

Глава 5.

Медицинская нейроинформатика

Д.А.Росси́ев

Отдел медицинской нейроинформатики, КрасГМА

Введение

В последнее время сильно возрастает значение информационного обеспечения самых разных медицинских технологий. Оно становится критическим фактором развития практически во всех областях знания, поэтому разработка и внедрение информационных систем является на сегодняшний день одной из самых актуальных задач.

Анализ применения персональных ЭВМ в медицинских учреждениях показывает, что наибольшее использование компьютеров идет для обработки текстовой документации, хранения и обработки баз данных, ведения статистики и финансовых расчетов. Отдельный парк ЭВМ используется совместно с различными диагностическими и лечебными приборами.

В большинстве этих областей использования ЭВМ применяют стандартное программное обеспечение - текстовые редакторы, системы управления базами данных, статистические пакеты и др. Однако некоторые из важнейших участков лечебно-диагностических технологий практически не используют возможности ЭВМ. Прежде всего это диагностика, назначение лечебных мероприятий, прогнозирование течения заболеваний и их исходов. Причины этого носят чрезвычайно сложный характер и постоянно дискутируются. Основные из них - недостаточно развитая техническая база, низкая компьютерная грамотность участников технологий. Большое значение имеет психологический аспект применения ЭВМ. Это серьезная причина, связанная с характером работы врача. Врач явля-

ется исследователем, и его работа носит творческий характер, однако он несет прямую ответственность за результат своей деятельности. Принимая решение о диагнозе или лечении, он опирается на знания и опыт, свои собственные и коллег, являющихся для него авторитетом. Очень важно при этом обоснование решения, если оно подсказывается со стороны. Результат, подсказанный компьютерной программой, работающей по алгоритму, созданному другим человеком, как показывает практика, во многом лишает исследователя творческой инициативы. Навязанное таким образом решение, даже будучи достоверным, психологически подвергается серьезному сомнению.

Существенную роль играют также особенности медико-биологической информации. Большинство медицинских данных имеют описательный характер, выражаются с помощью формализмов, подверженных крайней вариабельности. Данные, даже выражаемые с помощью чисел, также в большинстве случаев не могут быть хорошо упорядочены и классифицируемы, т.к. изменяются в зависимости от клинических традиций различных школ, геосоциальных особенностей регионов и даже отдельных учреждений, а также от времени.

Все задачи, решаемые человеком, с позиций нейроинформационных технологий можно условно классифицировать на две группы.

1. Задачи, имеющие известный и определенный набор условий, на основании которого необходимо получить четкий, точный, недвусмысленный ответ по известному и определенному алгоритму.
2. Задачи, в которых не представляется возможным учесть все реально имеющиеся условия, от которых зависит ответ, а можно лишь выделить приблизительный набор наиболее важных условий. Так как часть условий при этом не учитывается, ответ носит неточный, приблизительный характер, а алгоритм нахождения ответа не может быть выписан точно.

Для решения задач первой группы с большим успехом можно использовать традиционные компьютерные программы. Как бы ни был сложен алгоритм, ограниченность набора условий (входных параметров) дает возможность составления алгоритма решения и написания конкретной программы, решающей данную задачу. Нет никакого смысла в использовании нейроинформационных тех-

нологий для решения таких задач, так как в этом случае нейросетевые методы будут априорно хуже решать такие задачи. Единственным исключением является случай, когда алгоритм вычисления ответа слишком большой и громоздкий и время на решение конкретной задачи по этому алгоритму не удовлетворяет практическим требованиям; кроме того, при получении ответа не требуется абсолютная точность.

При решении задач второй группы применение нейротехнологии оправдывает себя по всем параметрам, при выполнении, однако, двух условий: во-первых, наличия универсального типа архитектуры и единого универсального алгоритма обучения (отсутствие необходимости в их разработке для каждого типа задач), во-вторых, наличия примеров (предыстории, фиксированного опыта), на основании которых производится обучение нейронных сетей. При выполнении этих условий скорость создания экспертных систем возрастает в десятки раз, и соответственно снижается их стоимость.

Практически вся медицинская и биологическая наука состоит именно из задач, относящихся ко второй группе, и в большинстве этих задач достаточно легко набрать необходимое количество примеров для выполнения второго условия. Это задачи диагностики, дифференциальной диагностика, прогнозирования, выбора стратегии и тактики лечения и др. Медицинские задачи практически всегда имеют несколько способов решения и “нечеткий” характер ответа, совпадающий со способом выдачи результата нейронными сетями.

Разработка математических методов решения медико-биологических задач ведется уже не одну сотню лет. Учеными предложено огромное количество способов проверки гипотез и продукции выводов. В 60-е годы были разработаны методы анализа, получившие некоторое распространение и вызвавшие волну публикаций. Общим признаком, объединяющим их, является наличие явных алгоритмов принятия решений. “Диагностический алгоритм включает в себя совокупность правил, определяющих порядок переработки медицинской информации с целью постановки диагноза” [1]. Несмотря на то, что наиболее популярные методы до сих пор активно используются в теоретической биологии и медицине, в клинической практике они не нашли широкого применения. Это связано, во-

первых, с тем, что методы, ориентированные на обработку групповых данных, слабо применимы к отдельным объектам, а во-вторых, с особенностями самой медико-биологической информации. Решения в медицинских и биологических задачах зависят от большого количества неодинаковых по значимости факторов. Поэтому, даже если удастся выстроить правила вывода, связывающие условия задачи с решением, метод, как правило, хорошо работает только на той группе объектов, на которой производились исследования. Естественно, создать универсальный алгоритм невозможно, и при использовании метода для другой подобной группы объектов его почти всегда приходится полностью переконструировать практически заново.

Многолетние исследования, проводимые с самыми различными явными алгоритмами, показали, что медицинские задачи, имеющие неявный характер, решаются явными методами с точностью и удобством, совершенно недостаточными для широкого практического использования в конкретных задачах диагностики, прогнозирования и принятия решений [2].

Поиски и изучение неявных алгоритмов, позволяющих автоматически накапливать и затем использовать опыт при обучении [3], продолжаются уже более 100 лет [4]. Однако первые серьезные попытки создания нейронных сетей были сделаны в 40-50-х годах, когда У.Маккалох и У.Питтс выдвинули основные положения теории работы головного мозга. С появлением дешевых ЭВМ произошел резкий скачок в этой области, которая в начале 80-х годов сформировалась в целую науку - нейроинформатику [5,6,7].

Неявные задачи медицины и биологии явились идеальным полем для применения нейросетевых технологий, и именно в этой области наблюдается наиболее яркий практический успех нейроинформационных методов.

Рассмотрим несколько наиболее интересных нейросетевых приложений для биологии и медицины, созданных различными авторами и школами.

Наибольший интерес для практического здравоохранения представляют системы для диагностики и дифференциальной диагностики заболеваний. При этом для принятия решений могут использоваться самые разнообразные данные - анамнез, клинический осмотр (создаются экспертные системы диагностики,

ограничивающиеся только этим набором [8]), результаты лабораторных тестов и сложных функциональных методов. Список областей медицины, в которых начали применяться новые технологии, чрезвычайно обширен и продолжает расти.

Одним из наиболее интенсивно развиваемых направлений является применение нейросетей в кардиологии.

В Италии разработана чрезвычайно интересная экспертная система для диагностики и лечения артериальной гипертензии [9]. Система включает в себя три нейросетевых модуля, причем ответы одних являются входными данными для других. В начале исследования больному проводят измерение систолического и диастолического давления каждые полчаса в течение суток. Данные за каждый час усредняются. Таким образом, образуется массив из 48 величин артериального давления (по 24 для систолического и диастолического). После этого первый модуль, состоящий из двух трехслойных нейросетей (в каждой из которых 2 входных, 4 "скрытых" и 24 выходных нейрона), на основании данных о поле и возрасте больного рассчитывает аналогичные "должные" величины и сравнивают их с реальными. Параллельно второй модуль (двухслойная нейросеть с 17 входными и 4 выходными нейронами) на основании клинических данных (симптоматика, анамнез) рассчитывает возможные сочетания гипотензивных лекарственных средств, которые могут быть использованы для лечения данного больного. Данные, снятые с выходов обоих модулей, вместе с клиническими данными подаются на вход последнего, третьего модуля (6-слойная нейросеть). Этот модуль оперирует 4 группами гипотензивных препаратов (диуретики, бетаадреноблокаторы, ингибиторы ангиотензина, блокаторы кальциевых каналов). Цель - назначить суточный (почасовой) график приема больным лекарств каждой (если требуется) из 4 групп. Поэтому этот модуль имеет 96 выходных нейронов (4 препарата x 24 часа). С каждого выходного нейрона снимается доза, соответствующая одному препарату, назначаемому на данный час суток. Естественно, что в реальной ситуации большинство выходных данных равны нулю. Таким образом, создается оптимальная для пациента схема лечения гипертензии. Нужно отметить, что система учитывает некоторые особенности приема препаратов больными, например, затруднение приема препаратов ночью (назначает

ночной прием только в крайних случаях), запрет на назначение мочегонных лекарств на ночь. Отличительной чертой системы является возможность пользователя (врача) передавать нейронной сети свой опыт. Для этого создателями программы предусмотрен специальный блок, который выводит на экран компьютера суточные кривые артериального давления и предлагает врачу ввести в компьютер суточную схему приема гипотензивных препаратов в необходимых, по его мнению, дозах. Введенный пример помещается в базу данных. В любое время можно инициировать доучивание нейронных сетей с новыми примерами.

В одной из работ приводится метод выявления атеросклеротических бляшек в артериях [10]. Для этого применяется нейросеть, интерпретирующая флюоресцентные спектры, получаемые при исследовании тканей с помощью лазера.

Аналогичным образом проводится диагностика заболеваний периферических сосудов [11], например, определение форм артериита [12].

Проводится комплекс исследований по использованию нейросетей для диагностики инфаркта миокарда [13,14,15]. Автор приводит данные по чувствительности (77,7%) и специфичности (97,2%) нейросетевого теста. В работе [16], кроме того, с помощью нейронной сети устанавливали диагностическую значимость клинических параметров при диагностике инфаркта миокарда.

Нейросетевой анализ акустических сигналов позволяет проводить диагностику клапанных шумов сердца [17], и оценивать систолическую и диастолическую фазы сердечного сокращения с постановкой предварительного диагноза [18].

Нейросети используются терапевтами для диагностики заболеваний печени по лабораторным данным исследования функций печени [19]; дифференциальной диагностики заболеваний печени [20] и желчного пузыря по УЗИ [21].

Нейропрограммы могут с успехом работать с медицинскими данными, относящимися к субъективным категориям, например, в психиатрии [22]. Оценка субъективных данных дает возможность распознавания психических симптомов] и диагностики и изучения некоторых психиатрических симптомокомплексов.

Актуальная проблема диагностики злокачественных новообразований, возможно, получит новый уровень осмысления с началом применения нейроалгоритмов. Так, в работе [23] показана 80%-я точность ранней диагностики меланом кожи - одного из самых злокачественных заболеваний.

Одним из серьезных направлений применения нейронных сетей является интерпретация медицинских данных. В последние годы идет бурное развитие новых средств диагностики и лечения. При этом наблюдается "вторая волна" изучения и использования древних, старинных методов, и, наоборот, применение последних технических новшеств. Нередко и те и другие методы при использовании предоставляют врачу массу самых разнообразных данных. При этом встает проблема их грамотной и корректной интерпретации. Поиск глубинных закономерностей между получаемыми данными и патологическими процессами начинает отставать от разработки все новых и новых методов, поэтому применение для этой цели нейросетей может оказаться чрезвычайно выгодным.

В одной из старинных методик диагностики по пульсу используются 14 характеристик пульса, измеряемых с нескольких точек. Распознавание и интерпретация данных требует огромного опыта врача, практически невозможного в современных условиях. Нейросеть была применена для "узкой" диагностики только по одной из точек [24], позволяющей оценивать состояние левой почки. Пульс считывается специальным датчиком, совмещенным с микрофоном. Полученная пульсовая кривая (сфигмограмма) передается в компьютер. Вначале программа анализирует несколько пульсовых волн и выстраивает "среднюю" волну. После этого по 5 точкам этой волны нейронная сеть оценивает состояние левой почки.

Классической проблемой в кардиологии является интерпретация электрокардиограмм, требующая значительного опыта врача. Сотрудники Университета Глазго (Великобритания) ведут исследования по применению нейросетей для ЭКГ-диагностики инфарктов миокарда [25]. Входными данными для сетей являются избранные параметры 12-канальной электрокардиограммы и 12-канальной векторкардиограммы (длины зубцов, расстояния между зубцами). Исследователи обучили огромное количество нейросетей (167 сетей для диагностики инфаркта

миокарда передней стенки и 139 сетей для инфаркта нижней стенки) на массиве данных из 360 электрокардиограмм. Обученные сети затем тестировали отдельную выборку с заранее известными ответами (493 случая). Одновременно для получения отдельной серии ответов на тестируемой выборке был использован логический метод (с заранее заданным алгоритмом). Затем сравнивались результаты тестирования выборки лучшими нейросетями и с помощью логического алгоритма. Сравнение показало, что во многих случаях чувствительность и специфичность нейросетевого теста оказались выше, чем у логического метода. Авторы делают справедливый вывод, что в случаях, когда логический алгоритм решения задачи все-таки можно выстроить, разумно комбинировать в экспертных системах оба подхода.

Интерпретация ЭКГ с помощью нейросетей была применена для диагностики злокачественных желудочковых аритмий [26]. Трехслойная сеть с 230 входными синапсами была обучена на 190 пациентов (114 с хронической сердечной недостаточностью и 34 с дилатационной миокардиопатией) различать наличие (у 71 пациента) и отсутствие (у 119 пациентов) желудочковой тахикардии. Результаты тестирования сравнивались с логическим методом интерпретации данных. Показано, что нейросетевой тест обладает большей чувствительностью (73% по сравнению с 70 для логического метода) и специфичностью (83 и 59%).

Интересная работа описывает моделирование применения нейросетей для работы электрокардиостимуляторов (искусственных водителей ритма) [27]. Выпускаемые за рубежом электрокардиостимуляторы задают ритм не жестко, а в зависимости от исходного ритма, генерируемого синусовым узлом сердца. Например, если синусовый узел при какой-либо патологии генерирует недостаточное количество импульсов, водитель ритма компенсирует ритм. Таким образом, электрокардиостимулятор представляет собой систему вход→преобразование→выход, где входом является ритм синусового узла, выходом - собственный ритм электрокардиостимулятора, а преобразование осуществляется по заданному логическому алгоритму. Авторы смоделировали замену логического преобразователя нейронной сетью, так как взаимоотношения между

генерацией импульсов в синусовом узле и требуемым ритмом не линейны и применяемые алгоритмы на практике не всегда эффективны. Нейросеть, обученная на 27 здоровых людях в ситуациях с различной физической нагрузкой, показала гораздо лучшую способность задавать ритм, чем логический алгоритм, применяющийся в электрокардиостимуляторе.

Одной из самых сложных задач для нейросетей в практической медицине является обработка и распознавание сложных образов, например рентгенограмм. В работе [28] описывается экспертная система интерпретации рентгенограмм груди у новорожденных с выбором одного и более диагнозов из 12.

Созданы нейросетевые экспертные системы для классификации опухолей молочной железы (определения, доброкачественная опухоль, или злокачественная) по данным маммографии (сканограмма молочной железы) [29]. По данным, которые приводят авторы, точность такого вывода до применения нейросети составляла не более 75%. При тестировании системы, нейросеть, анализирующая сканограмму, давала правильный ответ в 100% случаев. При тестировании изображение, получаемое в результате метода, представляется в виде матрицы точек размером 1024x1024 пиксела с 10-битовой шкалой яркости. Изображение подается на нейросеть, имеющую 2 входных, 80 "скрытых" и 2 выходных нейрона. При этом один из выходных нейронов "отвечает" за доброкачественную опухоль, другой за злокачественную. Диагноз определяется в зависимости от выходного нейрона, выдавшего больший по величине ответ. Столь высокий процент правильности распознавания, возможно, случаен, и объясняется недостаточным количеством примеров, использовавшихся при обучении и тестировании нейросети (по 10 примеров). Однако даже при такой малой обучающей выборке нейросеть выигрывала по сравнению с традиционным методом интерпретации сканограммы.

Несколько работ посвящены нейросетевой обработке лабораторных анализов и тестов. Приводится нейросетевой метод интерпретации лабораторных данных биохимического анализа крови [30]. В работе показаны преимущества нейронных сетей в сравнении с линейным дискриминантным анализом, которым параллельно обрабатывались данные.

Особое место среди нейросетевых экспертных систем занимают прогностические модели, применяемые, например, для прогнозирования исходов заболеваний.

В 1990 году американская фирма "Апачи Медикл Системз Инк." установила в реанимационном отделении одной из больниц штата Мичиган экспертную систему "Апачи - III" [31]. Ее цель - прогнозирование исхода заболевания у больных, находящихся в тяжелом состоянии. Для прогноза в компьютер необходимо ввести 27 параметров больного: первичный диагноз, симптомы, степень утраты сознания, наличие или отсутствие СПИД и других заболеваний. После этого система выдает вероятность выживания больного в диапазоне от 0 до 100 процентов. Ценность применения системы заключается в том, что она позволяет очень быстро оценить динамику изменения состояния больного, незаметную "на глаз". Например, можно получить ответ у системы до и после введения какого-либо лекарства, и, сравнив ответы, посмотреть, будет ли наблюдаться эффект от терапии. Без программы же изменение состояния иногда не удастся обнаружить в течение нескольких дней. Тестирование показало, что 95% прогнозов, которые делает программа, сбываются с точностью до 3%, что значительно точнее, чем у лучших врачей. Необходимо отметить, что система была обучена на данных, взятых из историй болезней 17448 пациентов, лечившихся в 40 больницах штата в 1989 году. Очевидно, что если качество работы системы обеспечивается таким большим объемом выборки, возможности перенастройки системы не слишком велики. Идеология авторов, создавших эту систему, заключается в как можно большем охвате различных примеров и вариантов (сбор данных в 40 больницах), а не в возможности индивидуализации системы к конкретной клинике. Поэтому данная система не способна к подучиванию в процессе работы, опыт "зашит" в нее жестко. Это может быть существенным недостатком при установке программы в регионы, резко отличающиеся по социально-географическим условиям от тех, где проводилось обучение. Кроме того, огромный массив примеров для обучения повышает стоимость программы.

Прогностические нейросетевые модели могут использоваться в демографии и организации здравоохранения. Создана экспертная система, предска-

вающая, умрет ли человек (в возрасте 55 лет и старше) в ближайшие 10 лет. Прогноз делается по результатам ответов на 18 вопросов анкеты. В анкету включены такие вопросы, как раса, пол, возраст, вредные привычки, семейное положение, семейный доход. 4 из 18 вопросов выявляют индекс массы тела (body mass index) в различные периоды жизни респондента. Индекс рассчитывается как отношение веса к квадрату роста (индекс более 27 кг/м считается тучностью). Повышенное внимание к этому показателю говорит о его значимости для прогноза жизни.

Развитие нейросетевых методов дает возможность их использования как инструмента научных исследований, с помощью которого можно изучать объекты и явления.

Судя по литературным данным, именно биологические научные исследования являются наиболее развиваемой областью применения нейросетей [32]. В последнее время биологи, знакомые с исследованиями в области нейроинформатики, приходят к выводу, что многие системы в живых организмах работают по принципам, сходным с алгоритмами нейронных сетей (или наоборот, нейронные сети работают по принципу биосистем). Таким образом, можно наблюдать "взаимное стимулирование" научных разработок в биологии и нейроинформатике. В работе [33] эндокринная система человека рассматривается как нейронная сеть из 30 элементов, которые представлены различными гормонами, взаимодействующими друг с другом с помощью прямых и обратных связей. Похожие исследования проводятся для иммунной системы [34].

