

ФИЗИЧЕСКИЕ ПРОЦЕССЫ В ЭЛЕКТРОННЫХ ПРИБОРАХ

УДК 621.396

ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ В ЗАДАЧЕ КЛАССИФИКАЦИИ ДЕФЕКТОВ ЭЛЕКТРОННЫХ УСТРОЙСТВ

© 2021 г. Л. К. Х. Фам^а, К. И. Бушмелева^{б, *}, А. С. Увайсова^а, С. У. Увайсов^а

^аМИРЭА – Российский технологический университет,
просп. Вернадского, 78, Москва, 119454 Российская Федерация

^бСургутский государственный университет,
ул. Энергетиков, 22, Сургут, 628408 Российская Федерация

*E-mail: bkiya@yandex.ru

Поступила в редакцию 16.09.2020 г.

После доработки 20.04.2021 г.

Принята к публикации 23.04.2021 г.

Рассмотрена проблема идентификации и классификации механических дефектов электронных устройств, предложено ее решению с помощью искусственной нейронной сети. Показано, как с использованием искусственных нейронных сетей можно получить более быстрые решения проблемы и при этом достичь требуемой достоверности. В результате экспериментов было построено несколько нейронных сетей при различных входных условиях, которые прошли обучения и были запущены в эксплуатацию.

DOI: 10.31857/S0033849421090047

ВВЕДЕНИЕ

Активное развитие технологий производства электронных устройств (ЭУ), постоянное стремление обеспечить их высокое качество и надежность способствуют появлению совершенных методов контроля и диагностики различных дефектов, которые неизбежно проявляют себя на производстве любой сложности.

Так как дефект в изделии негативно сказывается на его эксплуатационных параметрах и даже может привести к серьезным авариям, наука и производство всегда находятся в поисках способов повышения такого показателя как надежность. Одним из способов повышения надежности систем в целом и ЭУ в частности является техническая диагностика. Поскольку значительная доля дефектов электронных устройств имеет механическую природу, на сегодняшний день процессы контроля и диагностирования указанных изделий связаны прежде всего с отслеживанием их электрических характеристик, которые чаще всего поддаются выявлению в процессе эксплуатации. Однако важность механической диагностики состоит в том, что обнаружение некоторых конструктивных недостатков, как правило, возможно только в рамках приемочного контроля. Исследования показали, что наибольшее влияние на ЭУ оказывает сочетание вибрационных воздействий и одиночных ударов [1]. По этой причине данные виды испытаний имеют наибольший приоритет,

остальные нагрузки при этом можно отнести к дополнительным.

В данной работе принцип диагностического исследования состоит в том, что на ЭУ оказывается механическое, в частности ударное, воздействие. Реакции ЭУ на это воздействие являются исходными данными, используемыми при проведении диагностики в разрабатываемой системе. В механике удар – это некоторый кратковременный скачкообразный процесс. Во время удара в точке соприкосновения твердых тел возникают большие, но мгновенно действующие силы, вызывающие некоторый конечный импульс возбуждения, представляющий собой реакцию изделия. Конечный результат зависит от динамических параметров (масса, жесткость) исследуемого объекта.

Известно, что любая конструкция представляет собой анизотропную гетерогенную структуру, обладающую собственными механическими характеристиками, например, распределением массы в конструкции. При смене механической характеристики меняется картина отклика системы вследствие изменения распределения массы. Таким образом, проводя испытания на ударные воздействия электронных устройств, можно выделить определенные зависимости.

Предложенный в данной работе метод автоматизированной диагностики ЭУ основан на создании математической модели, показывающей ампли-

