**Лекция №10-11**

# МЕДИЦИНСКАЯ КОМПЬЮТЕРНАЯ ДИАГНОСТИКА

## Информационные основы диагностики

Познание окружающего мира неразрывно связано с изучением принципов организации и развития живой материи. Понимание процессов, протекающих в живом организме, опознание определенных функциональных состояний и прогнозирование развития представляется чрезвычайно важной задачей. С позиции медицинской практики, накопленные и осмысленные в этой области знания, позволяют формулировать определенные понятия о механизмах и принципах развития биологических систем, а в рамках большого числа математических концепций — создавать модели, описывающие поведение таких систем в реальных условиях окружающего мира.

В самом общем понимании медицинская диагностика строится на двух уровнях. Первый уровень представляют инструментальные методы исследования состояния живого организма. Перечень методов этого направления достаточно большой. На практике хорошо известны кардиографический, энцефалографический, рентгеновский, изотопный, ультразвуковой и другие методы исследования. Отдельную позицию в этом списке занимают методы клинического лабораторного анализа. Такие методы позволяют установить функциональное состояние живого организма на основе биохимических исследований крови, мочи, клеточных фрагментов биологической ткани.

На основе инструментальных методов исследования получают специальные пакеты данных, анализируя которые формируется диагностическое заключение. Форма представления данных в таких методах различна и определяется техническими параметрами медицинской аппаратуры. Однако на современном этапе все большее количество методов исследования ориентировано на использование компьютерной техники в целях проведения математического анализа получаемых данных.

Другой уровень формирования медицинского заключения строится на возможности врача сформировать некоторое словесное описание состояния организма. Это так называемая вербальная форма, характеризующая состояние биологической системы. Существуют специальные методы, которые такое описание также переводят в форму цифровой записи. Это намного упрощает процесс построения результирующего диагностического правила.

С точки зрения системного подхода, в медицинской практике два уровня представления медицинского решения о состоянии организма чрезвычайно важны. Именно объединение этих диагностических уровней после соответствующего анализа данных позволяет построить адекватное представление о текущем состоянии живого организма.

Медицинская диагностика — это своего рода классификационная задача. В простейшем случае рассматривается вопрос о нормальном состоянии организма или об имеющемся отклонении от нормы. По современным медицинским представлениям понятие “нормы” или “не нормы” формируются на основе большого количества реальных фактов. Это так называемый статистический показатель, который в отдельных случаях может значительно изменяться. В связи с этим возникает проблема описания промежуточных состояний организма между этими крайними категориями, которые с учетом индивидуальных особенностей организма могут быть использованы при формировании диагностического заключения.

Именно этот аргумент является главным при стремлении медицинского персонала активнее использовать средства вычислительной техники.

Персональный компьютер и соответствующие математические пакеты программ анализа данных позволяют не только повысить эффективность поиска и построения

диагностического решения, но и создают необходимые условия для понимания мельчайших деталей возникновения нарушения в организме.

Таким образом, представить исходные данные медицинского обследования в форме удобной для дальнейшего анализа с помощью компьютера сегодня не представляется сложной задачей. Практически любой материал, полученный при обследовании, может быть представлен в цифровом виде. Вопрос о форме отображения — в виде графика, фрейма или цифровой последовательности — решается отдельно, с учетом требований специалистов.

Анализ данных обследования неразрывно связан со временем. Время и информационные параметры, характеризующие функциональное состояние биологической системы, имеют неразрывную связь. Каждый информационный параметр имеет определенную значимость при формировании диагностического решения. Ранг — значимость параметра для диагностики состояния проявляется в определенное время, в течение некоторого периода обследования. Известно, что на относительно коротком интервале наблюдения некоторые информационные показатели состояния организма не могут быть замечены. Напротив при длительном наблюдении происходит накопление избыточной информации, в которой информационный параметр представляется в скрытой форме.

Прежде чем приступить к диагностической процедуре требуется найти информационные параметры изучаемого процесса. Отыскание информационных параметров происходит с использованием некоторой априорной информации, основу которой составляют анатомические, биофизические и биохимические данные, различные модели поведения изучаемой системы или процесса. На всем этом многоплановом поле возможных представлений о поведении системы необходимо построить свою модель выбора информационных параметров, указать момент времени, когда такие параметры становятся значимыми и могут быть обнаружены. Это достаточно сложная задача.

Обычно такую задачу начинают решать последовательно, этап за этапом. Вначале определяют продолжительность регистрации исследуемого процесса, изучение и анализ которого позволят в будущем построить диагностическое решение о состоянии живого организма. Одновременно с этим определяются требования к точности регистрируемого сигнала и форма — вид его представления исследователю. После того как эти проблемы оказываются решенными, приступают к построению алгоритма отыскания информационных параметров в зарегистрированном массиве данных.

В большинстве случаев на практике одиночное (разовое) исследование биологической системы позволяет сформировать так называемую выборку. Количество выборок, которое позволяет уверенно проводить построение диагностического решения, различно для каждого вида медицинского исследования.

Надо отметить, что формирование выборки происходит в течение некоторого промежутка времени. Каждое значение выборки оказывается элементом некоторой временной последовательности, которая, как утверждалось выше, формируется по заранее установленному правилу. Так например, временная последовательность амплитудных значений сигнала при кардиологическом исследовании образует исходную выборку. Это так называемый двухмерный массив данных. При радиоизотопном исследовании формируется сцинтиграфическое изображение, структура которого представляется временной последовательностью значений регистрируемого сигнала, но в трехмерном пространстве. Специфичность временного формирования исходной выборки должна приниматься во внимание при отыскании информационных параметров.

В процедуре формирования исходной выборки важно выделить элемент связи появления отдельного значения регистрируемого сигнала с текущим временем. Понятно, что произвольное изменение расположения во времени текущих значений выборки недопустимо, потому что может изменить наше представление о характере процессов,

протекающих в системе. Поэтому независимо от методов последующего анализа данных структура выборки должна оставаться неизменной.

Поиск информационных параметров на выборке проводят с использованием различных математических методов. Несмотря на широкую возможность выбора методов отыскания информационных параметров все они должны удовлетворять определенным условиям. Например, позволять исследовать случайные, непериодические или условно периодические сигналы.

Действительно, жизнедеятельность организма подтверждается наличием большого числа информационных показателей, которые непрерывно изменяются в определенных допустимых пределах. Это соответствует нашим представлениям о гомеостазисе. Поддержание своих существенных параметров в строго определенных пределах изменения характерно для биологических систем. Однако надо помнить, что область допустимых значений для каждого параметра может отражать как индивидуальные, так и возрастные особенности организма. Кроме этого, область допустимых значений того или другого параметра может существенно меняться от географических координат местности, времени суток, и, наконец, образа жизни. Принимая это во внимание, на практике оказывается трудной задачей сформировать набор очень похожих выборок. Другими словами, приступая к формированию некоторого однородного класса, например, выборок, характеризующих нормальное состояние организма, трудно получить устойчивые однотипные оценки. Эта трудная, но интересная задача привлекает многих исследователей.

