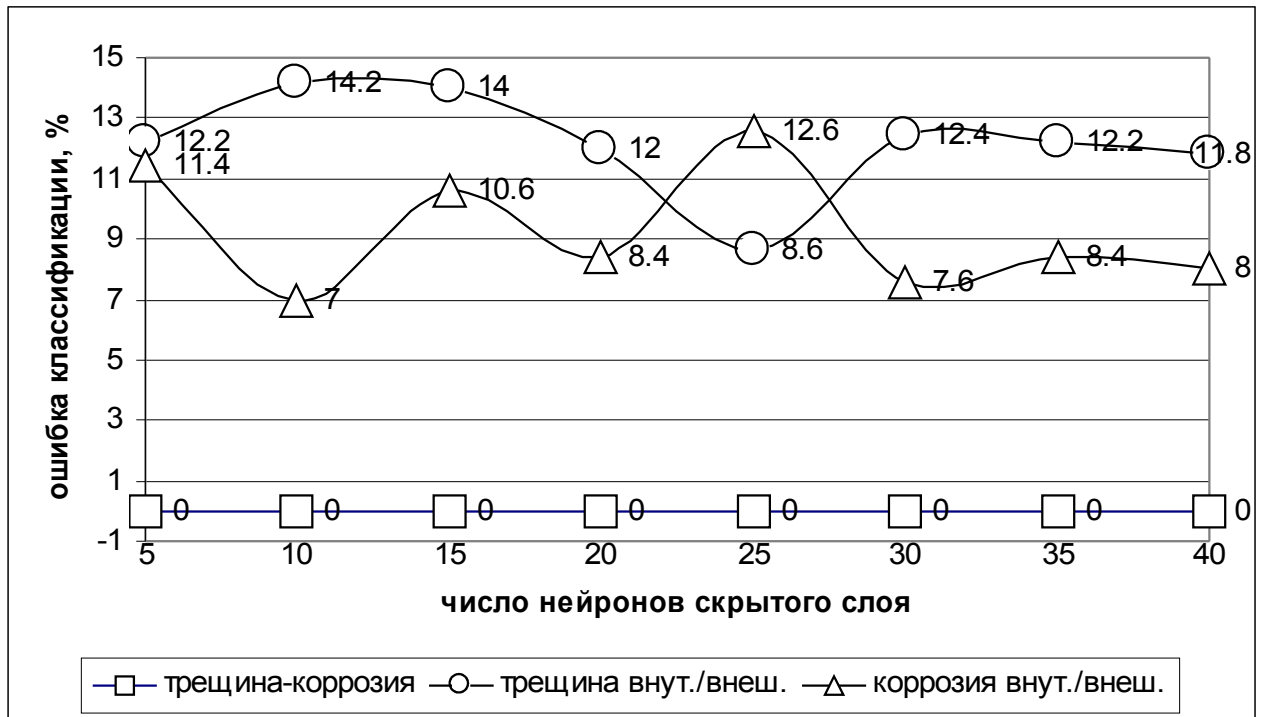


Рис. 3.3. Зависимость ошибки классификации от числа нейронов скрытого слоя для раздельной схемы

На рис. 3.3. приведены результаты классификации для раздельной схемы. Здесь нейронная сеть обучалась и тестировалась на модельных данных (500 векторов в обучающем наборе, 50 в проверочном, 500 в тестовом). Нейронная сеть, для которой показан лучший результат (20 нейронов в скрытом слое), обеспечила минимальную для совместной схемы ошибку классификации в 19.2%. Ошибка на выходе классификатора, построенного по раздельной схеме, составила 9.2% при условии равенства числа нейронов скрытого слоя каждой из сетей 20. Этот результат можно улучшить, если для каждой из сетей индивидуально выбрать размер скрытого слоя. Например, из вышеприведенного графика видно, что наилучший результат получен для 10 нейронов в скрытом слое для сети 2 и 25 для сети 3.

