

**МИНИСТЕРСТВО
ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ имени С.П. КОРОЛЕВА**

Киселев Ю.В.

**ОСНОВЫ ТЕОРИИ ТЕХНИЧЕСКОЙ
ДИАГНОСТИКИ**

**САМАРА
2004**

Киселев Ю.В.

Основы теории технической диагностики. - Самара : СГАУ, 2004. - 138 с. : Ил.42, табл. 20, список лит.: 42 наименования.

В данном учебном пособии рассматривается комплекс вопросов, связанных с теоретическими основами технической диагностики.

Дается трактовка основных понятий, терминов и определений, используемых в технической диагностике, описывается классификация методов диагностики и приводятся общие сведения о них.

Значительное место отведено математическим (диагностическим) моделям, которые используются для формализованного описания объектов диагностики и протекающих в них процессов.

В пособии нашли отражение вопросы, связанные с выбором диагностических признаков и оценкой их диагностической ценности.

Приведены сведения об основных математических методах распознавания технических состояний (диагнозов) объектов. Изложены принципы решения прогностических задач диагностирования.

В настоящее время особое внимание уделяется сбору диагностической информации непосредственно в полете, что и определило включение в пособие сведений о бортовых устройствах регистрации параметров.

Пособие предназначено для студентов, обучающихся по специальностям 130300 и 131000.

О Г Л А В Л Е Н И Е

ВВЕДЕНИЕ.....	3
1. ОСНОВЫ ТЕХНИЧЕСКОЙ ДИАГНОСТИКИ	4
1.1. Основные понятия, термины и определения	4
1.2. Место и роль технической диагностики в системе технической эксплуатации авиационной техники.....	7
1.3. Общие сведения о методах диагностики и их классификация.....	11
2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ В ТЕХНИЧЕСКОЙ ДИАГНОСТИКЕ.....	21
2.1. Основные понятия.....	21
2.2. Функциональные модели.....	25
2.3. Модели на основе графов причинно-следственных связей.....	31
2.4. Методы оптимизации диагностических тестов.....	37
2.5. Аналитические модели в виде дифференциальных уравнений.....	43
2.6. Аналитические модели на основе алгебраических уравнений.....	50
3. ДИАГНОСТИЧЕСКАЯ ЦЕННОСТЬ ПРИЗНАКОВ.....	61
3.1. Энтропия системы.....	62
3.2. Определение количества информации.....	65
3.3. Диагностический вес признака.....	69
4. МЕТОДЫ РАСПОЗНАВАНИЯ В ТЕХНИЧЕСКОЙ ДИАГНОСТИКЕ.....	74
4.1. Основные понятия.....	74
4.2. Метод Байеса.....	76
4.3. Методы принятия статистических решений.....	83
4.3. Методы распознавания в пространстве параметров.....	98
5. ОСНОВЫ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЯ АВИАЦИОННОЙ ТЕХНИКИ.....	112
5.1. Основные положения.....	112
5.2. Порядок обработки результатов измерений при прогнозировании.....	120

6. БОРТОВЫЕ УСТРОЙСТВА РЕГИСТРАЦИИ.....	122
6.1. Назначение бортовых устройств регистрации.....	122
6.2. Характеристики и устройство бортовых регистраторов.....	125
6.3. Краткие сведения о наземных средствах обработки полетной информации.....	130
 СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	 134

ВВЕДЕНИЕ

Техническая диагностика это молодая наука, которая возникла в связи с возрастающей ролью в народном хозяйстве сложных и дорогостоящих технических систем и предъявлении к ним повышенных требований по безопасности, безотказности и долговечности. Методы технической диагностики направлены на практическую реализацию этих требований путем определения технического состояния объектов и разработке оперативных рекомендаций по их дальнейшей эксплуатации.

Требование о необходимости иметь систему контроля технического состояния (ТС) воздушного судна содержится в таком основополагающем документе как Авиационные правила АП-25 «Нормы летной годности самолетов транспортной категории». Согласно этим требованиям, для систем воздушных судов, отказ которых может служить непосредственной причиной возникновения опасной ситуации в полете, должны быть предусмотрены контроль и диагностика их ТС.