В этой области существует немало перспективных направлений. Одним из таких направлений является метод структурного координатного анализа (СКА). В основу метода положены следующие постулаты:

1. Исходная выборка — последовательность значений исследуемого сигнала, представленная по заранее определенному правилу временных событий, происходящих в изучаемой системе.

2. Размерность выборки — количество элементарных событий, представленных соответствующими значениями сигнала, характеризует полный пространственно-временной континуум возможных изменений, происходящих в системе.

3. Информационные параметры, характеризующие состояние системы, могут быть найдены посредством соответствующего алгоритма, устанавливающего функциональную связь между отдельными фрагментами исходной выборки.

4. Размер фрагментов, оцениваемый по количеству элементов, и их местоположение в текущей выборке может меняться, но количество фрагментов сохраняется постоянным для однородного класса изучаемых событий.

5. Исходная выборка может быть эквивалентным образом представлена набором фрагментов разного размера, не искажающих информационное содержание изучаемого процесса.


**Рис. 3.1. Структурная схема построения оценки состояния организма**

Такую процедуру отыскания информационных параметров, а вслед за этим и построение диагностического решения, можно проиллюстрировать схемой, представленной на рис. 3.1.

На этапе СКА осуществляются процедуры:

- поиск информативных параметров в исходной выборке;
- поиск временных интервалов, содержащих информативные параметры;
- формирование фрагментов выборки.

На этапе формирования оценки — диагностического правила, осуществляются процедуры:

- построение функциональной зависимости фрагментов выборки;
- формирование диагностического решения о функциональном состоянии системы.

Таким образом, метод СКА позволяет сократить размерность исходной выборки посредством создания фрагментарного описания. С другой стороны, набор фрагментов, объединенных некоторым общим функциональным описанием, может представлять объект для дальнейшего дробления. Как будет описано ниже, это представляется исключительно полезным при использовании фрактальных методов анализа данных.

Поиск фрагментов выборки осуществляется на некоторой последовательности значений исходного сигнала. В свою очередь, временная последовательность значений изучаемого сигнала характеризуется некоторой формой представления, которая отражает состояние изучаемой системы. Взаимная связь формы сигнала и состояния системы представляется чрезвычайно важной при построении диагностического решения. При изучении различных процессов, протекающих внутри биологического организма, приходится иметь дело с сигналами очень сложной формы.

В таком понимании выборка может быть охарактеризована динамическим образом, отражающим поведение биологической системы на ограниченном интервале времени. Для того чтобы не потерять элементы такого образа при анализе поведения системы, требуется знать правило формирования последовательности исходных значений выборки. Это трудная задача, которая носит название “проблемы Гильберта”.

В математическом отношении формулировка этой проблемы имеет свою историю. В 1900 г. немецкий математик Давид Гильберт в своем докладе на Международном конгрессе математиков в Париже сформулировал 23 нерешенные задачи, которые он считал наиболее важными в математике того времени. Эти задачи получили название “проблемы Гильберта” и оказали огромное влияние на развитие всей математики 20 в. До сих пор не все проблемы Гильберта полностью решены, а многие из них побудили ученых

к созданию совершенно новых теорий. Как выяснилось в последние годы, теория нейронных сетей также связана с одной из этих проблем, а именно с тринадцатой.

Тринадцатая проблема Гильберта формулируется так: "... Верно ли, что существует непрерывная функция от трех переменных, которая не может быть представлена в виде композиции непрерывных функций от двух переменных?"

Под композицией функций понимается подстановка одной функции в качестве аргумента другой. Например, функция трех переменных $F(x, y, z) = xz + yz$ может быть представлена в виде композиции функций двух переменных:

$F(x, y, z) = S(M(x, z), M(y, z))$,(*)

где $M(x, z) = xz$, а $S(a, b) = a + b$.

Как нам сегодня известно, тринадцатая проблема Гильберта была решена в 1957 г. студентом мехмата МГУ, а ныне академиком Владимиром Игоревичем Арнольдом. Он показал, что любая непрерывная функция трех переменных представляется в виде композиции непрерывных функций двух переменных. Таким образом, гипотеза Гильберта была опровергнута.

В том же 1957 г. математик Андрей Николаевич Колмогоров доказал гораздо более сильную теорему.

**Теорема Колмогорова**: Любая непрерывная функция от $n$ переменных $F(x_1, x_2, ..., x_n)$ может быть представлена в виде

$$F(x_1, x_2, ..., x_n) = \sum_{j-1}^{2n+1} g_j \left( \sum_{i-1}^{n} h_{ij}(x_i) \right),$$

,(**)

где $g_j$ и $h_{ij}$ — непрерывные функции, причем $h_{ij}$ не зависят от функции $F$.

Эта теорема означает, что для реализации функций многих переменных достаточно операций суммирования и композиции функций одной переменной. Удивительно, что в этом представлении лишь функции $g_j$ зависят от представляемой функции $F$, а функции $h_{ij}$ универсальны.

Заметим, что формула (**) очень похожа на формулу (*). Если перевести эту теорему на язык нейронной сети, то она будет звучать так.

**Если известны функции $h_{ij}$, то любую непрерывную функцию от $n$ переменных можно точно реализовать с помощью простой нейросети на основе трехслойного персептрона. Для этого достаточно подобрать $2n+1$ передаточных функций $g_j$ нейронов скрытого слоя.**

Эта сеть не будет персептроном в строгом смысле, так как на входах второго слоя к сигналам необходимо применить функции $h_{ij}$, а не просто умножить их на веса.

К сожалению, при всей своей математической красоте теорема Колмогорова *малоприменима на практике*. Это связано с тем, что функции $h_{ij}$ — негладкие и трудно вычислимые; также неясно, каким образом можно подбирать функции $g_j$ для данной функции $F$. Роль этой теоремы состоит в том, что она показала принципиальную возможность реализации сколь угодно сложных зависимостей с помощью относительно простых автоматов типа нейронных сетей.

### Нейронная сеть и персептрон

Поиск возможных решений этой задачи может быть значительно упрощен, если обратиться к рассмотрению биологических систем. Живые организмы обладают развитой информационной сетью, которая носит название нейронной сети.

В течение последних лет уходящего столетия в этой области появилось много интересных научных работ, описывающих как внутреннее строение нейронных сетей, так и возможности практического использования накопленных знаний в различных областях.