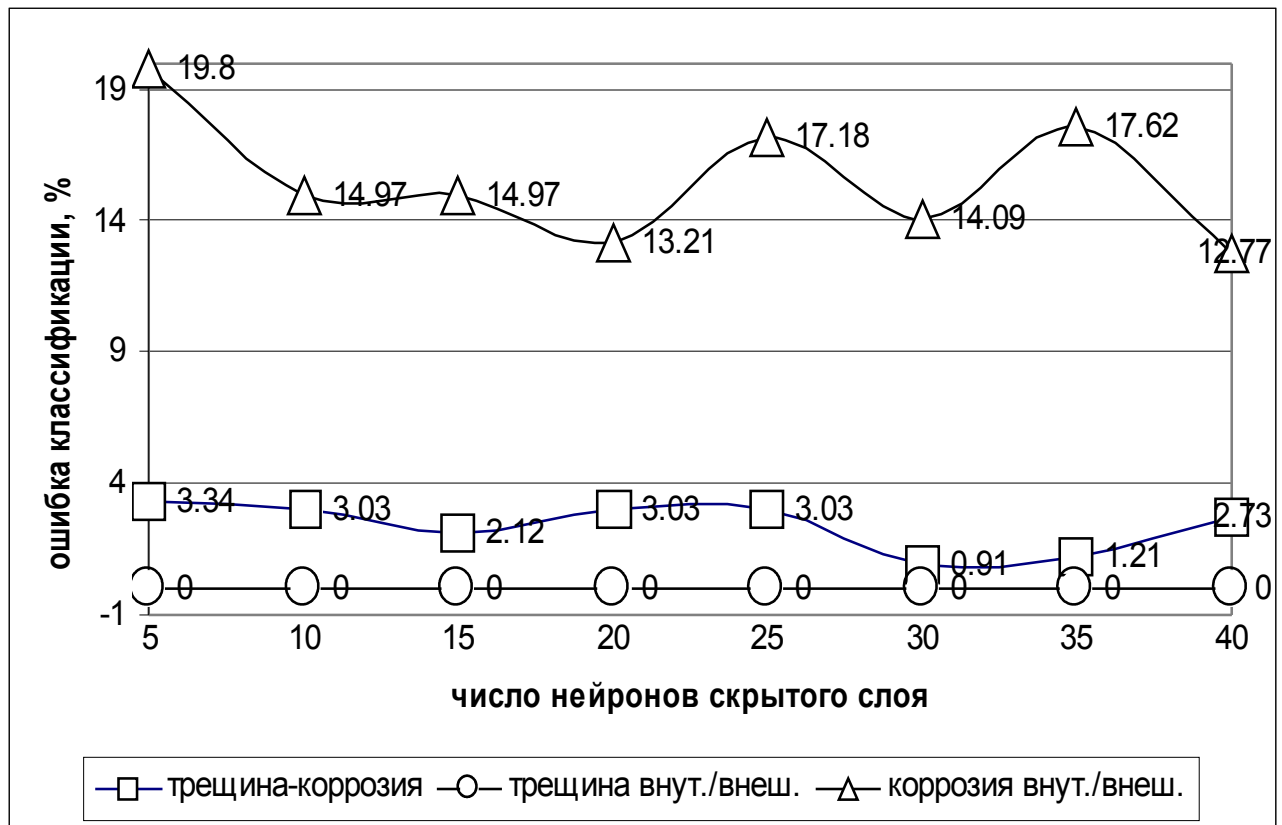


Рис. 3.4. Зависимость ошибки классификации от числа нейронов скрытого слоя для раздельной схемы

На рис. 3.4. приведена ошибка классификации данных от тестовой трубы нейронной сетью обученной на этих же данных (2/3 обучающий набор, 1/3 проверочный набор, всего 329 векторов). Ошибка на выходе классификатора составляет 13.6% при индивидуальном выборе размера скрытого слоя, тогда как для совмещенной схемы при обучении на том же наборе данных была достигнута ошибка 16.7%. Аналогичные результаты получены и при обучении сетей на модельных данных с шумом, при классификации модельных, возмущенных и данных от тестовой трубы.

Полученные результаты показывают, что применение раздельной схемы классификации позволяет снизить ошибку классификации.

4. Особенности определения параметров

Нейронная сеть типа многослойный персептрон с одним скрытым слоем с нелинейной функцией активации и выходным слоем с линейной функцией активации нейронных элементов есть универсальный аппроксиматор. Это означает, что при наличии достаточного количества векторов в обучающей выборке, а также количества нейронов в скрытом слое, нейронная сеть может быть обучена с заданной точностью аппроксимировать непрерывную функцию m переменных.

При наличии зависимости значений параметров дефектов от вектора их признаков вышеназванный эффект может быть использован для определения параметров дефектов:

- длина;
- ширина;
- глубина.

4.1. Схемы параметризации

Для определения нескольких параметров может использоваться как раздельная, так и совмещенная схема. Однако при использовании совмещенной схемы для того, чтобы нейронная сеть запомнила зависимости изменения параметров от вектора признаков, ей требуется гораздо большая емкость. Вследствие чего, резко увеличивается необходимое количество векторов в обучающей выборке. Более того, при использовании совмещенной схемы от сети требуется воспроизводить более сложную зависимость (вследствие увеличения числа выходных переменных), что может обусловить большую ошибку для каждого из определяемых параметров. В виду вышесказанного, очевидно, что следует использовать раздельную схему параметризации, когда выходной слой сети имеет только один нейрон, и сеть вычисляет только один параметр. Таким образом, для вычисления всех параметров необходимо обучить несколько сетей, каждая из которых будет определять только один параметр.

Так как существуют дефекты нескольких типов, то зависимости их параметров от вектора признаков могут отличаться от одного класса дефекта к другому. Есть несколько возможностей учета данной особенности:

- ввести в вектор признаков кодированное значение, обозначающее тип дефекта
- использовать несколько нейронных сетей, каждая из которых будет параметризовать дефекты только одного типа.

Первая схема обладает всеми описанными недостатками совмещенной схемы параметризации, поэтому следует использовать второй вариант. Однако, в виду того, что классификация дефектов на внутренний-внешний имеет большую погрешность, представляется необходимым при параметризации разделять дефекты только на трещины и коррозии.

Таким образом, для определения параметров дефектов необходимо обучить нейронные сети для решения следующих задач:

- определение глубины трещин;
- определение глубины коррозии;
- определение длины трещин;
- определение длины коррозии;
- определение раскрытия трещин;
- определение ширины коррозии.

4.2. Строение нейронной сети для решения задачи параметризации

При параметризации в отличие от классификации вычисляемые значения являются не кодированным признаком принадлежности к классу, а непосредственно значениями параметров. Так как диапазон значений параметров выходит за диапазон $[0; 1]$, то невозможно производить их вычисление при использовании только сигмоидальной функции активации нейронных элементов. Таким образом, появляется необходимость использовать линейную функцию активации для выходного слоя. Возможно также использование линейной функции для некоторых из скрытых слоев.