Применение нейросетей для исследований в области нейрофизиологии строится на похожих принципах функционирования нейросетей и нервных структур живых организмов [35]. С помощью нейросети осуществлена попытка моделирования простейшей нервной системы [36].

Сделана попытка применения нейросети для классификации живых организмов [37]: нередко биологам, открывающим новые виды организмов, требуется определить, к какому виду (классу, типу) относится тот или иной представитель флоры или фауны (как правило, это касается микроорганизмов и растений). Система способна работать при отсутствии некоторых входных данных. Это яв-

ляется существенным преимуществом, так как часто при изучении живых объектов не всегда возможно получить всю необходимую информацию.

Нейросети использованы для идентификации человеческих хромосом. В биологических исследованиях, а также в криминалистике, часто бывает нужно определить, к какой из 23 имеющихся у человека пар хромосом относится выделенная хромосома. Точность существующих методов достигала 75 - 85%. Нейро-классификатор, на вход которого подается 30 признаков изображения хромосомы, определяет ответ с точностью, приближающейся к 100% [38].

Анализ публикаций о применении нейросетевых технологий в медицине показывает, что практически отсутствуют какие-либо методологии разработки нейросетевых медицинских систем, о чем свидетельствует как отсутствие работ такого профиля, так и огромное разнообразие подходов к нейросетевым алгоритмам обучения и архитектурам нейронных сетей. Это подтверждает то, что медицинская нейроинформатика как наука находится еще, в основном, на стадии накопления фактического материала.

Нужно отметить, что все медицинские приложения нейронных сетей для практического здравоохранения (диагностика, лечение, прогнозирование) созданы зарубежными авторами. Большинство отечественных работ направлено на исследование самих нейронных сетей и моделирование с их помощью некоторых биологических процессов (в основном, функций нервной системы).

Общая черта, объединяющая приведенные примеры - отсутствие единой универсальной технологии создания таких приложений. В публикуемых разработках используются самые разнообразные архитектуры и алгоритмы функционирования нейронных сетей. Это приводит к тому, что для почти для каждой задачи разрабатывается своя собственная архитектура, и часто - уникальный алгоритм или уникальная модификация уже существующего. С точки зрения практического применения такие экспертные системы почти не отличаются от традиционных программ принятия решений; предложены даже методы преобразования традиционных экспертных систем в нейросетевые [39]. Их разработка требует участия специалистов по нейроинформатике, а возможности конструирования пользователем практически отсутствуют. Это делает такие системы чрезвычайно

дорогими и не очень удобными для практического применения, поэтому в публикациях авторы в основном сравнивают качество работы нейросетевых алгоритмов и традиционных систем, работающих по правилам вывода.

Что же можно предложить взамен? Об этом и пойдет речь ниже. На протяжении нескольких лет красноярская научная группа «НейроКомп» [40], объединяющая ученых-математиков и медиков (Красноярский ВЦ СО РАН, КГТУ, Красноярская Медицинская Академия и ряд клинических учреждений города), разрабатывает технологии создания нейросетевых экспертных систем, которые применяются в практической медицине и биологии на обычных IBM-совместимых компьютерах [41,42,43].

Наиболее важным отличием предлагаемого подхода является возможность конструирования экспертных систем самим врачом-специалистом, который может передать нейронной сети свой индивидуальный опыт, опыт своих коллег, или обучать сеть на реальных данных, полученных путем наблюдений. При использовании разработанного нами пакета программ MultiNeuron 2.0 [44] для конструирования экспертной системы не требуется участие специалистов-математиков и программистов, что делает создаваемые системы более дешевыми, а главное, адаптированными к конечному пользователю.

Далее мы рассмотрим основные принципы и особенности этих технологий.

Все неалгоритмируемые или трудноалгоритмируемые задачи, решаемые нейронными сетями, можно классифицировать на два принципиально различающихся типа в зависимости от характера ответа - задачи классификации и задачи предикции.

Задачи классификации - основная и очень обширная группа медико-биологических задач. Ответом в них является класс - выбор одного варианта из заранее известного набора вариантов. Классификация может быть бинарной (элементарная классификация) - в этом случае набор возможных ответов состоит из двух вариантов (классов), и n-арной, где число классов более двух. Примерами бинарной классификации могут служить как объективные категории (пол чело-

века - мужской или женский; характер опухоли - доброкачественный или злокачественный), так и субъективные категории (здоров человек или болен; наличие или отсутствие склонности к простудным заболеваниям). В некоторых случаях не представляется возможным отнесение ответа задачи к объективной или субъективной категории, и это не имеет принципиального значения для обучения и работы нейросетевой экспертной системы.

Важной чертой задачи классификации по определению является возможность выбора одного и только одного варианта решения (класса). Поэтому постановка диагноза не может считаться одной классификационной задачей, т.к. у одного человека может одновременно присутствовать несколько патологий. В случае невозможности выбирать один вариант ответа (множественности выбора) задача подразделяется на подзадачи, каждая из которых представляет собой классификационную задачу.

Другой вид задач для нейросетей - задачи предикции, или предсказания. Они подразделяются на предсказание числа (одномерная предикция) и вектора (векторная предикция, более общий случай). Отличие от классификационных задач заключается в том, что ответ в задачах предикции может быть дробным и принимать любые значения на каком-либо интервале.

Векторная предикция предполагает, что ответ может быть представлен в виде нескольких независимых друг от друга чисел, образующих точку (или вектор) в многомерном пространстве, размерность которого равно количеству предсказываемых чисел. Число координат вектора называется при этом размерностью вектора ответа.

При решении реальных задач возможны различные комбинации предикции и классификации, и постановка задачи должна быть сделана самим предметным специалистом.

Архитектура нейронной сети

Основой работы самообучающихся нейропрограмм является нейронная сеть, представляющая собой совокупность нейронов - простых элементов, связанных между собой определенным образом. Нейроны и межнейронные связи

задаются программно на обычном компьютере или могут иметь "материальную" основу - особую микросхему (нейрочип), которые применяются в специально созданных нейрокомпьютерах. Структура взаимосвязей между нейронами в нейрокомпьютере или нейропрограмме аналогична таковой в биологических объектах. Искусственный нейрон имеет коммуникации с другими нейронами через синапсы, передающие сигналы от других нейронов к данному (дендриты) или от данного нейрона к другим (аксон). Кроме того, нейрон может быть связан сам с собой. Несколько нейронов, связанных между собой определенным образом, образуют нейронную сеть.

Нейросеть, также как и биологический аналог, должна иметь каналы для связи с внешним миром. Одни каналы обеспечивают поступление информации из внешнего мира на нейросеть, другие выводят информацию из нейросети во внешний мир. Поэтому одни нейроны сети рассматриваются как входные, другие же - как выходные. Часть нейронов может не сообщаться с внешним миром, а взаимодействовать с входными, выходными и такими же нейронами ("скрытые" нейроны).

Очевидно, что существует огромное количество способов соединения нейронов, растущее с увеличением числа нейронов в сети. Наиболее употребительной является слоистая архитектура, в которой нейроны располагаются "слоями". В наиболее общем случае аксоны каждого нейрона одного слоя направлены к нейронам следующего слоя. Таким образом, нейроны первого слоя являются входными (принимающими информацию из внешнего мира), нейроны последнего слоя - выходными (выдающими информацию во внешний мир). Схема трехслойной сети изображена на рисунке 1.

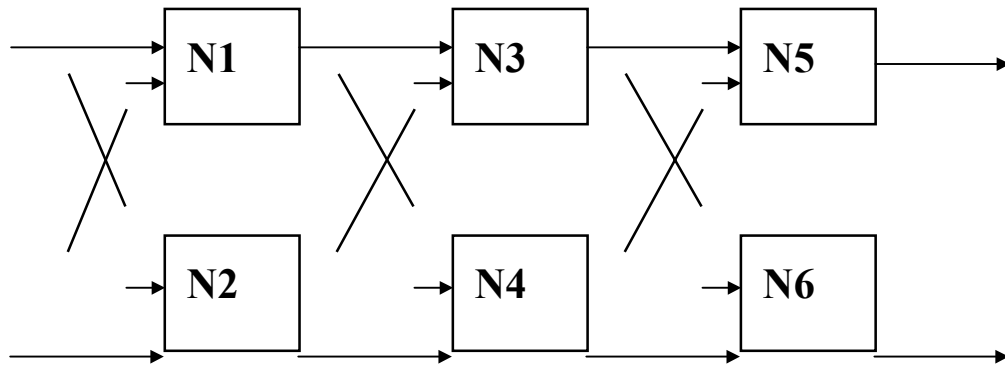


Рисунок 1. Трехслойная сеть с 6 нейронами

Другой вид архитектуры - полносвязная, когда каждый нейрон соединен с каждым, в том числе сам с собой. Пример простейшей нейросети из 3 нейронов показан на рисунке 2. Для удобства изображения из каждого нейрона выходит не один, а несколько аксонов, направленных на другие нейроны или во внешний мир, что аналогично присоединенным к одному аксону через синапсы нескольким дендритам.

Слоистые сети являются частными случаями полносвязных.

Для построения экспертных систем мы выбрали именно полносвязные нейросети, исходя из следующих соображений. Во-первых, при одинаковом числе нейронов полносвязные сети имеют большее количество межнейронных связей, что увеличивает информационную емкость сети. Во-вторых, полносвязная архитектура является намного более универсальной, что не требует экспериментов с вариациями схемы соединений для каждой задачи. В-третьих, в случае эмуляции сети на обычной ЭВМ полносвязные сети обладают серьезными преимуществами, прежде всего в скорости функционирования и простоте программной реализации без ущерба качеству обучаемости.

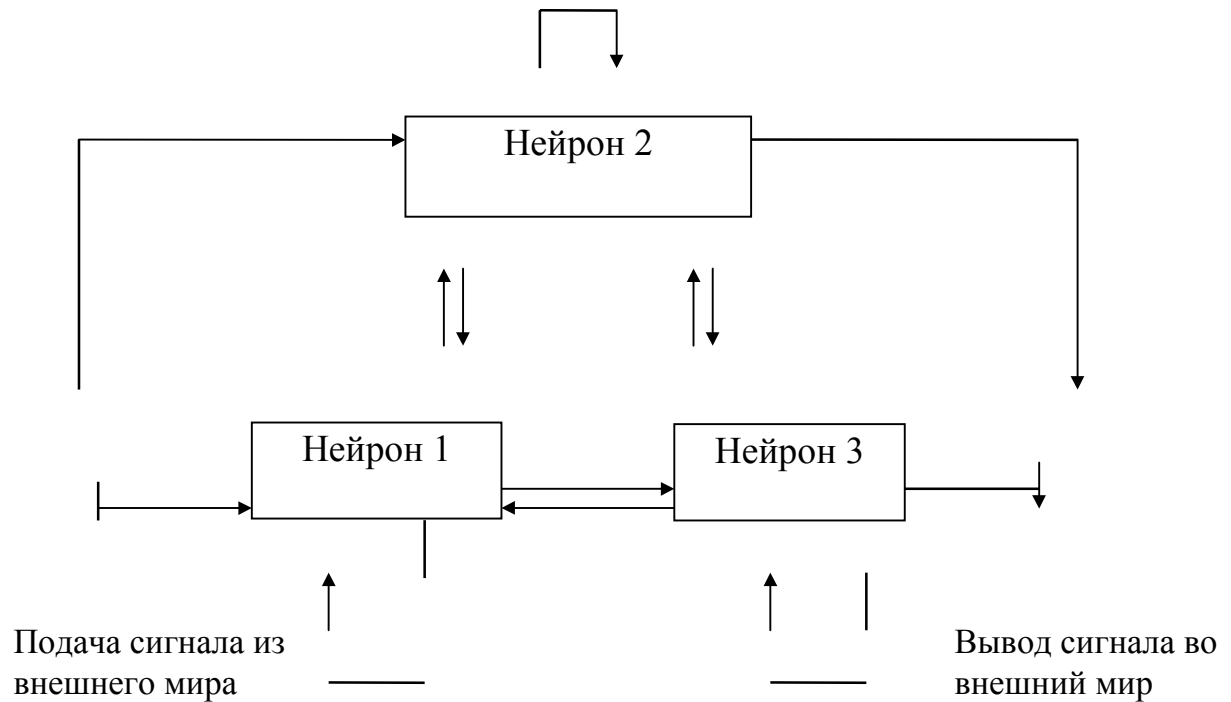


Рисунок 2. Схема простейшей нейронной сети из 3 нейронов. Сеть имеет 13 синапсов, 4 из которых служат для связи с внешним миром, а остальные соединяют нейроны между собой

Функционирование нейрона

Рассмотрим устройство и функционирование отдельного нейрона. Каждое соединение от нейрона к нейрону называется синапсом. На рисунке 3 представлен нейрон с группой синапсов, соединяющих нейрон либо с другими нейронами, либо с внешним миром. Для рассмотрения работы нейрона неважно, приходит ли сигнал к нейрону из внешнего мира или с другого нейрона, и неважно, куда отправляется сигнал с нейрона. В полносвязных сетях выходной сигнал направляется всем остальным нейронам.

Нейрон состоит из двух функциональных блоков: входного сумматора и собственно нейрона, или преобразователя.

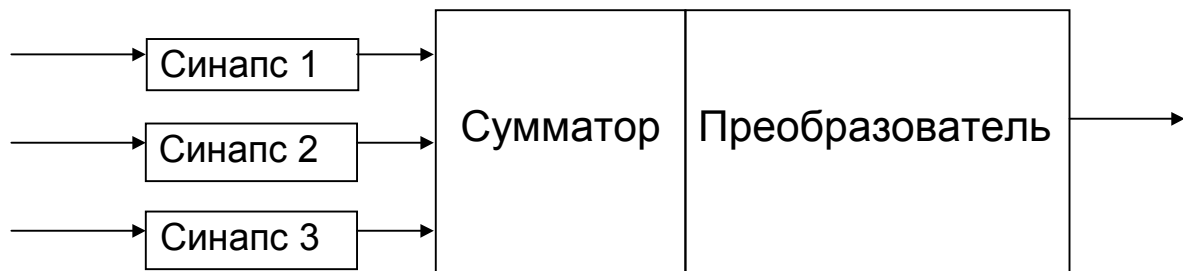


Рисунок 3. Схема нейрона

Функционирование нейрона происходит следующим образом.

В текущий момент времени через входные синапсы (на рисунке их 3) на нейрон направляются сигналы от других нейронов и/или из внешнего мира. Каждый синапс имеет параметр, называемый весом синапса, и представляющий какое-либо число. Сигнал, проходящий через синапс, умножается на вес этого синапса. В зависимости от веса, сигнал может быть усилен (модуль веса > 1) или ослаблен (модуль веса < 1) по амплитуде. Сигналы от всех синапсов, ведущих к данному нейрону, принимает сумматор.

Сумматор производит суммирование всех пришедших сигналов и подает на собственно нейрон (преобразователь) одно число - полученную сумму. Величина этого числа будет зависеть как от величин исходных сигналов, так и от весов синапсов. Нейрон, получивший это число, преобразует его согласно своей функции, в результате которой получается другое число, и отправляет его по "аксону" всем остальным нейронам через соответствующие синапсы. Последующие нейроны производят с полученными сигналами такие же операции, лишь с тем различием, что во-первых, веса их синапсов могут быть уже другими, во-вторых, другие нейроны могут иметь другой вид функции преобразования. В конструируемых нами нейронных сетях все нейроны имеют одну и ту же функцию. Эта функция, называемая характеристической, имеет вид:

$$f(X) = X / (C + |X|) \quad (1)$$

где X - сигнал, поступающий с сумматора, C - константа, называемая характеристикой нейрона. Экспериментальным путем мы получили, что оптимальный диапазон характеристики для решения подавляющего большинства задач составляет от 0,1 до 0,8. График характеристической функции представлен на рисунке 4.

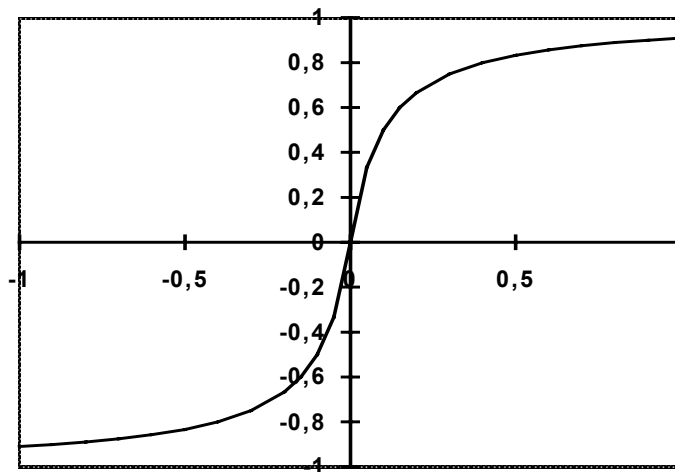


Рисунок 4. График характеристической функции

Выбор функции такого вида обусловлен тем, что она гладкая, непрерывная на всем диапазоне переменных X , диапазон значений всегда ограничен.

Функционирование нейросети

В случае эмуляции нейросети на обычном компьютере все математические операции осуществляет программа. Нейронная сеть при этом представляет собой массив синаптических весов. Этот массив может находиться либо на диске компьютера в виде файла определенного формата (при хранении нейросети) либо в оперативной памяти компьютера (когда нейросеть функционирует).

При создании новой нейросети в памяти компьютера отводится место под массив синаптических весов, называемый картой. Этот массив заполняется совершенно случайными числами из определенного диапазона. Поэтому каждая созданная сеть даже при одних и тех же параметрах (число нейронов, их характеристика) является уникальной. Уникальность сетей проявляется в том, что сети с одинаковыми параметрами, обучающиеся на одинаковых задачах, ведут себя неодинаково. Это касается времени обучения, качества обучения, уверенности в выдаваемых ответах при тестировании. В этом еще одно сходство сетей с био-объектами.

Рассмотрим работу сети безотносительно к процессу обучения или выдачи решения. После инициализации сети веса всех синапсов случайны. В начальный

момент времени на входные нейроны через входные синапсы (один или несколько) подается из внешнего мира вектор входных сигналов, представляющий набор чисел (или одно число). Далее этот сигнал начинает распространяться по всем связям между нейронами, изменяясь при прохождении через каждый нейрон согласно схеме функционирования нейрона. В конечном итоге, после одного прохода выходные нейроны выдадут во внешний мир какие-либо сигналы. Вся процедура однократного прохождения сигналов по нейронной сети является тактом функционирования сети. Можно не считывать сигналы с выходных нейронов после одного такта функционирования, а продолжить обмен сигналами еще несколько раз, не подавая сигналов на вход. Обычно количество тактов функционирования между подачей сигналов на вход и снятием сигналов с выхода фиксировано для данной сети. Как правило, этот параметр задается при инициализации сети и хранится в файле сети вместе с матрицей коэффициентов (как и некоторые другие параметры - число нейронов, характеристика нейронов и т.д.).

Таким образом, нейронная сеть, получающая на входе некоторый сигнал, способна после прохода его по нейронам выдавать на выходе определенный ответ, который зависит от весовых коэффициентов всех нейронов и от самого сигнала. Очевидно, что при осуществлении таких процедур на только что инициализированной сети мы будем получать на выходе сигналы, лишенные всякого смысла (весовые коэффициенты случайны). Чтобы добиться выдачи сетью требуемого результата, необходимо обучить ее.