Все более широкое внедрение в практику методов и средств технической диагностики предполагают наличие в эксплуатационных предприятиях инженерных кадров, имеющих соответствующую подготовку. Поэтому знание основ технической диагностики является обязательным требованием к специалистам, занимающимся эксплуатацией сложных технических систем, в том числе и авиационной техникой.

В данном пособии изложены теоретические основы технической диагностики. В нем нашли отражение следующие вопросы.

Первоначально дается трактовка основных понятий, терминов и определений технической диагностики, обсуждаются её роль и место в системе технической эксплуатации авиационной техники, рассматривается общая схема постановки и решения диагностических задач, описывается классификация методов диагностики и приводятся общие сведения о них.

Значительное место отведено математическим (диагностическим) моделям, которые используются при решении задач формализованного описания объектов диагностики и формировании описаний текущего или перспективного диагнозов.

В пособии нашли отражение вопросы, связанные с выбором диагностических признаков и оценкой их диагностической ценностью.

Большое внимание уделено математическим методам распознавания технических состояний (диагнозов) объектов. Рассмотрены метод Байеса, методы принятия статистических решений, методы распознавания в пространстве диагностических признаков. Изложены принципы решения прогностических задач диагностирования.

В настоящее время особое внимание уделяется сбору диагностической информации непосредственно в полете, что и определило включение в пособие сведений о бортовых устройствах регистрации параметров.

Пособие предназначено для студентов, обучающихся по специальностям 130300 и 131000.

1. ОСНОВЫ ТЕХНИЧЕСКОЙ ДИАГНОСТИКИ

1.1 Основные понятия, термины и определения

В технической диагностике используются понятия, термины и определения, значения которых установлены ГОСТами [ГОСТ 20911-89, ГОСТ 27.002-89, ГОСТ 27518-87]. Кроме того, имеется ряд терминов и понятий, которые не вошли в ГОСТы, но используются в научно-технической литературе [37]. Ниже приведены наиболее употребляемые термины и определения.

Техническая диагностика (ТД) - область знаний, охватывающая теорию, методы и средства определения технического состояния объекта.

В этом определении ключевым понятием является понятие «техническое состояние». В соответствии с новой редакцией ГОСТа 20911 это понятие определяется следующим образом.

Техническое состояние (ТС) объекта - это состояние, которое характеризуется в определенный момент времени, при определенных условиях внешней среды, значением параметров, установленных технической документацией на объект.

Недостатком этого определения является то, что понятие «состояние» определяется через одноименное понятие. Поэтому целесообразно напомнить определение ТС, которое дается в старой редакции этого ГОСТа.

Техническое состояние объекта - совокупность подверженных изменению в процессе производства или эксплуатации свойств объекта, характеризуемая в определенный момент времени параметрами, установленными технической документацией на этот объект.

Процесс определения технического состояния объекта авиационной техники называется *техническим диагностированием* или просто *диагностированием*.

В зависимости от того, для какого времени проводится диагностирование различают три типа диагностических задач. Первый тип это задачи определения технического состояния объекта в данный момент времени, которые называются *задачами диагностирования*. Второй тип это задачи предсказания технического состояния, в котором окажется объект в некоторый будущий момент времени - *задачи прогнозирования*. Третий тип это задачи определения технического состояния, в котором находился объект в некоторый момент времени в прошлом - *задачи генезиса*. Первый тип задач важен при выполнении технического обслуживания и принятии решения о дальнейшей эксплуатации. Второй тип задач важен для предсказания (прогнозирования) остаточного ресурса. Третий тип - для расследования авиационных происшествий.

Объект технического диагностирования это изделие и (или) его составные части подлежащие (подвергаемые) диагностированию. То есть имеется материальный объект, у которого определяется техническое состояние и, в результате, устанавливается *технический диагноз*. Для постановки диагноза необходимо оценить соответствие параметров состояния требованиям нормативно-технической документации.