Для определения наилучшей конфигурации сети для задачи определения дефектов были обучены с использованием алгоритма SARPROP следующие сети:

Таблица 1.

Исследованные структуры нейронных сетей

Номер слоя	1	2	3	4	5
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 20	выходной линейная 1		
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 10	выходной линейная 1		
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 5	выходной линейная 1		

Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 10	скрытый сигмоидальная 5	выходной линейная 1	
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 10	скрытый линейная 5	выходной линейная 1	
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 5	скрытый сигмоидальная 5	скрытый линейная 5	выходной линейная 1
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 5	скрытый линейная 5	скрытый линейная 5	выходной линейная 1
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 5	выходной сигмоидальная 3	выходной линейная 1	
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 5	выходной линейная 3	выходной линейная 1	

Наименьшая ошибка обобщения была достигнута нейронной сетью со следующей конфигурацией

Таблица 2

Наилучшая определенная конфигурация нейронной сети

Номер слоя	1	2	3	4
Функция слоя Функция активации Количество нейронов в слое	входной линейная 13	скрытый сигмоидальная 5	выходной линейная 3	выходной линейная 1

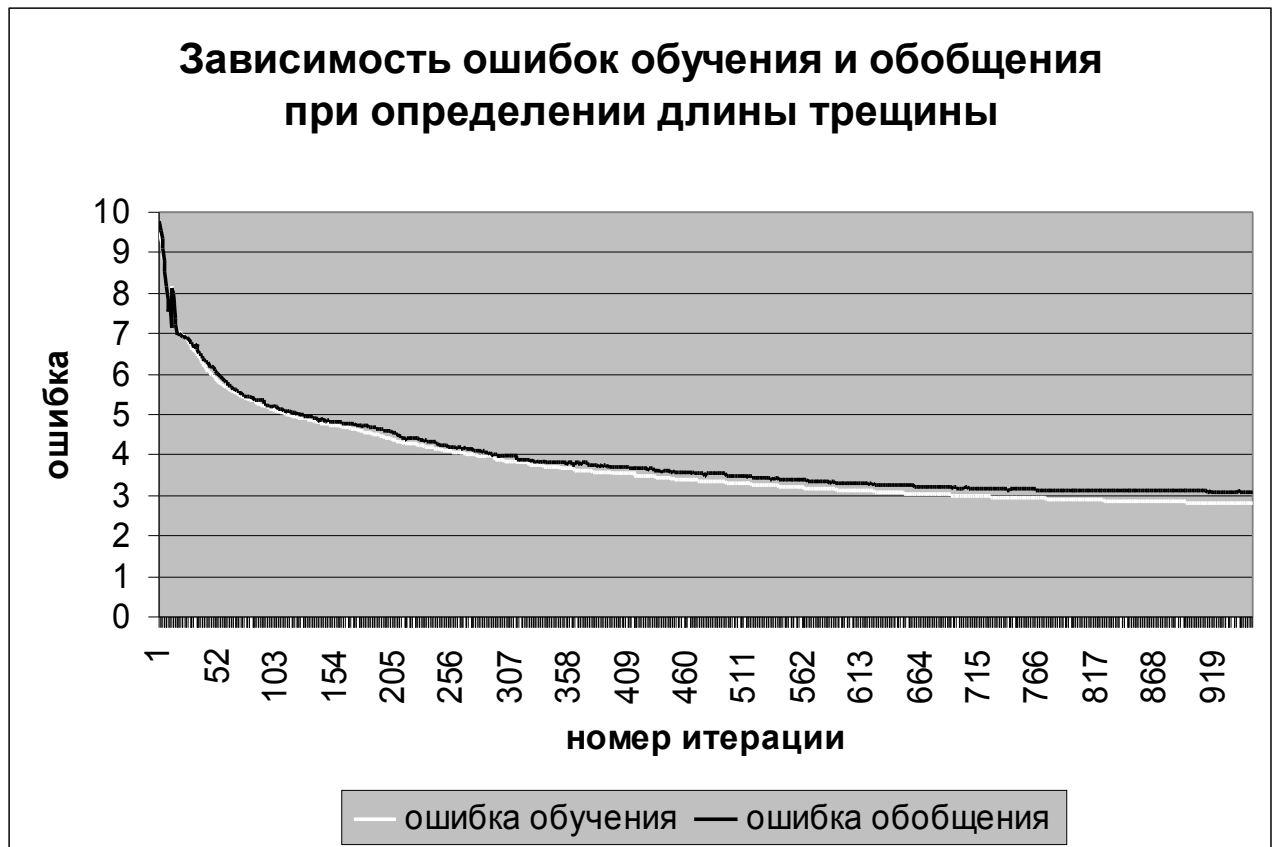


Рис. 4.1. Зависимость ошибок обучения и обобщения при определении параметров (модельные данные в проверочном наборе)