Общая схема обучения нейронной сети

Для обучения нейронной сети необходима обучающая выборка (задачник), состоящая из примеров. Каждый пример представляет собой задачу одного и того же типа с индивидуальным набором условий (входных параметров) и заранее известным ответом. Например, в качестве входных параметров в одном примере могут использоваться данные обследования одного больного, тогда заранее известным ответом в этом примере может быть диагноз. Несколько примеров с разными ответами образуют задачник. Задачник располагается в базе данных, каждая запись которой является примером.

Не останавливаясь на математических алгоритмах, подробно описанных в монографии [45], рассмотрим общую схему обучения нейросети.

1. Из обучающей выборки берется текущий пример (изначально, первый) и его входные параметры (представляющие в совокупности вектор входных сигналов) подаются его на входные синапсы обучаемой нейросети. Обычно каждый входной параметр примера подается на один соответствующий входной синапс.
2. Нейросеть производит заданное количество тактов функционирования, при этом вектор входных сигналов распространяется по связям между нейронами (прямое функционирование).
3. Измеряются сигналы, выданные теми нейронами, которые считаются выходными.
4. Производится интерпретация выданных сигналов, и вычисляется оценка, характеризующая различие между выданным сетью ответом и требуемым ответом, имеющимся в примере. Оценка вычисляется с помощью соответствующей функции оценки. Чем меньше оценка, тем лучше распознан пример, тем ближе выданный сетью ответ к требуемому. Оценка, равная нулю, означает что требуемое соответствие вычисленного и известного ответов достигнуто. Заметим, что только что инициализированная (необученная) нейросеть может выдать правильный ответ только совершенно случайно.
5. Если оценка примера равна нулю, ничего не предпринимается. В противном случае на основании оценки вычисляются поправочные коэффициенты для каждого синаптического веса матрицы связей, после чего производится подстройка синаптических весов (обратное функционирование). В коррекции весов синапсов и заключается обучение.
6. Осуществляется переход к следующему примеру задачника и вышеперечисленные операции повторяются. Проход по всем примерам обучающей выборки с первого по последний считается одним циклом обучения.

При прохождении цикла каждый пример имеет свою оценку. Вычисляется, кроме того, суммарная оценка множества всех примеров обучающей выборки.

Если после прохождения нескольких циклов она равна нулю, обучение считается законченным, в противном случае циклы повторяются.

Количество циклов обучения, также как и время, требующееся для полного обучения, зависят от многих факторов - величины обучающей выборки, количества входных параметров, вида задачи, типа и параметров нейросети и даже от случайного расклада весов синапсов при инициализации сети.

Методологические аспекты обучения нейросетей

Иногда (при решении медико-биологических задач - крайне редко) встречаются ситуации, когда сеть не может обучаться. Это происходит в том случае, когда на определенном этапе обучения исчерпываются дальнейшие возможности поиска закономерностей между обучающими параметрами и результатами. Простейшая ситуация - когда два примера с совершенно одинаковыми наборами параметров подаются сети как принадлежащие различным классам (в классификаторах) или имеющие различное значение ответа (в предикторах). Очевидно, оба этих примера всегда будут попадать в одну и ту же точку в пространстве, их невозможно будет отделить друг от друга, и процесс обучения остановится. Программа, управляющая нейросетями, сигнализирует об окончании процесса обучения, причем указывает, что дальнейшее обучение невозможно. Задача специалиста, обучающего нейросети - избежать таких ситуаций, для чего нужны четкая постановка задачи и тщательный контроль обучающей выборки.

Обученная нейросеть автоматически записывается на диск компьютера как обыкновенный файл и может храниться там, сколько необходимо. В любой момент времени можно считать сеть с диска и продолжить обучение решению данной задачи со старой или новой обучающей выборкой. Одна нейросеть обучается решать только одну задачу классификации или предикции, однако может использовать для обучения различные обучающие выборки. Они могут различаться по количеству примеров, но должны соответствовать друг другу по числу обучающих параметров, числу классов (в классификационной задаче), а главное, по смыслу.

Говоря об обучении нейросетей, следует рассмотреть еще один важный аспект этой темы. Мы уже знаем, что успех обучения во многом зависит от числа

нейронов в сети, или, точнее, от числа синапсов. Именно весовые коэффициенты синапсов хранят "опыт" сети. Теоретически, бесконечно увеличивая число нейронов и синапсов, всегда можно добиться полного обучения сети на данном задатнике, однако это ли является целью создателя экспертной системы? Очевидно, нет. Главное, чтобы обученная сеть хорошо распознавала примеры, как раз не входящие в задатник.

Проблема заключается в том, что сеть с заведомо большим (избыточным) числом синапсов (относительно данного задатника) может хорошо обучиться, просто "механически запомнив" имеющиеся примеры. Такая сеть обучится быстро (нет необходимости как можно более точной подстройки весов) за счет количества, а не качества.

Хорошим практическим выходом из данной затруднительной ситуации были бы сети, способные автоматически наращивать число нейронов при невозможности дальнейшего обучения, не теряя при этом уже имеющегося опыта. Последнее условие вызывает значительные трудности. Нейросеть представляет собой единое целое, и добавление нового нейрона к сети, работающей в рамках имеющейся сейчас концепции, приведет к необходимости полностью переучивать сеть. Это требует обращения к первоначальному задатнику, что во многих случаях неприемлемо.

Поэтому создателю самообучающихся систем приходится идти на компромисс: либо делать сеть с некоторым избытком нейронов, имеющую резерв для накопления опыта, но обладающую относительно низкой способностью к экстраполяции, либо обучить сеть с небольшим числом нейронов, которая вряд ли сможет набрать потом дополнительный опыт. Все это, конечно, зависит еще и от задатника - насколько тесные взаимосвязи имеются между обучающими параметрами и известными ответами примеров. Чем больше таких взаимосвязей, тем меньше необходимость в "механическом запоминании" примеров. Практика показывает, что большинство биологических и медицинских задач имеют достаточно хорошие взаимосвязи, конечно, при грамотной постановке задачи и выборе обучающих параметров. Однако в рамках предлагаемой методологии, с учетом высокой скорости обучения нейросетей разработаны стратегия и тактика

обучения, позволяющие обойти вышеуказанный компромисс, за счет, правда, большего времени, необходимого для обучения.

Тестирование примеров

При тестировании примеров необходимость в системе подстройки весов синапсов отпадает, поэтому при создании экспертных систем блок программы, содержащий алгоритмы обучения, может не включаться в программу в случае, если не предполагается доучивать сети в процессе работы экспертной системы. Тестирование примеров нейросетью может проводиться с различными целями:

1. Проверка того, как обучилась нейросеть;
2. Решение конкретных задач.
3. Моделирование.

В первом случае осуществляется тестирование выборки с заранее известными ответами примеров. Таким образом можно проверить, правильно ли сеть определяет ответы примеров и насколько уверенно она это делает. Определенный сетью ответ примера сравнивается с заранее известным. Как правило, сначала тестирование проводится на той выборке, на которой сеть обучалась. Если сеть обучилась полностью, то при тестировании той же самой обучающей выборки ответы всех примеров будут определяться правильно. Гораздо больший интерес представляет тестирование аналогичной выборки с заранее известными ответами, но примеры которой не участвовали в обучении сети. Неправильное определение ответов некоторых примеров может быть вызвано несколькими причинами:

1. Выборка, по которой обучалась нейросеть, недостаточно полно отражает картину соответствия ответов обучающим параметрам, иначе говоря, в обучающей выборке слишком мало примеров.
2. Выборка, по которой обучалась нейросеть, составлена тенденциозно. Это означает, что для обучения подбирались примеры, которые, по мнению исследователя, "являются самыми яркими представителями своего класса или группы". Это серьезная ошибка. При такой выборке, конечно, нейросеть будет обучаться намного лучше, но способность ее к тестированию других примеров

существенно падает. В действительности необходимо обучать сеть на реальных данных, какими бы они противоречивыми ни были. Если сеть не сможет обучиться полностью, можно применить некоторые меры, которые будут рассмотрены ниже.

3. Обучающая выборка имеет недостаточное количество обучающих параметров и сеть не может найти закономерности между входными сигналами и ответами.
4. При создании сети не оптимально были выбраны некоторые сетевые параметры, например, число нейронов, число тактов функционирования или характеристика сети. Ниже мы подробно остановимся на методологии оптимального выбора этих значений.
5. Задана неверная классификационная модель (при обучении нейросетей-классификаторов). Возможно, на самом деле примеры группируются в 3 класса, а пользователь задает только 2. Далее будет рассмотрен метод коррекции классификационной модели.

При решении конкретных задач сети подаются примеры, ответ которых неизвестен. В этой ситуации программа не может проверить правильность решения.

В классификационных задачах при ответе нейросеть не только выдает результат - класс тестируемого примера. Как уже говорилось, в отличие от большинства экспертных систем, работающих по четким правилам, решение задачи на основе опыта всегда имеет “нечеткий” характер. Поэтому кроме класса тестируемого примера сеть вычисляет коэффициент уверенности в данном решении. Коэффициент уверенности зависит от заданного уровня надежности и рассчитывается по формуле:

$$КУ = (Max1 - Max2) / R \times 100\%, \quad (2)$$

где Max1 - ответ выходного нейрона, отвечающего за класс-"победитель", Max2 - ответ выходного нейрона, выдавшего следующий по максимальной величине сигнал, R - уровень надежности. Судя по формуле ясно, что уверенность сети зависит от того, насколько наибольший из выходных сигналов превышает второй по величине сигнал. Естественно, если КУ получается более 100%, он при-

равнивается к этому числу. Из того, что в знаменателе правой части формулы стоит уровень надежности, следует, на первый взгляд, парадоксальный вывод: сеть, обученная лучше (уровень надежности больше) даст меньшую уверенность в ответе, чем сеть, обученная хуже. Однако при внимательном рассмотрении проблемы делается ясно, что при тестировании примера первой сетью, выражение, стоящее в числителе, также будет больше (ответ одного нейрона будет намного больше ответа другого) вследствие лучшей обученности. Кроме того, КУ выражает все же уверенность конкретной сети, которая во многом зависит от того, насколько тестируемый пример близок к примерам, на которых обучалась эта сеть.

Если пример отличается достаточно сильно, лучше обученная сеть будет и сомневаться больше, чем сеть с меньшим "опытом".

Один из показателей качества обучения - определение прогностической способности нейросети - состоит в подсчете процента правильно распознанных примеров. При сравнении качества обучения двух нейросетей, в случае, когда обе сети дают одинаковую прогностическую способность, можно подсчитывать средний процент уверенности при тестировании выборки. Он рассчитывается как средняя арифметическая процентных величин уверенности, полученных при тестировании каждого примера за известный результат.

Иногда необходимо знать, к каким еще классам, кроме найденного, близок тестируемый пример. Это можно сделать несколькими способами, из которых достаточно трудно выбрать наилучший, однако самый оптимальный, на наш взгляд, заключается в том, чтобы просто сравнить сигналы, полученные со всех выходных нейронов (их можно выразить в процентах от максимально возможного). Понятно, что когда все выходные сигналы близки друг к другу, сеть затрудняется дать уверенный ответ.

Из вышесказанного вытекает очень полезный для практики вывод. Изменяя в различных направлениях значения параметров примера и повторяя его тестирование, можно видеть, что и на сколько нужно изменить, чтобы пример стал принадлежать к требуемому классу. Это может быть полезным для медицинской диагностики и прогнозирования. Предположим, что сеть обучена дифференци-

ровать больных и здоровых людей по набору клинических параметров. Изменяя на компьютере эти параметры, можно добиться, чтобы пример, определяемый как "больной", стал принадлежать классу "здоровый". Таким образом, станет ясно, какие клинические параметры подлежат изменению для улучшения состояния больного.

Создание медицинских нейросетевых экспертных систем

Любая экспертная система должна состоять условно из четырех блоков: интерфейс с пользователем, база знаний, вычислительный блок, блок объяснений, позволяющий пользователю проследить "ход рассуждений" системы в конкретном случае. Связующим элементом между этими блоками является метод, с помощью которого экспертная система в ответ на запрос пользователя выдает результат (заключение). Такие методы можно разделить на три основные группы:

1. Методы логических правил "в чистом виде", когда формализация правил получения результата осуществляется специалистом;
2. Те же методы, однако формализация правил осуществляется исследователем, наблюдающим за работой специалиста со стороны;
3. Методы, основанные на принципе "смотри и учись".

Создание даже простых экспертных систем, основанных на методах 1 и 2, представляет собой нелегкую задачу, прежде всего потому, что требует совместной работы специалистов различного профиля. Традиционные экспертные системы, основанные на базах знаний и логических правилах, требуют для создания довольно большого времени и средств. Создание традиционной экспертной системы можно условно разделить на несколько этапов.

1. Постановка задачи: определение целей работы экспертной системы, набора входных данных и формы представления ответа.
2. Сбор данных: набор репрезентативного материала для статистических исследований и его структурирование - разделение на подгруппы по разнообразным признакам.

3. Статистическая обработка: выявление закономерностей, связывающих входные данные с ответом - расчет средних и относительных величин, их сравнение, корреляционный, регрессионный, факторный анализы и т.д.
4. Создание базы знаний: оформление логических правил, по которым должна работать экспертная система.
5. Программирование алгоритмов: перенесение логических правил на язык программирования.
6. Создание интерфейса системы: разработка средств взаимодействия системы с пользователем - формы ввода данных, вывода ответа и т.п.
7. Отладка и тестирование: проверка работы программы и испытание в реальных условиях.

При создании логических экспертных систем наибольшую часть времени занимают 3, 4 и 5 этапы, требующие совместной работы как предметных специалистов, так и программистов и математиков. Несмотря на появление компьютерных средств проектирования экспертных систем, основная работа все равно возложена на специалистов. При этом возникают сразу несколько серьезных проблем.

Первая из них состоит в том, что при решении сложных реальных задач (экономика, проектирование, инженерия, биология) число логических правил значительно увеличивается. Часто возникает настолько сложная система взаимосвязей между ними, что ее просто не удастся осмыслить. Разбивка задачи на блоки также не всегда помогает: во-первых, это тоже не всегда просто сделать, во-вторых, при разбивке иногда могут теряться некоторые взаимосвязи.

Вторая, еще более серьезная проблема состоит в том, что далеко не всегда удастся выразить вычислительный процесс логическими правилами. Это может быть связано как со сложностью самой задачи, так и с особенностями деятельности предметного специалиста. Особенно ярко это проявляется в медицине, где процесс принятия решения во многом опирается на интуицию и опыт врача, не являющегося экспертом в области собственного мышления. Во всех этих случаях говорят, что задача не поддается алгоритмизации. Кроме того, даже если создателям удастся разработать алгоритм, никогда нет достаточной гарантии, что он

будет корректно работать в реальных условиях, а это можно проверить только после окончания всех работ по созданию системы.

В создании самообучающейся системы также можно выделить несколько этапов, часть из которых совпадает с этапами создания традиционных систем.

1. Постановка задачи. То же, что и для традиционных систем плюс выбор оптимальной структуры нейронной сети и методов обучения (для большинства задач структура и методы стандартны).
2. Сбор обучающих данных. Набор примеров для обучения сети, каждый из которых представляет массив входных данных и соответствующий ему заранее известный ответ.
3. Создание и обучение нейросети. Данный этап не требует проведения статистических вычислений, а если задача укладывается в стандартную схему (в большинстве случаев), то и программистской работы. Если задача нестандартная, требуется адаптация структуры нейросети и метода вычисления оценки при обучении. Обучение нейросети в большинстве стандартных случаев представляет собой автоматический процесс, который только после его окончания требует участия специалиста для оценки результатов. Естественно, часто может требоваться корректировка, создание дополнительных сетей с другими параметрами и т.д., однако всегда есть возможность оценить работу системы на любом этапе обучения, протестировав контрольную выборку. Разрабатывая методологию создания нейросетевых экспертных систем, мы исходили из возможности разработки наиболее индивидуализированных (рассчитанных на одного конкретного пользователя-специалиста) систем самим этим специалистом. Конечно, ничто не мешает объединять в одной системе индивидуальный опыт нескольких специалистов. Отсутствие "математических" этапов реализует такие возможности. Предметный специалист в состоянии самостоятельно поставить задачу, более того, никто, кроме него, не сможет сделать это лучше. Сбор материала также должен осуществлять предметный специалист. Схемы постановки задач, способы представления данных и способы продукции ответа нейросетью разработаны таким образом, что большинство задач во многих областях укладываются в эти стандартные схемы. Поэтому при наличии хо-

рошо продуманных инструментальных программных средств работы с нейронными сетями и документации к ним большинство специалистов способны самостоятельно разрабатывать не очень сложные нейросетевые приложения.

4. Создание интерфейса. То же, что и для традиционных экспертных систем.
5. Отладка и тестирование. Этап включает в основном отладку работы программы, т.к. тестирование часто проводится в процессе обучения сетей.
6. Доучивание. Этап, характерный только для обучающихся систем. При создании нейроэкспертных программ довольно редко возможно сразу собрать достаточное количество данных для хорошего обучения сети. Поэтому, создавая нейросистему, исследователи определяют наилучшие параметры сетей и проводят стартовое обучение. В последующем пользователи доучивают систему в условиях реальной работы и реальных данных, передавая ей опыт. Более того, коренное отличие методологии создания нейросетевых систем от традиционных состоит именно в том, что система никогда не создается сразу готовой, и более того, никогда не является полностью законченной, продолжая накапливать опыт в процессе эксплуатации.

Резюмируем имеющиеся преимущества нейросетевых экспертных систем перед обычными, которые, как уже говорилось, проявляются только при решении трудноалгоритмируемых задач.

1. Нейросети принимают решения на основе опыта, приобретаемого ими самостоятельно. "Самостоятельно" в данном случае означает то, что создателю экспертной системы не требуется устанавливать взаимосвязи между входными данными и необходимым решением, затрачивая время на разнообразную статобработку, подбор математического аппарата, создание и проверку математических моделей.
2. Решение, принимаемое нейросетью, не является категоричным. Сеть выдает решение вместе со степенью уверенности в нем, что оставляет пользователю возможность критически оценивать ее ответ.
3. Нейросеть позволяет моделировать ситуацию принятия решения.

4. Нейросети дают ответ очень быстро (доли секунды), что позволяет использовать их в различных динамических системах, требующих незамедлительного принятия решения.
5. Возможности нейросетей (коррекция классификационной модели, минимизация обучающих параметров и др.) позволяют упрощать процесс создания экспертных систем, определять направления научного поиска.

Главным критерием работы нейросетевых экспертных систем должна быть практика - многократные испытания и проверки в самых различных условиях.

Определенным препятствием использования нейросетей может является все же некоторая ограниченность задач, решаемых ими. Иногда в блоке трудно-алгоритмируемых задач, решаемых с помощью самообучающейся экспертной системы, могут присутствовать элементы четких правил. В таком случае совершенно логично комбинировать в одной экспертной системе несколько нейросетей или даже обычные математические методы и строить из них иерархические блоки, одни из которых используют для своих действий результаты работы других. Следует подчеркнуть, что применение неявных алгоритмов не противоречит и не отменяет использование формальных методов, а может дополняться ими при необходимости. Например, если с помощью нейросети определяется оптимальная комбинация лекарственных препаратов для лечения пациента, и имеется совершенно четкое противопоказание к назначению определенного препарата, в экспертную систему может быть введен простой логический блок, препятствующий назначению этого лекарства независимо от решения нейросетей.

Ниже мы рассмотрим принципы и особенности создания нейросетевых экспертных систем для биологии и медицины.

Постановка задачи

Постановка задачи всегда является прерогативой предметных специалистов.