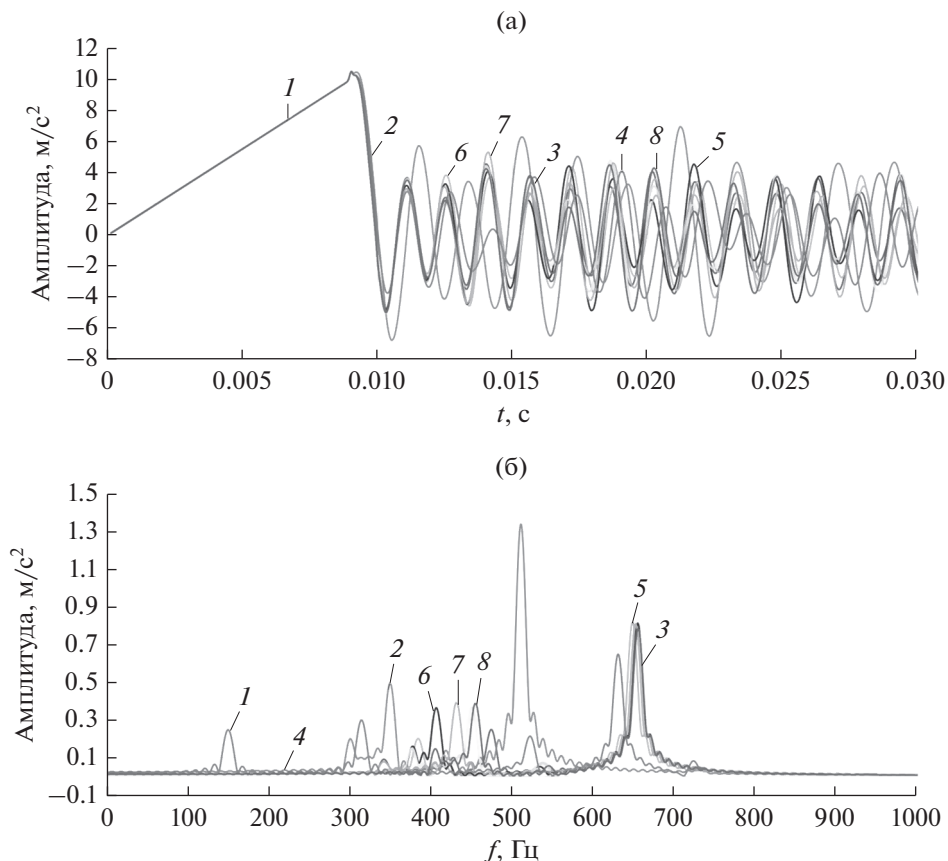


Рис. 1. Амплитудно-временная (а) и амплитудно-частотная (б) характеристики реакции на ударное воздействие: 1 – деформация электрорадиоизделий, 2 – изменение материала платы, 3 – норм, 4 – ослабление крепежа, 5 – изменение компоновки платы, 6 – наличие в ЭС посторонних элементов, 7 – образование трещины в конструкции, 8 – отсутствие электрорадиоизделий.

тудно-временную зависимость отклика системы от ударного воздействия. При этом сравнение результатов, полученных после ударного воздействия, с базой данных механических характеристик, полученных ранее для печатного узла [2, 3], позволяет классифицировать и идентифицировать дефекты. Однако при проведении данных исследований возникает проблема в сборе и обработке информации, полученной датчиками после реакции ЭУ на ударное воздействие. Применение общих математических методов для анализа полученных характеристик, сравнения их и получения последующих выводов о дефектах очень сложный процесс и отнимает много времени.

В данной работе предложено использовать искусственную нейронную сеть (ИНС), которая в свою очередь позволит более точно решить указанную проблему и потребует на это меньше времени.

Сегодня нейронные сети используются для решения различных сложных задач, которые требуют аналитических вычислений, подобных процессам, протекающим в мозге человека. Самым распространенным применением нейронных сетей яв-

ляется классификация, предсказание, распознавание.

Основная идея разрабатываемой ИНС состоит в анализе подобных ударных воздействий на печатные узлы ЭУ. Поскольку реакция или отклик узла будет различным в зависимости от наличия тех или иных конструктивных дефектов, задача нейронной сети будет состоять в том, чтобы научиться классифицировать результаты ударных воздействий, выделяя ряд признаков, которые характеризуют конкретный дефект либо его отсутствие. Так как для обучения классификации нейронной сети всегда нужен учитель, то в данном случае в качестве учителя выступает массив размеченных объектов или образов (текст, изображение) с указанием того, к какому классу относится каждый объект.