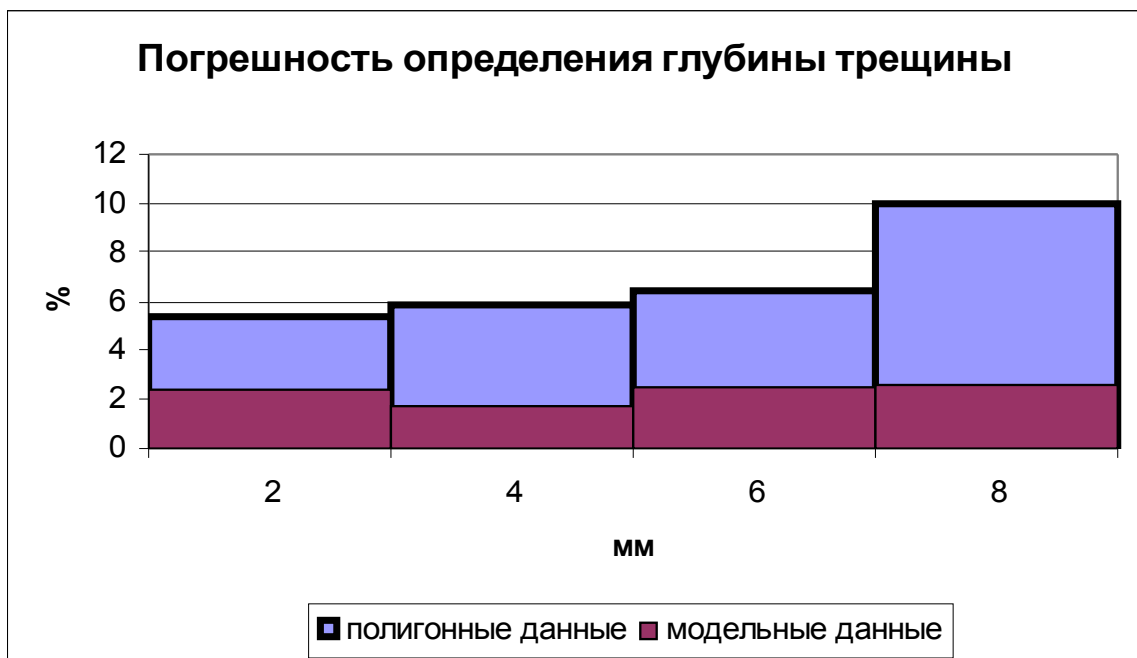


Рис. 4.2. Ошибка определения глубины трещины

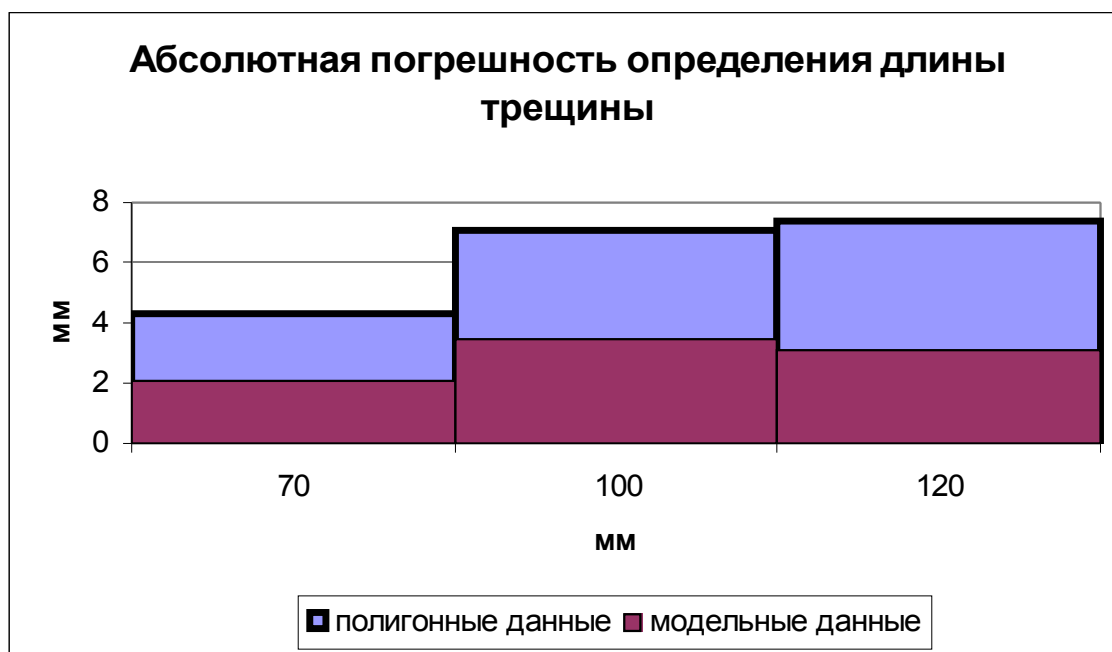


Рис. 4.3. Ошибка определения длины трещины

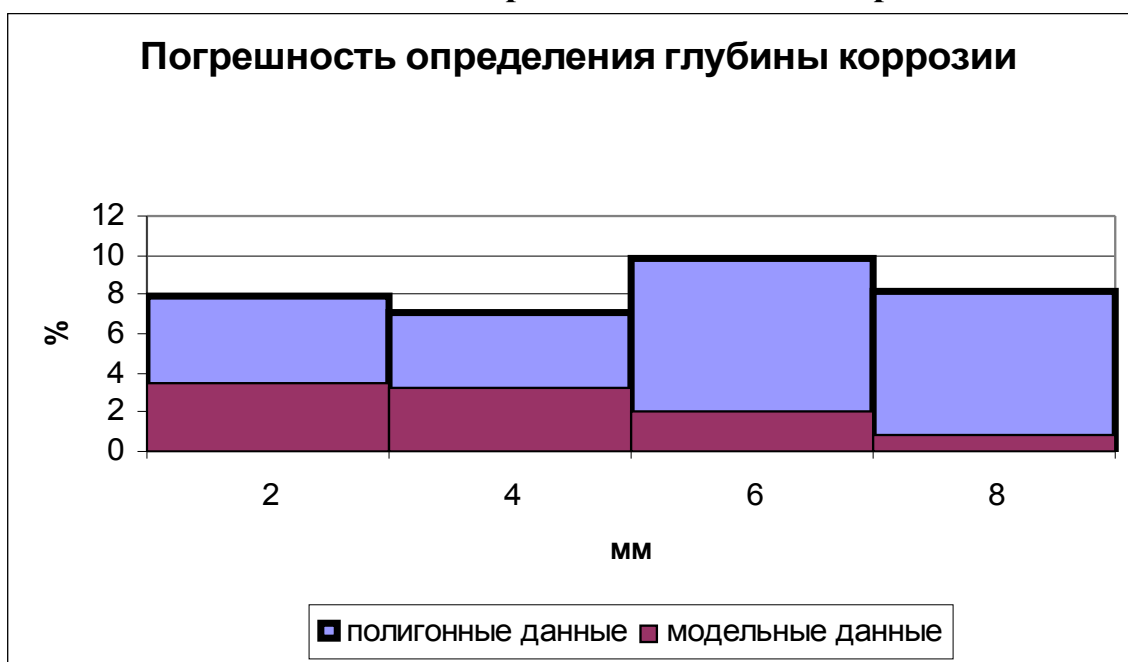


Рис. 4.4. Ошибка определения глубины коррозии

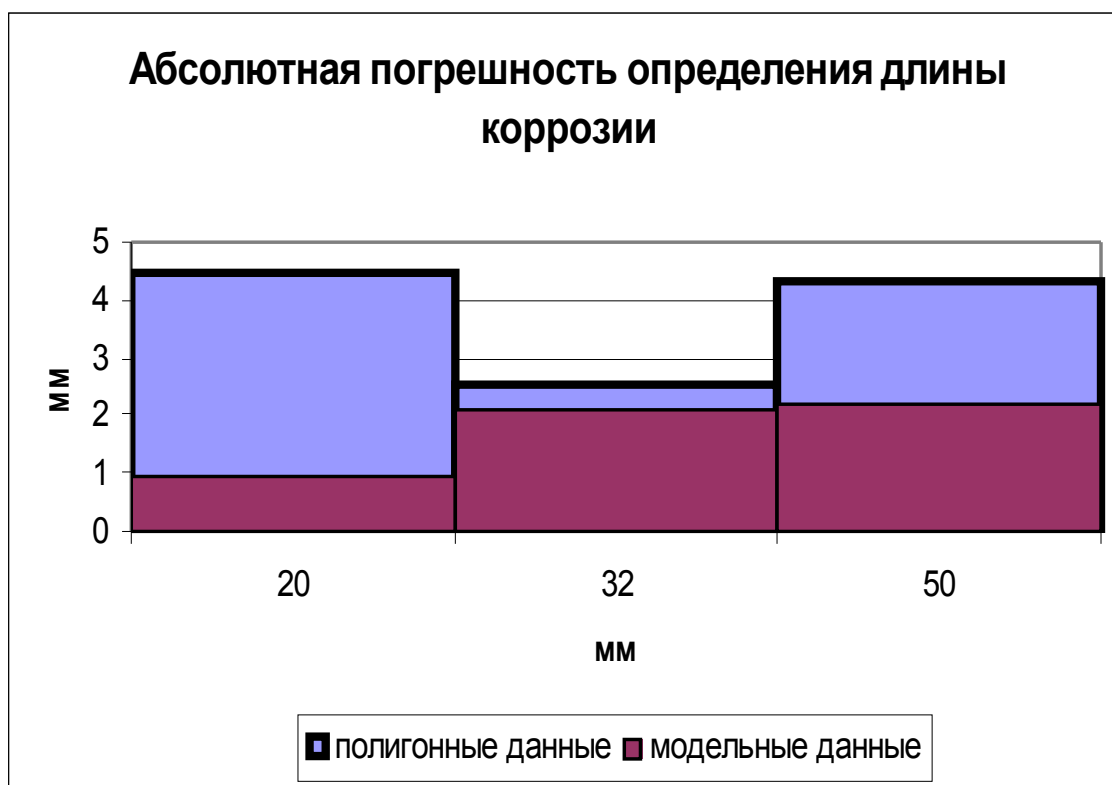


Рис. 4.5. Ошибка определения длины коррозии

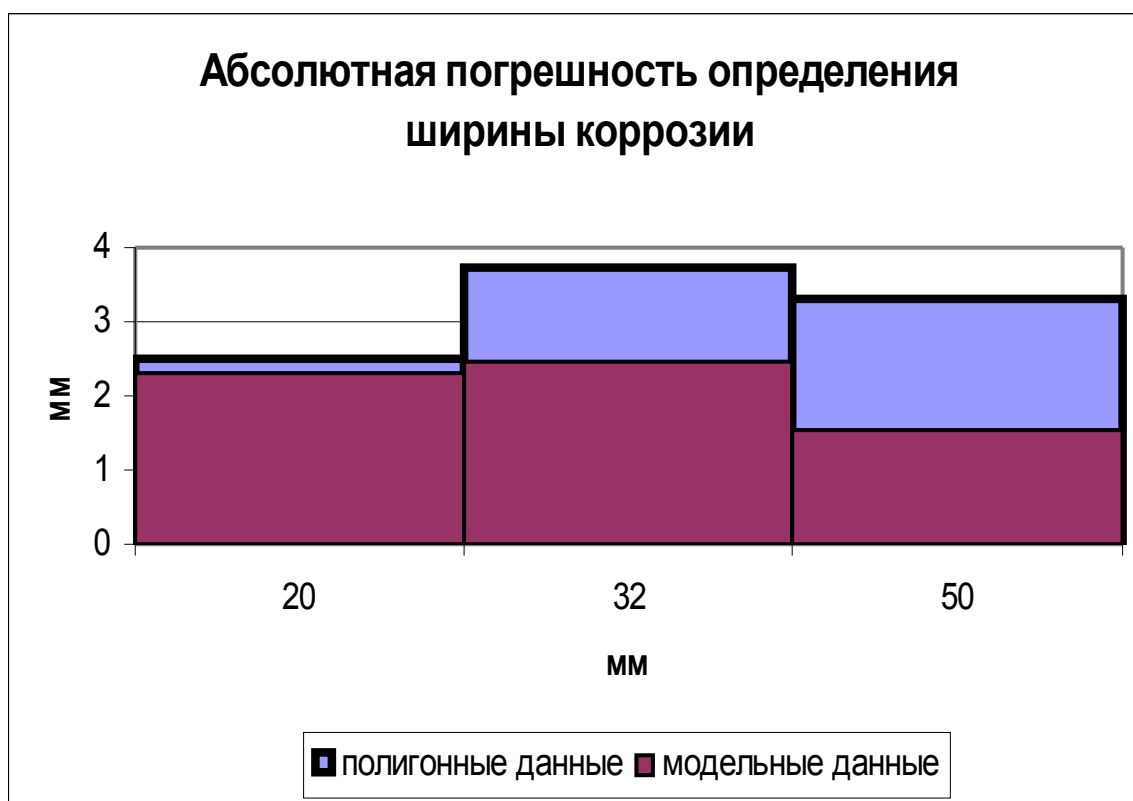


Рис. 4.6. Ошибка определения ширины коррозии

Приведенные выше результаты были получены при использовании в обучающем наборе 1500 векторов признаков модельных дефектов и 500 в проверочном наборе.

Значительная погрешность определения глубины трещины и коррозии для полигонных данных вызвана тем, что толщина стенки трубы здесь лежит в диапазоне 7.8-10 мм, в то время как база модельных дефектов включает признаки, рассчитанные только для толщины стенки трубы 8 мм. В результате этого нейронная сеть дает заниженную оценку глубины дефекта. Для устранения данного недостатка необходимо дополнительно включить в базу дефекты на трубе с толщиной стенки 10 мм.

