

На правах рукописи



СИЧИНАВА Зураби Иродиевич

**НЕЙРОСЕТЕВЫЕ АЛГОРИТМЫ АНАЛИЗА ПОВЕДЕНИЯ
РЕСПОНДЕНТОВ**

**Специальность: 05.13.01 – Системный анализ, управление и
обработка информации (в технике и технологиях)**

А в т о р е ф е р а т
диссертации на соискание ученой степени
кандидата технических наук

ПЕНЗА 2014

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Пермский государственный гуманитарно-педагогический университет».

Научный руководитель – доктор технических наук, профессор
ЯСНИЦКИЙ Леонид Нахимович

Официальные оппоненты: **ШКОДЫРЕВ Вячеслав Петрович**,
доктор технических наук, профессор,
ФГБОУ ВПО «Санкт-Петербургский
государственный политехнический
университет»,
заведующий кафедрой «Системы и технологии
управления»

ИСТОМИНА Татьяна Викторовна,
доктор технических наук, профессор,
ФГБОУ ВПО «Пензенский государственный
технологический университет»,
заведующая кафедрой «Информационные
технологии и менеджмент в медицинских
и биотехнических системах»

Ведущая организация: Федеральное государственное автономное
образовательное учреждение высшего
профессионального образования
«Московский физико-технический институт
(государственный университет)»

Защита диссертации состоится 15 мая 2014 года, в 14 часов, на заседании диссертационного совета Д 212.186.04 в Федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Пензенский государственный университет» по адресу: 440026, г. Пенза, ул. Красная, 40.

Диссертация размещена на сайте Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Пензенский государственный университет»: <http://science.pnzgu.ru/page/13778>.

Автореферат разослан

19 марта 2014 г.

Ученый секретарь
диссертационного совета



Косников Юрий Николаевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Психофизиологические исследования с использованием полиграфа в правоохранительных органах, а также в ряде других организаций проводятся в России с 1976 года, в мире – с 1950-х гг. При этом постоянно возникает вопрос об адекватной оценке результатов, полученных в ходе этих исследований. Существующие полиграфные аппараты не отличаются высокой точностью. Так, система экспертной оценки (диагностики) полиграмм широко применяемого в органах МВД России полиграфного аппарата ЭПОС-7 имеет погрешность 30%, вследствие чего в судебно-следственной практике случаются ошибки.

Создание надежных полиграфных аппаратов является актуальным также и в связи с необходимостью предотвращения терроризма, борьбы с коррупцией, раскрытия экономических, политических и другого рода преступлений.

По мнению известных полиграфологов А.М. Петрова, А.П. Сошникова, А.В. Четина и других причины низкой точности современных полиграфных аппаратов состоят в следующем:

1. Большинство полиграфных аппаратов работает по жестко детерминированным алгоритмам и поэтому не учитывает индивидуальных физиологических особенностей опрашиваемого человека (респондента).

2. Многие полиграфные аппараты требуют трудоемкой настройки на респондентов, выполняемой специалистом-полиграфологом, что осложняет практическое применение этих аппаратов и также может приводить к человеческим ошибкам, оказывающим влияние на точность полиграфных аппаратов.

3. При работе на полиграфных аппаратах специалисты-полиграфологи обычно вынуждены перепроверять результаты работы существующих алгоритмов оценки (диагностики) полиграмм (т.к. традиционные алгоритмы оценки полиграмм имеют погрешность около 30%), привлекая различные психологические методики, свой собственный опыт и интуицию. Поэтому результат полиграфного опроса во многом зависит от квалификации и моральных качеств полиграфолога, т.е. от так называемого «человеческого фактора».

Актуальность диссертационной работы заключается в создании алгоритмов оценки полиграмм, направленных на устранение указанных недостатков.

Целью исследования является разработка нейросетевых алгоритмов анализа информации, получаемой в результате опроса на полиграфном аппарате, которые обеспечивают минимальную погрешность заключений при минимальном времени полиграфного опроса.

Для достижения этой цели необходимо было решить следующие задачи:

1. Разработать алгоритм конвертации данных, считываемых с датчиков традиционного полиграфного аппарата, в формат, используемый при проектировании нейронных сетей.

2. Разработать алгоритм обнаружения и исключения выбросов из статистической информации, снимаемой с датчиков полиграфного аппарата.

3. С целью снижения трудоемкости применения нейросетевых полиграфных аппаратов разработать алгоритм выявления и исключения параметров, не оказывающих существенного влияния на степень достоверности заключений.

4. Разработать нейросетевые алгоритмы оценки полиграмм и исследовать их эффективность для решения конкретных классов задач.

5. Реализовать все разработанные алгоритмы в виде программного пакета, предназначенного для моделирования нейронных сетей и поддержки принятия решений в задачах инструментальной детекции лжи.

Объектом исследования является система данных, получаемых при экспертном опросе респондентов и поступающих с датчиков полиграфного аппарата и анкет.

Предметом исследования являются методы и алгоритмы обработки информации, снимаемой с респондента, предназначенные для поддержки принятия решений об истинности или ложности его ответа.

Методы исследований: методы системного анализа, методы математической статистики, теории эксперимента, а также методы графического представления и обработки результатов вычислительных экспериментов.

Соответствие паспортам специальностей. Результаты исследования соответствуют пунктам 4, 5, 13 паспорта научной специальности 05.13.01.

Научная новизна работы заключается в следующем.

1. Разработан нейроэкспертный алгоритм последовательного обнаружения и исключения выбросов из статистической информации, отличающийся от известных своей нейросетевой базой, а потому применимый в тех случаях, когда не выполняется закон нормального распределения статистических данных. Применение алгоритма не только позволяет повысить точность нейросетевых моделей, но и расширяет круг задач, для которых возможно успешное применение нейросетевых технологий.

2. Разработан нейросетевой алгоритм оценки полиграмм, названный индивидуально настраиваемым. Его принципиальное отличие от традиционных алгоритмов оценки полиграмм состоит в том, что в его основе лежит нейронная сеть, обучаемая непосредственно на опрашиваемом респонденте. Это позволяет учитывать индивидуальные особенности именно его организма, в результате чего обеспечивается низкая погрешность полиграфных заключений. Алгоритм предназначен для особо ответственных случаев, когда требуется высокая точность полиграфных заключений, а время полиграфного опроса значения не имеет.

3. Разработан нейросетевой алгоритм оценки полиграмм, названный универсальным. Его отличие от предыдущего алгоритма состоит в том, что

обучение нейронной сети осуществляется предварительно на примерах, сформированных на множестве различных респондентов, и потому при опросе конкретного респондента не требуются дополнительных настроек, за счет чего сокращается время полиграфного опроса, однако при этом увеличивается погрешность полиграфных заключений. Алгоритм предназначен для быстрых скрининговых обследований больших масс респондентов.