В данной работе база данных (БД), или массив размеченных объектов [4], представляет собой набор графиков, амплитудно-временных и амплитудно-частотных характеристик реакции ЭУ на ударное воздействие, соответствующих определенному типу дефекта (рис. 1а и 1б), так как каждый тип дефекта имеет свой уникальный гра-

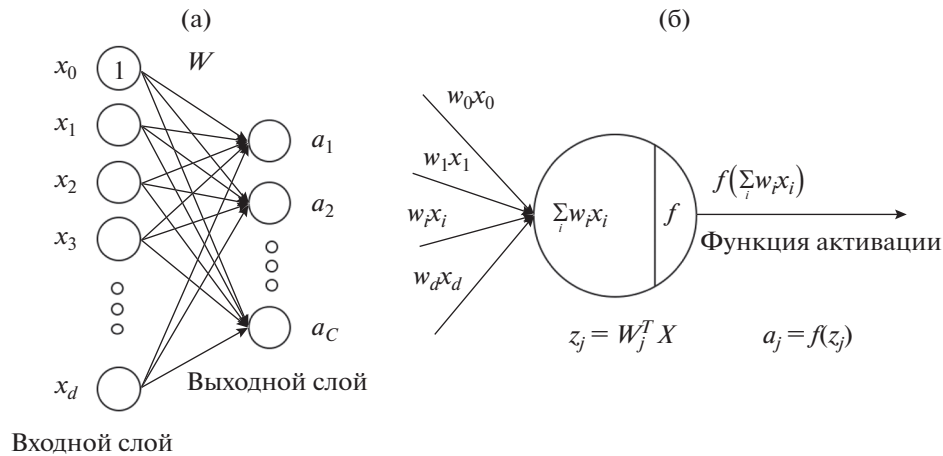


Рис. 2. Модель нейронной сети: один входной и один выходной слой (а), математические действия в нейроне (б).

фик, который отличается от других. В свою очередь, когда появляется новый график, его нужно сравнить с имеющимися в БД после чего возможно классифицировать тип дефекта.

Таким образом, в работе поставлена задача разработать ИНС на основе созданной БД для выявления и классификации дефектов.

Для реализации поставленной задачи на первом этапе необходимо нормализовать графики, приведя их к соответствующим координатным точкам, и далее представить эти точки в виде набора чисел. Эти наборы чисел будут помечены в соответствии с predetermined типом дефекта в БД, после чего их будут использовать для обучения ИНС [5–7].

1. МОДЕЛЬ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

При формировании информации для входного слоя ИНС необходимо иметь совокупность точек, при этом следует обратить внимание на следующие моменты: если разделить график на несколько точек, то эти точки не будут представлять полную характеристику графика; если мы разделим график на слишком много точек, то процесс расчета усложнится и отнимет много времени. Поэтому необходимо разделить график на такое количество точек, которое позволит наиболее полно отображать характеристику графика. В результате было предложено разделить график на 1000 точек и занести эти данные в БД. Таким образом, входной слой ИНС будет иметь 1000 нейронов ($d = 1000$) [8].

В результате может быть создана ИНС, представленная на рис. 2, где входная матрица будет содержать совокупность точек X , выходная матрица Y , а весовая матрица будет содержать W . При этом данные для матриц X и Y известны, стоит за-

дача – найти матрицу W , это можно сделать следующим образом:

$$XW = Y, \quad (1)$$

$$X^{-1}XW = X^{-1}Y, \quad (2)$$

$$W = X^{-1}Y. \quad (3)$$

Однако в данном случае применить этот метод невозможно, так как матрица X не всегда является квадратной и, как следствие, матрицу X^{-1} невозможно найти. Кроме того, если матрица X слишком велика, то для нахождения матрицы X^{-1} требуется много вычислений.

При начальной инициализации допустим, что матрица W будет матрицей случайных чисел, вычислим XW и сравним ее с матрицей Y , после чего на основе выявленных ошибок можно постепенно корректировать весовую матрицу W .

Функция активации – это способ нормализации входных данных, т.е. если на входе у нас будет большое число, то, пропустив его через функцию активации, можно получить выходные данные в нужном диапазоне.

Для данного исследования необходимо задать вероятность такой, чтобы для каждого входа X выход $a_i = f(Y)$ представлял собой вероятность того, что вход попадает в класс i . Таким образом, условие будет состоять в том, что каждое a_i является положительным числом, а их сумма была равна 1. Чтобы выполнить данное условие, необходимо оценить каждое значение $z_i = w_i^T X$ и на основе отношения между этими z_i вычислить значение a_i . В дополнение к условию, что $a_i > 0$ и их сумма равна 1, добавим еще одно условие: чем больше значение z_i , тем выше вероятность того, что данные попадут в класс i . Последнее условие

Таблица 1. Кодирование дефектов

Тип дефектов	Десятичный код	Одно горячее кодирование
Деформация электрорадиоизделия	1	1 0 0 0 0 0 0 0
Изменение компоновки платы	2	0 1 0 0 0 0 0 0
Изменение материала платы	3	0 0 1 0 0 0 0 0
Наличие в ЭУ посторонних элементов	4	0 0 0 1 0 0 0 0
Норма	5	0 0 0 0 1 0 0 0
Образование трещины в конструкции	6	0 0 0 0 0 1 0 0
Ослабление крепежа	7	0 0 0 0 0 0 1 0
Отсутствие электрорадиоизделия	8	0 0 0 0 0 0 0 1

указывает на то, что в данном случае нужно использовать возрастающую функцию.