Следует отметить, что именно благодаря высокому приоритету данного направления и огромному материалу, накопленному в этой области знаний, в сфере

профессионалов, работающих по данной теме, имеются существенные различия во взглядах о перспективности нейросетевых технологий.

С позиции математики, такие технологии могут значительно повысить эффективность работы технических систем, ориентированных на сбор и обработку данных. Оговоримся сразу: нейросетевые технологии, как и в живом организме, ориентированы на анализ информационных потоков. По сути, это в некотором роде попытка создания машинных — компьютерных методов анализа данных, построенных на принципах нейронной сети, хорошо известной в живом организме. В таком понимании нейронные технологии представляют некоторое обобщенное название группы технических средств и математических методов по анализу информационных потоков, применение которых позволяет получить желаемые результаты.

Примем во внимание тот факт, что на практике существуют некоторые современные концепции, выраженные в математической форме и определенный экспериментальный материал, полученный на основе многочисленных медицинских исследований. Две составляющие этого базиса представляют разные творческие коллективы, сотрудничающие на одном научном поле познания процессов анализа информационных потоков. В историческом плане такое сотрудничество идет с переменным успехом для каждой из сторон. В настоящее время математики утверждают, что нейронные технологии подобны процессам анализа информации в живых системах. Это справедливо, но только отчасти и не полностью. Такие утверждения отражают общие надежды математиков к данному подходу, но совсем не связаны с теоретическими основами нейросетевого подхода. Такое несоответствие общих взглядов на существующую проблему порождается в результате пропуска существенных, пока еще до конца не решенных проблем, выраженных в специальных понятиях и терминах.

Рассмотрим некоторые вопросы, имеющие существенное значение в работе математиков и практиков, обращающихся к анализу информационных потоков. Не акцентируя внимание на специальных вопросах, дадим общее представление об этой проблеме.

Начнем с простых понятий. На практике, например в медицине, существуют задачи по сбору данных о состоянии организма. На основании таких данных можно построить диагноз — дать характеристику функционального состояния организма. В простейшем случае такое построение имеет всего две ступени различия: норма и патология.

Если задана функция изменения информационного параметра, то можно построить прогноз о состоянии организма на какой-то отдаленный промежуток времени. Будем говорить лишь об одном, наиболее строгом, способе выражения зависимости, а именно в виде функции от нескольких переменных. Например, если начальный размер живого организма составлял $S$ сантиметров, а увеличение размера тела — $h$ процентов в месяц, то через $N$ лет можно ожидать, что этот показатель составит $K = (1 + h)^N \cdot S$ сантиметров, — вот типичный пример зависимости, выраженной в виде функции от трех переменных $K(S, h, N)$. Очевидно, что мы получим приближенный результат, поскольку сам функционал записан в общем, не полном виде. Однако приближенные функциональные зависимости подобного вида очень часто присутствуют в реальных задачах биометрии.

Ясно, что эти функционалы (функции) невозможно определить априори. Здесь нельзя воспользоваться предыдущим опытом, знаниями, накопленными в процессе длительных экспериментов. Более того, выражаясь математическими терминами, можно утверждать, что эти функции записываются достаточно сложными формулами неизвестного вида, в которых участвуют все переменные. В данном примере имеются в виду переменные, определяющие рост живого организма.

Имеет смысл решать задачу подбора подходящей функции, если известны значения этой функции в некоторых точках. Именно такой подход, именно такую точку зрения имеет современный научный мир. В медицинской практике это означает, что первым этапом является сбор данных — биометрия, а вторым — построение некоторого

решающего правила путем подбора соответствующих функций. И вот на этом этапе возникает большое число проблем.

Для справедливости отметим, что похожими задачами занимается статистика, однако существенный прогресс при применении этих методов достигнут лишь в некоторых ограниченных случаях, например при анализе *линейных функций*. И это понятно. По канонам современных знаний, в соответствии с теоремой И. Пригожина — живой организм представляется открытой термодинамической системой, в которой нелинейные процессы обмена вещества, энергии и информации составляют значительную часть. Именно этим во многом объясняется столь повышенный интерес к нейронным технологиям анализа информационных потоков, в которых просматривается возможность решения задач с нелинейными функциями, описывающими поведения системы в целом.

Отметим, что такие представления о биологических системах были известны давно. Однако распространение этих взглядов на научную исследовательскую деятельность в предыдущее время было затруднено из-за отсутствия соответствующих технических средств, реализующих сложные и многоэтапные вычислительные процедуры. С развитием технических средств — компьютерной техники, наметился резкий прорыв и в области математики. Дальнейшее прогрессивное развитие этого направления в ближайшее время возлагается на тот момент, когда будут сформулированы общие концептуальные взгляды специалистов из разных смежных областей науки.

Таким образом, рассматриваемая проблема приобретает определенную направленность. Если в прошлом поиск зависимостей среди элементов информационного потока мы осуществляли с помощью собственного мозга – интеллекта, а теперь хотим хотя бы частично автоматизировать этот процесс, то, естественно, возникает идея — смоделировать деятельность мозга в некотором автомате и заставить его работать на наших числовых данных.

Рассмотрим основные принципы построения таких систем, или, применяя выше введенные термины, принципы построения автоматов. Эта тема хорошо известна. Рассмотрим основные особенности биологических нейронов и принципы построения решающего правила. При этом, а это самое главное, не будем забывать, что принципы — принципами, а возможность построения реальных алгоритмов анализа — отдельная и непростая задача!

Вот эти принципы.

1. Основную роль в деятельности нервной системы и мозга животных играют специальные клетки — нейроны, связанные между собой нервными волокнами. Нейроны могут посылать друг другу электрические импульсы — сигналы различной силы и частоты.

2. Нейрон состоит из дендритов (по ним принимаются сигналы от других нейронов), тела нейрона (оно обеспечивает жизнедеятельность всей клетки) и аксона (это длинная нить, по ней нейрон может передавать сигналы другим нейронам). Аксон контактирует с дендритами других нейронов посредством специальных образований — синапсов, которые влияют на силу сигнала. Таким образом, синапсы можно считать входами нейрона.

3. Сигналы, полученные нейроном от нескольких других нейронов одновременно, суммируются. Если сила суммарного сигнала превышает некоторое пороговое значение (важна также длительность сигнала), то нейрон возбуждается, генерирует собственный импульс и передает его по аксону.

Биологическая нейронная теория очень развита и сложна. Чтобы построить математическую модель процессов, происходящих в мозгу, мы вынуждены принять несколько предположений:

1. Будем считать, что каждый нейрон обладает некоторой передаточной функцией, определяющей условия его возбуждения в зависимости от силы полученных сигналов. Предполагается, что передаточные функции не зависят от времени.

2. При прохождении синапса сигнал меняется линейно, т. е. сила сигнала умножается на некоторое число. Это число мы будем называть весом синапса или весом соответствующего входа нейрона.