Прежде всего, необходимо определить, что представляет собой акт работы экспертной системы, начинающийся со ввода данных (условия задачи) пользователем и заканчивающийся выдачей ответа. Например, в медицинских диагностических задачах акт работы системы чаще всего начинается с введения в ЭВМ

данных о пациенте (симптоматика, анамнез, результаты анализов и т.д.) и заканчивается выдачей возможного диагноза и/или прогноза. Часто одна задача komponуется из нескольких подзадач. В этом случае каждой подзадаче может соответствовать отдельный акт работы системы, хотя для пользователя это может быть совершенно незаметным. Например, по данным о пациенте требуется установить диагноз и назначить соответствующее лечение. Установление диагноза на базе введенных параметров - первая подзадача, назначение терапии - вторая, при этом для назначения терапии нужны не только исходные данные, но и результат решения предыдущей задачи - диагноз.

Полезно сразу же для каждой подзадачи определить ее тип (классификация, предикция, векторная предикция).

Обычно каждая подзадача решается одной нейросетью или несколькими нейросетями, объединенными в один функциональный блок (малые эксперты [46]). При этом введенные данные подаются последовательно на каждую нейросеть блока и каждая нейросеть выдает ответ. Ответы могут различаться, поэтому в такой ситуации требуется разработать способ получения единственного ответа. Это можно сделать двумя способами:

1. Путем логических правил. Например, если 5 нейросетей выдали ответ "здоров", а 2 - "болен", то общее решение - "здоров", т.к. за него проголосовало большее число нейросетей-экспертов. Если не удастся формализовать правило, можно принимать решение на основании степеней уверенности каждой из сетей.
2. Путем надстройки над блоком малых экспертов нейросети-"супервизора", которая обучена принимать решение по результатам работы этих малых экспертов.

После определения акта работы системы и разбивки (если требуется) задачи на подзадачи следует разработка схемы обучающих примеров для каждой из подзадач. Схема примера включает список входных и выходных параметров для данной подзадачи.

Определение списка входных данных - квалифицированная работа предметного специалиста, требующая знания изучаемой области и ориентировочной

важности тех или иных параметров, необходимых для получения ответа. Желательно в начале работы над проектом задать некоторую избыточность списка входных данных. В дальнейшем “лишние” параметры можно будет легко исключить из работы системы, добавить же новые несколько труднее, прежде всего потому, что потребуется вновь обращаться за этими параметрами к источникам данных.

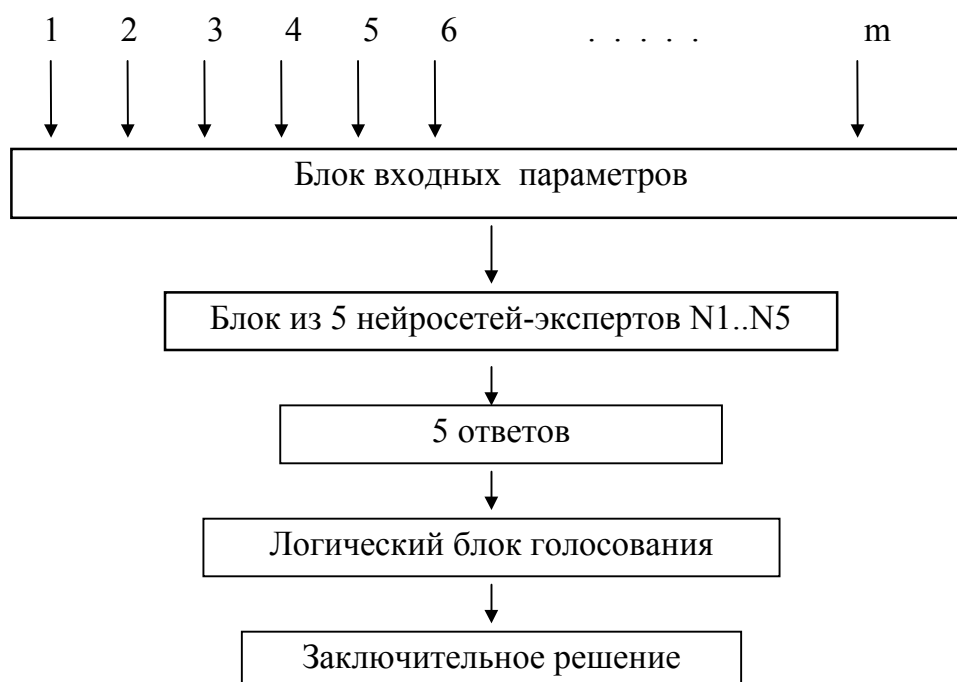


Рисунок 5. Схема движения информации для экспертной системы с 5 нейросетями-экспертами, решающими одну и ту же задачу и логическим блоком, осуществляющим голосование для выдачи заключительного решения

Далее желательно составить наглядную схему движения информации в экспертной системе начиная со ввода данных из внешнего мира и заканчивая выводом ответа (ответов) пользователю. Пример такой схемы приведен на рисунке 5.

Необходимо определить ориентировочные размеры обучающих выборок. Так, в классификационных задачах (подзадачах) желательно, чтобы каждый класс ответа был представлен достаточным количеством примеров. Необяза-

тельно, но желательно, чтобы количества примеров каждого класса не различались слишком сильно (более, чем на порядок).

Предварительное обучение и тестирование нейросетей можно проводить на небольшом количестве примеров с целью определения дальнейшей стратегии и тактики обучения, а также уточнения общего размера обучающей выборки.

Представление обучающих данных

Следующий этап работы над проектом - разработка представления обучающих данных. Нейросети оперируют с информацией, представленной только в виде чисел. Числа подаются на входные синапсы нейросети; ответы, снимаемые с выходных нейронов, также представляют собой числа, поэтому для оценки примера заранее известный ответ также должен быть представлен в виде числа (чисел). Информация же, на основании которой нейросеть должна давать ответ, может быть самого разнообразного вида: термины, описывающие какие-либо ситуации, числа различного вида и величины, динамические ряды, массивы, графики, динамические кривые, двух- и трехмерные изображения и т.д. Поэтому возникает необходимость корректного представления этой информации в виде чисел, сохраняющих смысл и внутренние взаимосвязи в данных.

Существует огромное количество способов представления информации для различных целей.

Для работы нейронных сетей с медико-биологическими данными мы предлагаем по возможности полную классификацию данных, с которыми может столкнуться создатель медицинских экспертных систем, и оптимальные способы их представления в численном виде, учитывающие специфику работы нейросетевых систем. Однако это не готовые рецепты, поэтому в каждом конкретном случае все же требуется квалифицированное решение специалиста.

1. Число с "плавающей точкой". Один из самых распространенных типов данных. Данные такого типа могут принимать любые значения, дробные или целые. Чаще всего они положительны, но могут быть и отрицательными. Немаловажно, что для работы нейронных сетей практически не имеет значение, подчиняется или нет вариационный ряд этих данных закону нормального распределения. Как правило, данные такого вида располагаются на каком-либо интервале с

нечеткими границами. Примером может послужить большинство данных лабораторных анализов. Данные в виде чисел с "плавающей точкой" не требуют каких-либо перерасчетов и могут использоваться в готовом виде. Следует особо отметить, что способы кодирования таких показателей путем разбивки на интервалы и присвоения каждому интервалу порядкового номера нежелательны для нейронных сетей.

2. Взаимоисключающие варианты. Один из наиболее сложных типов данных, требующих продуманного представления. Информация при этом представлена в виде одного и только одного варианта из заранее известного и ограниченного набора вариантов и не может принимать вид дробного числа. Простейшим примером может служить пол человека - мужской или женский. Такая информация требует численного кодирования. В приведенном примере можно закодировать мужской пол как 1, женский - как 2 или наоборот. Однако далеко не всегда можно сделать такое простое и произвольное кодирование. Для обучения нейросетей подобную информацию логично подразделять на 3 основных подтипа.

А. Неупорядоченные варианты - приведенный пример с полом человека. Их можно кодировать произвольным способом. Часто к этому типу относятся данные, представляемые всего двумя вариантами (да - нет, согласен - не согласен, болел - не болел и т.д.).

Б. Упорядоченные варианты. Такие данные находятся в определенных взаимосвязях друг с другом, которые могут быть выражены отношениями типа "больше", "меньше". Примером может служить степень тяжести заболевания (I,II,III). В любом случае варианты располагаются в определенном порядке - как правило, возрастания или убывания.

В. Частично упорядоченные варианты. Способ упорядочивания не очевиден, однако его можно найти, если это необходимо для решения задачи.

3. Совместимые варианты. Информация может быть представлена одним или одновременно несколькими вариантами из известного и ограниченного набора вариантов. Примером может являться наличие у обследуемого каких-либо заболеваний или заболеваний, перенесенных в детстве. В таких случаях имеется два различных подхода к кодированию данных.

А. Если варианты неупорядоченные, наилучший (но, к сожалению, не всегда удобный) способ состоит в том, чтобы разбить признак на несколько признаков, количество которых равно количеству вариантов и каждый из них кодировать отдельно. Каждый подпризнак в таком случае делается самостоятельным признаком и подается на отдельный входной синапс нейросети.

Б. Если варианты упорядоченные, можно применить принцип битовой маски.

4. Дата, время. Очень часто медицинские данные содержат даты различных событий в жизни обследуемых. Для численного представления временных точек необходимо в каждом признаке выбирать соответствующую точку отсчета и, отталкиваясь от нее, выражать временной интервал в удобных единицах (секунды, часы, сутки, годы). Например, если у больного указана дата возникновения какого-либо заболевания, удобнее перевести ее в возраст, который соответствовал этому событию. Сделав это для всех обследуемых, специалист приведет показатель к единой шкале. Если требуется другой подход, можно посчитать, сколько времени прошло с момента возникновения заболевания до момента настоящего обследования. В каждом случае предметный специалист должен выбрать способ представления, наиболее хорошо отражающий смысл параметра.

5. Графики. Разработка способа представления графиков часто требует довольно большого времени.

А. Наиболее простой, но не всегда возможный способ - дискретное разложение графика. Однако для подачи такой информации потребуются большее количество входных синапсов нейросети.

Б. Можно измерять на графике специально выбранные величины, отражающие наиболее важные характеристики явления. Например, для ЭКГ - это ширина и высота различных зубцов, наличие отрицательных зубцов (да, нет) и т.д. Такой способ может оказаться и экономнее, и эффективнее, однако требует хорошего знания изучаемого явления. Во многих случаях приходится прибегать к совмещению обоих методов, особенно в экспериментальных работах при изучении новых явлений с помощью нейросетей.

6. Произвольные изображения. Для подачи изображения на нейросеть оно представляется в виде большого массива чисел (иногда состоящего из сотен тысяч элементов). Каждое число при этом отражает соответствующую точку изображения, разделенного на маленькие фрагменты. Чем больше фрагментов, тем точнее представление изображения, но тем больший массив для этого требуется. Стандартное графическое изображение компьютеров IBM AT 286 имеет размер 640*350 точек и может быть представлено массивом из 224000 чисел. Кодировка каждой точки может быть различной. Если изображение монохромное и не допускает полутонов, каждая точка может быть отображена значениями 0 (белая) или 1 (черная) или наоборот. Для хранения такого значения достаточно 1 бита информации, поэтому для представления всего изображения потребуется 28000 байт, что вполне приемлемо. Если же изображение цветное или имеет полутона, то каждая точка должна выражаться числом, разрядность которого достаточна для представления любого цвета или полутона.

После разработки списка обучающих параметров для каждой подзадачи и определения способа представления каждого параметра и ответа можно приступать к формированию обучающих выборок - сбора информации и помещения ее в базы данных.

Инициализация нейронной сети

После формирования обучающей выборки производится инициализация нейронной сети (сетей), для которой определяются стартовые параметры. Данная подглава посвящена выбору стартовых параметров нейросети в зависимости от обучающей выборки и цели обучения.

Разработанная нами методика создания нейросетевых экспертных систем включает фиксированный набор стартовых параметров, выбор которых может осуществляться предметным специалистом. Однако даже этот небольшой набор может быть минимизирован путем автоматизации установки достаточно очевидных параметров.

Список стартовых параметров включает следующие пункты (параметры, выбор которых может быть автоматизирован, помечены звездочками):

- 1*. Выбор типа нейронной сети;

- 2.* Имя файла нейронной сети;
- 3. Схема подачи обучающих данных, которая определяет количество входных сигналов и соответствующих им входных синапсов, а также ответ;
- 4*. Включение или выключение нормирования входных сигналов;
- 5*. Количество нейронов;
- 6*. Параметр плотности (количество нейронов, на которые подается один входной сигнал);
- 7*. Время отклика нейросети (число тактов функционирования нейросети при прохождении одного примера с момента подачи входных сигналов до выдачи ответа);
- 8*. Величина характеристики нейронов.

Как видно, в минимизированном варианте списка параметров специалист-пользователь должен всего лишь поставить задачу, не задумываясь о выборе остальных параметров.

Рассмотрим методику задания каждого параметра, а также вариант, который принимается при автоматической установке.

Следует заметить, что в рамках принятой нами идеологии создание нейросетей должно осуществляться при наличии стартовой обучающей выборки, находящейся в базе данных, что значительно облегчает выбор параметров пользователем и дает возможность автоматизации выбора параметров.

1. Выбор типа нейронной сети. Определяется типом задачи (классификация, предикция, векторная предикция). Перед выбором необходимо убедиться в соответствии обучающей выборки типу нейросети. При автоматическом выборе программа, после определения пользователем набора обучающих параметров и выбора поля базы данных, в котором содержится ответ, осуществляет анализ этого поля. Если данные в нем представлены целыми числами небольшой размерности или фиксированным набором чисел с “плавающей точкой”, устанавливается классификационный тип нейронной сети. Если же данные представлены числами с “плавающей точкой” с достаточно широким варьированием, устанавливается предикционный тип нейросети.

2. Имя файла нейронной сети. При автоматической установке имя файла связывается с установленной задачей.

3. Схема подачи обучающих данных и ответа не может быть автоматизирована, так как изначальный набор обучающих параметров и параметры, содержащие ответ, могут быть заданы только самим специалистом. При наличии базы данных пользователь указывает, какие поля базы рассматриваются как обучающие, и в каких полях содержатся ответы. Каждая запись базы является примером, если ячейка в поле, содержащем ответ, не пуста, и если пользователь специально не указал, что данная запись не должна участвовать в обучении. Каждый входной (обучающий) параметр можно при обучении и тестировании подавать на один (стандартно) или несколько (при необходимости) входных синапсов нейросети. Вторым вариантом можно использовать в тех случаях, если пользователь заранее считает, что какой-либо параметр должен играть особое значение для получения ответа и его влияние необходимо усилить. Для примера рассмотрим две схемы. Первая (Таб. 1) задается программой по умолчанию (стандартно) и не меняется, если создателю нейросети нет необходимости переопределять форму подачи данных. Вторая схема (Таб. 2) является результатом направленных действий создателя нейросети.

Таблица 1.

Схема подачи обучающих параметров и определение ответа, устанавливаемые стандартно

Поля базы данных	Количество входных синапсов, на которые подается обучающий параметр	Номера входных синапсов, на которые подается обучающий параметр
1. Диагноз	1	- (Ответ)
2. Возраст	1	1
3. Пол	1	2
4. Количество эритроцитов	1	3
5. Количество лейкоцитов	1	4
6. Гемоглобин	1	5

В первой схеме база данных содержит 6 полей, содержащих данные об обследуемых. Первое поле содержит ответ (диагноз), остальные поля являются обучающими. Если поле, содержащее ответ, первое в базе данных, не требуется переопределять ответ. В приведенном примере 5 обучающих полей, данные каждого из них будут подаваться на один соответствующий входной синапс.

Во втором примере поле, содержащее ответ, стоит последним; кроме того, имеются поля, содержащие код обследуемого и его фамилию. В этой ситуации пользователь должен указать, что ответом является именно поле “Диагноз”.

Подавать на вход нейросети код обследуемого не имеет никакого смысла, поэтому данное поле выключается из схемы подачи входных сигналов, несмотря на то, что оно может иметь числовой тип. Поле, содержащее фамилию обследуемого, также выключается из схемы, т.к. оно является не числовым, а символьным. Поля, содержащее данные о количестве эритроцитов и гемоглобине, подаются соответственно на 2 и 3 входных синапса создаваемой нейросети, что будет определять их относительно большее влияние на выдачу ответа у обученной се-

ти. Таким образом, несмотря на то, что число обучающих параметров равно по прежнему 5, число входных синапсов будет увеличено до 8.

Таблица 2.

Схема подачи обучающих параметров и определение ответа, заданные создателем нейросети

Поля базы данных	Количество входных синапсов, на которые подается обучающий параметр	Номера входных синапсов, на которые подается обучающий параметр
1. Код пациента	-	-
2. Ф.И.О. пациента	-	-
3. Возраст	1	1
4. Пол	1	2
5. Количество эритроцитов	2	3, 4
6. Количество лейкоцитов	1	5
7. Гемоглобин	3	6, 7, 8
8. Диагноз	-	- (Ответ)

Подачу некоторых входных параметров на несколько входных синапсов можно рекомендовать и в случаях, когда общее число обучающих параметров мало (единицы).

4. Включение или выключение нормирования входных сигналов. Нормирование входных сигналов представляет собой один из видов предобработки и является исключительно важным в методологии создания нейросетевых экспертных систем. При нормировании на входной синапс подается не величина параметра (для данного примера), а ее эквивалент, полученный путем пересчета по определенной схеме. Мы применяем нормирование входных сигналов на диапазон $[-1...1]$. Система нормирования имеет два аспекта - методологический и

технологический. Первый состоит в том, что нормирование позволяет унифицировать представление информации внутри “черного ящика” нейросетевой экспертной системы и поэтому отменяет необходимость создателю и/или пользователю экспертной системы контролировать диапазоны числовых значений медико-биологических данных. Технологический аспект заключается в следующем. Разработанная технология обучения нейронных сетей предусматривает универсальную структуру и алгоритмы обучения для медико-биологических данных любого характера. Однако в результате проводимых нами многочисленных экспериментов было установлено, что наиболее универсальная и быстрообучающаяся архитектура полносвязной сигмоидной (имеющей характеристическую функцию нейронов) нейросети оптимально работает при нахождении входных сигналов в диапазоне приблизительно $[-1...1]$. При расширении этого диапазона происходит сначала практически незаметное, плавное, а затем быстро нарастающее снижение качества обучения, особенно ярко проявляющееся при сильно различающихся диапазонах у разных входных сигналах.

Еще в большей степени нормирование влияет на экстраполяционные возможности нейросети. Исследования показали, что процент правильно распознанных примеров, не входящих в обучающую выборку, максимален также при нормировании входных данных на диапазон $[-1...1]$.

Учитывая результаты экспериментов, разработанная методология предусматривает нормирование всех входных параметров на диапазон $[-1...1]$. Каждый входной сигнал перед подачей на синапс пересчитывается по формуле

$$Y_i = 2 * (x_i - \min_i) / (\max_i - \min_i) - 1 \quad (3),$$

где x - исходный сигнал, Y_i - получаемый нормированный сигнал, \min и \max - соответственно минимальное и максимальное значения интервала входных параметров в поле, подаваемом на синапс i .

Диапазон входных параметров рассчитывается для каждого поля, подаваемого на синапсы нейросети. Расчет происходит при создании нейросети. После создания нейросеть хранит минимум и максимум диапазона в своем файле. Расчет диапазонов происходит автоматически исходя из значений, заданных в примерах обучающей выборки. Поэтому желательно, чтобы обучающая выборка

содержала примеры с возможными крайними значениями всех обучающих параметров. Если впоследствии, например при дообучении или тестировании, какое-либо значение будет находится вне промежутка $\min..max$, пересчитанное значение будет соответственно лежать за пределами диапазона $[-1...1]$. Если "выпадение" из интервала будет небольшим (десятые доли), это практически не повлияет на качество дообучения и тестирование.