Параметр состояния это величина, количественно характеризующая одно из основных свойств объекта или процесса, протекающего в объекте. В качестве параметров состояния могут приниматься масса, коэффициент трения, геометрические размеры, зазоры, электрическое сопротивление и т. п. Эти параметры еще называют *первичными*. Часто на практике невозможно произвести непосредственную оценку (измерение) параметров состояния.

В технической диагностике вводится понятие *диагностических параметров* (ДП), под которыми понимаются параметры объекта, используемые в процессе диагностирования. В качестве ДП могут использоваться как параметры технического состояния (первичные), так и параметры, которые характеризуют различные процессы, протекающие в объекте диагностирования, и лишь косвенным образом связанные с параметрами состояния. Последние в технической литературе иногда называют вторичными параметрами.

Таким образом, оценка технического состояния выполняется на основе информации о величинах диагностических параметров. Следовательно, должна быть известна связь между диагностическими параметрами и параметрами состояния. Указанная связь устанавливается с помощью так называемых математических (диагностических) моделей.

Математическая модель - формализованное описание объекта, необходимое для решения задач диагностирования.

Математическое описание (модель) объекта диагностирования может быть выполнена как с использованием диагностических параметров, так и с использованием *диагностических признаков*. Отличие этих двух понятий заключается в следующем. Диагностические признаки формируются (выбираются) на основе диагностических параметров, они образуют дискретное множество, а появление их конкретных значений непосредственно связано с нахождением объекта диагностики в соответствующем *классе технического состояния (диагнозе)*. Например, пусть диагностический параметр выражает температуру газов за турбиной авиационного двигателя. Распределение этого параметра непрерывно. Пусть определены три класса технического состояния двигателя, которые характеризуются соответственно пониженной ($< 450\text{ }^{\circ}\text{C}$), нормальной ($450 - 600\text{ }^{\circ}\text{C}$) и повышенной ($> 600\text{ }^{\circ}\text{C}$) температурой. Попадание значения температуры газов в один из этих интервалов и есть появление диагностического признака, соответствующего двигателю с пониженной, нормальной или повышенной температурой. Отметим, что принципиальных отличий при описании объекта с помощью параметров или признаков нет.

Назначение математических (диагностических) моделей (ММ) не только в установлении связи между параметрами состояния и диагностическими параметрами. ММ позволяют составлять алгоритмы технического диагностирования и выбирать диагностическое обеспечение.

Алгоритм ТД - совокупность предписаний, определяющих последовательность действий при проведении диагностирования.

Диагностическим обеспечением (ДО) называется комплекс взаимосвязанных правил, методов, алгоритмов и средств, необходимых для осуществления диагностирования на всех этапах жизненного цикла изделия.

Задачи диагностирования, на заключительном этапе, неразрывно связаны с *задачами классификации*, поскольку по имеющейся информации необходимо установить диагноз, то есть указать класс технического состояния, к которому относится наш объект диагностирования. Для классификации необходимо знать классы (диагнозы), которые устанавливаются до начала диагностирования на основе анализа самого объекта, его функций и отказов, возникающих в нем. В принципе, количество классов может быть бесконечно. Но существуют достаточно общие принципы классификации, известные из теории надежности. В теории надежности введены понятия для четырех видов технического состояния:

1. Работоспособное; 2. Неработоспособное;
3. Исправное; 4. Неисправное.

В технической диагностике вводится понятие о еще двух видах технического состояния [37]:

5. Состояние правильного функционирования;
6. Состояние неправильного функционирования.

Состояние правильного функционирования - означает, что объект в текущий момент времени выполняет предписанный ему алгоритм функционирования.

Состояние неправильного функционирования - означает, что объект в текущий момент времени не выполняет предписанный ему алгоритм функционирования.

Если в результате процесса диагностирования определяется один из шести видов ТС, то такая операция называется операцией контроля ТС.