3. Деятельность нейронов синхронизирована, т. е. время прохождения сигнала от нейрона к нейрону фиксировано и одинаково для всех связей. То же самое относится к времени обработки принятых сигналов. Заметим, что веса синапсов могут меняться со временем — это принципиальная особенность. Именно изменение этих весов отвечает за возможность различной реакции организма на одни и те же условия в разные моменты времени, т. е. возможность обучения.

Нужно признать, что все эти предположения достаточно сильно огрубляют биологическую картину. Например, время передачи сигнала напрямую зависит от расстояния между нейронами и может быть достаточно большим. Тем удивительнее, что при этих огрублениях полученная модель сохраняет некоторые важные свойства биологических систем, в том числе адаптивность и сложное поведение. Построим математическую модель нейрона (далее мы будем называть ее нейроном).

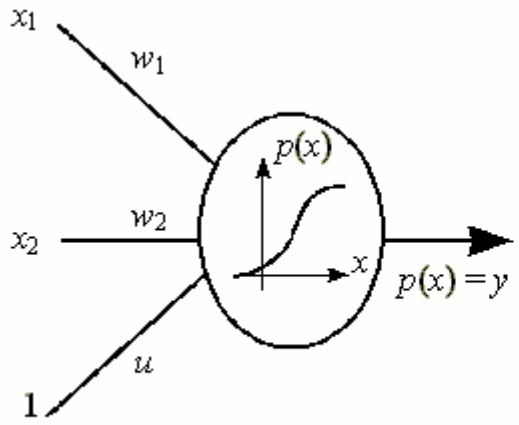


Рис. 3.2. Математический нейрон

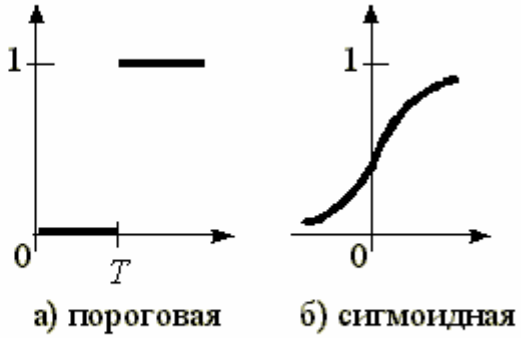


Рис. 3.3. Передаточные функции

С точки зрения математического описания работы такого сложного органического элемента как нейрон, не все так просто. Некоторые первоначальные представления о работе “простейшего” автомата поменялись, другие — получили дополнение и развитие в современных устройствах. Входные сигналы для такого устройства, как правило, представляют в цифровой форме. Бинарный входной сигнал позволяет несколько упростить понимание процессов преобразования информации. Последуем этому правилу и рассмотрим нейрон.

Нейрон — это несложный автомат, преобразующий входные сигналы в выходной сигнал. Сигналы силы $x_1$, $x_2$, ..., $x_n$, поступая на синапсы, преобразуются линейным образом, т. е. к телу нейрона поступают сигналы силы $w_1x_1$, $w_2x_2$, ..., $w_nx_n$ (здесь $w_i$ — веса соответствующих синапсов). Для удобства к нейрону добавляют еще один вход (и еще один вес $u$), считая, что на этот вход всегда подается сигнал силы 1. Нейрон суммирует эти сигналы, затем применяет к сумме некоторую фиксированную функцию $p$ и выдает на выходе сигнал силы $y = p(w_1x_1 + w_2x_2 + ... + w_nx_n + u)$.

Эта модель была предложена Маккалоком и Питтсом еще в 1943 г. При этом использовались пороговые передаточные функции (рис. 3.3,а), и правила формирования выходного сигнала $y$ выглядели особенно просто:

если $w_1x_1 + w_2x_2 + ... + w_nx_n + u > T$, то $y = 1$, иначе $y = 0$.

В 1960 г. на основе таких нейронов Розенблатт построил первый в мире автомат для распознавания изображений букв, который был назван “персептрон” (*perception* — восприятие). Этот автомат имел очень простую однослойную структуру и мог решать только относительно простые (линейные) задачи. С тех пор были изучены и более сложные системы из нейронов, использующие в качестве передаточных любые непрерывные функции.

Одна из наиболее часто используемых передаточных функций называется сигмоидной (или логистической) и задается формулой $s(x) = 1/(1 + e^{-x})$ (см. рис. 3.3,б).

Нейронная сеть — это набор нейронов, определенным образом связанных между собой. В качестве основного примера рассмотрим сеть, которая достаточно проста по структуре и в то же время широко используется для решения прикладных задач, — трехслойный персептрон с $n$ входами и одним выходом.

Как следует из названия, эта сеть состоит из трех слоев, изображенных на рис. 3.4. Собственно нейроны располагаются во втором (скрытом) и в третьем (выходном) слое. Первый слой только передает входные сигналы ко всем $H$ нейронам второго слоя (здесь $H = 4$). Каждый нейрон второго слоя имеет $n$ входов, которым приписаны веса $w_{i1}, w_{i2}, ..., w_{in}$ (для нейрона с номером $i$). Получив входные сигналы, нейрон суммирует их с соответствующими весами, затем применяет к этой сумме передаточную функцию и пересылает результат на один из входов нейрона третьего слоя. В свою очередь, нейрон выходного слоя суммирует полученные от второго слоя сигналы с некоторыми весами $y_i$. Для определенности будем предполагать, что передаточные функции в скрытом слое являются сигмоидными, а в выходном слое используется функция $p(x) = x$, т. е. взвешенная сумма выходов второго слоя и будет ответом сети.