4. Разработан нейросетевой алгоритм оценки полиграмм, названный универсально-анкетным. Отличие универсально-анкетного алгоритма от предыдущих состоит в том, что информация в нейронную сеть подается не только с датчиков, снимающих физиологические параметры респондента, но и с заранее подготовленных анкет, характеризующих особенности его организма. Данная модификация нейросетевого алгоритма позволила получить синергетический эффект: существенно сократить время полиграфного обследования при незначительном увеличении погрешности заключений полиграфного аппарата. В связи с этим данный алгоритм рекомендуется для применения в большинстве случаев полиграфного тестирования как альтернатива традиционным алгоритмам.

Практическая значимость состоит в создании программного пакета, используемого для интеллектуальной поддержки принятия решений, который позволяет повысить точность экспертных заключений, уменьшает роль человеческого фактора, позволяет применять полиграфные аппараты пользователям, не имеющим специальной квалификации.

На защиту выносятся:

1. Нейроэкспертный алгоритм последовательного обнаружения и исключения посторонних выбросов из статистической информации при проектировании нейронных сетей.

2. Индивидуально настраиваемый нейросетевой алгоритм оценки полиграмм.

3. Универсальный нейросетевой алгоритм оценки полиграмм.

4. Универсально-анкетный нейросетевой алгоритм оценки полиграмм.

5. Программная реализация всех алгоритмов – программный пакет, предназначенный для моделирования нейронных сетей и поддержки принятия решений в задачах инструментальной детекции лжи.

Реализация результатов работы. Результаты диссертационной работы использованы Центром прикладной психофизиологии (г. Москва) для совершенствования методик психофизиологических опросов, в частности, выявления значимых физиологических параметров, оказывающих влияние на точность и качество заключений при использовании полиграфов.

Кроме того, основные результаты и положения диссертационной работы внедрены в учебный процесс Пермского государственного национального исследовательского университета и Пермского государственного гуманитарно-педагогического университета.

Акты об использовании результатов диссертационной работы помещены в приложении к диссертации.

Получено свидетельство о регистрации электронного ресурса «Программный продукт, предназначенный для проектирования нейронных сетей "ЗС"», выданное Сичинава З.И. Объединенным фондом электронных ресурсов «Наука и образование» Института научной и педагогической информации Российской академии образования, зарегистрированное под № 17926 20.02.2012.

Достоверность научных положений, выводов и практических результатов, сформулированных в диссертации, обеспечивается корректным использованием теории нейронных сетей, а также подтверждается регистрацией разработанного программного пакета и результатами тестирования нейронных сетей на реальных данных – результатах полиграфных опросов заключенных следственных изоляторов Пермского края.

Апробация работы. Основные положения и результаты диссертационной работы докладывались и обсуждались на российских и международных научно-технических конференциях: Международной научно-методической конференции, посвященной 90-летию высшего математического образования на Урале, «Актуальные проблемы математики, механики, информатики» (Пермь, 2006 г.); Первой Всероссийской конференции студентов, аспирантов и молодых ученых «Искусственный интеллект: философия, методология, инновации» (Москва, 2006 г.); Международной научно-практической конференции «Перспективные технологии искусственного интеллекта» (Пенза, 2008 г.); Третьей Всероссийской конференции студентов, аспирантов и молодых ученых «Искусственный интеллект: философия, методология, инновации» (Москва, 2009 г.); Всероссийской научно-практической конференции «Современные проблемы математики и ее прикладные аспекты» (Пермь, 2010 г.); Всероссийской научно-практической конференции «Актуальные проблемы механики, математики, информатики» (Пермь, 2010 г.); Всероссийской научно-практической конференции с международным участием, посвященной 50-летию юбилею механико-математического факультета ПГУ, «Актуальные проблемы механики, математики, информатики» (Пермь, 2010 г.); Международной научно-практической конференции «Интеллектуальные технологии в образовании, экономике и управлении» (Воронеж, 2010 г.).

Публикации. По теме диссертации опубликовано 17 работ, из них 4 статьи в журналах, входящих в перечень ВАК РФ, одна монография. Зарегистрирован электронный ресурс в Объединенном фонде электронных ресурсов «Наука и образование» Института научной и педагогической информации Российской академии образования.

Все результаты, составляющие основное содержание диссертации, и выносимые на защиту положения получены и сформулированы диссертантом самостоятельно. Работы [1–4, 6–8, 9, 10, 12–14, 16] опубликованы в соавторстве с научным руководителем, которому принадлежат постановка задачи и разработка концепций решения проблем. Работы [2, 3, 6, 8, 14] опубликованы в соавторстве с главным полиграфологом Пермского края, полковником

МВД А.М. Петровым и заместителем директора Центра прикладной психофизиологии (г. Москва) А.П. Сошниковым, оказавшими помощь по организации экспериментов и интерпретации их результатов. Монография [1] опубликована совместно с аспирантом Ф.М. Черепановым, с помощью компьютерной программы которого была произведена вербализация обученных нейронных сетей и получены математические формулы, облегчающие процесс переноса готовых алгоритмов из одной программной среды в другую. Работы [6, 8, 13] опубликованы совместно с аспирантами А.Н. Зибатовой и С.Л. Ясницким, осуществлявшими сбор статистической информации и её первоначальную обработку.

Программный пакет «ЗС» [17], реализующий все предлагаемые в диссертации алгоритмы, разработан диссертантом самостоятельно.

Структура и объем работы. Диссертация состоит из введения, трех глав, заключения, списка литературы из 109 наименований и приложения. Основное содержание диссертации включает текст и 30 рисунков общим объемом 101 с. Список литературы и приложения занимают 14 с.

СОДЕРЖАНИЕ РАБОТЫ

Во введении содержится обоснование актуальности проблемы, описываются объект и предмет исследования, формулируются цель и задачи диссертационной работы, определяются методы исследования, дается краткое содержание диссертации по главам, приводятся основные положения диссертационной работы, выносимые на защиту.

В первой главе описан анализ предметной области, выявлены проблемы, возникающие при проведении тестирования на полиграфном аппарате, и предложено направление их решения.

Глава начинается с определения полиграфа, краткой истории его возникновения и развития, дается расшифровка некоторых профессиональных терминов. Описываются принцип действия современного полиграфного аппарата и классическая схема (порядок) работы с ним. Делается попытка классифицировать алгоритмы обработки снимаемой с респондентов информации:

- сервисные алгоритмы первого уровня, реализованные в виде компьютерных программ, позволяющих считывать сигналы, обрабатывать их статистическими методами, представлять в удобной графической форме – в виде полиграмм;

- сервисные алгоритмы второго уровня, реализованные в виде компьютерных программ, вычисляющих интегральные характеристики, на основании которых полиграфологами делается экспертная оценка – заключение об истинности ответа респондента;

- алгоритмы, позволяющие делать экспертные оценки автоматически – на основании результатов вычисления по данным полиграмм.