Приложение

Классификация

Результаты классификации для различных наборов обучающих, проверочных, тестовых данных и схем при использовании алгоритма RPROP.

Схема классификации	Данные обучающего набора	Данные проверочного набора	Число нейронов в скрытом слое	Ошибка классификации, %		
				модельные	модельные с шумом	реальные
совмещенная	модельные с шумом	данные от тестовой трубы	5	30	30	23.4043
			10	31	32.2	24.3161
			15	28.4	31.4	24.924
			20	27.4	30.6	21.8845
			25	27.4	29	23.1003
			30	29	30.6	23.4043
			35	28	30.6	23.4043
			40	27.8	30.4	24.6201
совмещенная	модельные	данные от тестовой трубы	5	32.6	29.4	21.8845
			10	32.4	29	22.7964
			15	25.8	28	23.4043
			20	23.2	28	22.7964
			25	33.2	29.2	22.7964
			30	33.2	29.6	24.924
			35	32.2	30.4	24.924
			40	33.6	30.4	24.6201
совмещенная	модельные	модельные	5	26.4	29.2	27.0517
			10	19.8	17.4	34.6505
			15	21.6	21.2	26.4438
			20	19.2	20.4	30.0912
			25	19.8	21.8	31.003
			30	21.2	20.2	31.003
			35	21.8	25	30.0912
			40	21.4	20.8	32.8267