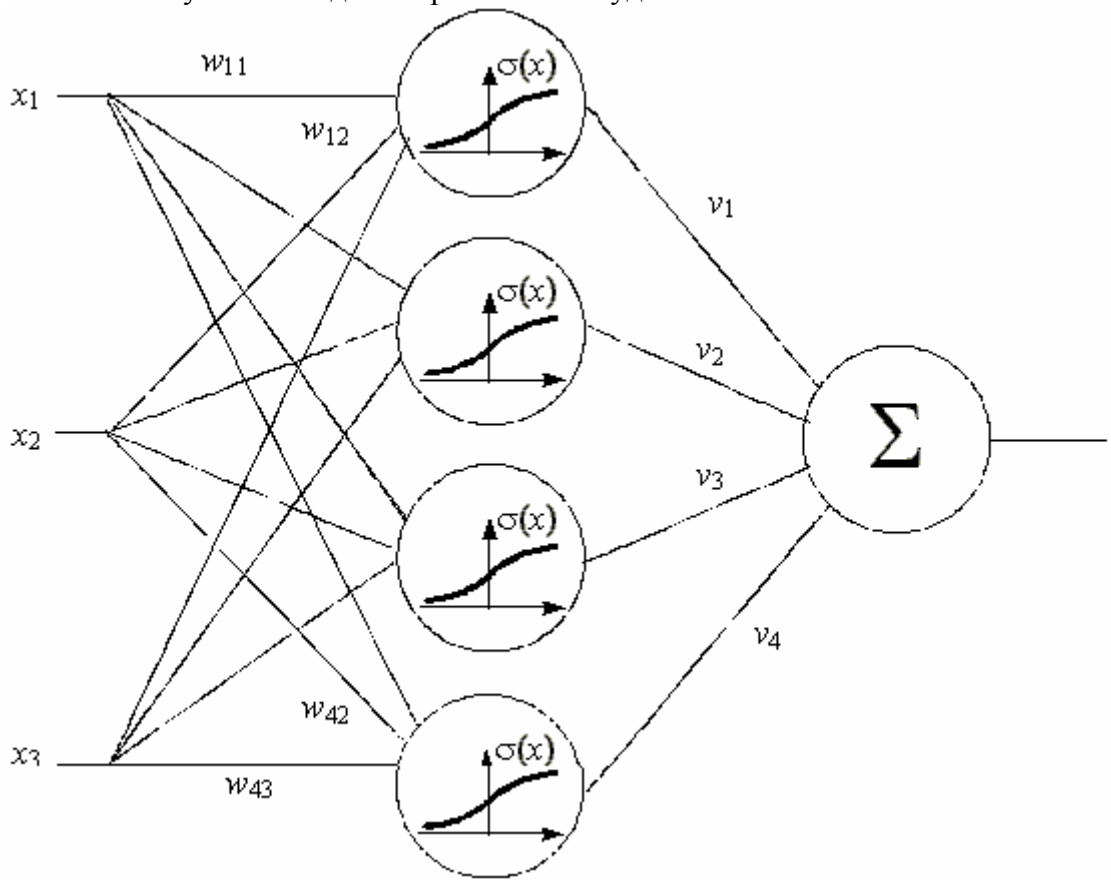


Рис. 4.3. Трёхслойный персептрон

Итак, подавая на входы персептрона любые числа $x_1, x_2, ..., x_n$, мы получим на выходе значение некоторой функции $F(x_1, x_2, ..., x_n)$, которое является ответом (реакцией)

сети. Очевидно, что ответ сети зависит как от входного сигнала, так и от значений ее внутренних параметров — весов нейронов. Выпишем точный вид этой функции:

$$F(x_1, x_2, ..., x_n) = \sum_{i=1}^{H} v_i \cdot \sigma(w_{i1}x_1 + w_{i2}x_2 + ... + w_{in}x_n + u_i).$$

Как видно из этого равенства, такой персептрон реализует только функции определенного вида, а именно суммы значений сигмоидных функций, где в качестве аргументов подставляются линейные комбинации входных сигналов. Например, функцию $F(x_1, x_2) = x_1x_2$ не удается с ходу представить в таком виде.

Естествен вопрос: а может ли персептрон реализовать достаточно сложную функцию? Например, когда в процедуре биометрии фиксируется сигнал, характеризующий изменение значений температуры на поверхности тела человека? Этот вопрос чисто математический — о возможности представить одни функции посредством других. Но и этот вопрос достаточно сложен. На практике решение этого вопроса занимает немало времени и в значительной мере влияет на результат работы всей нейронной сети.

Так как задача очень просто сформулирована, неудивительно, что и занимались ею ученые достаточно долго. Удивительно то, что ответ на нее был получен сравнительно недавно — в 1989 г., но для этого пришлось ввести некоторые упрощения. Это способствовало значительному росту работ в этом направлении, но не внесло достаточного понимания в саму проблему построения нейронной сети.

На практике представляет интерес аппроксимация реальных сигналов. Чтобы получить более значимые для практики результаты в этом направлении, приходится ослабить требования. Во-первых, для нас не принципиально найти точное представление данной функции — достаточно иметь приближенное. Во-вторых, мы можем по необходимости увеличивать число нейронов в скрытом слое, насколько это требуется. Новый вариант теоремы Колмогорова, обладающий этими особенностями, был открыт только в 1989 г. одновременно несколькими авторами.

Для определенности сформулируем теорему в новом представлении. Пусть $F(x_1, x_2, ..., x_n)$ — любая непрерывная функция, определенная на ограниченном множестве, и $e > 0$ — любое сколь угодно малое число, означающее точность аппроксимации. Через “сигму” мы обозначаем сигмоидную функцию, определенную выше.

**Теорема**. Существуют такое число $H$, набор чисел $w_{ij}$, $u_i$ и набор чисел $v_i$, таких, что функция

$$F(x_1, x_2, ..., x_n) = \sum_{i=1}^{H} v_i \cdot \sigma(w_{i1}x_1 + w_{i2}x_2 + ... + w_{in}x_n + u_i).$$

приближает данную функцию $F(x_1, x_2, ..., x_n)$ с погрешностью не более $e$ на всей области определения.

Тем, кто внимательно читал этот текст, легко заметить, что эта формула полностью совпадает с предшествующим выражением для функции, реализуемой персептроном. В терминах теории нейронной сети эта теорема формулируется так. **Любую непрерывную функцию нескольких переменных можно с любой точностью реализовать с помощью обычного трехслойного персептрона с достаточным количеством нейронов в скрытом слое.**

Действительно, это возможно, но, какие использовать методы реализации, как и прежде остается нерешенной задачей. Можно утверждать, что в какой-то мере произвольность выбора решений, именно в данном контексте понимания проблемы, порождает бесконечность научных поисков.

Для справедливости можно отметить, что все не так безнадежно. Использование метода СКА позволяет преодолеть указанные трудности. Для этого надо применить

алгоритм фрагментного представления исходной выборки, а затем построить функционал объединения этих фрагментов.

**Классификация и идентификация**

В задачах медицинской диагностики, использующей новейшие методы построения диагностического правила, часто применяются и типовые решения.

Положим, что имеется метод, обеспечивающий классификацию данных различной природы. Математические основы такой процедуры хорошо известны. Задача классификации, скажем нарушений в живом организме, и задача идентификации нарушений могут быть представлены как сопряженные математические процедуры. Если использовать термин “идентификация” в смысле построения поисковой процедуры, направленной на обнаружение заданного объекта — “нарушения”, “патологии”, личности, то нетрудно увидеть близость этих задач. В таком понимании процедура классификации может отождествляется с “верхним” уровнем, а идентификация — с “нижним”. Действительно, понимая процедуру идентификации как процесс отыскания объекта — личности, обладающей заданным типовым набором показателей состояния организма, мы решаем диагностическую задачу. Результатом такой процедуры является отнесение объекта всего к двум категориям: “норма” или “патология”. Или в терминах идентификационной процедуры личности: “свой” или “чужой”. Нормальное состояние индивидуального организма в терминах идентификационной процедуры означает “Свой”, а патология, с различными проявлениями, которые в последующем будут уточняться, может быть охарактеризована термином “Чужой”.