Отмечается, что полиграфологи используют в своей работе первые две группы алгоритмов и программ, а третья группа алгоритмов, предназначен-

ная для автоматической (компьютерной) оценки полиграмм, практически не используется. Более того, в наиболее распространенной классической схеме (порядке) работы с полиграфным аппаратом, приведенной во многих руководствах и инструкциях, алгоритмами автоматической оценки полиграмм пользоваться не рекомендуется.

На этом основании в диссертации делается вывод о том, что причиной столь негативного отношения полиграфологов к математическим методам и алгоритмам диагностики (оценки) полиграмм является их высокая погрешность, являющаяся следствием устаревшей методологической основы. Причем в качестве недостатков существующих алгоритмов оценки полиграмм указываются их субъективизм, т.е. зависимость результатов полиграфных обследований от профессионализма, опыта и интуиции полиграфолога, а также высокая трудоемкость полиграфного обследования.

В качестве выхода из сложившейся ситуации в диссертации предлагается применение более современной технологической основы для построения алгоритмов автоматической диагностики полиграмм, а именно – аппарата нейросетевых технологий. Обоснованием этого предложения является, во-первых, анализ литературы, который показал, что в ряде иностранных источников уже имеются сообщения о подобных попытках. Идея применения нейросетевых технологий в полиграфном деле «витают в воздухе», но до серьезных научных исследований и, тем более, до практического применения дело пока не доходит. Во-вторых, и это главное, обширный мировой опыт создания алгоритмов принятия решений на базе нейронных сетей (А.И. Галушкин, А.В. Чечкин, А.Н. Горбань, Е.М. Миркес, В.В. Круглов, В.В. Борисов, Г.М. Алакоз, Л.С. Куравский, С.Д. Кулик, В.И. Горбаченко, А.И. Иванов, С.А. Филист, С. Хайкин, С. Рассел, С. Осовский, Дж. Ф. Люгер и др.) показывает, что нейросетевые алгоритмы исключительно объективны и во многих случаях позволяют добиваться высокой точности моделирования объектов, явлений и процессов.

Во второй главе описываются материально-техническая, методологическая и программно-инструментальная базы, используемые для разработки и создания нейросетевых алгоритмов анализа поведения респондентов.

Описывается алгоритм применения метода нейросетевого математического моделирования. Отмечается, что успешное применение этого алгоритма в значительной степени осложнено наличием большого количества выбросов, неизбежно присутствующих среди сигналов, снимаемых с датчиков полиграфного аппарата. Как известно, под посторонними выбросами понимаются примеры поведения предметной области, по каким-либо причинам выпадающие из общих закономерностей этой предметной области. Другими словами, это наблюдения, не удовлетворяющие закономерностям, которым подчиняется подавляющее большинство примеров поведения предметной области. Причинами появления посторонних выбросов могут быть: недостаточно чисто проведенный эксперимент (полиграфный опрос), ошибка изме-

рений, сбой приборов и оборудования, искажение информации при формировании множества обучающих примеров.

С целью преодоления указанных трудностей автором предложен и реализован специальный алгоритм, названный нейроэкспертным алгоритмом последовательного обнаружения и исключения посторонних выбросов из статистических выборок. Идея алгоритма основана на том факте, что если выбросов в обучающем множестве сравнительно немного и если нейронная сеть имеет сравнительно небольшое число степеней свободы (небольшое количество синаптических весов), то после применения процедуры обучения нейронная сеть на примерах, которые являются выбросами, как правило, показывает более высокую погрешность обучения, чем на примерах, не являющихся выбросами.

С учетом этого эмпирического факта алгоритм обнаружения и исключения посторонних выбросов предлагается в следующем виде:

1. Разбить множество примеров поведения предметной области на обучающее (L), тестирующее (T) и подтверждающее (P) подмножества в отношении 80% : 15% : 5%.

2. Для множества примеров $L \cup T$, пользуясь формулами следствия теоремы Арнольда – Колмогорова – Хехт-Нильсена, рассчитать минимальное и максимальное количество скрытых нейронов двухслойного персептрона (используем самый простой из всех допустимых вариантов персептронов, удовлетворяющих условиям теоремы существования Арнольда – Колмогорова – Хехт-Нильсена):

$$N_{w\min} = \frac{N_y Q}{1 + \log_2(Q)},$$

$$N_{w\max} = N_y \left(\frac{Q}{N_x} + 1 \right) (N_x + N_y + 1) + N_y,$$

$$N_{\min} = \frac{N_{w\min}}{N_x + N_y},$$

$$N_{\max} = \frac{N_{w\max}}{N_x + N_y}.$$

Здесь: N_{\min} и N_{\max} – минимальное и максимальное количество скрытых нейронов; $N_{w\min}$ и $N_{w\max}$ – минимальное и максимальное количество сил синаптических связей; N_x – количество нейронов входного слоя; N_y – количество нейронов выходного слоя; Q – число элементов множества $L \cup T$.

3. Рассчитать число скрытых нейронов двухслойного персептрона, предназначенного для выявления посторонних выбросов, с помощью предлагаемой автором эмпирической формулы:

$$N = N_{\min} + \xi(N_{\max} - N_{\min}),$$

в которой ξ – эмпирический коэффициент, значение данного коэффициента в первом приближении принимается, например, равным 0,1.

4. Обучить нейронную сеть на множестве $L \cup T$ и выявить пример, для которого ошибка обучения нейросети ε_0 имеет максимальное значение.

5. Предоставить информацию о выявленном примере специалисту в исследуемой предметной области и согласовать с ним вопрос о возможности удаления данного примера из множества $L \cup T$.

6. В зависимости от решения эксперта либо удалить выявленный пример из множества $L \cup T$ и перейти к п.7, либо пометить его как не подлежащий удалению, выявить следующий по величине ε_0 пример и перейти к п. 5.

7. Разбить очищенное множество $L \cup T$ на обучающее L и тестирующее T подмножества в отношении 85% : 15%.

8. Обучить и протестировать нейросеть, вычислив ошибку тестирования ε_T на множестве T . Результат тестирования изобразить графически, как показано на рисунке 1.

9. Повторять пп. 2–8, пока кривая на рисунке 1 не перестанет снижаться.

10. Вычислить ошибку сети ε_p на подтверждающем множестве P .