Тем не менее, рекомендуется после создания нейросети (до начала всякого обучения!) внимательно просмотреть диапазоны значений, соответствующие каждому входному синапсу, и если в будущем предполагается подача значений вне данного диапазона, вручную изменить его. Ни в коем случае нельзя менять диапазоны после обучения нейросети, т.к. это приведет к последующей некорректной обработке входных сигналов, сбоем при дообучении и некорректному тестированию. При автоматическом задании параметров сети всегда устанавливается нормирование, автоматический подсчет диапазонов значений и расширение их на заданную относительную величину. Отмена нормирования предусматривается в основном для проведения экспериментов и исследований, как возможность дополнительной настройки.

Кроме нормирования входных сигналов, устанавливается нормирование значений ответов у нейросетей-предикторов. Проведенные эксперименты показали безусловное преимущество такого подхода. Нормирование значений ответов проводится по схеме, аналогичной нормированию входных сигналов.

5. Количество нейронов. Этот параметр определяет суммарное количество синаптических связей нейронной сети. Для полносвязной нейросети оно равно квадрату числа нейронов. Если сеть использует адаптивную матрицу входных сигналов (в нашей методике это делается всегда), общее число подстраиваемых связей равно квадрату числа нейронов плюс произведение числа входных синапсов и плотности подачи входных сигналов (на сколько нейронов подается каждый входной сигнал).

Оптимальное количество нейронов во многом зависит от решаемой задачи. Большее число нейронов повышает гарантию успешного обучения, но увеличивает размер нейросети, а значит, время ее срабатывания при тесте и время ее

загрузки с диска компьютера в память. В общем случае, число нейронов может быть равно числу обучающих параметров. В классификаторах без адаптивной матрицы входных сигналов количество нейронов устанавливается автоматически и равно сумме числа обучающих параметров и количества классов в задаче. Если набор входных параметров избыточен (предполагается, что для обучения можно обойтись лишь частью этого набора) можно задавать число нейронов меньшим.

Не представляется возможным корректно исследовать зависимость качества обучения от числа нейронов, т.к. невозможно создать сети с одинаковыми стартовыми картами и разным числом нейронов. При сравнении сетей с различными картами на качество будет влиять (хотя и не сильно) индивидуальность сети, определяемая случайными стартовыми значениями синаптических весов.

При автоматическом задании числа нейронов предлагается следующая технология. Инициализируется и обучается нейросеть с небольшим количеством нейронов (например, 2). При невозможности обучения, в зависимости от суммарной оценки и количества входных синапсов рассчитывается новое, увеличенное количество нейронов. Инициализируется и обучается новая сеть, повторяющая параметры предыдущей, но имеющая заново установленное большее число нейронов. Процесс повторяется до тех пор, пока какая-либо из сетей не обучится полностью. Недостатком метода является увеличение времени, затрачиваемого на обучение. Преимущества - создателю экспертной системы нет необходимости экспериментировать с различным количеством нейронов, на что в конечном итоге может потребоваться гораздо больше времени. Кроме того, предлагаемый способ позволяет избежать создания заранее слишком избыточной сети, которая может быстро и хорошо обучиться на обучающей выборке, однако будет обладать низкой экстраполяционной способностью.

6. Параметр плотности. Плотность подачи входных сигналов определяет, на сколько нейронов будет одновременно подаваться сигнал с каждого входного синапса (не путать подачу некоторых, определяемых создателем сети, сигналов на несколько входных синапсов). При числе нейронов, равном количеству входных параметров, и плотности, равной единице, каждый входной сигнал подается на один соответствующий нейрон. Если в этом же случае плотность будет равна

числу нейронов, то каждый входной сигнал будет подаваться на все нейроны. Такое "размазывание" сигналов по нейронам обеспечивает адаптивная матрица входов. В общем случае, это всегда приводит к улучшению качества обучения, хотя и замедляет срабатывание сети и незначительно увеличивает ее размер. При автоматическом задании плотности она устанавливается равной числу нейронов.

7. Время отклика. Определяет число тактов функционирования сети с момента подачи входных сигналов до момента снятия выходных сигналов. Эксперименты показали, что в подавляющем числе медико-биологических задач следует задавать этот параметр равным 2 (редко 3). Это связано с относительно тесными взаимосвязями (часто приближающимися к линейным), присутствующими в медицинских данных. Большие значения времени отклика, как правило, приводят к замедлению обучения и снижению прогностической способности нейросети. Несколько большие значения времени отклика можно порекомендовать для сетей-предикторов.

Эксперименты показали, что увеличение времени отклика приводит также к увеличению общего числа тактов функционирования, необходимого для полного обучения сетей практически на всех медицинских задачах, обсуждаемых ниже. Кроме того, увеличение реального времени, требуемого на один такт работы нейросети при большом времени отклика, увеличивает еще и машинное время, требуемое для обучения. Поэтому при автоматическом задании времени отклика оно принимается равным 2.

8. Величина характеристики нейронов. Как уже говорилось, при прохождении сигнала через нейроны он пересчитывается по функции нейрона, имеющей заданный параметр характеристики. В результате величина сигнала всегда уменьшается, причем чем больше характеристика нейрона, тем существеннее уменьшение одного и того же сигнала (Рисунок 6).

Кроме того, разность двух поданных сигналов (уже нормированных) после их преобразования по характеристической функции будет различной в зависимости от характеристики и зон, в которых находятся поданные сигналы. Функция с малой характеристикой обеспечивает относительно большую разность между преобразованными сигналами, которые до преобразования находились или на

значительном расстоянии друг от друга, или группировались в зоне наиболее крутого участка функции (около нуля). Наоборот, функция с большой характеристикой лучше разделяет сигналы, попадающие в крайние зоны и находящиеся на небольшом расстоянии друг от друга. Так как довольно значительная часть медико-биологических параметров подчиняется закону нормального распределения или же носят дискретный характер, нейронные сети с малой характеристикой нейрона обучаются на этих данных гораздо быстрее, особенно, если в задаче, например, классификации, классы достаточно легко разделимы в пространстве входных параметров. Однако сети с большей характеристикой обладают значительно лучшими экстраполяционными способностями, хотя и хуже обучаются.

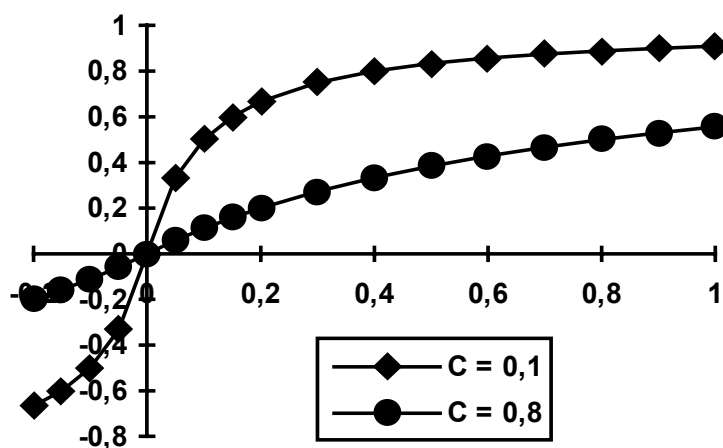


Рисунок 6. Схема изменения сигнала при прохождении через нейрон в зависимости от величины характеристики

Если характеристика устанавливается пользователем, рекомендуется вначале обучать сеть с небольшой характеристикой, а при успешном обучении генерировать и обучать новые сети, увеличивая этот параметр. Нельзя менять характеристику у обучающейся или обученной сети, т.к. это приведет к некорректной

обработке входных сигналов. Мы рекомендуем диапазон значений характеристики от 0,1 до 0,8.

При автоматическом задании этого параметра характеристика вначале устанавливается небольшой, а затем, в случае успешного обучения нейросети, постепенно повышается.

Стратегия и тактика обучения нейронных сетей

Под стратегией обучения нейросетей понимается общий план обучения, включающий разбивку задачи на подзадачи, определение типа и количества нейросетей, обучаемых по каждой из них, предварительные параметры сетей, планирование действий в случае возможных затруднений.

Тактика представляет собой возможное вмешательство оператора в процесс обучения с целью его ускорения, подстройки некоторых параметров, а также оптимальные действия при невозможности дальнейшего обучения сети.

Сформулируем цель обучения, на достижение которой направлены стратегия и тактика. Целью будем считать полное на данном задатнике обучение нейросети (набора нейросетей-экспертов) с минимально возможным числом подстраиваемых связей, максимально возможной характеристикой и максимально возможным уровнем надежности (минимальным уровнем отклонения). Дополнительным условием может быть минимальное (или определенное пользователем) число входных сигналов.

Критерием достижения цели будем считать результат тестирования набора примеров с известными ответами, не входящих в обучающую выборку. Требуемый результат определяется пользователем.

Рассмотрим, как влияют на качество и длительность обучения некоторые параметры. В таблицу 3 сведены зависимости между обучением (требуемым временем и качеством) сети, ее параметрами, особенностями задачи и используемого компьютера.

Таблица 3.

Зависимость времени и качества обучения нейросетей от их параметров, особенностей задачи и применяемого компьютера. Стрелки в двух правых колонках показывают, как изменяются длительность и качество обучения при изменении параметра, указанного слева

Параметр	Длительность обучения	Качество обучения
Увеличение тактовой частоты компьютера	↓	-
Вывод информации в процессе обучения на экран в графическом режиме (по сравнению с текстовым режимом)	↑	-
Увеличение количества примеров	↑	↑
Увеличение количества обучающих параметров	↑	↑
Увеличение количества классов (в классификаторах)	↑	↓
Увеличение размерности вектора ответа (в предикторах)	↑	↓
Близкое количество примеров каждого класса (в классификаторах)	↓	↑
Повышение уровня надежности (в классификаторах)	↑	↑
Повышение уровня отклонения (в предикторах)	↓	↓
Увеличение числа нейронов	↑	↓
Увеличение параметра плотности	↑	↑
Увеличение времени отклика сети	↑	↓
Увеличение характеристики	↑	↑

Разработанная методология включает две стратегии: с участием пользователя и без него. Автоматизированное выполнение стратегии может проводиться в двух вариантах - повторять принципы “ручной” стратегии, или использовать другие правила. Кроме того, стратегии автоматического обучения подразделяются еще на два класса - стратегии решения подзадачи и стратегии решения единой задачи, состоящей из взаимосвязанных (параллельно или иерархически) подзадач. Мы заранее отказались от “ручной” стратегии решения единой комплексной задачи, так как, во-первых, они встречаются не так уж часто, во-вторых, при их решении (особенно при достаточно большом количестве связанных подзадач)

число подстраиваемых параметров, за которым необходимо следить пользователю, может очень сильно вырасти.

Примером может служить минимизация набора обучающих параметров в системе из нескольких иерархически связанных нейросетей, одни из которых используют для выдачи результата ответы других как входные параметры. Следует заметить, что пользователь вовсе не должен обладать знаниями этих стратегий. Именно для такого случая и разработаны автоматизированные стратегии обучения. Рассмотрим вначале принципы “ручной” стратегии для решения одной подзадачи (сведенной к элементарной - классификации или предикции).

Как правило, первый стратегический шаг для решения подзадачи - создание нейросети с параметрами, максимально выгодными для подтверждения факта принципиальной возможности обучения. В этом случае задается число нейронов и плотность, равные исходному количеству обучающих параметров (в случае классификационных задач эти параметры можно еще увеличить на число классов в задаче). Задаются минимальные значения времени отклика нейросети, характеристика и уровня надежности (для предикторов - требуемый уровень допустимого отклонения). Если такая нейросеть обучается успешно, можно считать, что задача имеет закономерности, связывающие входные параметры и результат.

Далее желательно найти максимальный уровень надежности (в классификаторах) или минимальный уровень отклонения (для предикторов), с которым может обучиться сеть. Для этого нет необходимости генерировать новую нейросеть. Разработанный нами программный инструментарий (программы “MultiNeuron 1.0” и “MultiNeuron 2.0 Professional”) позволяет задать режим доучивания с автоматическим ступенчатым повышением уровня надежности. Это означает, что после каждого успешного полного обучения уровень надежности будет автоматически повышаться на заданную пользователем величину (шаг), и обучение будет продолжаться, пока не перестанет быть возможным. После окончания процесса предыдущий уровень надежности, при котором сеть обучилась полностью, и является максимально возможным для данной сети.

В случае предикторов используется понятие уровня отклонения, который определяет максимальное расхождение между требуемым (заранее известным) и

вычисляемым сетью ответами, при котором пример считается правильно распознанным. При этом качество обучения тем выше, чем меньший уровень отклонения позволяет сети тестировать все примеры правильно, т.е. увеличивается точность ответа сети. Автоматический режим при этом снижает уровень отклонения после каждого полного обучения. Минимально возможный уровень отклонения равен нулю (абсолютное совпадение ответов) и практически не может быть достигнут.

Один из принципов стратегии - создание и обучение для каждой подзадачи не одной, а нескольких нейросетей. Желательно создать несколько сетей с разными характеристиками и после их обучения сравнить их прогностические способности. Весьма вероятно, что при тестировании контрольной выборки разными нейросетями у каждой из них будут свои собственные ошибки. Решение в таком случае принимается на основании голосования и сравнения коэффициентов уверенности каждой из сетей.

Как правило, классификаторы, распознающие несколько классов одновременно, обучаются несколько хуже, чем бинарные классификаторы. В наших исследованиях встречались ситуации, когда задача с 5 классами не могла быть решена одним 5-классовым классификатором, а система из нескольких бинарных классификаторов прекрасно решала задачу. В подобных ситуациях есть несколько способов решения проблемы.

Изменение весов классов.

При обучении классификатора каждому классу задачи можно присвоить вес - "степень внимания", которая будет уделяться нейросетью примерам данного класса. Исходно веса всех классов равны единице. При увеличении веса какого-либо класса оценка каждого примера, принадлежащего этому классу, будет умножаться на его вес. Таким образом, нейросеть будет прилагать особые усилия для распознавания примеров этого класса. При задании нулевого веса оценки всех примеров класса будут равны нулю и эти примеры будут считаться распознанными. Практически это эквивалентно исключению всех примеров класса из обучения.

Изменение весов классов - элемент как стратегии, так и тактики. В стратегическом плане можно спроектировать систему "малых экспертов". Если задача включает, например, 3 класса, можно использовать 3 нейросети, каждая из которых обучается с повышенным весом "своего" класса. В итоге такая нейросеть будет различать все 3 класса, но "свой" класс - особенно хорошо. При тестировании примера решение принимается опять таки голосованием с учетом особого мнения "экспертов".

Разделение задачи.

Если одна нейросеть не может обучиться различать несколько классов или делает это недостаточно хорошо, можно разделить задачу между несколькими разными сетями (принцип, основанный на дихотомии). Это делается следующим образом. Предположим, задача имеет 4 класса. Создается 3 бинарных классификатора. Первый обучается отличать примеры 1 класса от примеров 2,3 и 4 классов, объединенных в один класс. Второй бинарный классификатор обучается отличать примеры 2 класса от примеров 3 и 4 класса, также объединенных в один класс. Третья сеть обучается различать примеры 3 и 4 класса. Тестирование примера при этом осуществляется так: пример подается первой сети. Если она относит его к 1 классу, ответ и есть 1 класс, тестирование прекращается. Если сеть относит пример ко 2 классу, это означает, что на самом деле он может принадлежать 2,3 или 4 классам. Поэтому пример подается второй сети. В зависимости от ее ответа тест может остановиться или продолжиться третьей сетью. В такой системе пример в конце концов оказывается распознанным.

Для предикционных задач подобный способ неприемлем по причине отсутствия классов, поэтому простейший способ - вычисление среднего ответа у нескольких нейросетей. Более сложный метод - привлечение классификаторов для решения предикционных задач. В случае одномерной предикции можно разбить интервал значений ответа на несколько участков (необязательно равных) и обучить классификатор распознавать, к какому участку принадлежит пример. Если точности ответа недостаточно, для каждого участка создается и обучается свой предиктор, умеющий определять ответ только у примеров, отнесенных классификатором к этому участку. Здесь мы наблюдаем иерархическую систему

нейросетей, в которой предикторы подчинены "центральному" классификатору. Подобная система может работать и в задачах классификации, когда число классов слишком велико.

После успешного обучения по вышеприведенным схемам желательно переобучить сети с повышением характеристики, особенно в случае недостаточно удовлетворительных результатов контрольного тестирования.

В общем случае, уменьшение количества связей нейросети (путем контрастирования существующей сети или обучения новых сетей с меньшим числом нейронов и плотностью) лучше всего производить в последнюю очередь, после того, как были настроены предыдущие параметры и результаты тестирования по-прежнему высоки. То же касается и минимизации входных параметров, которая, так же как и принципы автоматической стратегии, будут рассмотрены ниже.

Тактика обучения.

В подавляющем большинстве случаев при обучении сети не требуется никакого вмешательства в этот процесс. Однако все же бывают ситуации, когда сеть не может обучаться. Здесь требуются некоторые действия пользователя, которые мы называем "тактикой обучения". Как правило, нейросеть перестает обучаться не сразу, а через некоторое время после начала процесса. При этом нейросетевая программа сигнализирует оператору о невозможности дальнейшего обучения и в простейшем случае останавливает процесс.

Первое, что необходимо сделать - проверить корректность заданных параметров нейросети. Самая простая причина - недостаточное количество нейронов, малая плотность или слишком высокая характеристика. Если же все параметры заданы оптимально, можно применить специальное средство, называемое "удар".

Удар может помочь в ситуации, когда сеть попадает в локальный минимум при минимизации целевой функции обучения. Когда производится удар, в карту синаптических весов вносится случайный вклад, максимальная величина которого задается заранее. Новое значение синаптического веса вычисляется по формуле

$$x = x_0 + B * (\text{Random} - 0,5) \quad (4)$$

где x - новое значение синаптического веса, x_0 - исходное значение синаптического веса, B - заданный уровень удара (в диапазоне 0..1), Random - случайно сгенерированное для данного синаптического веса число (в диапазоне 0..1).

После удара обучение продолжается. Проведенные эксперименты показывали, что особенно эффективен удар у сетей с адаптивной матрицей входных сигналов. Задание оптимального уровня удара вручную может потребовать некоторого опыта пользователя. Для этого желательно наблюдать за динамикой снижения средней оценки на протяжении нескольких десятков циклов до момента остановки обучения. Если за этот период оценка снижалась быстро, лучше задавать уровень удара в пределах 0,1-0,2. Если оценка на протяжении многих циклов "зависла" около какого-либо значения, следует задать небольшой уровень удара (0,0001 - 0,001). Задание максимального уровня удара (1,0) практически полностью обновляет карту весов, что равнозначно инициализации новой сети.

Частое применение удара небольшого уровня может быть полезным для создания помехоустойчивых нейросетей.

В разработанных нами нейросетевых программах используется как ручной, так и автоматический режимы удара. В последнем случае после остановки обучения программа производит удар и продолжает обучение. Только если после трех (или количества, определенного пользователем) попыток удара обучение по-прежнему невозможно, программа останавливает процесс.

Если и применение удара не приносит эффекта, причину невозможности обучения следует искать в обучающей выборке. Как уже говорилось, самая частая причина - наличие в выборке идентичных по обучающим параметрам примеров, имеющих различный ответ. В принципе, это почти тупиковая ситуация, и единственным выходом из нее является увеличение количества обучающих параметров, при котором добавленные параметры будут различными у таких примеров. В самом деле, очевидно, что различные ответы в такой ситуации определяются факторами, не нашедшими отражение в обучающих параметрах. В простейшем случае можно исключить один из идентичных примеров, однако в таком случае в экспертную систему заранее закладывается возможность ошибки.