совмещенная	данные от тестовой трубы	данные от тестовой трубы	5			22.7964
			10			20.0608
			15			19.7568
			20			20.9726

			25			17.0213
			30			19.1489
			35			16.7173
			40			24.924
трещина-коррозия	модельные с шумом	данные от тестовой трубы	5	0	0	2.43161
			10	0	0	2.73556
			15	0	0	2.73556
			20	0	0	2.73556
			25	0	0	2.73556
			30	0	0	2.73556
			35	0	0	2.73556
			40	0	0	2.43161
трещина-коррозия	модельные	данные от тестовой трубы	5	0	0	2.43161
			10	0	0	2.12766
			15	0	0	2.43161
			20	0	0	2.73556
			25	0.4	0	2.12766
			30	0	0	3.03951
			35	0	0	2.73556
			40	0	0	2.73556
трещина-коррозия	модельные	модельные	5	0	0	2.73556
			10	0	0	3.34347
			15	0	0	3.34347
			20	0	0	3.34347
			25	0	0	3.95137
			30	0	0.2	3.34347
			35	0	0	4.25532
			40	0	0	3.03951
трещина-коррозия	данные от тестовой трубы	данные от тестовой трубы	5			3.34347
			10			3.03951
			15			2.12766
			20			3.03951
			25			3.03951
			30			0.911854
			35			1.21581
			40			2.73556
трещина-внутренняя-внешняя	модельные с шумом	данные от тестовой трубы	5	33	32.6	14.7059
			10	31.8	33.6	12.7451