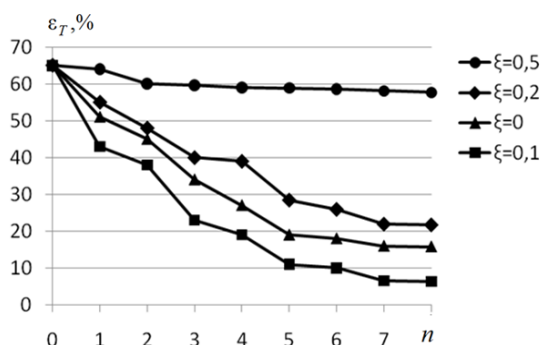


Рисунок 1 – Примерные зависимости погрешности тестирования ε_T от эмпирического коэффициента ξ и от числа итераций n по пп. 2–8 предлагаемого алгоритма

Как видно из рисунка 1, качество получаемой таким способом нейронной сети зависит от значения коэффициента ξ . В приведенном на рисунке примере оптимальное значение коэффициента ξ оказалось равным 0,1. Это значит, что любое отклонение от этого оптимального значения коэффициента ξ в сторону увеличения или в сторону уменьшения приводит к росту ошибки тестирования сети ε_T (а также ε_p). Как показали численные эксперименты, значение коэффициента $\xi = 0,1$ является оптимальным только для рассматриваемого в диссертационной работе класса задач. Для других же предметных областей оптимальное значение коэффициента ξ может отличаться

от 0,1, однако, как правило, оно не выходит за рамки интервала $[0; 0,2]$. В любом случае оно может быть уточнено путем построения кривых, аналогичных кривым рисунка 1.

Применение предлагаемого нейроэкспертного алгоритма обнаружения и исключения посторонних выбросов при создании нейросетевого детектора лжи позволило снизить погрешность нейронных сетей на величину от 20 до 80% в зависимости от используемой статистической выборки и используемого нейросетевого алгоритма.

Необходимо отметить, что попытки построения нейросетевых моделей без предварительного обнаружения и исключения посторонних выбросов из статистической информации в большинстве случаев вообще не приводили к положительным результатам, т.е. погрешности нейронных сетей не удавалось снизить до сколько-нибудь приемлемых для практического применения значений.

На рисунке 2 представлена схема взаимодействия элементов исследуемой в диссертационной работе системы. Как показано на схеме, информация снимается с респондента двумя путями: с помощью датчиков полиграфного аппарата и путем предварительного заполнения анкет. Информация, полученная с датчиков, обрабатывается алгоритмами первого и второго уровней полиграфного аппарата, после чего подается на входы нейронной сети. Информация с анкет сразу подается на входы нейронной сети. Нейронная сеть вычисляет результат, используемый специалистом-полиграфологом для интеллектуальной поддержки решения об истинности или ложности ответа респондента. Необходимо обеспечить минимальную погрешность заключений при минимальном времени полиграфного обследования респондента. То есть критериями оценки алгоритма являются: 1) погрешность заключений и 2) время полиграфного обследования респондента.

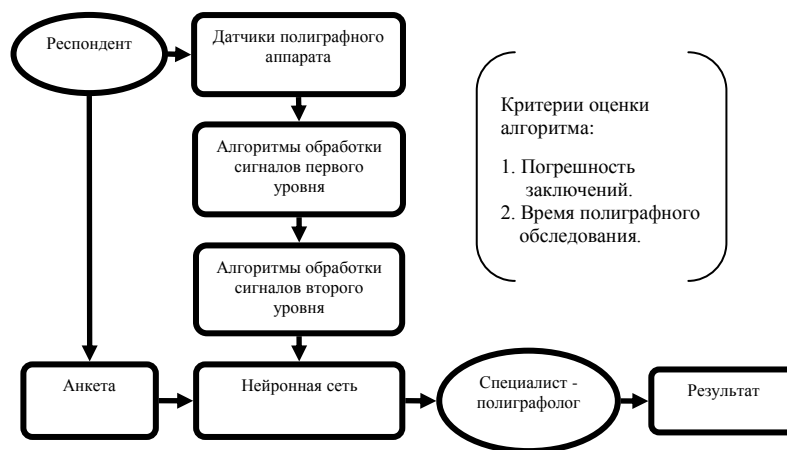


Рисунок 2 – Структура исследуемой системы

В третьей главе разрабатываются три нейросетевых алгоритма оценки полиграмм, названные:

- индивидуально настраиваемым,
- универсальным,

- универсально-анкетным.

Принцип действия индивидуально настраиваемого нейросетевого алгоритма является традиционным для метода нейросетевого математического моделирования. Согласно этому принципу сначала формируется множество обучающих примеров, которое в данном случае создается в результате предъявления обследуемому человеку-респонденту стимуляционных тестов. Полиграфолог задает вопросы, ответы на которые ему известны. С респондента во время ответов с помощью штатных датчиков полиграфного аппарата ЭПОС-7 снимаются физиологические показатели, которые формируют входной вектор X_q :

- x_1 – верхнее дыхание (грудное), число вдохов в минуту;
- x_2 – нижнее дыхание (брюшное), число вдохов в минуту;
- x_3 – амплитуда дыхания, амплитуда;
- x_4 – кожно-гальванические реакции 1-го уровня, число колебаний в минуту;
- x_5 – кожно-гальванические реакции 2-го уровня, число колебаний в минуту;
- x_6 – кожно-гальванические реакции, амплитуда;
- x_7 – фотоплетизмограмма, частота;
- x_8 – фотоплетизмограмма, амплитуда;
- x_9 – кровенаполнение, скорость кровенаполнения сосудов.

Требуемый выходной сигнал нейронной сети d формирует полиграфолог, сопоставляя ответы респондента с тем, что ему заранее известно: $d = 1$, если ответ правдивый, и $d = 0$, если ответ ложный.

Накопив достаточное количество примеров, обучив и протестировав нейронную сеть, можно задать опрашиваемому человеку вопрос, ответ на который заранее не известен. Нейронная сеть выдаст сигнал y , кодирующий степень истинности ответа опрашиваемым человеком: если сигнал y окажется близким к единице, то ответ считается правдивым, а если к нулю – то ложным. Причем поскольку нейронная сеть обучена на физиологических реакциях, формируемых самим опрашиваемым человеком, то ее заключение объективно учитывает индивидуальные особенности организма именно этого человека.

В ходе работ по созданию индивидуально настраиваемого алгоритма была выполнена оптимизация нейронной сети. В качестве критерия оптимальности использовалось условие минимума среднеквадратичной ошибки тестирования ε_T . Минимизация функции ε_T осуществлялась путем варьирования количества скрытых слоев, количества нейронов в скрытых слоях, типов активационных функций нейронов. Примеры кривых, позволяющих выбрать оптимальное количество нейронов скрытого слоя, приведены на рисунке 3. Результатом оптимизации явилась слоистая нейронная сеть с девятью нейронами входного слоя, одним нейроном выходного слоя и с одним скрытым слоем, включающим пять сигмоидных нейронов.