Контроль ТС - это проверка соответствия значений параметров объекта требованиям технической документации и определение на этой основе одного из заданных видов технического состояния. Например, если в результате оценки технического состояния делается заключение о том, что объект неисправен.

Термин *техническое диагностирование* применяется тогда, когда основной задачей процесса диагностирования является поиск места и определения причины отказа, неисправности или неправильного функционирования.

Существует еще ряд терминов, о которых следует упомянуть.

Метод контроля (диагностики) - правила применения к объекту контроля (диагностики) определенных принципов и средств контроля (диагностики). Выбор метода контроля или диагностики базируется на анализе физических особенностей протекания рабочих процессов и развития неисправностей в диагностируемом объекте.

Средства технической диагностики (контроля) - аппаратура и программное обеспечение, с помощью которых осуществляется определение величины диагностических параметров, их обработка и разделение объектов на классы.

Система технической диагностики (контроля) - совокупность средств, объекта и исполнителей, необходимая для проведения диагностирования (контроля) по правилам, установленным в нормативно-технической документации.

Существует *тестовое ТД* - это диагностирование, при котором на объект подается тестовое воздействие.

Рабочее (функциональное) ТД - это диагностирование, когда на объект подается только рабочее воздействие, то есть диагностирование осуществляется в процессе непосредственного использования объекта по назначению.

1.2 Место и роль технической диагностики в системе технической эксплуатации авиационной техники

В технической эксплуатации авиационной техники существует такое понятие как стратегия технического обслуживания (ТО) [34]. В соответствии с выбранной стратегией ТО определяются принципы и правила назначения сроков и объемов ТО.

В соответствии с действующей нормативно технической документацией различают следующие стратегии:

1. ТО по наработке, при котором перечень и периодичность выполнения операций по обслуживанию определяется значением наработки изделия.
2. ТО по состоянию, при котором перечень и периодичность выполнения операций по обслуживанию определяется фактическим техническим состоянием изделия. Различают ТО по состоянию с контролем параметров и ТО по состоянию с контролем уровня надежности.

Реализация любой стратегии предполагает использование систем контроля и диагностирования. При ТО по наработке параметром, который необходимо контролировать, является наработка. То есть требуются счетчики наработки. Кроме того, при периодическом ТО, выполняемом в соответствии с данной стратегией через определенные интервалы времени, проводится дефектация элементов и систем авиационной техники. Дефектация проводится с помощью средств ТД.

При ТО с контролем параметров нужны средства измерения параметров и оценки их соответствия нормативным значениям. То есть в этом случае напрямую используются подходы, характерные для технической диагностики.

При ТО с контролем уровня надежности необходим первичный материал об отказах для расчета характеристик надежности. Этот материал невозможно получить без регулярного диагностирования технического состояния.

Следовательно, системы контроля и диагностики в современных условиях являются неотъемлемой частью технической эксплуатации авиационной техники при любой стратегии ТО [41]. Поэтому в разработанных в настоящее время сертификационных требованиях к организациям, занимающимся технической эксплуатацией и ремонтом авиационной техники, указывается на необходимость наличия в их составе подразделения, занимающегося технической диагностикой.

Кроме того, эффективность ТО напрямую зависит от эффективности систем контроля и диагностирования. Известно, что 90% времени при ТО занимает процесс обнаружения неисправностей. Чем выше эффективность средств диагностики, тем быстрее проводится ТО, тем больше можно обеспечить

годовой налет. Таким образом, техническая диагностика является важнейшей составляющей системы технической эксплуатации современных воздушных судов.

Средствами технической диагностики и контроля на всех этапах эксплуатации авиационной техники (в полете, при оперативном ТО, при периодическом ТО, при расследовании авиационных происшествий и др.) решаются следующие задачи (рис. 1.1).