			15	32	33.4	14.7059
			20	31.8	32.6	11.7647
			25	32	32.8	11.7647
			30	31.4	32.6	14.7059
			35	33.8	34.4	11.7647
			40	45.4	44.8	11.7647
трещина внутренняя- внешняя	модельные	данные от тестовой трубы	5	26.2	28.8	19.6078
			10	25.4	30	11.7647
			15	23.8	27	14.7059
			20	24.4	27	10.7843
			25	22	26	10.7843
			30	23.6	28.6	14.7059
			35	24	25.8	11.7647
			40	24.4	27.8	12.7451
трещина внутренняя- внешняя	модельные	модельные	5	12.2	18.4	47.0588
			10	14.2	15.2	43.1373
			15	14	17.2	34.3137
			20	12	15.2	51.9608
			25	8.6	17.2	49.0196
			30	12.4	16.8	34.3137
			35	12.2	16.2	38.2353
			40	11.8	15.8	44.1176
трещина внутренняя- внешняя	данные от тестовой трубы	данные от тестовой трубы	5			0
			10			0
			15			0
			20			0
			25			0
			30			0
			35			0
			40			0
коррозия внутренняя- внешняя	модельные с шумом	данные от тестовой трубы	5	20	22.2	23.7885
			10	22.4	22.8	22.9075
			15	18.8	21	22.9075
			20	20.8	20.6	24.6696
			25	16	15.8	21.5859
			30	20	21.8	23.7885
			35	19.8	21.6	22.467

			40	14.8	12.4	24.6696
коррозия внутренняя- внешняя	модельные	данные от тестовой трубы	5	15.2	17.2	21.1454
			10	17.4	20	21.5859
			15	15.8	17.2	19.3833
			20	15	16.8	19.8238
			25	17.2	19.4	20.2643
			30	18.4	20.6	22.467
			35	16	18.4	21.5859
			40	17.8	18.8	20.2643
коррозия внутренняя- внешняя	модельные	модельные	5	11.4	12.8	28.6344
			10	7	10	28.1938
			15	10.6	10.8	29.0749
			20	8.4	9.6	27.3128
			25	12.6	13.4	30.837
			30	7.6	7.6	32.5991
			35	8.4	10.4	33.0396
			40	8	7.8	30.837
коррозия внутренняя- внешняя	данные от тестовой трубы	данные от тестовой трубы	5			19.8238
			10			14.978
			15			14.978
			20			13.2159
			25			17.1806
			30			14.0969
			35			17.6211
			40			12.7753

Количество векторов в обучающем, проверочном и тестовом наборах для различных схем.

Тип набора	Схема классификации	Данные набора*	Количество векторов в наборе
обучающий	совмещенная	модельные	500
		от тестовой трубы	110
		модельные с шумом	500
	трещина-коррозия	модельные	500
		от тестовой трубы	110
		модельные с шумом	500
	трещина внутренняя-внешняя	модельные	500
		от тестовой трубы	102

		модельные с шумом	500
	коррозия внутренняя-внешняя	модельные	500
		от тестовой трубы	34
		модельные с шумом	500
проверочный	совмещенная	модельные	250
		от тестовой трубы	110
	трещина-коррозия	модельные	250
		от тестовой трубы	110
	трещина внутренняя-внешняя	модельные	250
		от тестовой трубы	102
	коррозия внутренняя-внешняя	модельные	250
		от тестовой трубы	34
тестовый	совмещенная	модельные	1000
		от тестовой трубы	329
	трещина-коррозия	модельные	1000
		от тестовой трубы	329
	трещина внутренняя-внешняя	модельные	1000
		от тестовой трубы	102
	коррозия внутренняя-внешняя	модельные	1000
		от тестовой трубы	227

* Вектора для наборов из модельных данных и возмущенных данных выбирались случайным образом из базы признаков дефектов содержащей 5193 вектора для коррозии и 7200 векторов для трещины.

Результаты классификации для различных наборов обучающих, проверочных, тестовых данных и схем при использовании алгоритма SARPROP.

Схема классификации	Данные обучающего набора	Данные проверочного набора	Число нейронов в скрытом слое	Ошибка классификации, %	
				модельные	полигонные и данные от тестовой трубы
трещина-коррозия	модельные	модельные	20	0	5.66
	модельные с шумом	данные от тестовой трубы			5.44
трещина внешняя-внутренняя	модельные	модельные	20	10.06	23.38

	модельные с шумом	данные от тестовой трубы	20		10.48
коррозия внутренняя-внешняя	модельные	модельные	20	3.88	33.73
	модельные с шумом	данные от тестовой трубы	20		25.97

Определение параметров

Результаты определения параметров дефектов *.

		Диапазон		Модельные данные		Полигонные данные	
Тип дефекта	Параметр	минимум, мм	максимум, мм	Абсолютная погрешность, мм	Относительная погрешность, %	Абсолютная погрешность, мм	Относительная погрешность, %
трещина	глубина	0.4	2		2.375		5.375
		2	4		1.75		5.875
		4	6		2.5		6.375
		6	8		2.625		10
	длина	40	70	2.04		4.33	
		70	100	3.45		7.09	
		100	120	3.08		7.35	
коррозия	глубина	0.4	2		3.5		7.875
		2	4		3.25		7.125
		4	6		2		9.875
		6	8		0.875		8.125
	длина	6	20	0.94		4.48	
		20	32	2.11		2.56	
		32	50	2.23		4.34	
	ширина	6	20	2.31		2.5	
		20	32	2.48		3.75	
		32	50	1.53		3.29	

ЗАКЛЮЧЕНИЕ

Результатом работы стало предложение нескольких схем работы с нейронными сетями типа многослойный персептрон для классификации дефекта по вектору признаков и определения его параметров. Оптимальной является отдельная схема классификации и параметризации, дающая наименьшую погрешность.