Приведенный способ создания нейросетевого алгоритма оценки поли-

грамм, который был назван индивидуально настраиваемым, обладает серьезным недостатком, состоящим в высокой трудоемкости, а именно: необходимо настраивать нейросеть на каждого опрашиваемого респондента, что занимает от 12 до 20 часов. Причина высокой трудоемкости заключается в том, что для каждого обследуемого человека приходится заново формировать множество обучающих примеров (порядка 20–30 вопросов), а затем обучать нейронную сеть на каждом таком множестве. После этого нейронная сеть пригодна для тестирования только одного человека, на которого она была настроена.

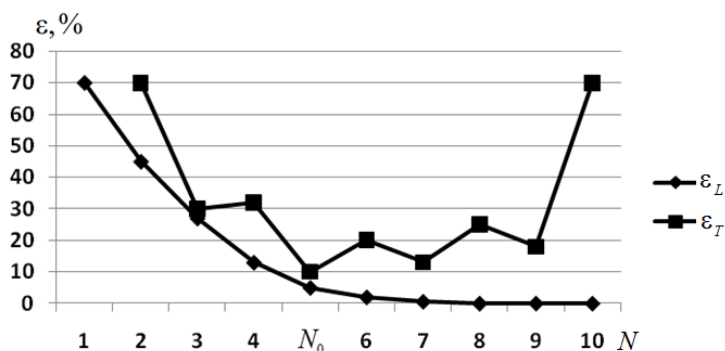


Рисунок 3 – Пример кривых зависимости ошибки обучения ε_L и ошибки обобщения (тестирования) ε_T от количества нейронов скрытых слоев персептрона

Обучающее множество для индивидуально настраиваемого алгоритма состояло из девяти сот примеров (по тридцать обучающих примеров для каждого респондента) и трехсот тестирующих примеров (по одному на каждого респондента). Среднее значение погрешности, полученное в результате тестирования индивидуально настраиваемого алгоритма, составило 3,37%.

Таким образом, предлагаемый способ оценки полиграмм, с одной стороны, позволяет проводить обследование с учетом индивидуальных особенностей организма респондента, что снижает погрешность с 30% (такую погрешность имеет система экспертной оценки штатного полиграфного аппарата ЭПОС-7, применяемого в МВД России) до 3,37%, а с другой - требует значительного времени полиграфного обследования: 12–20 часов на каждого респондента. Поэтому такой алгоритм можно рекомендовать только для особо ответственных случаев, когда затраты на длительное обучение нейронной сети оправданны.

С целью уменьшения времени полиграфного обследования предлагается другой способ построения нейросетевого алгоритма оценки полиграмм, названный универсальным. В отличие от индивидуально настраиваемого алгоритма, множество обучающих примеров формируется не на конкретном опрашиваемом человеке, а предварительно – на большом количестве различных людей. Таким образом, нейронная сеть настраивается на некоторого «среднего» респондента и перестает учитывать индивидуальные особенности отдельных респондентов. Зато она становится пригодной для тестирования любого респондента без каких-либо дополнительных обучений, за счет чего достигается эффект снижения времени полиграфного обследования респон-

дента.

В проведенных экспериментах использовались достоверные данные, полученные экспертом-полиграфологом, причем половина из них с заключением «Правда» и половина – «Ложь». Всего было представлено 300 вопросов-ответов, составляющих множество примеров предметной области. Согласно сложившейся практике обучения и тестирования нейронных сетей множество примеров предметной области разбивалось на обучающее и тестирующее, причем объем тестирующего множества составлял примерно 10% от обучающего. Сеть подвергалась многократному обучению и тестированию по методике *multifold cross-validation* (многократной перекрестной проверки).

При разработке универсального алгоритма использовалась нейронная сеть структуры, аналогичной той, которая была использована для индивидуально настраиваемого алгоритма, а именно: слоистая нейронная сеть с девятью нейронами входного слоя, одним выходным нейроном и с одним скрытым слоем, включающим девять сигмоидных нейронов. Как и ранее, сеть была оптимизирована по методике, описанной выше, т.е. из условия обеспечения минимального значения ошибки тестирования ε_T .

Среднее значение погрешности, полученное в результате тестирования универсального алгоритма, составило 19,8%, при этом время полиграфного обследования сократилось до 40 минут на одного респондента.

Зафиксированное в экспериментах повышение погрешности (по сравнению с погрешностью индивидуально настраиваемого алгоритма) объясняется тем, что эта нейронная сеть, предварительно обученная на большом количестве различных людей, получилась ориентированной на некоторого «среднего» человека и, естественно, перестала учитывать индивидуальные особенности конкретного респондента. Использование нейронной сети стало менее трудоемким, но это привело к существенному снижению надежности заключений. Поэтому универсальный алгоритм оценки полиграмм можно рекомендовать в тех случаях, когда требуется быстро обследовать большое количество людей, например, выполнить скрининговое тестирование сотрудников крупной фирмы, пассажиров авиалайнера и т.п.

Выше были описаны и проанализированы два нейросетевых алгоритма оценки полиграмм, предназначенные для использования в полиграфных аппаратах: первый (индивидуально настраиваемый) учитывает индивидуальные особенности организма опрашиваемого человека, за счет чего обладает высокой надежностью заключений, но требует больших трудозатрат в применении; второй (универсальный) не учитывает индивидуальные особенности организма опрашиваемого человека и поэтому не обладает высокой надежностью заключений, но и не требует больших трудозатрат при применении. Перед диссертантом встала задача создания алгоритма, обладающего достоинствами двух предыдущих алгоритмов, т.е. учитывающего индивидуальные особенности организма опрашиваемого человека и не требующего больших трудозатрат при применении.

Эта задача была решена следующим образом. В схему нейронной сети, применяемой для индивидуально настраиваемого алгоритма, были добавлены дополнительные входные нейроны. Эти нейроны предназначены для ввода дополнительных параметров респондента, получаемых с предварительно заполненной анкеты, включающей следующие сведения: 1. Пол. 2. Возраст, лет. 3. Знак зодиака. 4. Телосложение (полное, среднее, худощавое). 5. Вид работы (интеллектуальный, физический, смешанный). 6. Рост, м. 7. Вес, кг. 8. Занимался ли человек профессионально спортом. 9. Судимость (не судим, судим, судим неоднократно). 10. Тревожность (высокая, обычная, слабая). 11. Курение (регулярное, редкое, не курит). 12. Тип характера (экстраверт, интроверт). 13. Употребление алкоголя (регулярное, умеренное, редкое, не употребляет). 14. Употребление наркотиков (регулярное, редкое, не употребляет). 15. Состоит (состоял ли) на учёте у психиатра. 16. Состоит (состоял ли) на учёте у невролога. 17. Состоит (состоял ли) на учёте у нарколога. 18. Наличие беременности. 19. Наличие хронических заболеваний организма. 20. Наличие хронических заболеваний сердечно-сосудистой системы. 21. Наличие хронических заболеваний системы дыхания. 22. Наличие физического или психического истощения. 23. Наличие травм головы. 24. Наличие травм позвоночника. 25. Было ли обморожение кистей рук.