Необходимо также отметить, что z_i может получать как положительные, так и отрицательные значения. Простая гладкая функция, безусловно, может превратить z_i в положительное значение, кроме этого, данное преобразование может произвести возрастающая функция $\exp(z_i)$. Конечное условие, т.е. когда сумма $a_i = 1$, может быть гарантировано, если:

$$a_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (4)$$

2. ВЫБОР ФУНКЦИИ ДЛЯ РЕАЛИЗАЦИИ МОДЕЛИ

В указанном сетевом представлении каждый выход больше не будет значением, соответствующим каждому классу, но будет вектором с одним элементом, равным 1, остальные элементы равны 0. Элемент равен 1 в позиции, соответствующей определенному классу, это в свою очередь указывает на то, что рассматриваемая точка попадает в класс с вероятностью 1. Способ кодирования вывода представлен в табл. 1 [9].

При использовании модели регрессии softmax для каждого входа x мы имеем прогнозируемый результат $a = \text{softmax}(x, W)$. Между тем фактический результат – это вектор y , представленный в виде одного горячего кодирования.

Функция потерь будет построена таким образом, чтобы минимизировать разницу между прогнозируемым выходом a и фактическим выходом y . В результате будет получено следующее выражение:

$$J(W) = \sum_{i=1}^N (a_i y_i)^2 \quad (5)$$

Тем не менее это не лучший вариант решения, существуют более эффективные, например, перекрестная энтропия, которая между двумя распределениями p и q определяется в виде

$$H(p, q) = E_p[\lg q]. \quad (6)$$

Поскольку p и q дискретны, как u и a в реализуемой задаче, то эта формула может быть записана в следующем виде:

$$H(p, q) = \sum_{i=1}^C p_i \lg q_i. \quad (7)$$

Чтобы лучше понять преимущества функции перекрестной энтропии по сравнению с функцией квадрата расстояния, рассмотрим пример на рис. 3, когда $C = 2$, а p_1 принимает значения 0.5, 0.1 и 0.8.

Из анализа рис. 3 можно сделать два важных вывода:

- минимальное значение обеих функций достигается при $q = p$ в координатах зеленых точек;

- функция перекрестной энтропии получает очень высокое значение (т.е. очень большие потери), когда q находится далеко от p . Между тем разница между потерями вблизи или вдали от решения функции квадрата расстояния $(p - q)^2$ незначительна. Оптимальным является то, что функция перекрестной энтропии дает решение, близкое к p , потому что удаленные решения строго наказываются.

Для оптимизации функции потерь в работе использован стохастический градиентный спуск (8). В этом алгоритме одновременно вычисляется производная функции потерь на основе только одной точки данных x_i , после чего обновляется W на основе этой производной. Подобные вычисления производятся для каждой точки всех полученных данных, затем описанный выше процесс повто-