ЭКСПЛУАТАЦИОННЫЕ ЗАДАЧИ ТЕХНИЧЕСКОЙ ДИАГНОСТИКИ

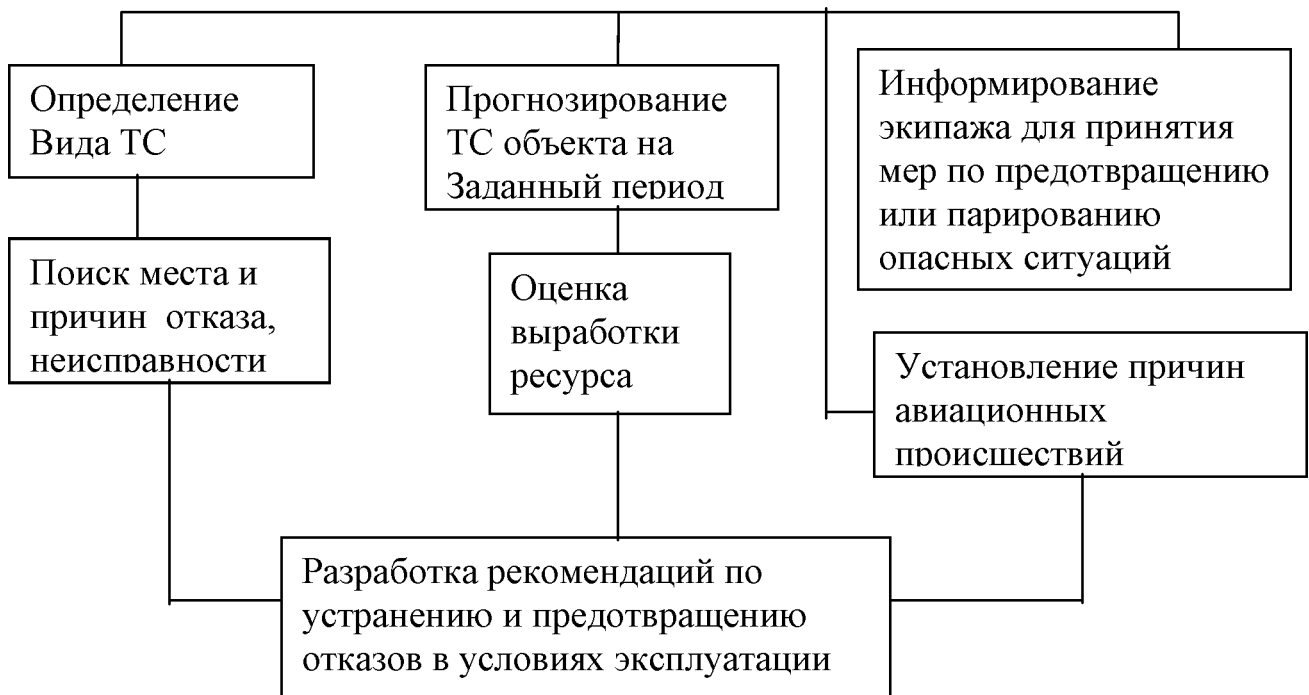


Рис. 1.1 Задачи технической диагностики

Перечисленные задачи тесно взаимосвязаны и их решение требует привлечения большого количества высококвалифицированных специалистов из различных областей науки и техники.

Внедрение в практику систем технической диагностики и контроля позволяет получить как технико-экономический, так и социальный эффекты.

Технико-экономический эффект достигается за счет:

- сокращения трудоемкости и времени на ТО за счет более быстрого выявления изделий требующих замены или ремонта;
- обнаружения развивающихся неисправностей, запрещение к вылету ВС с такими неисправностями, что направлено на предотвращение возвратов с линии, предотвращение простоев в промежуточных портах, предотвращение потерь ВС;
- более полного использования ресурса каждого изделий в следствие регулярной оценки его фактического состояния и возможности применения методов эксплуатации по состоянию;

- экономия ГСМ и ресурса за счет сокращения количества и времени наземного опробования авиационной техники.

Социальный эффект достигается за счет:

- повышением безопасности полета;
- повышением регулярности полетов;
- внедрением систем контроля и диагностики способствуют повышению квалификации ИТР.