Принимая во внимание эти понятия, рассмотрим типовую задачу. Положим, имеется некоторая выборка, полученная в процессе измерения величины физического параметра, регистрируемого с поверхности тела человека. В общем случае такая процедура носит название биометрии. Будем использовать основные элементы процедуры построения решающего правила для идентификации личности. Пусть имеется техническая система, посредством которой осуществляется измерение физического параметра.

Будем считать, что система осуществляет измерение вектора $\boldsymbol{v}=(v_1,v_2,...,v_k)$, состоящего из $k$ существенно коррелированных биометрических параметров.

Положим, что личность на этапе идентификации предъявила $N$ своих динамических образов и, соответственно, мы имеем $N$ реализаций векторов $\boldsymbol{v_i}$. Проанализировав имеющиеся реализации векторов биометрических параметров, мы можем найти характерный для личности интервал изменения каждого конкретного параметра $[\min(v_j), \max(v_j)]$.

Если теперь при попадании параметра $v_j$ в интервал $[\min(v_j), \max(v_j)]$ присваивать $e_j = 0$, а при выпадении $v_j$ из интервала $[\min(v_j), \max(v_j)]$ присваивать $e_j = 1$, то мы получим вектор Хемминга. Для “Своего” этот вектор должен состоять практически из одних нулей. Для “Чужого”, предъявляющего иные биометрические параметры, вектор Хемминга будет иметь много несовпадений (много единиц).

Для рассматриваемого случая биометрическим эталоном, зафиксированным при обучении, являются значения минимумов и максимумов измеряемых параметров. Тогда абсолютное значение расстояния Хемминга — $E_x$ до биометрического эталона следует определить как общее число выпадений измерений за интервалы допустимых значений биометрического эталона. Расстояние Хемминга — $E_x$ всегда положительно и может изменяться от 0 до $k$ (где $k$ — это число контролируемых биометрических параметров).

Число примеров при обучении биометрической системы может существенно варьироваться. Как правило, биометрические системы, анализирующие динамические образы, опираясь на меру Хемминга, способны удовлетворительно работать при их обучении на 5...7 примерах, однако для их хорошей работы требуется порядка 20...30

примеров. Использование более 50...60 примеров для обучения системы перестает приносить ощутимые преимущества.

Следует отметить, что задание в биометрическом эталоне интервалов допустимых значений измеряемых параметров может осуществляться различными способами. Используя понятия СКА, формирование интервалов, а точнее говоря, фрагментов, лучше проводить с использованием заранее определенного функционала, который регламентирует построение системы интервалов.

На малых обучающих выборках целесообразно осуществлять прямое вычисление минимума и максимума измеренных значений контролируемых параметров. При объеме обучающей выборки в 5 и более примеров становится целесообразным вычисление математического ожидания значений параметров $m(v_j)$ и их дисперсий $s(v_j)$. В этом случае значение минимальной и максимальной границ принято вычислять следующим образом:

$\min(v_j)=m(v_j) - t(N,(1-P_1))Ч\ s(v_j)$(3.1)

$\max(v_j)=m(v_j) + t(N,(1-P_1))Ч\ s(v_j)$(3.2)

где $N$ — число использованных при обучении примеров, $P_1$ — заданное значение вероятности ошибок первого рода (в этих операциях $P_1$ принимают равным 0,1), $t(N,(1 - P_1))$ — коэффициенты Стьюдента.

При вычислении математического ожидания контролируемого параметра может использоваться формула:

$$m(v_j) \approx \frac{1}{N}\sum_{i=1}^{N} v_{ij}$$ .(3.3)

Известно, что с ростом $n$ оценка (3.3) приближается к точному значению математического ожидания в смысле Ляпунова и Чебышева.

Основным недостатком (3.3) является то, что при обучении приходится помнить значения всех измеренных ранее параметров. Эта проблема усугубляется тем, что в неопределенном будущем может понадобиться дообучение биометрической системе и, следовательно, при использовании (3.3) приходится хранить все данных обучении неопределенно долго. Более удобным для реализации является рекуррентное вычисление математического ожидания:

$$m_i(v_j) \approx \frac{i-1}{i}\cdot m_{i-1}(v_j) + \frac{1}{i}\cdot v_{ji}$$ .(3.4)

При использовании (3.4) приходится помнить только общее число уже использованных примеров и текущее значение математического ожидания. На каждом последующем шаге появляется новое значение математического ожидания и запоминается $i$ — число учтенных примеров.

Аналогичная ситуация возникает и при вычислении дисперсии контролируемых параметров. Если хранятся все значения измеренных параметров, то может быть использована обычная форма вычисления:

$$\sigma^2(v_j) \approx \frac{1}{N-1}\sum_{i=1}^{N}(v_{ij} - m(v_j))^2$$ .(3.5)

При необходимости экономии памяти компьютера и времени вычислений используется рекуррентное вычисление дисперсии:

$$\sigma_i^2(v_j) \approx \frac{i-2}{i-1}\cdot\sigma_{i-1}^2(v_j) + \frac{1}{i-1}(v_{ij} - m(v_j))^2$$ .(3.6)

После того как сформирован биометрический эталон, возможна реализация процедур аутентификации зарегистрированного пользователя. При осуществлении процедур аутентификации “Свой” пользователь достаточно редко ошибается и, соответственно, мера Хемминга оказывается малой. Иначе обстоит дело при попытках

аутентифицироваться “Чужих”. Для “Чужого” ошибки оказываются гораздо более частыми. Эта типовая ситуация иллюстрируется рис. 3.5, где приведен пример гистограмм “Свой” – “Чужой” биометрической системы контроля динамики регистрируемого сигнала.