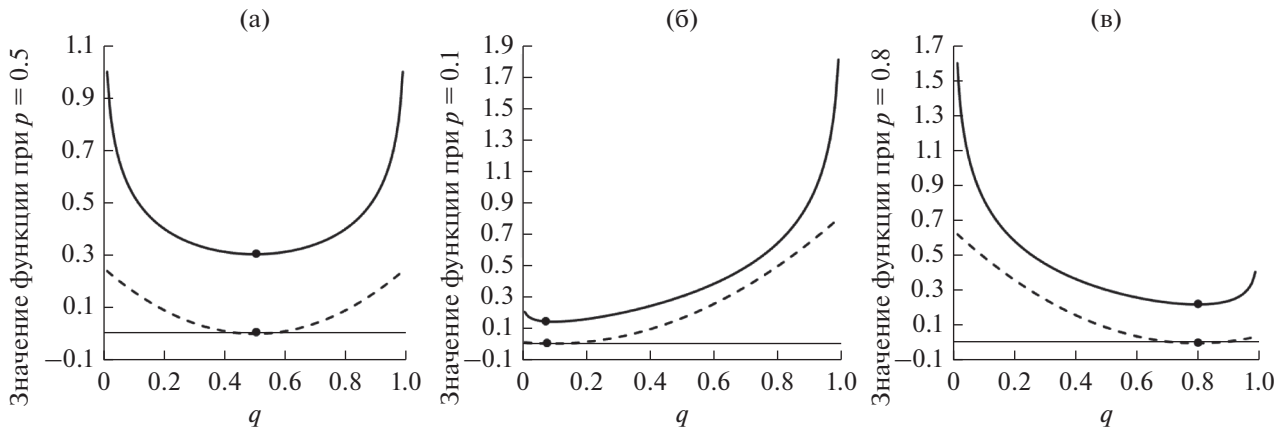


Рис. 3. Сравнение функций перекрестной энтропии (сплошная кривая) и квадрата расстояния (штриховая) при $p = 0.5$ (а), 0.1 (б) и 0.8 (в).

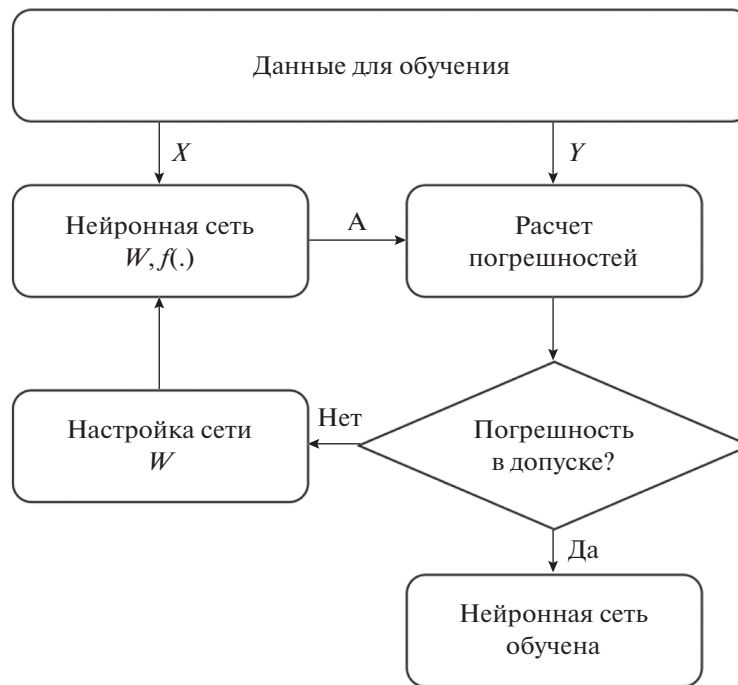


Рис. 4. Алгоритм процесса обучения искусственной нейронной сети.

руется. При этом формула для расчета градиентов имеет вид

$$\frac{\partial J(W)}{\partial W} = X (AY)^T. \quad (8)$$

Обновленная формула для расчета весовой матрицы W имеет вид

$$W = W + \eta x_i (y_i a_i)^T, \quad (9)$$

где η – скорость обучения.

3. ОБУЧЕНИЕ, ОЦЕНКА И ВЫБОР ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

Обучение ИНС можно разделить на два типа: обучение с учителем и обучение без учителя. В нашем случае при решении поставленной задачи был выбран вариант с учителем. Полученная БД включает в себя информацию для сравнения и метки, соответствующие этим данным. Процесс обучения ИНС представлен на рис. 4.

Обычно данные в этом случае делятся на две части, одна используется для обучения, другая –