Невозможность дальнейшего обучения в классификаторах иногда может быть спровоцирована неверной постановкой задачи, точнее, заданной классификационной моделью. Например, при бинарной классификации примеры одного из классов представляют на самом деле два различных подкласса, группирующихся по параметрам "по разные стороны" от другого класса задачи. В этом случае можно использовать нейросеть для коррекции классификационной модели путем кластеризации примеров.

Существует несколько методов нейросетевой кластеризации, наиболее простой из которых заключается в последовательном исключении из процесса обучения наиболее трудных примеров с помощью "слабых" нейросетей, настройка параметров которых заранее не дает возможности для полного обучения (например, слишком малое число нейронов). В процессе прохождения каждого цикла обучения программа отслеживает самый трудный пример на этом цикле - пример с наибольшей оценкой. После захода сети в тупик можно исключить этот пример из дальнейшего обучения, пометив его, и продолжить процесс. В конце концов может накопиться некоторое количество исключенных примеров, после чего сеть обучится полностью. Часто получается, что все исключенные примеры принадлежат только одному из классов задачи, что является свидетельством неоднородности класса. Это можно проверить путем обычных статистических методов, сравнив группу исключенных примеров как с оставшимися примерами этого же класса, так и с примерами других классов. Различия могут послужить поводом для изменения классификационной модели.

Это не самый лучший способ кластеризации, однако он может с успехом применяться как при создании экспертных систем, так и в научных исследованиях для изучения различных явлений. По сравнению с многочисленными существующими способами кластеризации он имеет определенное преимущество, заключающееся в возможности "наведенной" кластеризации, т.е. такой, к которой пользователь "подталкивает" программу, не зная заранее точных критериев - параметров, по которым ее следует проводить, и имея лишь интуитивное представление о возможном результате. В практике врача-исследователя такое может встречаться довольно часто, например, когда он интуитивно относит какие-либо

2 примера к разным классам, однако не может формализовать принцип разделения. Применение нейросети может помочь в этом случае найти закономерность и экстраполировать ее на большой набор данных.

Как уже говорилось, изменение весов классов может считаться тактическим действием, так как может осуществляться прямо в процессе обучения. Поводом для изменения веса класса обычно служит его средняя оценка, значительно (в несколько раз) превосходящая оценки остальных классов.

Дополнительные возможности нейросетей и автоматизированные стратегия и тактика обучения

В этом разделе будут рассмотрены несколько дополнительных возможностей нейросетей, разработанных в рамках построения теории и методологии создания нейросетевых экспертных систем, и полезных для практического применения.

Обученная нейросеть не только умеет распознавать примеры, но и хранит достаточно важную информацию, которую можно получить, грамотно сформулировав вопрос для сети. Возможность нейросетей содержательно отвечать на поставленные вопросы позволяет использовать их не только как "решающие" устройства, но и как инструменты для проведения научных исследований.

Вычисление информативности обучающих параметров и минимизация их набора.

Врач, ставящий диагноз больному, всегда старается анализировать весь комплекс сведений о пациенте - жалобы, анамнез болезни и жизни, данные клинического осмотра и лабораторных анализов. При этом одна часть параметров имеет принципиальное значение для принятия решения, другая же не столь важна и даже может просто игнорироваться.

Нейросеть, принимающая решение на основе опыта, поступает сходным образом. Поэтому исследователю очень важно знать, какие из параметров важны для сети, а какие - не очень. Следует сказать, что абсолютно незначимых параметров практически не бывает, и каждый входной сигнал в большей или меньшей степени влияет на принятие решения.

Разработанный нами метод позволяет выявить относительную значимость входных параметров, которая может представлять самостоятельный интерес или служить для последующей минимизации их набора [47].

Для этого включаются несколько циклов обучения нейронной сети с заранее завышенным уровнем надежности и внесением в матрицу синапсов случайного вклада после прохождения каждого цикла. При прохождении циклов используется метод подсчета градиентов по входным сигналам, подаваемым на нейронную сеть, и последующего их сравнения, осуществляемого двумя различными методами по выбору пользователя программы.

Первый метод насчитывает значимость каждого параметра как максимальное значение модуля градиента по всем примерам. Этот режим полезно применять в тех случаях, когда в обучающей выборке имеются примеры, "выбывающиеся" из общей массы, и которые могут существенно влиять на принятие решения об ответе. Данный метод будет прежде всего учитывать наличие таких примеров.

Второй метод насчитывает значимость каждого параметра как среднюю величину модуля градиента по всем примерам обучающей выборки. Результат применения этого метода показывает среднюю значимость параметров по всей обучающей выборке. Если в обучающей выборке имеются примеры, "выбывающиеся" из общей массы, и которые могут существенно влиять на принятие решения об ответе, то влияние таких примеров на значимость параметров будет нивелироваться. Данный метод полезно применять в тех случаях, когда обучающая выборка достаточно однородна и необходимо, чтобы возможные "выбывающиеся" примеры существенно не влияли на оценку значимости параметров. Результатом применения метода является набор значений, каждое из которых соответствует определенному входному параметру и может сравниваться с остальными.

Применение метода имеет ценность как для научных исследований, так и при практическом применении нейросетей для экспертных систем, так как позволяет уменьшать набор входных параметров, подаваемых нейросети для получения ответа. Действительно, часто для диагностики и прогноза больному проводят сложные и дорогостоящие методы исследования, порой небезвредные для

здоровья. Однако во многих случаях представляется возможным получить ответ и без них.

В нейросетевых программах предусмотрен как ручной, так и автоматический режимы минимизации входных параметров, причем последний может применяться как самостоятельно, так и в случае задания автоматического определения стратегии обучения. Использование “ручной” минимизации параметров имеет смысл только после того, как нейросеть достаточно хорошо обучена, в противном случае исключение параметров может привести к ухудшению экстраполяционных возможностей сети и затруднению последующего обучения.

Если режим минимизации включен, он начинает работать после полного обучения нейронной сети. При этом осуществляется вычисление относительных показателей значимости и входной синапс, имеющий минимальную значимость. После этого этот синапс отключается от приема сигналов, и обучение продолжается. Предполагается, что нейросеть сможет обучиться без данного параметра. Если сеть вновь обучилась, после повторного вычисления значимости выключается следующий синапс и т.д. В итоге сеть перестанет обучаться, когда набор входных параметров будет слишком мал для полноценного обучения. После этого можно проанализировать различные варианты исключения параметров и выбрать их оптимальный набор. Эксперименты с самыми разными медицинскими задачами показывали, что во многих случаях приблизительно четверть обучающих параметров может быть исключена совершенно безболезненно. Это подтверждает гипотезы об априорной избыточности данных в задачах медицинской диагностики и прогнозирования, решаемых человеком.

Подстройка параметров примера для получения требуемого ответа нейросети.

При применении обученной нейронной сети может возникнуть обратная задача, относящаяся к области нейросетевого моделирования - выяснить, в каком направлении и на какие величины необходимо изменить параметры примера, чтобы сеть выдала требуемый ответ. Приведем простой гипотетический пример. При тестировании примера сетью, разделяющей больных и здоровых, получен ответ "больной". Как нужно изменить параметры этого примера, чтобы он "пе-

решил" в класс "здоровый"? Как уже говорилось, можно вручную изменять параметры, каждый раз тестируя пример после такого изменения. Однако это не самый быстрый и удобный способ, особенно, если число параметров велико, а решение принимается несколькими нейросетями. Поэтому нами был разработан специальный автоматизированный метод подстройки параметров для получения требуемого ответа нейросети.

При решении данной проблемы прежде всего необходимо разделить все параметры примера на подстраиваемые и неподстраиваемые. Например, невозможно изменить возраст пациента, поэтому такой параметр является неподстраиваемым. После определения набора подстраиваемых параметров для каждого из них задается диапазон возможных изменений. Необходимо помнить, что в случае каждого примера эти диапазоны могут диктоваться конкретной ситуацией (одному больному допустимо снизить частоту сердечных сокращений до 50, другому же - только до 70).

После этого для каждой нейросети (если их несколько) необходимо задать требуемый ответ - тот, которого должна добиться эта сеть. Важно отметить, что выполняемая затем автоматическая процедура подстройки выполняется одновременно несколькими нейросетями, так как подстройка, выполненная одной сетью, может резко ухудшить ответ других. Таким образом, каждая сеть, как правило, идет на некоторый компромисс.

В итоге вырабатывается оптимальное решение, одинаково хорошо (или плохо) удовлетворяющее все сети, принявшие участие в процедуре. Исследователь может проанализировать решение, сравнив ответы каждой нейросети до и после подстройки примера [48].

Автоматическое задание стратегии обучения.

Рассмотрим полностью автоматизированное задание стратегии построения решающего блока. Прежде всего определяется список подзадач, причем в каждой подзадаче определяется список обучающих параметров (он может быть общим для всех подзадач) и ответы. Затем задается, сколько нейросетей-экспертов будут решать каждую подзадачу (в простейшем случае - по одному). Далее определяется, будет ли проводится минимизация обучающих параметров. Если да, то

можно задать произвольную минимизацию (нахождение минимально возможного для решения задачи списка параметров) или определяется набор параметров, который желательно оставить в списке или, наоборот, исключить из списка. Эту операцию можно провести отдельно для каждой подзадачи, для групп подзадач или для всей задачи в целом. Как дополнительный параметр, можно указать для каждой подзадачи контрольные выборки, на которых будут проверяться результаты обучения и требуемое качество обучения (в процентах правильно распознанных примеров (для классификаторов) или максимально допустимого отклонения (для предикторов). Это все, что должен сделать пользователь при автоматизированном построении решающего блока.

Непосредственно обучение проводится исходя из следующей стратегии. Обучение каждого эксперта в каждой подзадаче начинается с автоматической инициализации нейросети минимальной конфигурации (число нейронов - 2, плотность - 1, характеристика - 0,1, уровень надежности - 0,1 (для классификаторов), уровень отклонения для предикторов - требуемый или вычисленный исходя из характера значений ответа (в самом общем случае берется $1/3$ минимальной разности между значениями ответа), время отклика - 2. Проводится обучение нейросети. Дальнейшие действия определяются результатом обучения. В случае успешного обучения проводится тест указанной контрольной выборки и в случае неудовлетворительного результата инициализируется новая сеть с увеличенной характеристикой (шаг увеличения - 0,1). При невозможности обучения инициализируется новая сеть с большим числом нейронов и плотностью.

Таким способом проводится обучение сетей по всем подзадачам. Обученные сети запоминаются. Если задана минимизация параметров, создаются рабочие копии обученных нейросетей и проводится вычисление значимости параметров для каждой сети, после чего следует минимизация параметров у каждой нейросети. Если требуется связанная минимизация (для нескольких нейросетей одновременно), окончательное отключение минимально значимого на данном этапе обучения параметра производится только после того, как все нейросети в связанной группе успешно обучились с исключением данного параметра. В процес-

се обучения могут автоматически применяться тактические методы (удар, изменение весов классов).

Исходя из приведенного описания, управление стратегией осуществляется по заранее заданному алгоритму. При необходимости пользователь может корректировать стратегию в процессе построения решающего блока на любом его этапе.

Экспертные системы, созданные на основе нейросетевой технологии группой «НейроКомп»

Создание каждой ЭС проводилось согласно разработанной технологии и включало: изучение проблемы; постановку задачи; набор обучающих данных и тестирующих примеров; обучение нейросетей; определение оптимальной схемы ЭС; проведение дополнительных экспериментов; разработку и создание интерфейса программы; подключение к ней обученных нейросетей; испытание системы на примерах, не входящих в обучающую выборку; доучивание системы на этих примерах.

Прогнозирование осложнений инфаркта миокарда.

Поиски возможностей прогнозирования осложнений, которые могут возникнуть в госпитальный период инфаркта миокарда, очень актуальны и являются одной из наиболее сложных задач кардиологии. Прогнозирование необходимо осуществлять при поступлении больного в стационар, сразу же после проведения стандартных методов обследования. Оно должно быть быстрым, проводиться неоднократно в процессе наблюдения за больным по мере поступления новых данных о его состоянии.

Получение прогноза осложнений позволяет врачу целенаправленно проводить профилактику, усилить наблюдение за больным, скорректировать режим физической активности пациента (особенно в предполагаемые сроки возникновения осложнения). Прогноз может определять более длительное и интенсивное лечение антикоагулянтами при угрозе тромбоэмболических осложнений, антиаритмическими препаратами - для профилактики аритмий.

При создании базового ядра ЭС прогнозирования осложнений инфаркта миокарда выбрано 4 вида осложнений, достаточно частых и/или довольно опасных [49]. Это фибрилляция предсердий, тромбоэмболические осложнения, перикардит и возникновение/усугубление хронической СН. При постановке задачи мы исходили из необходимости прогнозировать возможность возникновения каждого из четырех выбранных осложнений в отдельности и возможного срока его появления, считая от момента поступления больного в стационар. Таким образом, задача разбивается на 8 подзадач, четыре из которых решаются нейросетями-классификаторами (возникновение осложнений), и четыре - нейросетями-предикторами (сроки возникновения осложнений).

Были выбраны 32 обучающих параметра, отражающие клиническое состояние больного инфарктом миокарда на момент поступления в клинику, данные анамнеза и результаты лабораторных и функциональных исследований: возраст, пол, глубина и локализация инфаркта (по данным электрокардиографии), количество инфарктов в анамнезе, характеристика предшествующей стенокардии, наличие и тяжесть гипертонической болезни и сердечной недостаточности, наличие в анамнезе нарушений сердечного ритма и проводимости, эндокринных заболеваний, тромбоэмболий, хронического бронхита, концентрация калия и натрия в крови, частота сердечных сокращений, характеристика выбранных показателей ЭКГ, размеры отделов сердца по данным эхокардиографии.

Исследованы 300 клинических примеров (Таб. 4). Для экспериментов из общей выборки отдельно для каждого типа нейросетей были выделены обучающая группа (250 человек) и контрольная (тестирующая) группа (50 человек). Разделение выборки производилось случайным образом. Для прогнозирования возникновения каждого осложнения создавались несколько нейросетей, составляющих консилиум. Все нейросети вначале обучались на 250 пациентах обучающей выборки, а затем тестировались на контрольной группе. Результаты теста определялись голосованием в каждом консилиуме. Результаты теста контрольной выборки для прогнозирования фибрилляции предсердий приведены в Таб. 5.

Таблица 4.

Количество больных инфарктом миокарда по классам в каждой из четырех выделенных по осложнениям подгрупп

Осложнение		Количество больных
Фибрилляция предсердий	нет	217
	пароксизмальная форма	60
	постоянная форма	23
Перикардит	нет	172
	есть	128
Тромбоэмболические осложнения	нет	242
	есть	58
Возникновение или усугубление сердечной недостаточности	нет	144
	есть	156

Таблица 5.

Результаты тестирования консилиума нейросетей, прогнозирующих возникновение и форму фибрилляции предсердий (ФП), проведенные на контрольной выборке из 50 примеров, не участвующих в обучении

Известный класс	Вычислено как...		
	Класс 1	Класс 2	Класс 3
Нет ФП - 30 примеров	28	2	-
Пароксизмальная форма ФП - 12 примеров	-	11	1
Постоянная форма ФП - 8 примеров	-	1	7

Процент правильно распознанных примеров в тестирующей выборке при прогнозировании возникновения перикардита составил 76%, тромбоэмболий - 82%, возникновения/усугубления сердечной недостаточности - 78%.

Отдельно были созданы и обучены нейросети-предикторы для прогноза сроков возникновения осложнений (начиная с момента поступления больного в клинику).

Система назначения оптимальной стратегии лечения больных облитерирующим тромбангиитом и прогнозирования его непосредственных исходов.

Облитерирующий тромбангиит (болезнь Бюргера) - тяжелое воспалительное заболевание сосудов мелкого и среднего калибра, сопровождающееся тромбозом и нарушением их проходимости. Этиология этого заболевания до настоящего времени остается неизвестной. Подавляющее большинство больных тромбангиитом - мужчины молодого возраста (18 - 45 лет).

Лечение облитерирующего тромбангиита - трудная, далеко еще не решенная задача. В начальных стадиях заболевания обычно ограничиваются терапевтическими мероприятиями - назначением дезагрегантов, противовоспалительных и антигистаминных препаратов. Однако заболевание часто носит злокачественный характер и быстро приводит к ампутации конечности.

В 1990 - 1994 гг. в отделении хирургии сосудов Краевой Клинической Больницы № 1 г. Красноярска под наблюдением находилось 130 больных облитерирующим тромбангиитом. На каждого больного заполнялась анкета, состоящая из 3 разделов.

Первый раздел (104 пункта) включает вопросы, касающиеся анамнеза жизни и данного заболевания, сопутствующих заболеваний; состояние органов и систем, подробное описание имеющегося у больного тромбангиита с характеристикой состояния сосудов конечностей, данные лабораторных и инструментальных методов исследования, характеристику проводившегося ранее лечения. Другими словами, этот раздел отражает исходный статус больного на момент его поступления в стационар.

Второй раздел (11 пунктов) характеризует проведенное в стационаре лечение (консервативное и/или оперативное).

Третий раздел (4 пункта) содержит сведения о непосредственных исходах проведенного лечения.

Существующие методы лечения тромбангиита часто малоэффективны и процент выполняемых ампутаций остается высоким. Перед врачом стоит задача подобрать оптимальное сочетание методов лечения, действующих на ведущие

звенья патогенеза у конкретного больного. Целью проводимого исследования стало создание нейросетевой ЭС для прогноза непосредственных исходов заболевания и выбора оптимального сочетания терапевтических и хирургических воздействий. Соответственно этому ЭС подразделяется на два функциональных блока, каждый из которых решает свой круг задач [50].

Один блок (блок "И" - исходы) прогнозирует непосредственные исходы заболевания, которые зависят от двух групп параметров. Первая группа - исходный статус больного, фиксированные параметры, отражающие состояние больного на момент поступления в клинику, а также данные анамнеза. Однако исход заболевания зависит не только от исходных параметров, но и от проводимого в клинике лечения. Поэтому вторая группа параметров, необходимая для прогноза - примененные методы лечения. Эти параметры неизвестны при поступлении, на этот момент их можно только предполагать. Однако при обучении нейросети используются уже пролеченные больные с известным набором терапевтических и хирургических воздействий. Обучив нейросети прогнозировать исходы тромбангиита в зависимости от исходного статуса и проведенного лечения, можно моделировать результат, оставляя неизменными фиксированные параметры и подстраивая предполагаемые методы лечения.

Другой блок ЭС обучается прямому выбору наиболее оптимальных методов лечения (блок "Л"), используя только первый, фиксированный набор параметров (Рис. 6).

С учетом поставленной задачи были сформированы подгруппы примеров для обучения нейросетей. Для решения подзадач блока "И" примеры были сгруппированы по классам четыремя способами (Таб. 6), для блока "Л" - 11 способами (Таб. 7).

Для тестирования обученных нейросетей использовались 35 клинических примеров обследованных и пролеченных больных с известными исходами заболевания. Эти примеры не входили в обучающую выборку. Тест каждого примера проводился следующим образом. Сначала тестировались 4 нейросети, прогнозирующие исходы заболевания, причем в качестве параметров лечения, проведенного в клинике, использовались данные о реально назначенном хирур-

гами лечения. Затем пример тестировался нейросетями, назначающими оптимальный набор методов лечения.

После этого проводился повторный тест нейросетями, прогнозирующими исходы, но теперь в пример подставлялись предполагаемые методы лечения, назначенные нейросетями.