Внедрение систем контроля и диагностики способствует повышению общей культуры технической эксплуатации, поскольку это внедрение требует выполнения ряда условий.

Во первых, авиационная техника должна обладать достаточной контролепригодностью [34, 41]. Под *контролепригодностью* понимается свойство изделия АТ, характеризующее его приспособленность к измерению параметров самого изделия и его составных частей с целью решения задач контроля и диагностики технического состояния АТ. В настоящее время разработана нормативно-техническая документация, задающая требования к контролепригодности АТ и методы ее определения (ГОСТы 26656-85, В200436-88, В20570-88, В23743-88).

Следующим условием внедрения систем контроля и диагностики является наличие методик и алгоритмов, включающие указания на периодичность контроля, данные о предельных значениях параметров, а также об алгоритмах принятия решения по эксплуатации АТ в соответствии с результатами контроля и диагностики.

Необходимым условием внедрения систем контроля является также налаженное производство средств диагностики и контроля, обеспеченность эксплуатационных предприятий этими средствами и соответствующей документацией.

Важнейшим условием внедрения является наличие в эксплуатирующих предприятиях подготовленных кадров.

Учитывая важность и актуальность задач, стоящих перед системами контроля и диагностики, сложность их внедрения, к этим системам предъявляется ряд специфических требований [41]:

1.Комплектность:

- система контроля и диагностирования должна охватывать как можно большее количество систем ВС;
- система должна обеспечивать одновременное выполнение функций, связанных с формированием решений по ТО, с выдачей команд экипажу, с обменом информацией с наземными средствами и с использованием при расследовании авиационных происшествий.

2.Возможность адаптации к различным типам ВС и к их модификациям.

3.Автоматизация обработки информации на борту и на земле.

4. Оптимизация распределения функций между бортовыми и наземными системами диагностики и контроля.

5. Возможность накопления и хранения информации.

6.Наличие собственного автоматизированного контроля этих систем.

Процесс внедрения систем диагностики от постановки задачи и до получения первых результатов практического использования можно разбить на три этапа:

I - этап физического анализа;

II - этап математического анализа;

III - этап технической реализации.

Каждый из этапов имеет свою специфику. Физический этап предполагает знание объекта и физической сущности процессов, протекающих в нем. Математический этап требует знания формализованных методов построения диагностических моделей, формирования диагностических признаков, разработки алгоритмов классификации технических состояний. Этап технической реализации (технический этап) требует выбора аппаратурных и методических решений. Перечень наиболее характерных для каждого этапа задач представлен в таблице 1.2.

Таблица 1.1. Общая схема решения задач ТД

I. Физический этап	II. Математический	III. Технический
1. Определение объекта контроля и диагностики. 2. Анализ существующих и возможных отказов и неисправностей. 3. Классификация отказов и неисправностей по степени опасности. 4. Определение отказов и неисправностей, подлежащих диагностированию. 5. Составление логической или словесной модели, устанавливающей связь отказов с параметрами изделия. 6. Определение диагностических параметров, подлежащих измерению. 7. Выбор метод диагностирования по его физической сущности. 8. Определение контроле-пригодности изделия.	1. Разработка диагностической модели, устанавливающей связь между параметрами технического состояния и измеряемыми диагностическими параметрами. 2. Выбор или разработка диагностических признаков. 3. Разработка методик и алгоритмов диагностирования (классификации). 4. Разработка требований к диагностической и контрольной аппаратуре.	1. Выбор или создание средств ТД и контроля. 2. Разработка метрологического, методического и информационного обеспечения системы ТД.

Несмотря на различие эти три этапа существенно зависят друг от друга. Весь ход решения диагностической задача определяется физической сущность объекта и выбором метода диагностики.

В настоящее время для оценки технического состояния авиационной техники используется широкий спектр методов диагностики. Эти методы весьма разнообразны и сильно отличаются друг от друга физической сущностью и своими характеристиками. Знакомство с методами диагностики будем проводить с учетом существующей в настоящее время системой их классификации.