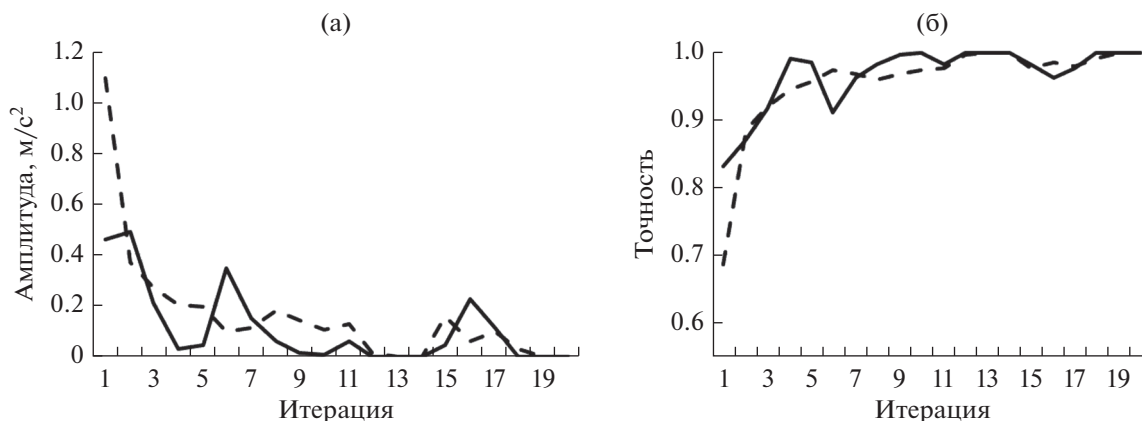


Рис. 5. Погрешность (а) и точность (б) проверки правильности работы ИНС (сплошная кривая) и обучения ИНС (штриховая) (скрытый слой отсутствует).

для обзора процесса обучения. Однако многие ИНС после прохождения обучения и ее последующей оценки не работают должным образом, делая неверные прогнозы для новых случаев (данное явление называется переобучением), а если БД была недостаточно велика возникает недообучение. Чтобы избежать указанных проблем, разделим полученные данные на три части. В части обучения выделим две меньшие части, одну будем использовать для обучения, а другую для проверки правильности работы ИНС непосредственно в процессе обучения.

Известно, что хорошая модель нейронной сети — это модель, которая имеет процесс как обучения,

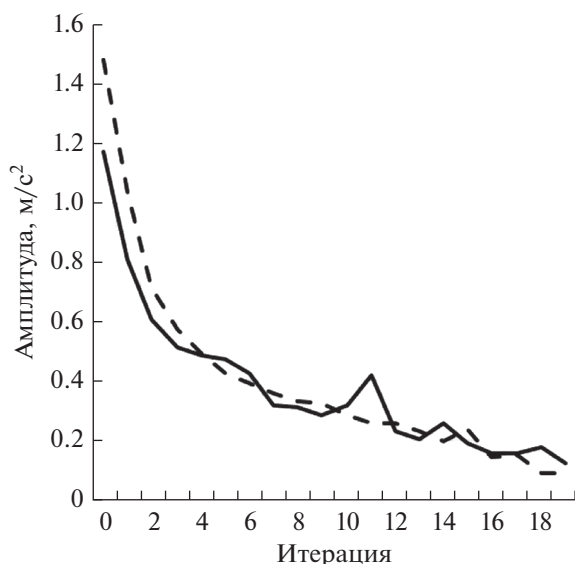


Рис. 6. Погрешность проверки правильности работы ИНС (сплошная линия) и обучения ИНС (штриховая линия) (скрытый слой с восьмью нейронами).

так и проверки правильности работы ИНС, при этом:

- погрешность уменьшается и стремится к 0;
- точность увеличивается и достигает 1;
- результат на финальном тесте равен почти 100%.

Указанная модель называется стабильной, если процессы обучения и проверки правильности работы ИНС не являются зигзагообразными и далеко не отклоняются.

В результате реализации первой модели (один входной и один выходной слой нейронов, отсутствуют скрытые слои) и последующего обучения ИНС были использованы следующие данные: амплитудно-временная характеристика реакции на ударное воздействие (общее количество — 4848; для обучения — 2375; для проверки правильности работы ИНС — 1018; для тестирования — 1455). Полученные результаты представлены на рис. 6.

Однако несмотря на то, что результат тестирования равен 100%, из рис. 5 видно, что с одним слоем входных нейронов и одним слоем выходных нейронов модель работала, но не очень стабильно. Хотя процесс обучения и проверки правильности работы ИНС стремится к высокой точке, однако он идет медленно, проходит много шагов, а зигзаг силен и нестабилен.

Для реализации следующей модели ИНС был добавлен скрытый слой, количество слоев которого равно количеству выходных нейронов, не менее 8. Так как функция softmax предназначена только для выходного слоя, то для реализации скрытого слоя часто используется функция Sigmoid.