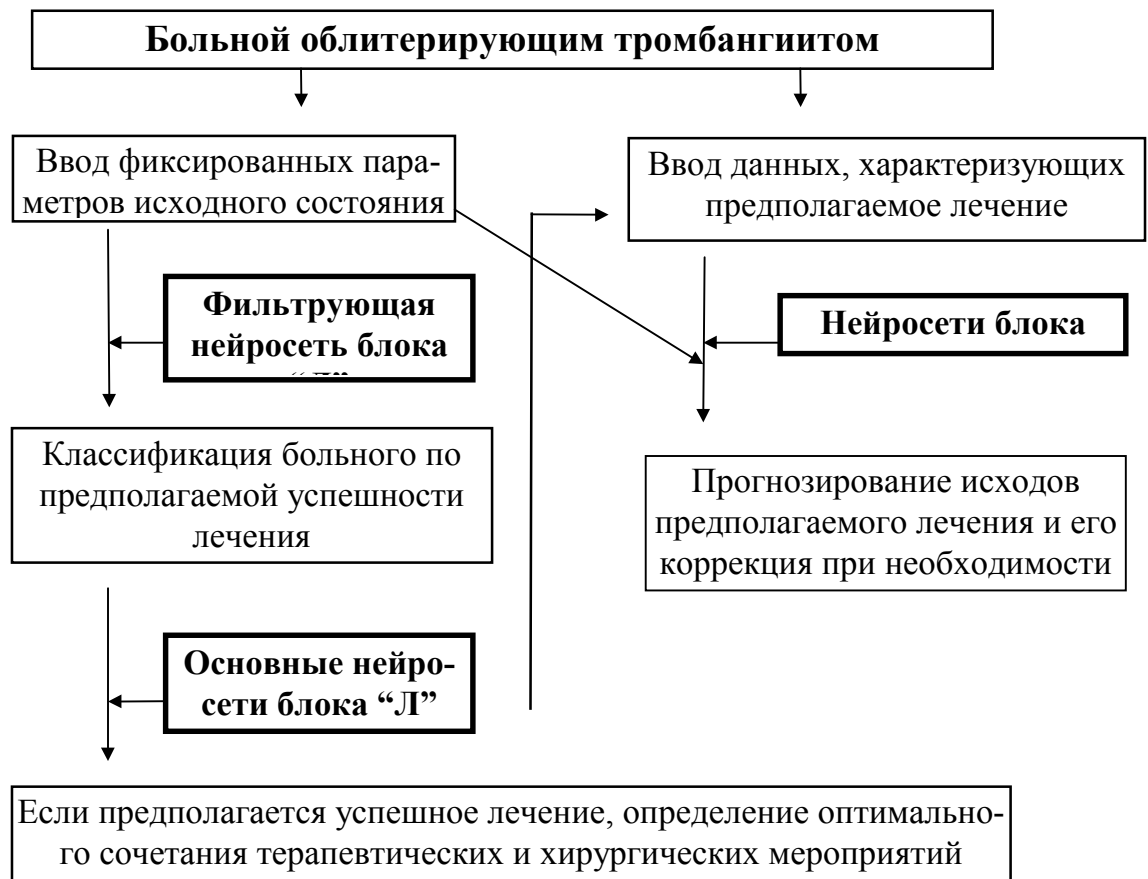


Рисунок 6. Схема функционирования ЭС

Таблица 6.

Подзадачи первого (“И”) блока ЭС

Подзадача	Классы	Число примеров
1. Прогноз динамики ишемии	1. Уменьшение ишемии 2. Без изменений 3. Усиление ишемии	85 38 7
2. Прогноз динамики трофических расстройств	1. Динамики нет 2. Уменьшение трофических расстройств	50 80
3. Прогноз исчезновения болей в покое	1. Боли остались 2. Боли исчезли	57 73
4. Ампутация	1. Не производилась 2. Производилась	123 7

Таблица 7.

Подзадачи второго (“Л”) блока ЭС

Подзадача	Классы	Число примеров
Прогнозирование возможности успешного лечения (“фильтрующая” нейросеть)	1. Больные с неблагоприятным исходом после лечения 2. Больные с благоприятным исходом и больные со смоделированным лечением	24 106
1. Назначение вазопростана	1. Не назначать 2. Назначать	122 8
2. Плазмаферез, иммунопротекторы	1. Не назначать 2. Назначать	111 19
3. Пульстерапия	1. Не проводить 2. Проводить	126 4
4. Симпатэктомия	1. Не проводить 2. Односторонняя 3. Двухсторонняя	104 21 5
5. Деструкция надпочечников	1. Не проводить 2. Односторонняя 3. Двухсторонняя	84 25 21
6. Эмболизация надпочечников	1. Не проводить 2. Односторонняя 3. Двухсторонняя	82 41 7
7. Компактотомия или РОТ	1. Не проводить 2. Проводить	120 10
8. Ревизия	1. Не проводить 2. Проводить	127 3
9. Ампутация	1. Не проводить 2. Проводить	125 5
10. Реконструктивная операция	1. Не проводить 2. Проводить	126 4
11. Некрэктомия	1. Не проводить 2. Проводить	125 5

Рассмотрим результаты первого теста. Нейросеть, прогнозирующая динамику ишемии, сделала правильное заключение в 29 случаях из 35 (82,9%). Регресс трофических расстройств прогнозировался правильно в 14 из 17 случаях имеющихся трофических расстройств (только у 17 больных имелись трофические расстройства, поэтому в остальных случаях прогноз не имел смысла). Это составило 82,4%. Нейросеть, прогнозирующая динамику болевого синдрома,

сделала правильное предсказание в 22 из 25 случаев (88%). Ответы сети, отвечающей за предполагаемую ампутацию, совпали в 33 случаях из 35 (94,3%).

Всего в тестирующей выборке имелось 9 больных, у которых не совпал хотя бы один прогноз. У остальных пациентов все прогнозы делались 4 нейросетями правильно. Из этих 9 больных 7 были отнесены нейросетью-”фильтром” к категории “трудных”.

Интерес представляют два формально неверных результата нейросети, прогнозирующей ампутацию. У данных больных ампутация не была проведена в сроки госпитализации, однако была выполнена в первые 2 месяца после выписки. Нейросеть правильно предсказала отсроченные исходы, поэтому эти два случая нельзя считать явными ошибками.

Рассмотрим результаты теста нейросетей, назначающих лечение больным. У 12 примеров из 25 (35 примеров минус признанные “трудными” нейросетью-”фильтром”) лечение, назначенное нейросетями, полностью совпало с реально назначенным хирургами, и только в одном случае прогноз исходов оправдался лишь в более поздние сроки. Рассмотрим этот случай.

Больной С., 42 лет, болен в течение 5 лет, поступил с критической ишемией (IV степени). Больному была назначена консервативная терапия и вазопро-стан. На фоне такой терапии 3 из 4 нейросетей прогнозировали улучшение состояния. Однако на момент выписки эффекта от лечения не наблюдалось. Тем не менее, при дальнейшем наблюдении за больным отмечался регресс ишемии, регресс трофических расстройств и купирование болей в покое.

В остальных случаях наблюдались некоторые расхождения. При этом у 10 больных нейросети назначили другое лечение, которое (по результатам анализа хирургов), можно было действительно назначить данным больным. Исходы, прогнозируемые после такого лечения, были такими же, как и в случае реально назначенных лечебных мероприятий. В 3 примерах лечение, назначенное нейросетями, было похоже на реально назначенное, однако превосходило его по мощности. В этих случаях прогнозировались исходы, гораздо лучшие, чем наблюдавшиеся в реальности.

Система дифференциальной диагностики “острого живота”.

Задача дифференциальной диагностики "острого живота", несмотря на огромное количество разработок в этой области, для практической хирургии остается одной из самых актуальных проблем. Свыше 90% больных хирургических стационаров - больные с острыми хирургическими заболеваниями органов брюшной полости; оперативные вмешательства при этой патологии составляют от 20 до 40% всей оперативной деятельности хирургических стационаров.

При постановке диагноза и планировании лечения от хирурга требуется принятие единственно правильного решения, от которого часто зависит жизнь больного. Решение должно быть принято как можно в более сжатые сроки, причем в условиях, зачастую не позволяющих получить развернутые данные лабораторных и инструментальных исследований.

Опыт практической хирургии показывает, что наиболее часто встречающимися острыми заболеваниями брюшной полости являются следующие: острый аппендицит (ОА), острый холецистит (ОХ), острый панкреатит (ОП), осложнения язвенной болезни желудка и двенадцатиперстной кишки, острая кишечная непроходимость (ОКН), дифференциальная диагностика которых представляет в некоторых случаях значительные сложности, особенно при поступлении больных в клинику в отдаленные сроки от начала заболевания или с изначально атипичным или стертым процессом.

Изучены 216 историй болезней из архивов клинических больниц г. Красноярска, специализирующихся на приеме экстренной хирургической патологии - ГКБ N 7, ККБ N1, ГБСМП, что позволило работать с уже верифицированными диагнозами, подтвержденными комплексными клинико-лабораторными исследованиями, данными гистологического исследования и операционных находок. Данные еще о 49 пациентах были получены в результате непосредственного клинического наблюдения от момента поступления в хирургический стационар и до постановки окончательного диагноза.

Всего проанализировано 265 клинических примеров (ОА - 63, ОХ - 64, ОП - 52, ОКН - 27, КЯБ - 37, ПЯБ - 22).

При определении входных параметров, необходимых для работы ЭС, проанализирована структура историй болезней различных клиник, данные литературы, возможности лабораторно-инструментальных баз хирургических стационаров. Таким образом, набор входных параметров отражает клинический минимум обследования пациента, используемый врачами хирургических отделений в практической деятельности.

Для сбора клинических данных была разработана анкета, содержащая набор входных параметров и диагноз, являющийся ответом (всего 131 показатель).

При постановке задачи для обучения нейросетей мы исходили из того, что ЭС должна выбирать один или несколько предполагаемых диагнозов из заданного набора (6 диагнозов) на основании 131 параметра пациента при поступлении в клинику. Для экспериментов создавались наборы из 7 нейросетей. В каждом наборе одна нейросеть представляла собой шестиклассовый классификатор, выдающий в качестве ответа один диагноз из шести, а остальные шесть нейросетей - бинарные классификаторы, обучающиеся отличать каждый из рассматриваемых диагнозов от всех остальных.

Для обучения нейросетей были взяты 216 примеров, данные которых были взяты из историй болезни пациентов с уже подтвержденными диагнозами.

Остальные 49 примеров больных, наблюдавшихся в клинике, были оставлены для тестирования ЭС.

Результаты клинической проверки приведены в таблице 8.

Прогностическая способность ЭС после стартового обучения составила 83,7%. Однако мы провели тщательный разбор всех случаев ошибок, который выявил следующие закономерности.

Основное количество ошибок возникало при диагностике острого холецистита и острого панкреатита, причем в этих случаях нейросистема путала эти диагнозы между собой. Более того, при этом правильный диагноз набирал вес (сумму выходных сигналов нейронов нейросетей, ответственных за данный класс), находящийся на втором месте. Это означает, что в случае диагностики острого холецистита (наибольший вес) острый панкреатит оказывался на втором месте (следующий по величине вес) и наоборот. Как известно, при развитии па-

тологического процесса в одном из органов панкреато-дуоденальной системы другие органы, как правило, не остаются интактными, и клинически решить вопрос о преобладании патологии в одном из них зачастую не представляется возможным. В таких случаях больному выставляется диагноз холецистопанкреатит. 5 из 6 примеров, которые при тестировании распознавались ошибочно (острый холецистит как панкреатит и наоборот) относились, судя по анализу клинических данных, как раз именно к такой группе. Поэтому результаты, выданные нейросистемой для этих примеров, не являются явными диагностическими ошибками, так как в этих случаях оба диагноза занимали первые два места по набранному весам.

Таблица 8.

Результаты теста 49 примеров консилиумом из 14 нейросетей

Диагноз	Число примеров	Число правильно распознанных примеров
ОА	19	19 (100%)
ОХ	7	4 (57%)
ОП	12	10 (84%)
ОКН	2	1 (50%)
КЯБ	3	3 (100%)
ПЯБ	6	4 (66%)
Всего	49	41 (83,7%)

Для нейросистемы представлял определенные трудности дифференциальный диагноз между острым панкреатитом и перфоративной язвой (2 ошибки), что имеет место и в клинической практике - при остром панкреатите с выраженным болевым синдромом в 4 случаях из 265 проанализированных клинических примеров, врачами приемного покоя выносились оба диагноза в качестве предварительного и предпочтение одному из них отдавалось уже в процессе динамического наблюдения за состоянием пациента. Также следует учитывать, что достаточно небольшое количество примеров 6 класса (перфоративная язва - 22), как показали результаты теста, несомненно, недостаточно для качественного обучения нейросетей.

В одном случае ЭС распознавала пример 4 класса (ОКН) как острый аппендицит. При анализе примера выявлено, что оба диагноза вынесены врачом

приемного покоя в качестве предварительного и в дальнейшем, при изучении данных истории болезни выяснилось, что больная после проведения консервативной терапии была выписана с диагнозом "кишечная колика". Данная ошибка достаточно серьезная и говорит о необходимости расширения списка диагнозов, определяемых ЭС.

При создании ЭС мы исходили из необходимости выведения полной информации о весах всех диагнозов, выдаваемых системой в каждом случае. Так как диагнозы холецистопанкреатита подразумевают наличие одновременно 2 ответов (два класса), случаи, когда веса этих двух ответов находятся рядом, не могут рассматриваться как ошибки работы системы, при этом наиболее вероятный диагноз формулируется как холецистопанкреатит. С учетом этой поправки, прогностическая способность системы составляет 93,9% (3 ошибки из 49 примеров).

Проведенный анализ значимости обучающих данных показал, что тремя наиболее значимыми параметрами являются боль в правом подреберье, симптом Ортнера и наличие камней в желчевыводящих путях. В число наиболее значимых параметров попали основные характеристики болевого синдрома, важные диагностические признаки (симптомы Кохера, Щеткина - Блумберга, Ровзинга, Воскресенского) и большинство основных характеристик клинического состояния пациента. Минимально допустимым для работы ЭС является набор из 29 параметров, при этом необходимо учитывать, что увеличение числа вводимых параметров существенно повышает диагностическую точность системы.

Нейросети для изучения иммунореактивности

Недостаток четких лабораторных критериев иммунологической недостаточности, большая вариабельность различных показателей состояния иммунной системы [51], особенно при разных патологиях, определяет необходимость поиска простых и эффективных методов, позволяющих дифференцировать состояния иммунореактивности. Наибольшей информативностью обладают показатели субпопуляционного состава лимфоцитов крови, а также параметры их внутриклеточного метаболизма [52]. Однако трудно делать вывод о состоянии иммун-

ной системы, основываясь на величинах отдельных параметров. Необходима комплексная оценка, учитывающая многообразие гомеостатических состояний иммунной системы.

Мы попытались применить нейросети для диагностики вторичной иммунологической недостаточности - иммунодефицита (ИД) [53]. Причинами его являются многие факторы в различном сочетании (другие заболевания, вредные привычки, наследственная предрасположенность, экологические и социальные условия, питание, авитаминоз и др.). В отличие от первичного ИД, причиной которого являются явные, легко выявляемые генетические дефекты, диагноз вторичного ИД весьма сложен.

При иммунологическом обследовании в крови больного определяют различные показатели, главные из которых - количество лимфоцитов (клеток иммунной системы) и абсолютные и относительные (соотношения) количества различных классов лимфоцитов. Кроме того, полезно оценить активность основных внутриклеточных ферментов лимфоцитов и содержание в крови иммуноглобулинов. На основании полученной иммунограммы нужно установить, имеется ли у обследуемого ИД. Сложность заключается в том, что даже у здорового человека могут наблюдаться существенные сдвиги отдельных параметров, а яркий больной может иметь незначительные изменения. Все это связано во-первых, с индивидуальными особенностями иммунной системы, а во-вторых, иммунная система может на короткое время среагировать на какие-либо внешние воздействия (стресс, переохлаждение, переутомление) весьма непредсказуемым образом. Поэтому необходимо оценивать весь комплекс параметров сразу во всех имеющихся взаимосвязях, что на глаз сделать довольно трудно даже для достаточно опытного врача.

Первоначально нейросети была поставлена задача научиться по набору параметров давать простой ответ: имеется ИД у больного или нет. Была сформирована обучающая выборка, состоящая из тщательно обследованных людей, диагноз которым был установлен на основании клинического и иммунологического анализа. Она состояла из двух классов; первый - здоровые люди (51 человек), второй - с диагнозом ИД (42 человека). У обследуемых изучалась активность ос-

новых внутриклеточных ферментов лимфоцитов [54], проводились иммунологические тесты первого уровня [55]. Определялось содержание в крови лейкоцитов, иммуноглобулинов различных классов (IgA, IgM, IgG, IgE), циркулирующих иммунных комплексов (ЦИК), относительное и абсолютное количество лимфоцитов и их субпопуляций: тотальных розеткообразующих клеток (т-РОК), теофиллинрезистентных (ТФР-РОК), теофиллинчувствительных (ТФЧ-РОК), ранних (р-РОК), и стабильных (с-РОК) [56]. Эти показатели служили обучающими параметрами.

Несмотря на кажущуюся простоту задачи и многочисленные попытки, нейросеть не смогла обучиться решать задачу в поставленном виде. Тогда было решено постепенно исключать из обучения самые трудные примеры (имеющие максимальную оценку), чтобы добиться полного обучения на оставшихся. Программа MN_TRAIN делала это автоматически, исключая труднейший пример каждый раз, когда нейросеть заходила в локальный минимум и не могла обучаться далее. Естественно, в конце концов сеть обучилась полностью, исключив из обучающей выборки 30 примеров. После обучения было обнаружено, что все до одного исключенные примеры относятся ко 2 классу (больные с ИД). Таким образом, группа примеров 2 класса оказалась разделенной нейросетью на 2 подгруппы (исключенную и оставшуюся). Далее была проведена статистическая обработка полученных групп.

Общая группа индивидуумов с диагнозом ИД достоверно отличалась от группы здоровых только по одному показателю клеточного метаболизма (фон НАДФ-МДГ, Таб. 9) и одному показателю клеточного иммунитета (соотношение ТФР-РОК/ТФЧ-РОК, Таб. 10). Оставшаяся же группа лиц достоверно отличалась от здоровых людей уже по 8 исследуемым параметрам (причем все они превышали аналогичные параметры у здоровых людей).

Таблица 9.

Активность НАД(Ф)-зависимых дегидрогеназ лимфоцитов крови у здоровых людей, лиц с иммунодефицитом и в группах, выделенных нейросетью (мкЕ + m)

Параметры	Здоровые n=51 1	Группы лиц с ИД		
		общая n=42 2	оставшаяся n=12 3	исключенная n=30 4
Фон Г6ФДГ	0,03±0,002	0,03±0,003	0,03±0,006	0,03±0,003
Фон ГЗФДГ	5,15±0,71	5,12±0,78	7,49±2,32	4,17±0,59
Фон ЛДГ	2,51±0,44	2,44±0,33	3,50±0,59	2,01±0,38 P ₃ <0,05
Фон МДГ	5,10±0,77	4,51±0,69	6,97±1,65	3,52±0,66 P ₃ <0,05
Фон НАДФ-МДГ	0,11±0,01	0,08±0,007 P ₁ <0,05	0,09±0,01	0,08±0,008 P ₃ <0,05
Фон НАДФ-ГДГ	0,44±0,11	0,48±0,12	1,05±0,35 P ₁ <0,05	0,26±0,08 P ₃ <0,01
Фон НАД-ГДГ	3,16±0,42	3,67±0,40	5,24±0,90 P ₁ <0,05	3,04±0,40 P ₃ <0,05
Г6ФДГ	4,75±1,05	5,73±0,90	9,75±2,19 P ₁ <0,05	4,12±0,78 P ₃ <0,01
ГЗФДГ	6,38±1,09	6,38±1,22	11,65±3,15	4,27±1,00 P ₃ <0,01
ЛДГ	8,06±1,03	8,29±0,85	9,33±1,43	7,87±1,05
МДГ	14,77±1,59	14,53±1,40	17,02±2,80	13,53±1,64
НАДФ-МДГ	0,92±0,26	1,78±0,50	2,54±1,17 P ₁ <0,05	1,47±0,54
НАДФ-ГДГ	1,53±0,51	1,75±0,57	2,53±1,02	1,44±0,69
НАД-ГДГ	6,68±0,68	7,16±0,83	9,77±1,93	6,11±0,84 P ₃ <0,05
Обратная ЛДГ	38,90±9,10	35,60±9,80	75,30±28,40	19,69±6,9 P ₃ <0,001
Обратная МДГ	151,00±19,60	109,20±15,20	181,10±35,10	80,4±13,7 P ₃ <0,01
ГР	6,38±0,86	6,55±0,86	8,53±1,13	5,76±0,87

Таблица 10.