1.3. Общие сведения о методах диагностики и их классификация

Классификация методов диагностики проводится на основе различных принципов [19, 24, 26, 37]. О некоторых из них говорилось выше. Например, при классификации по глубине выявления отказа или неисправности различают контроль технического состояния (работоспособности) и диагностирование (поиск места и причины неисправности). При классификации по воздействию на объект различают тестовое и функциональное (рабочее) диагностирование.

В зависимости от физической природы диагностических параметров и способа их измерения различают физические и параметрические методы диагностирования [37].

Физические методы основаны на использовании различных физических явлений, сопутствующих работоспособному или неработоспособному состоянию объекта.

Физические методы различают на две группы:

1. Диагностирование и контроль объектов в нерабочем состоянии. С помощью этих методов, как правило, обнаруживают наличие (отсутствие) скрытых механических повреждений и дефектов (например: сквозных и несквозных микротрещин, внутренних раковин, надломов и т. д.) в материалах, из которых изготовлены объекты. Эти методы в силу своей специфики еще называют методами неразрушающего контроля (МНК).
2. Диагностирование и контроль объектов в рабочем состоянии. Они основаны на анализе характеристик (параметров) физических явлений, сопровождающих работу механизма, но не связанных на прямую с его функциональным назначением. Эти методы обеспечивают, например, выявление повреждений и износов в сопряжениях подвижных деталей механизмов.

Обзор методов диагностирования начнем с физических методов неразрушающего контроля, поскольку эти методы наиболее широко используются в эксплуатации авиационной техники [15, 19, 27, 28, 30, 38, 39, 42]. Дадим краткую характеристику этим методам.

Опτικο-визуальный метод контроля основан на применении оптических приборов для осмотра наружных и внутренних поверхностей деталей и агрегатов. Этим методом обнаруживают следующие дефекты: коррозия, раковины, трещины, вмятины, местный наклеп, подгары, окисления, электронная эрозия деталей

электрических узлов и т.д. Техническими средствами, которые используются для реализации этого метода являются:

- оптические лупы (налобные, телескопические, обзорные);
- механические управляемые зеркала (с изменяемым углом наклона);
- механические эндоскопы (жесткие и гибкие).

С помощью эндоскопов производится осмотр внутренних поверхностей емкостей, трубопроводов, проточной части авиационных двигателей.

Капиллярные методы предназначены для обнаружения поверхностных дефектов и основаны на проникновении специальных индикаторных жидкостей в несплошности объектов контроля (трубопроводы, корпусные детали, лопатки компрессоров и турбин авиационных двигателей и т. п.). Сущность метода заключается в следующем:

- контролируемая поверхность детали очищается от покрытий или загрязнения и обезжиривается;
- на обработанную поверхность наносят слой индикаторной жидкости, которая проникает в полость дефекта;
- через некоторое время избыток индикаторной жидкости удаляется;
- на поверхность наносят специальный проявитель;
- проявитель обладает свойством вытягивать индикаторную жидкости из полости дефекта (индикаторная жидкость сорбируется проявителем);
- в результате на поверхности возникает рисунок дефекта.

В зависимости от применяемой индикаторной жидкости или проявителя повреждение видно невооруженным глазом или под воздействием ультрафиолетового излучения.

Метод обеспечивает обнаружение поверхностных трещин глубиной от 0,01мм и шириной раскрытия от 0,001мм.

Метод магнитного контроля применяется для обнаружения повреждений в деталях из ферромагнитных сплавов. Это наиболее распространенный метод, суть которого в следующем. Деталь, которая подвергается дефектации, намагничивается. Если на пути магнитных силовых линий детали имеется дефект (трещина, раковина, немагнитное включение), то в этом месте силовые линии искривятся, огибая место дефекта. На поверхности детали создаются локальные магнитные полюса с соответствующими полями рассеивания. Это локальное рассеивание магнитного поля над дефектом можно обнаружить с помощью магнитного порошка. Искривление магнитных силовых линий вызывает соответствующую концентрацию порошка в месте расположения дефекта. При этом образуется рисунок дефекта - магнитограмма. Метод позволяет выявить трещину глубиной от 0,005мм и шириной раскрытия от 0,0001мм.