Функция Sigmoid представляет входные данные как действительное число и преобразует его в значения в диапазоне (0; 1). Ввод очень маленького отрицательного действительного числа приведет к асимптотике, равной на выходе 0, в противном случае, если на входе большое положительное дей-

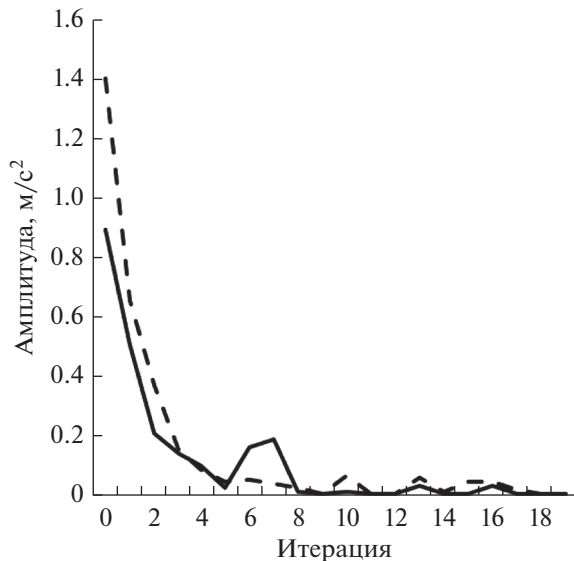


Рис. 7. Погрешность проверки правильности работы ИНС (сплошная кривая) и обучения ИНС (штриховая) (скрытый слой с 16 нейронами).

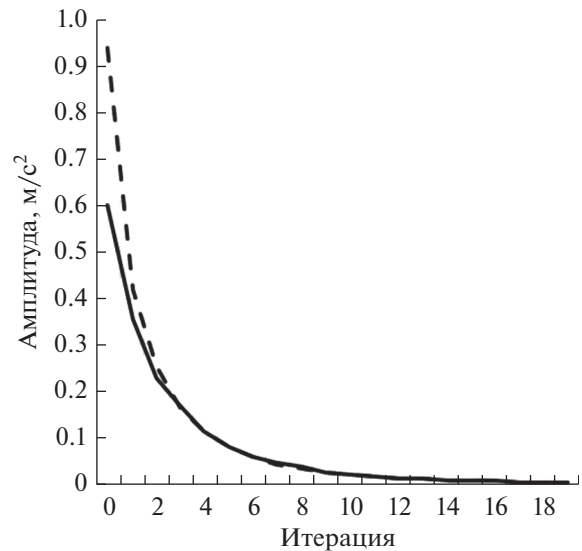


Рис. 8. Погрешность проверки правильности работы ИНС (сплошная кривая) и обучения ИНС для обработанных данных (штриховая) (скрытый слой отсутствует).

ствительное число, выход будет асимптотическим и равен 1.

Однако в настоящее время функция Sigmoid используется редко из-за того, что она отбрасывает и устраняет градиенты. Альтернативой является функция \tanh , которая принимает входные данные как действительные числа и преобразует их в значения в диапазоне $(-1; 1)$. Как и в случае с функцией Sigmoid, функция \tanh насыщена на обоих концах.

В последние годы при обучении ИНС довольно часто используется функция ReLU, которая позволяет легко фильтровать значения меньше нуля. Данная функция по сравнению с Sigmoid и \tanh имеет следующие преимущества:

- скорость сходимости намного быстрее. Функция ReLU имеет скорость сходимости в шесть раз больше, чем \tanh . Это связано с тем, что функция ReLU не насыщена на обоих концах, как в случае с Sigmoid и \tanh ;
- рассчитать данную функцию можно проще и быстрее. Функции \tanh и Sigmoid используют функцию $\exp(x)$, и расчет данной формулы намного сложнее, чем у ReLU, поэтому для ее вычисления и построения требуется больше времени.

При реализации следующей модели ИНС и ее последующего обучения были использованы данные: один входной и один выходной слой нейронов, один скрытый слой, содержащий восемь нейронов.

Добавив скрытый слой с восьмью нейронами, было обнаружено (рис. 6), что процессы обучения и проверки правильности работы ИНС имеют меньше выбросов на графике, при этом эти два

процесса происходят близко друг к другу, однако все еще содержат множество нечетких шагов, а конечная точка не является оптимальной (погрешность стремится к нулю, а точность — к единице).

Следующий эксперимент (рис. 7) проводился, когда количество нейронов в скрытом слое было увеличено до 16, это привело к тому, что модель ИНС работала стабильно, а процесс обучения и проверки правильности работы ИНС быстро достиг оптимальной точки.