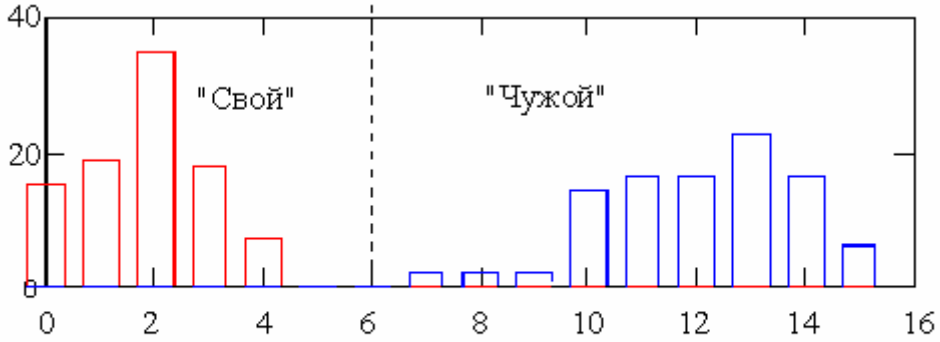


Рис. 3.5. Пример гистограмм распределения значений меры Хемминга

Гистограмма получена для меры Хемминга, построенной на контроле 80 параметров. Из рис. 3.5 видно, что для “Своего” наиболее вероятным отклонением является значение $E_x$=2 и зафиксированное значение меры Хемминга не превосходит 5. Для “Чужого”, пытающегося подстроить динамику сигнала, мера Хемминга принимает значения от 7 и выше. В качестве разделяющего порога областей “Свой” “Чужой” может быть принято непопадание в заданные интервалы 6 контролируемых параметров. Отметим, что здесь понятие “интервала” очень близко совпадает по смыслу с понятием “фрагмент”, которое было рассмотрено в методе СКА.

Этот пример соответствует хорошему разделению областей “Свой” и “Чужой” с достаточно примитивным решающим правилом. Для того чтобы задать порог, разделяющий эти области, необходимо вычислить значения математического ожидания меры Хемминга для области “Свой”, дисперсию для этой же области. Тогда можно воспользоваться следующими соотношениями:

“Свой” если $E_x < m(E_x^{\text{Свой}}) + t\,(N,(1-P_1))\text{Ч}\; s\,(E_x^{\text{Свой}})$,(3.7)

**“**Чужой” если $E_x > m(E_x^{\text{Свой}}) + t\,(N,(1-P_1))\text{Ч}\; s\,(E_x^{\text{Свой}})$,(3.8)

где $t(N,(1-P_1))$ — коэффициенты Стьюдента, $N$ — число примеров при обучении, $P_1$ — заданная вероятность ошибки первого рода для системы в целом (обычно принимается $P_1$ =0,01).

Полученные результаты позволяют увидеть хорошую взаимосвязь задач классификации и идентификации. Одновременно с этим отчетливо наблюдается связанность статистического и структурного координатного анализа.

Следует подчеркнуть, что (3.7), (3.8) корректны только для закона распределения значений меры Хемминга, близкого к нормальному закону. Нормализация этого закона осуществляется при условии достаточно большого числа контролируемых параметров $k$>20 и выборе достаточно узких диапазонов допустимых значений параметров биометрического эталона.

**Вэйвлет-анализ**

Как мы теперь видим, распознавание различных нарушений в организме человека и опознание конкретного человека являются близкими задачами. И в том и другом случае приходится анализировать некоторые массивы данных. Развитие средств телекоммуникации позволяет по-новому взглянуть на решение этих проблем. Предоставляя доступ к информационным ресурсам большому числу абонентов, можно проводить массовые медицинские обследования населения и одновременно заботиться об идентификации абонентов, о конфиденциальности медицинской информации. И в том и

другом случае можно строить диагностические алгоритмы – решения, используя типовые математические процедуры. При этом экспресс-диагностика в сети Интернет позволит значительно понизить требования к вычислительным ресурсам периферийных пользователей, не понижая качество диагностических заключений.

Рассмотрим одно из перспективных направлений построения диагностических решений на основе так называемого Вэйвлет-анализа данных. Отметим, что использование этого метода впервые было апробировано в США в задаче идентификации личности по графическому образу — отпечатку пальца. Столкнувшись с необходимостью хранить отпечатки пальцев 30 млн человек по 600 килобайт на каждый палец или 6 мегабайт на запись, государство должно было потратить 200 млн долларов только на хранение информации. Сотрудники Лос-Аламосской лаборатории предложили использовать для сжатия отпечатков вэйвлет-преобразование.

Вэйвлеты (*wavelet*) и вэйвлет-преобразование — это новый способ обработки и исследования сигналов, теория которого разработана совсем недавно, с появлением быстродействующих компьютеров, так как требует большого объема вычислений. Вэйвлет — это в дословном переводе “маленькая волна” (рис. 3.6). За основу обычно берется один из простейших графиков. Видно, что график быстро затухает по краям.

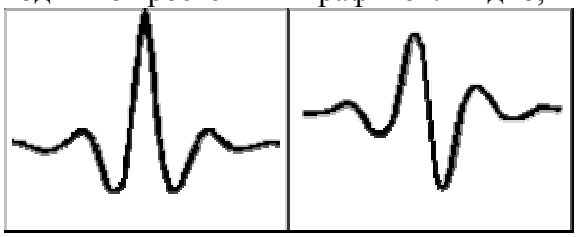

Рис. 3.6. Вэйвлет — “маленькая волна”

Вэйвлет можно считать трехмерным спектром, где по оси $X$ — время, по оси $Y$ — частота, а по оси $Z$ — амплитуда гармоники с данной частотой в данный момент времени (рис. 3.7).

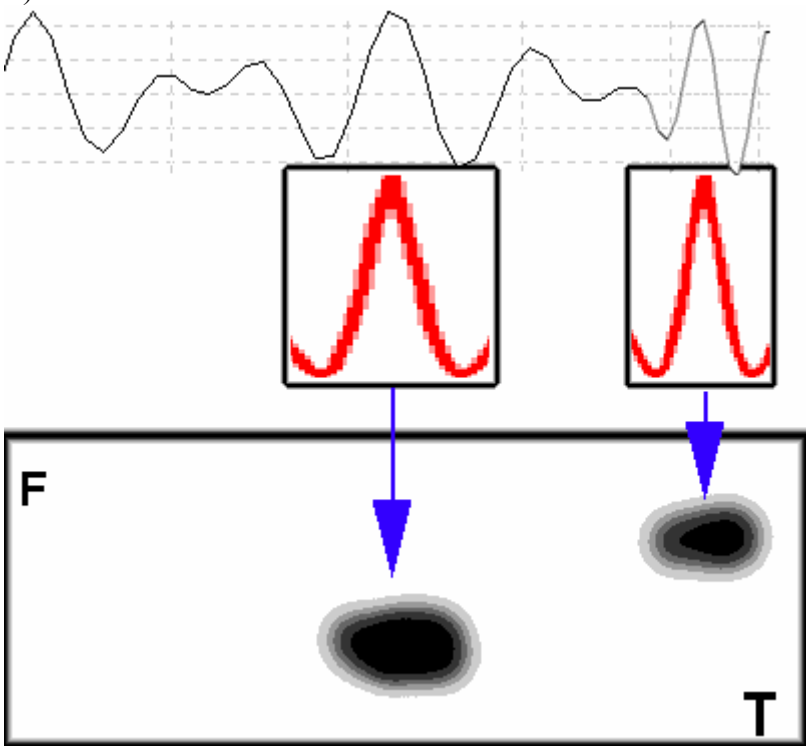


Рис. 3.7. Результат Вэйвлет- анализа

Обычно на двухмерной плоскости (на экране, на бумаге) ось $Z$ отображают в виде градаций черного цвета. При этом черный цвет на графике характеризует максимальную, а белый — минимальную амплитуду сигнала. Тогда самые интересные места сразу видны

по черным пятнам Далее этот вэйвлет прикладывается к сигналу (делается свертка), при этом его можно растягивать (т. е. менять частоту) и передвигать по временной оси (т. е. менять время). Мы получаем двухмерный массив амплитуд в зависимости от частоты и времени.