Иммунологические показатели крови у здоровых людей, лиц с иммунодефицитом и в группах, выделенных нейросетью

Параметры	Здоровые (1) n=51	Группы лиц с ИД		
		общая (2) n=42	оставшаяся (3) n=12	исключенная (4) n=30
Лейкоциты	5,97±0,34	6,38±0,42	7,72±1,11 P ₁ <0,05	5,85±0,38 P ₃ <0,05
Лимфоциты %	44,7±1,2	43,5±1,5	41,8±3,2	44,2±1,7
Лимфоциты абс.	2,60±0,14	2,80±0,18	3,28±0,37 P ₁ <0,05	2,60±0,20
т-РОК %	62,6±2,0	68,1±1,9	69,5±4,0	67,5±2,3
т-РОК абс.	1,74±0,12	1,91±0,13	2,24±0,21	1,78±0,15
ТФР-РОК %	40,9±1,7	45,3±1,9	39,8±3,4	47,3±2,3 P ₁ <0,05
ТФР-РОК абс.	1,11±0,09	1,26±0,11	1,23±0,23	1,27±0,13
ТФЧ-РОК %	24,0±1,5	24,6±2,3	32,1±5,5 P ₁ <0,05	21,7±2,3 P ₃ <0,05
ТФЧ-РОК абс.	0,67±0,07	0,68±0,08	0,95±0,16	0,58±0,08 P ₃ <0,05
р-РОК %	46,3±1,5	45,3±1,8	44,7±3,8	45,6±2,1
р-РОК абс.	1,29±0,1	1,28±0,10	1,49±0,22	1,19±0,11
с-РОК %	8,1±1,1	8,5±1,0	9,3±1,8	8,1±1,2
с-РОК абс.	0,23±0,03	0,28±0,04	0,39±0,09 P ₁ <0,05	0,23±0,04
ТФЧ / ТФР	2,07±0,17	2,94±0,4 P ₁ <0,05	1,68±0,38	3,42±0,52 P ₁ <0,01, P ₃ <0,05
IgA	2,17±0,45	1,54±0,08	1,56±0,13	1,54±0,10
IgM	0,57±0,04	0,67±0,09	0,92±0,26 P ₁ <0,05	0,57±0,06
IgG	8,97±0,41	8,80±0,37	9,43±0,59	8,53±0,47
IgE	15,76±3,03	15,34±3,47	11,82±3,72	16,74±4,65
ЦИК	35,68±5,29	31,43±4,71	34,55±8,05	30,00±6,12

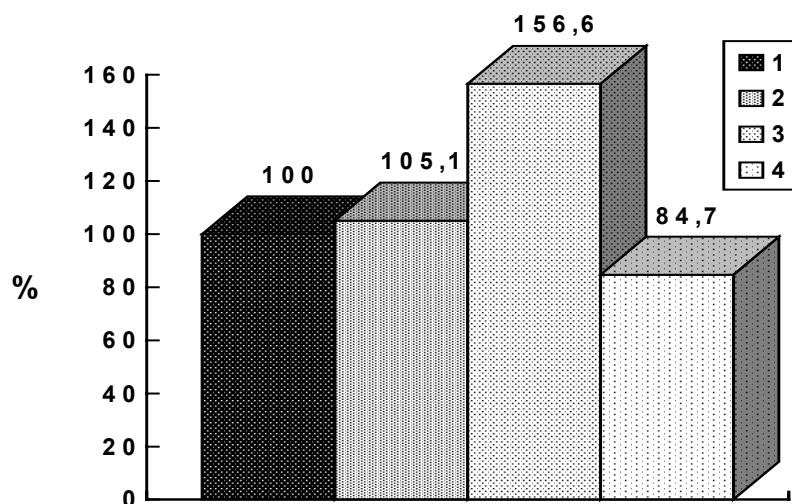


Рисунок 7. Иллюстрация сдвигов клеточного метаболизма лимфоцитов (приведены суммарные относительные величины) у больных с иммунодефицитными состояниями в сравнении со здоровыми людьми.

Уровень клеточного метаболизма лимфоцитов:

1. В группе здоровых людей (принят за 100%)
2. В группе больных иммунодефицитными состояниями (рассчитанный на всю группу)
3. В подгруппе больных, примеры которых остались в процессе обучения нейросети

В подгруппе больных, примеры которых были исключены в процессе обучения нейросети Исключенная группа достоверно отличалась от группы здоровых по 2 показателям иммунного статуса. Однако наибольшее количество достоверных различий наблюдалось между исключенной и оставшейся группами - 10 по параметрам клеточного метаболизма и 4 по параметрам иммунного статуса.

Анализ направлений параметров в полученных подгруппах показал противоположную направленность сдвигов некоторых из них по сравнению с группой здоровых. Наиболее яркие противонаправленные сдвиги наблюдались почти по всем параметрам активности внутрилимфоцитарных ферментов (Рис. 7). Интерес представляет также достоверный разнонаправленный сдвиг соотношения ТФР-РОК/ТФЧ-РОК, указывающий на различную направленность дифференцировки лимфоцитов в подгруппах, выделенных нейросетью.

Анализируя примеры больных, попавших в разные подгруппы, можно предположить, что нейросеть, используя заданные обучающие параметры, разделила общую выборку лиц с ИД на две группы с различными состояниями иммунореактивности.

В оставшейся группе при депрессии иммунного ответа наблюдается компенсаторная реакция иммунной системы, проявляющаяся прежде всего увеличением активности внутриклеточных ферментов.

В исключенных примерах большинство метаболических показателей приближается к аналогичным параметрам здоровых индивидуумов и, вероятно, наблюдается компенсаторная реакция, выражающаяся в активации иммунореактивности [57].

Таким образом, в данном случае нейронная сеть, обучавшись на неверно заданной классификационной модели (что сначала не было известно исследователям), стала источником гипотезы о разнородности одного из классов.

С учетом скорректированной классификационной модели были инициализированы 10 новых нейросетей-экспертов для классификации теперь уже всех 3 полученных групп. Нейросети имели параметры, аналогичные предыдущей. Все они полностью обучилась различать заданные 3 новых класса. При анализе значимости обучающих параметров самыми информативными оказались соотношение ТФР-РОК/ТФЧ-РОК и активности дегидрогеназ лимфоцитов.

Таким образом, новые 3-классовые нейросети накопили определенный опыт дифференцировки трех состояний иммунореактивности - характерного для здоровых людей (1 класс) и две фазы измененной иммунореактивности, характерной для состояний вторичного иммунодефицита (2 и 3 классы).

Литература

1. Парин В.В., Баевский Р.М. Медицина и техника.- М.: Знание, 1968.- С.36-49.
2. Переверзев-Орлов В.С. Советчик специалиста. Опыт разработки партнерской системы // М.: Наука, 1990.- 133 с.
3. Fu H.C., Shann J.J. A fuzzy neural network for knowledge learning // Int. J. Neural Syst.- 1994.- V.5, N.1.- P.13-22.
4. Масалович А.И. От нейрона к нейрокомпьютеру // Журнал доктора Добба.- 1992.- N.1.- С.20-24.
5. Stefanuk V.L. Expert systems and its applications // The lectures of Union's workshop on the main problems of artificial intelligence and intellectual systems. Part 2, Minsk, 1990.- P.36-55.
6. Горбань А.Н., Фриденберг В.И. Новая игрушка человечества // МИР ПК, 1993, № 9.
7. Горбань А.Н. Нейрокомпьютер или аналоговый ренессанс // МИР ПК, 1994, № 10.
8. Rozenbojm J., Palladino E., Azevedo A.C. An expert clinical diagnosis system for the support of the primary consultation // Salud. Publica Mex.- 1993.- V.35, N.3.- P.321-325.
9. Poli R., Cagnoni S., Livi R. et al. A Neural Network Expert System for Diagnosing and Treating Hypertension // Computer.- 1991.- N.3.- P.64-71.
10. Gindi G.R., Darken C.J., O'Brien K.M. et al. Neural network and conventional classifiers for fluorescence-guided laser angioplasty // IEEE Trans. Biomed. Eng.- 1991.- V.38, N.3.- P.246-252.
11. Allen J., Murray A.. Development of a neural network screening aid for diagnosing lower limb peripheral vascular disease from photoelectric plethysmography pulse waveforms // Physiol. Meas.- 1993.- V.14, N.1.- P.13-22.
12. Astion M.L., Wener M.H., Thomas R.G., Hunder G.G., Bloch D.A. Application of neural networks to the classification of giant cell arteritis // Arthritis Reum.- 1994.- V.37, N.5.- P.760-770.

13. Baxt W.G. A neural network trained to identify the presence of myocardial infarction bases some decisions on clinical associations that differ from accepted clinical teaching // *Med. Decis. Making.*- 1994.- V.14, N.3.- P.217-222.
14. Baxt W.G. Complexity, chaos and human physiology: the justification for non-linear neural computational analysis // *Cancer Lett.*- 1994.- V.77, N.2-3.- P.85-93.
15. Baxt W.G. Use of an artificial neural network for the diagnosis of myocardial infarction // *Ann. Intern. Med.*- 1991.- V.115, N.11.- P.843-848.
16. Baxt W.G. Analysis of the clinical variables driving decision in an artificial neural network trained to identify the presence of myocardial infarction // *Ann. Emerg. Med.*- 1992.- V.21, N.12.- P.1439-1444.
17. Guo Z., Durand L.G., Lee H.C. et al. Artificial neural networks in computer-assisted classification of heart sounds in patients with porcine bioprosthetic valves // *Med. Biol. Eng. Comput.*- 1994.- V.32, N.3.- P.311-316.
18. Barschdorff D., Ester S., Dorsel T et al. Phonographic diagnostic aid in heart defects using neural networks // *Biomed. Tech. Berlin.*- 1990.- V.35, N.11.- P.271-279.
19. Okamoto Y., Nakano H., Yoshikawa M. et al. Study on decision support system for the interpretation of laboratory data by an artificial neural network // *Rinsho. Byori.*- 1994.- V.42, N.2.- P.195-199.
20. Maclin P.S., Dempsey J. Using an artificial neural network to diagnose hepatic masses // *J. Med. Syst.*- 1992.- V.16, N.5.- P.215-225.
21. Rinast E., Linder R., Weiss H.D. Neural network approach for computer-assisted interpretation of ultrasound images of the gallbladder // *Eur. J. Radiol.*- 1993.- V.17, N.3.- P.175-178.
22. Modai I., Stoler M., Inbar-Saban N. et al. Clinical decisions for psychiatric inpatients and their evaluation by a trained neural network // *Methods Inf. Med.*- 1993.- V.32, N.5.- P.396-399.
23. Ercal F., Chawla A., Stoeker W.V. et al. Neural network diagnosis of malignant melanoma from color images // *IEEE Trans. Biomed. Eng.*- 1994.- V.41, N.9.- P.837-845.

24. Lee H.-L., Suzuki S., Adachi Y. et al. Fuzzy Theory in Traditional Chinese Pulse Diagnosis // Proceedings of 1993 International Joint Conference on Neural Networks, Nagoya, Japan, October 25-29, 1993.- Nagoya, 1993.- V.1.- P.774-777.
25. Yang T.-F., Devine B., Macfarlane P.W. Combination of artificial neural networks and deterministic logic in the electrocardiogram diagnosis of inferior myocardial infarction // Eur. Heart J.- 1994.- V.15.- Abstr. Supplement XII-th World Congress Cardiology (2408).- P.449.
26. Hoher M., Kestler H.A., Palm G. et al. Neural network based QRS classification of the signal averaged electrocardiogram // Eur. Heart J.- 1994.- V.15.- Abstr. Supplement XII-th World Congress Cardiology (734).- P.114.
27. Nakajima H., Anbe J., Egoh Y. et al. Evaluation of neural network rate regulation system in dual activity sensor rate adaptive pacer // European Journal of Cardiac Pacing and Electrophysiology.- Abstracts of 9th International Congress, Nice Acropolis - French, Rivera, June 15-18, (228), 1994.- Rivera, 1994.- P.54.
28. Gross G.W., Boone J.M., Greco-Hunt V. et al. Neural networks in radiologic diagnosis. II. Interpretation of neonatal chest radiographs // Invest. Radiol.- 1990.- V.25, N.9.- P.1017-1023.
29. Floyd C.E.Jr., Lo J.Y., Yun A.J. et al. Prediction of breast cancer malignancy using an artificial neural network // Cancer.- 1994.- V.74, N.11.- P.2944-2948.
30. Reinbnerger G., Weiss G., Werner-Felmayer G. et al. Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees // Proc. Natl. Acad. Sci., USA.- 1991.- V.88, N.24.- P.11426-11430.
31. Шварц Э., Трис Д. Программы, умеющие думать // Бизнес Уик.- 1992.- N.6.- C.15-18.
32. Aynsley M., Hofland A., Morris A.J. et al. Artificial intelligence and the supervision of bioprocesses (real-time knowledge-based systems and neural networks) // Adv. Biochem. Eng. Biotechnol.- 1993.- N.48.- P.1-27.

33. Budilova E.V., Teriokhin A.T. Endocrine networks // The RNNS/IEEE Symposium on Neuroinformatics and Neurocomputers, Rostov-on-Don, Russia, October 7-10, 1992.- Rostov/Don, 1992.- V.2.- P.729-737.
34. Varela F.J., Coutinho A., Dupire B. et al. Cognitive networks: immune, neural and otherwise // Teoretical immunology. Ed. by Perelson A.- Addison Wesley, 1988.- Part 2.- P.359-375.
35. Levine D.S., Parks R.W., Prueitt P.S. Methodological and theoretical issues in neural network models of frontal cognitive functions // Int. J. Neurosci.- 1993.- V.72, N.3-4.- P.209-233.
36. Van Leeuwen J.L. Neural network simulations of the nervous system // Eur. J. Morphol.- 1990.- V.28, N.2-4.- P.139-147.
37. Senna A.L., Junior W.M., Carvallo M.L.B., Siqueira A.M. Neural Networks in Biological Taxonomy // Proceedings of 1993 International Joint Conference on Neural Networks, Nagoya, Japan, October 25-29, 1993.- Nagoya, 1993.- V.1.- P.33-36.
38. Sweeney J.W.P., Musavi M.T., Guidi J.N. Probabilistic Neural Network as Chromosome Classifier // Proceedings of 1993 International Joint Conference on Neural Networks, Nagoya, Japan, October 25-29, 1993.- Nagoya, 1993.-V.1.- P.935-938.
39. Korver M., Lucas P.J. Converting a rule-based expert system into a belief network // Med. Inf. Lond.- 1993.- V.18, N.3.- P.219-241.
40. Gorban A.N., Rossiev D.A., Gilev S.E. et al. "NeuroComp" group: neural-networks software and its application // Russian Academy of Sciences, Krasnoyarsk Computing Center, Preprint N 8.- Krasnoyarsk, 1995.- 38 p.
41. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере // Новосибирск: Наука, 1996.- 276 с.
42. Горбань А.Н., Россиев Д.А., Коченов Д.А. Применение самообучающихся нейросетевых программ. Раздел 1. Учебно-методическое пособие для студентов специальностей 22.04 и 55.28.00 всех форм обучения // Красноярск: Сибирский Технологический Институт, 1994.- 24 с.
43. Гилева Л.В., Гилев С.Е., Горбань А.Н., Гордиенко П.В., Еремин Д.И., Коченов Д.А., Миркес Е.М., Россиев Д.А., Умнов Н.А. Нейропрограммы. Учебное

пособие: В 2 ч. // Красноярск, Красноярский государственный технический университет, 1994.- 260 с.

44. Gorban A.N., Rossiev D.A., Gilev S.E. et al. Medical and physiological applications of MultiNeuron neural simulator // Proceedings of World Congress on Neural Networks-1995 (WCNN'95).- Washington, 1995.- V.1, P.170-175.

45. Горбань А.Н. Обучение нейронных сетей. М.: Параграф, 1990.- 160 с.

46. Гилев С.Е., Горбань А.Н., Миркес Е.М. Малые эксперты и внутренние конфликты в обучаемых нейронных сетях // Доклады Академии Наук СССР.- 1991.- Т.320, N.1.- С.220-223.

47. Гилев С.Е., Коченов Д.А., Миркес Е.М., Россиев Д.А. Контрастирование, оценка значимости параметров, оптимизация их значений и их интерпретация в нейронных сетях // Доклады III Всероссийского семинара “Нейроинформатика и ее приложения”.- Красноярск, 1995.- С.66-78.

48. Коченов Д.А., Миркес Е.М., Россиев Д.А. Автоматическая подстройка входных данных для получения требуемого ответа нейросети // Проблемы информатизации региона. Труды межрегиональной конференции (Красноярск, 27-29 ноября 1995 г.). Красноярск, 1995.- С.156.

49. Rossiev D.A., Golovenkin S.E., Shulman V.A., Matjushin G.V. Neural networks for forecasting of myocardial infarction complications // The Second International Symposium on Neuroinformatics and Neurocomputers, Rostov-on-Don, Russia, September 20-23, 1995.- Rostov-on-Don, 1995.- P.292-298.

50. Мызников А.В., Россиев Д.А., Лохман В.Ф. Нейросетевая экспертная система для оптимизации лечения облитерирующего тромбангиита и прогнозирования его непосредственных исходов // Ангиология и сосудистая хирургия.- 1995.- N 2.- С.100.

51. Лебедев К.А., Понякина И.Д. Иммунограмма в клинической практике. - М.: Наука, 1990.- 224 с.

52. Подосинников И.С., Чухловина М.Л. Метаболизм и функция лимфоцитов при первичных иммунодефицитных состояниях // Иммунология.- 1986.- N.4.- С.30-38.

53. Rossiev D.A., Savchenko A.A., Borisov A.G., Kochenov D.A. The employment of neural-network classifier for diagnostics of different phases of immunodeficiency // Modelling, Measurement & Control.- 1994.- V.42.- N.2. P.55-63.
54. Савченко А.А., Сунцова Л.Н. Высококчувствительное определение активности дегидрогеназ в лимфоцитах периферической крови человека биолюминесцентным методом // Лаб. дело.- 1989.- N.11.- С.23-25.
55. Лозовой В.П., Кожевников В.С., Волчек И.А. и др. Методы исследования Т-системы иммунитета в диагностике вторичных иммунодефицитов при заболеваниях и повреждениях.- Томск, 1986. 18 с.
56. Кожевников В.С., Волчек И.А. Функциональная активность субклассов Т-лимфоцитов и их роль в патогенезе вторичных иммунодефицитов // Функциональные характеристики циркулирующего пула иммунокомпетентных клеток: Сб. науч. тр. - Новосибирск, 1984.- С.11-17.
57. Лазарев И.А., Русаков В.И., Ткачева Т.Н. Иммунопатологические изменения у больных язвенной болезнью желудка и двенадцатиперстной кишки и их иммунорекция // 4 Всесоюзный Съезд гастроэнтерологов.: Тез. докл. - Москва, Ленинград, 1990.- Т.1.- С.702-703.