Хотелось отметить, что БД, которая была использована в указанных выше экспериментах, являлась необработанной, полученной из экспериментального процесса. При этом если эти данные немного обработать, прежде чем помещать на обучение в ИНС, и представить в виде амплитудно-частотной характеристики реакции на ударное воздействие, то результат будет следующим. Рассмотрим модель с одним входным, одним выходным слоем и без скрытых слоев (рис. 8). Оценивая полученные результаты, можно сказать, что они довольно хорошие, процессы обучения и проверки правильности работы ИНС сокращаются, практически не содержат выбросы, стабильны. Однако чтобы ускорить процесс обучения ИНС, при следующем эксперименте в модель был добавлен скрытый слой с восьмью нейронами (рис. 9).

Анализируя полученные данные в виде погрешности и точности проверки правильности работы ИНС и обучения ИНС, можно сделать

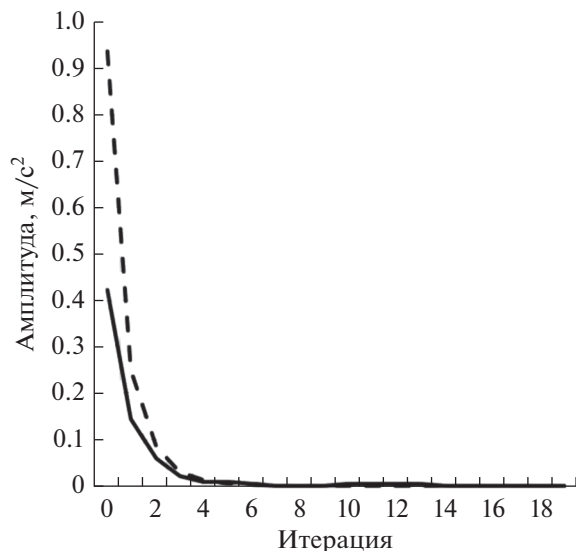


Рис. 9. Погрешность проверки правильности работы ИНС (сплошная линия) и обучения ИНС для обработанных данных (штриховая линия) (скрытый слой с восьмью нейронами).

вывод, что была получена очень хорошая и стабильная модель нейронной сети.

ЗАКЛЮЧЕНИЕ

Рассмотрена проблема классификации отказов в электронных устройствах на основе ИНС. Хотя метод использования ИНС не является методом, который дает наиболее точные результаты для диагностики, однако это помогает достаточно

быстро получать качественные и высокоточные результаты, которые в дальнейшем могут быть использованы в автоматизированных системах обработки данных.

СПИСОК ЛИТЕРАТУРЫ

1. Тихонов А.Н., Увайсов С.У., Иванов И.А., Лышов С.М. // Прикаспийский журн.: управление и высокие технологии. 2016. № 4. С. 144.
2. Bushmeleva K.I., Avdeuk O., Uvaysova S.U., Uvaysova A.S. // Proc. Int. Seminar on Electron Devices Design and Production (SED). Prague. 23–24 Apr. 2019. N.Y.: IEEE, 2019. P. 8798419.
3. Kokovin V.A., Diagilev V.I., Uvaysov S.U., Uvaysova S.S. // Proc. Int. Seminar on Electron Devices Design and Production (SED). Prague. 23–24 Apr. 2019. N.Y.: IEEE, 2019. 8798455.
4. Увайсова А.С., Увайсов С.У., Черновверская В.В., Рожков С.С. // Инновационные, информационные и коммуникационные технологии. 2018. № 1. С. 422.
5. Тадеусевич Р., Боровик Б., Гончарж Т., Леннер Б. // Элементарное введение в технологию нейронных сетей с примерами программ. М.: Горячая линия—Телеком, 2011.
6. Барский А.Б. Нейронные сети: распознавание, управление, принятие решений. М.: Финансы и статистика, 2004.
7. Ben-David Sh., Shalev-Shwartz Sh. Understanding Machine Learning: From Theory to Algorithms. Cambridge: Univ. Press, 2014.
8. Занин А.С., Бушмелева К.И. // Вестн. кибернетики. 2018. № 4. С.177.
9. Бушмелева К.И., Бажаев А.Б., Увайсов С.У., Бушмелев П.Е. // Вестн. кибернетики. 2018. № 1. С. 72.