Реально мы получаем комплексный массив (как и в спектре), но на рисунках изображены только амплитуды. Фазовая картинка выглядит очень красиво, но требует дополнительных пояснений. Существует алгоритм быстрого вэйвлет-преобразования, подобного Быстрому Преобразованию Фурье (БПФ) для спектров, время вычисления для которого значительно меньше. Также есть обратное вэйвлет-преобразование для восстановления формы сигнала.

Вэйвлет-анализ разработан для решения задач, оказавшихся “очень сложными” для традиционного анализа Фурье, и находит все более широкое применение в исследовании и прогнозе временных рядов, например медицинской диагностики. Это может быть запись ЭКГ сигнала или рентгеновское изображение одного из органов. Вэйвлет-анализ применяется при обработке данных медицинского обследования, сжатии изображений и распознавании образов и речи, задачах связи, теоретической биофизике и математике.

В простом изложении вэйвлет-анализ представляется обычной процедурой, понять которую несложно. Хорошо известно, что любой электрический сигнал можно разложить в сумму гармоник (синусоид) разной частоты. Но синусоидальное описание волны обеспечивается гладкими функциями, которые не очень-то отслеживают мельчайшие изменения сигнала во времени. Чтобы уловить эти изменения, вместо синусоидальных функций (волн) можно взять короткие “всплески” — совершенно одинаковые, но разнесенные по времени. Оказывается, этого недостаточно: надо добавить еще их всевозможные растянутые и сжатые копии. Вот теперь сигнал можно разложить в сумму таких всплесков разного размера и местоположения. Это и есть вэйвлет-анализ. По своему содержанию такой анализ имеет много общего с классическим Фурье анализом. Но он имеет явные отличия, которые выражаются в некоторой произвольной процедуре выбора “всплесков”. Однако если воспользоваться понятием фрагмента, которое используется в СКА, то “всплески” можно выбирать уже целенаправленно, а значит, строить диагностическое заключение более совершенно.

Коэффициенты разложения — важная информация об эволюции сигнала. Они зависят от выбора изначального всплеска. Для каждой прикладной задачи можно подобрать наиболее приспособленный (именно для нее) всплеск. Он как раз и называется Вэйвлетом.

Математическая сторона вэйвлет-анализа — вещь довольно тонкая, можно сказать сложная, хотя и весьма наглядная. А вот прикладная сторона вэйвлетов очень простая. Применение вэйвлетов для исследования биомедицинских сигналов, которые имеют развертку во времени или меняют свою частоту, началось сравнительно недавно. Например, медицинская диагностика — изучение скорости кровотока, скорости пульсовой волны или частоты дыхания активно использует алгоритмы вэйвлет-анализа.

Мода на использование дискретного Wavele*t*-преобразования (DWT или иначе WT) в задачах биометрии условно начинается с середины 90-х гг. Следует отметить, что *WT* не является “полноценным” преобразованием сигнала из временной в частотную область. Чтобы было легче представить, что это такое, отметим, что результат похож на фильтрацию в октавных полосах частот. Не рассматривая детали этой процедуры, отметим, что фактически здесь также приходится вводить понятие фрагмента — октавы. Такое понятие опять позволяет связать уже известные представления СКА и Wavelet-преобразования. Связь между частотной полосой осуществляется через параметр масштаба, который влияет на сжатие/растяжение базовой вэйвлет-функции. Количество частотных полос для дискретного WT определяется как log2($N$), где $N$=2$m$ опять размер блока в элементах, определяющих размер фрагмента или октавы. При $N$=2048 получим 11 частотных полос. Самая высокочастотная полоса простирается от 1/4 до 1/2 от частоты

дискретизации. Предыдущая частотная полоса — от 1/4 до 1/8 от частоты дискретизации и т. д. до самых низких частот.

В первом приближении можно говорить о WT как о полосовых фильтрах. Однако это сходство очень обманчиво, поскольку *WT* осуществляется принципиально иначе по сравнению с той же цифровой фильтрацией. Массив “однополосного” результата WT часто называется коэффициентами WT. Здесь следует пояснить, что для фильтрации сигналов с помощью WT просто обнуляются коэффициенты, соответствующие частотной полосе, в которой присутствуют нежелательные компоненты, а затем выполняется обратное преобразование, что дает отфильтрованный сигнал во времени.

Окно фильтрации как бы сжимается во времени при переходе к более высоким частотам. Преимущества WT проявляются в спектральном анализе нестационарных сигналов. Используя WT, можно получить импульсные функции отклика на разных частотах. В отличие от традиционных аналоговых и цифровых фильтров, WT имеет короткий по времени переходной процесс на импульсные воздействия. Математический аппарат WT не так прост, как может показаться на первый взгляд, а настройка параметров еще более трудоемка, чем для БПФ. Изменив тип базового вэйвлета, можно получить иной результат в зависимости от специфики сигнала. Именно поэтому интерпретация результатов *WT* доступна опытным экспертам и в настоящее время является самостоятельной задачей при выборе этого направления.

Анализ данных посредством WT может быть сравним с обычными классическими статистическими методами. Для получения спектральной оценки первоначально вычисляют среднеквадратичное значение по каждой группе коэффициентов, в противном случае результаты WT следует интерпретировать как подобие спектрограммы. Таким образом, само по себе WT — скорее масштабно-временное преобразование сигнала с разложением в различных частотных полосах, нежели преобразование из временной области в частотную область представления сигнала, как в случае БПФ.

Разложение исходного сигнала осуществляется обычно с помощью вэйвлет-функций, вид которых определяется второй производной от функции распределения Гаусса. Впрочем, возможны и другие функции, например, обладающие фрактальными свойствами. В этом смысле WT-анализ близок СКА. Возможность использования фрактального представления данных WT-анализа сегодня представляется интересным направлением. Однако переход от WT-анализа, результатом которого является вычисление некоторых показателей исследуемого процесса, к построению фрактальных кривых осуществляется достаточно сложно. Реализация такого перехода очень часто сопровождается использованием эвристических алгоритмов, которые затрудняют интерпретацию результата, но порождает интересные графические образы.