Такие параметры были первоначально приняты в результате консультаций с авторитетным специалистом-полиграфологом полковником МВД А.М. Петровым. Наша же задача состояла в том, чтобы выявить, какие из этих параметров действительно влияют на точность заключений нейронной сети. Нейронная сеть тоже обучалась на большом количестве различных людей (150 человек), с которых во время опросов снимались не только физиологические параметры, но и приведенные выше анкетные данные, характеризующие индивидуальные особенности их организма.

Первоначально были взяты все 34 входных параметра (9 физиологических параметров, снимаемых с помощью штатных датчиков полиграфного аппарата системы ЭПОС-7, и 25 анкетных данных), но результаты получались хуже на 15–20%, чем те, которые были получены с помощью универсального полиграфного аппарата. Кроме того, на заполнение анкет уходило некоторое время.

Поэтому было принято решение выявить и исключить из анкетных данных наименее значимые параметры, причем выявление наименее значимых параметров производилось путем исследования нейросетевой модели согласно следующему алгоритму:

1. Поочередное введение параметров. Для этого был проведен ряд экспериментов с обучением и тестированием нейронной сети, в каждом из которых использовались девять физиологических параметров, снимаемых с датчиков, и поочередно – один из анкетных параметров. Эксперименты повторялись по три раза (увеличение числа повторов до пяти и десяти раз приводило к изменению погрешности не более чем на 0,1%). В результате получили гистограмму распределения погрешности тестирования полиграфного ап-

парата в зависимости от номера используемого анкетного параметра (рисунок 4).

2. Из векторов множества обучающих примеров были исключены те анкетные параметры, для которых погрешность тестирования нейросети согласно гистограмме (рисунок 4) оказалась больше пороговой погрешности $\varepsilon_s = 20\%$ (ниже показано, что это и есть оптимальное значение, обеспечивающее максимальный эффект). В итоге остались следующие анкетные параметры: 1. Пол. 2. Возраст, лет. 3. Вид работы (интеллектуальный, физический, смешанный). 4. Рост, м. 5. Вес, кг. 6. Тип характера (экстраверт, интроверт). 7. Употребление алкоголя (регулярное, умеренное, редкое, не употребляет).

Было вновь проведено создание нейронной сети с использованием оставшихся параметров. Результирующая нейросеть содержала шестнадцать нейронов входного слоя, а именно: девять нейронов для ввода параметров, получаемых с датчиков полиграфного аппарата, и семь нейронов для ввода параметров, получаемых из анкеты. Как и ранее, сеть была оптимизирована по методике, описанной выше, т.е. из условия обеспечения минимального значения ошибки тестирования ε_T . Результатом оптимизации явилась слоистая нейронная сеть с одним скрытым слоем из пяти сигмоидных нейронов.

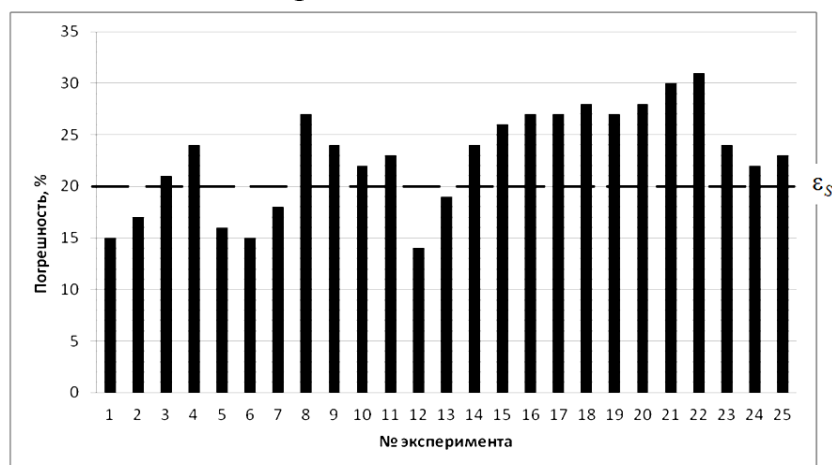


Рисунок 4 – Погрешность тестирования нейросети при поочередном использовании анкетных параметров. ε_s – пороговое значение погрешности

Всего было представлено триста вопросов-ответов, составляющих множество примеров предметной области. Нейронная сеть подвергалась тестированию по методике, аналогичной той, которая была использована для предыдущего алгоритма. Среднее значение погрешности для этого алгоритма составило 9,6%.

Таким образом, можно сделать вывод, что использование отобранных в результате экспериментирования анкетных данных привело к понижению погрешности предлагаемого здесь алгоритма по сравнению с рассмотренным ранее универсальным алгоритмом в среднем на 10,2%. Этот последний алгоритм был назван универсально-анкетным алгоритмом.

Как указывалось выше, данный результат был получен при исключе-

нии тех анкетных параметров, для которых погрешность тестирования нейросети согласно гистограмме (рисунок 4) оказалась больше пороговой $\varepsilon_s = 20\%$. Попытки исключения других наборов параметров, соответствующих другим значениям погрешности ε_s в интервале от 14 до 30% с шагом 1% на гистограмме (рисунок 4), не привели к уменьшению итоговой погрешности универсально-анкетного алгоритма. Таким образом, можно считать, что исключение избыточных анкетных параметров универсально-анкетного алгоритма было выполнено оптимальным образом, т.е. решена оптимизационная задача: найден минимум целевой функции ε_T при варьировании ε_s .

Можно констатировать, что предлагаемый в диссертационной работе универсально-анкетный алгоритм сочетает в себе достоинства двух предыдущих алгоритмов (индивидуально настраиваемого и универсального), учитывает индивидуальные особенности организма респондента и его применение не требует больших затрат времени. Налицо проявление синергетического эффекта.

Заметим еще раз, что идея и попытки создания детектора лжи с использованием нейросетевых технологий упоминаются в ряде иностранных источников. При этом имеются в виду индивидуально настраиваемый и универсальный алгоритмы. Идея же универсально-анкетного нейросетевого алгоритма автором диссертационной работы предложена и реализована впервые. Идея этого наиболее эффективного из трех рассмотренных здесь алгоритмов работы нейросетевого детектора лжи была впервые опубликована автором диссертационной работы в 2006 г., причем до сих пор упоминаний о существовании такого алгоритма в научной литературе не обнаружено.