При наличии достаточного количества данных от тестовой трубы с заранее известными параметрами дефектов следует включить эти данные в обучающий и проверочный наборы для достижения минимальной ошибки при использовании нейронной сети.

Следует использовать улучшенные схемы обучения нейронной сети, например групповую стратегию обучения и алгоритм устойчивого обратного распространения ошибки (RPROP, SARPROP). А также критерии ранней остановки процесса обучения.

Важным моментом является улучшение базы модельных дефектов с целью отражения в ней зависимости значений признаков от влияния мешающих факторов, а также исключение тех признаков, в значении которых главную роль играют неинформативные параметры.

Библиографический список

1. А. А. Абакумов, А. А. Абакумов (мл.), Магнитная диагностика газонефтепроводов. - М.: Энергоатомиздат, 2001, 440 с.
2. В. Г. Герасимов, А. Д. Покровский, В. В. Сухоруков, Неразрушающий контроль. В 5 кн. Кн. 3. Электромагнитный контроль. - М.: Высш. шк., 1992, 312 с.
3. В. А. Головкин, Нейронные сети: обучение, организация и применение. - М.: ИПРЖ, 2001, 256 с.
4. С. Осовский, Нейронные сети для обработки информации, Финансы и статистика, 2002, 344 с.
5. Yann Le Cun, Efficient BackProp. - Holmdel, 1996, 75 с.
6. Donald R. Tvetter, Backpropagator's Review, 1996
7. Martin Riedmiller, Advanced Supervised Learning in Multi-layer Perceptrons - From Backpropagation to Adaptive Learning Algorithms - Karlsruhe, 1994, 10 с.
8. Merten Joost, Speeding up backpropagation algorithms by using cross-entropy combined with pattern normalization, 1997, 20 с.
9. Lutz Prechelt, Automatic early stopping using cross validation: quantifying the criteria. - Karlsruhe, 1997, 12 с.
10. Martin Riedmiller, RPROP - A fast adaptive learning algorithm. - Karlsruhe, 1992, 12 с.
11. Martin Riedmiller, RPROP - Description and implementation details. - Karlsruhe, 1994, 16 с.
12. W. Shiffman, Comparison of optimized backpropagation algorithms. - Koblenz, 1993, 8 с.
13. Stewe Lawrence, What size neural network gives optimal generalization? Convergence properties of backpropagation. - Oxford, 1996, 37 с.
14. B. D. Ripley, Statistical ideas for selecting network architectures. - Princeton, 1995, 8 с.
15. N.K. Treadgold, T.D. Gedeon, THE SARPROP ALGORITHM: A SIMULATED ANNEALING ENHANCEMENT TO RESILIENT BACK PROPAGATION. - Sydney, 1998, 7 с.
16. P. A. Castilo, SA-Prop: Optimization of multilayer perceptron parameters using simulated annealing. - Granada, 1998, 10 с.

Содержание

Аннотация.....	3
ВВЕДЕНИЕ.....	4
1. Схема обработки данных.....	5
1.1. Источник входных данных.....	5
1.2. Программная система.....	14
1.2.1. Командный интерпретатор.....	16
1.2.2. Подсистема работы с нейронными сетями.....	16
1.2.3. Использование объектной технологии	17
2. Общие вопросы использования нейронных сетей.....	18
2.1. Строение сети.....	18
2.2. Модель нейрона.....	20
2.3. Функция активации.....	22
2.4. Инициализация параметров нейронной сети.....	23
2.5. Кодирование данных.....	23
2.5.1. Кодирование входных данных.....	23
2.5.2. Кодирование выходных данных.....	24
2.6. Стратегия обучения.....	25
2.7. Алгоритм обучения.....	25
2.7.1. Прямое распространение сигнала.....	26
2.7.2. Обратное распространение ошибки.....	26
2.7.3. Изменение весов.....	27
2.7.3.1. Алгоритм RPROP.....	27
2.7.3.2. Алгоритм SARPROP.....	28
2.7.4. Остановка процесса обучения.....	30
2.7.4.1. Переобучение.....	31
2.7.4.2. Ранняя остановка процесса обучения.....	32
2.8. Выбор числа нейронов скрытого слоя.....	33
2.8. Выбор данных для обучения.....	38
2.8.1. Использование шума в обучающем наборе.....	38
2.8.2. Использование в обучающем наборе данных с тестовых участков труб.....	40
2.9. Результаты тестирования.....	40
2.9.1. Алгоритм обучения.....	40
2.9.2. Выбор числа нейронов скрытого слоя.....	41
2.9.3. Использование тестовых данных при обучении.....	44
3. Особенности классификации.....	47
3.1. Определение типа дефекта.....	47
3.2. Оптимизация схемы распознавания.....	47

4. Особенности определения параметров.....	51
4.1. Схемы параметризации.....	51
4.2. Строение нейронной сети для решения задачи параметризации.....	52
Приложение.....	58
Классификация.....	58
Определение параметров.....	63
ЗАКЛЮЧЕНИЕ.....	64
Библиографический список.....	65