Автором диссертации разработано специальное программное обеспечение – программный пакет «ЗС», предназначенный для анализа информации, поступающей с датчиков полиграфного аппарата ЭПОС-7. Программный пакет «ЗС» представляет собой специализированный нейроимитатор. Как и большинство известных нейроинструментов (NeuroSolutions, NeuralWorks, Neural Network Toolbox Matlab и др.), программный пакет «ЗС» позволяет создавать нейронные сети, выполнять их обучение, тестирование, вычисления и прогнозы. Его отличительной особенностью является ориентация на материально-техническую базу полиграфного аппарата ЭПОС-7 с его штатным программным обеспечением первого и второго уровней. В частности, программный пакет «ЗС» предоставляет возможность конвертации данных, считанных с датчиков полиграфного аппарата ЭПОС-7 и обработанных его штатным программным обеспечением первого и второго уровней. Дело в том, что данные в полиграфном аппарате ЭПОС-7 хранятся в специальном формате QT1. Кроме того, они не структурированы и хранятся в разных таблицах и файлах среди другой информации. Для упрощения и автоматизации работы с данными полиграфного аппарата ЭПОС-7 разработан модуль-конвертер. Этот модуль автоматически находит и собирает нужные данные, структурирует их и переводит в формат Excel. Окончательный выходной сигнал в программном пакете «ЗС» формируется в результате взаимодейст-

вия входных и выходных сигналов искусственных нейронов, соединенных согласно схемам, в зависимости от используемого нейросетевого алгоритма.

Отличительной особенностью программного пакета «ЗС» является также наличие блока, реализующего предложенный автором нейроэкспертный алгоритм последовательного обнаружения и исключения посторонних выбросов из статистической информации.

Необходимо отметить, что предпринимавшиеся ранее попытки создания нейросетевых полиграфных аппаратов на базе использования NeuroSolutions, NeuralWorks, Neural Network Toolbox Matlab без применения программного пакета «ЗС» к успеху не приводили из-за высокой трудоемкости ручной подготовки и обработки данных, получаемых с датчиков полиграфного аппарата, а также из-за большого количества посторонних выбросов в статистических выборках.

В приложении представлены документы, подтверждающие внедрение результатов диссертационной работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Разработан нейроэкспертный алгоритм последовательного обнаружения и исключения выбросов из статистической информации, отличающийся от известных тем, что в его основе лежит нейронная сеть, спроектированная с использованием формул следствия из теорем Арнольда – Колмогорова – Хехт-Нильсена, а также предложенной в диссертации эмпирической формулы. За счет своей нейросетевой основы, в отличие от существующих алгоритмов, предлагаемый алгоритм применим в тех случаях, когда не выполняется закон нормального распределения статистических данных. Применение алгоритма не только позволяет снизить погрешность нейросетевых моделей, но и расширяет круг задач, для решения которых успешное применение нейросетевых технологий вообще возможно.

2. Разработан алгоритм, позволяющий с помощью нейронной сети выявлять наиболее значимые параметры респондента, оказывающие наибольшее влияние на погрешность заключений полиграфного опроса. В отличие от существующих, предлагаемый алгоритм, за счет своей нейросетевой основы, нелинеен по своей природе и работает в тех случаях, когда не выполняется закон нормального распределения статистических данных. Применение разработанного алгоритма (в совокупности с другими разработанными алгоритмами) позволило спроектировать нейронные сети, обеспечивающие приемлемую погрешность полиграфных заключений при минимуме временных затрат.

3. Разработан нейросетевой алгоритм оценки полиграмм, названный индивидуально настраиваемым. Его отличие от существующих алгоритмов, традиционно использующих принципы построения экспертных систем, состоит в том, что лежащая в его основе нейронная сеть обучается на ответах

опрашиваемого человека, а потому учитывает индивидуальные физиологические особенности именно его организма. В связи с этим применение алгоритма позволило обеспечить погрешность заключений нейросетевого полиграфного аппарата 3,37%, что на 26,63% ниже погрешности заключений применяемого в органах МВД России полиграфного аппарата ЭПОС-7. Однако предлагаемый алгоритм обладает недостатком, состоящим в высокой трудоемкости, а именно - необходимостью настройки нейросети на каждого опрашиваемого респондента, что занимает от 12 до 20 часов.

4. Разработан нейросетевой алгоритм оценки полиграмм, названный универсальным, отличающийся от алгоритма п. 3 тем, что лежащая в его основе нейронная сеть обучается на ответах не одного, а большого количества респондентов (100 человек), в результате чего применение этого алгоритма, в отличие от алгоритма п. 3, не требует предварительного трудоемкого обучения нейросетей на каждом респонденте, в результате чего время полиграфного обследования сокращается до 40 минут на одного респондента. Недостаток алгоритма состоит в его сравнительно высокой погрешности, которая составляет 19,8%.

5. Разработан нейросетевой алгоритм оценки полиграмм, названный универсально-анкетным, отличающийся от алгоритмов п. 3 и п. 4 тем, что информация в нейронную сеть подается не только с датчиков, снимающих физиологические параметры респондента, но и с заранее подготовленных анкет, характеризующих особенности организма респондента. Благодаря этому погрешность полиграфных заключений не превышает 9,6%, а время, необходимое для настройки нейронной сети на опрашиваемого респондента, составляет не более 60 минут. Таким образом, данный алгоритм сохраняет достоинства алгоритмов п. 3 и п. 4, устраняя их недостатки.

6. Для каждого из предложенных нейросетевых алгоритмов оценки полиграмм указаны их области применения: универсально-анкетный алгоритм рекомендуется для применения в большинстве случаев полиграфного тестирования как альтернатива традиционным алгоритмам оценки полиграмм; индивидуально настраиваемый нейросетевой алгоритм рекомендуется применять в особо ответственных случаях, когда требуется низкая погрешность полиграфных заключений, а время полиграфного опроса значения не имеет; универсальный алгоритм предназначен для быстрых скрининговых обследований больших масс респондентов, например, при приеме на работу, повышении по службе и др.

7. Все три предложенных нейросетевых алгоритма оценки полиграмм формализованы, в результате чего получены математические формулы. Помимо теоретических полученные математические формулы представляют и практический интерес, например, для построения индивидуальных микродетекторов лжи на базе электронных микросхем. Формулы можно занести в память микросхем, не выполняя более трудоемкого и ресурсоемкого процесса кодирования нейронных сетей. Они также облегчают процесс переноса готовых алгоритмов из одной программной среды в другую.

8. Все разработанные алгоритмы реализованы в виде программного пакета «ЗС», предназначенного для моделирования нейронных сетей. В отличие от других известных нейропакетов, программный пакет «ЗС» адаптирован для взаимодействия с полиграфным аппаратом ЭПОС-7 и предназначен для поддержки принятия решений на основании анализа информации, снимаемой с датчиков этого полиграфного аппарата.

9. Установлено, что применение разработанных в диссертационной работе алгоритмов и программного пакета «ЗС» для интеллектуальной поддержки принятия решений в практике инструментальной детекции лжи повышает точность экспертных заключений, уменьшает роль человеческого фактора, снижает необходимость привлечения опытных специалистов-полиграфологов и позволяет применять полиграфные аппараты пользователям, не имеющим специальной квалификации.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Монография

1. Сичинава, З.И. Нейросетевой детектор лжи: принципы построения и опыт разработки / Л.Н. Ясницкий, З.И. Сичинава, Ф.М. Черепанов. – Saarbrücken (Germany): LAP LAMBERT Academic Publishing GmbH & Co. KG., 2012. – 115p.

Публикации в изданиях, рекомендованных ВАК РФ

2. Сичинава, З.И. Сравнительный анализ алгоритмов нейросетевого детектирования лжи / Л.Н. Ясницкий, А.М. Петров, З.И. Сичинава // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2010. – №1(13). – С. 64–72.

3. Сичинава, З.И. Технологии построения детектора лжи на основе аппарата искусственных нейронных сетей / Л.Н. Ясницкий, А.М. Петров, З.И. Сичинава // Информационные технологии. – 2010. – № 11. – С. 66–70.

4. Сичинава, З.И. Нейросетевые алгоритмы анализа поведения респондентов / Л.Н. Ясницкий, З.И. Сичинава // Нейрокомпьютеры: разработка, применение. – 2011. – № 10. – С. 59–64.

5. Сичинава, З.И. Нейроэкспертный алгоритм последовательного обнаружения и исключения посторонних выбросов из статистической информации при построении нейросетевых математических моделей / З.И. Сичинава // Современные проблемы науки и образования. – 2013. – № 2; URL : <http://www.science-education.ru/108-8903> (дата обращения: 19.04.2013).

Публикации в других изданиях

6. Сичинава, З.И. Интеллектуальный полиграф / А.Н. Зибатова, А.М. Петров, З.И. Сичинава, А.П. Сошников, Л.Н. Ясницкий // Российский полиграф. – 2006. – № 1. – С. 76–83.

7. Сичинава, З.И. Нейросетевой детектор лжи / А.Н. Зибатова, З.И. Сичинава // Искусственный интеллект: философия, методология, инновации: материалы Первой Всероссийской конференции студентов, аспирантов и молодых ученых. – М.: Интел, 2006. – С. 351–353.

8. Сичинава, З.И. Этапы создания интеллектуального детектора лжи / А.Н. Зибатова, А.М. Петров, З.И. Сичинава, Л.Н. Ясницкий // Актуальные проблемы математики, механики, информатики: материалы Международной научно-методической конференции, посвященной 90-летию высшего математического образования на Урале. – Пермь: Изд-во ПГУ, 2006. – С. 125–126.

9. Сичинава, З.И. Нейросетевой детектор лжи / З.И. Сичинава, Л.Н. Ясницкий // Перспективные технологии искусственного интеллекта: материалы Международной научно-практической конференции «Перспективные технологии искусственного интеллекта». – Пенза: Изд-во Пенз. ун-та, 2008. С. 81–84.

10. Сичинава, З.И. Анкетный способ построения нейросетевого детектора лжи / З.И. Сичинава, Л.Н. Ясницкий // Искусственный интеллект: философия, методология, инновации: материалы III Всероссийской конференции студентов, аспирантов и молодых ученых. – М.: Связь-Принт, 2009. – С. 318–320.

11. Сичинава, З.И. Алгоритмы анализа поведения респондентов на основе обработки информации с использованием нейросетевых технологий / З.И. Сичинава // Интеллектуальные технологии в образовании, экономике и управлении: сборник материалов VII Международной научно-практической конференции. – Воронеж: Наука-Юнипресс, 2010. – С. 299–302.

12. Сичинава, З.И. Анкетный способ построения нейросетевого детектора лжи / З.И. Сичинава, Л.Н. Ясницкий // Современные проблемы математики и ее прикладные аспекты: материалы Всероссийской научно-практической конференции. – Пермь: Изд-во ПГУ, 2010. – С. 58.

13. Сичинава, З.И. Искусственный интеллект против коррупции / З.И. Сичинава, С.Л. Ясницкий, Л.Н. Ясницкий // Актуальные проблемы механики, математики, информатики: тезисы докладов Всероссийской научно-практической конференции. – Пермь: Изд-во ПГУ, 2010. – С. 203.

14. Сичинава, З.И. Анкетный способ построения нейросетевого детектора лжи / А.М. Петров, З.И. Сичинава, Л.Н. Ясницкий // Вестник Пермского университета. Математика. Механика. Информатика. – 2010. – Вып. 1. – С. 84–87.

15. Сичинава, З.И. К вопросу о создании алгоритма принятия решений полиграфным аппаратом, предназначенным для борьбы с коррупцией / З.И. Сичинава // Вестник Пермского университета. Математика. Механика. Информатика. – 2011. – Вып. 2. – С. 108–110.

16. Сичинава, З.И. Варианты построения нейросетевого детектора лжи / Л.Н. Ясницкий, З.И. Сичинава // Нейрокомпьютеры и их применение: тезисы докладов X Всероссийской научной конференции. – М: Изд-во МГППУ, 2012. – С. 53–55.

Регистрация программного продукта

17. Свидетельство о регистрации электронного ресурса № 17926. Программный продукт, предназначенный для проектирования нейронных сетей «ЗС» / З.И. Сичинава. – Зарегистрировано Объединенным фондом электронных ресурсов «Наука и образование» Института научной и педагогической информации Российской академии образования 20.02.2012.

Научное издание

СИЧИНАВА Зураби Иродиевич

НЕЙРОСЕТЕВЫЕ АЛГОРИТМЫ АНАЛИЗА ПОВЕДЕНИЯ
РЕСПОНДЕНТОВ

Специальность: 05.13.01 – Системный анализ, управление и обработка информации (в технике и технологиях)

Редактор *М. Г. Коровушкина*
Технический редактор *В. С. Худяков*
Компьютерная верстка *В. С. Худяков*

Подписано в печать 06.03.2014. Формат 60x84¹/16.
Усл. печ. л. 1,16. Заказ № 102. Тираж 100.

Полиграфический салон «Форвард-С».
614066, Пермь, Стахановская 54 лит М.
Тел./факс: (342) 205-54-41, (342) 205-54-42; e-mail: forward-s2011@yandex.ru