Метод вихревых токов позволяет обнаружить повреждение в деталях из магнитных и немагнитных токопроводящих материалов. Повреждения могут обнаруживаться под слоем окислов и лакокрасочных покрытий. Глубина

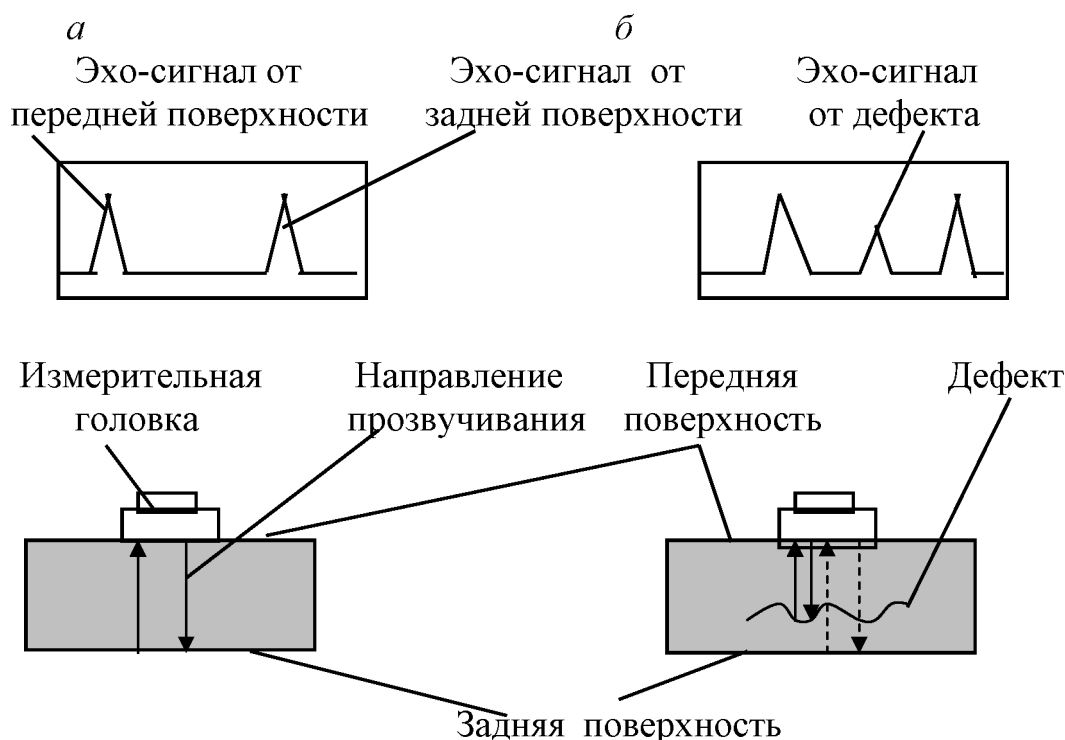
расположения дефекта (трещин, пустот, неметаллических включений) 1 – 3мм. Физическая сущность метода состоит в изменении характера распределения вихревых токов в контролируемом объекте в зависимости от его сплошности. На контролируемый участок поверхности детали изделия воздействуют переменным магнитным полем. Источником этого поля является катушка, помещенная в измерительную головку и питаемая током высокой частоты. Под действием этого поля в детали возникают вихревые токи, образующие свое магнитное поле. В результате суммарный магнитный поток (Φ_{Σ}) получается как разность магнитных потоков, создаваемых соответственно измерительной головкой ($\Phi_{и.г.}$) и вихревыми токами ($\Phi_{в.т.}$)

$$\Phi_{\Sigma} = \Phi_{и.г.} - \Phi_{в.т.}$$

Уменьшение суммарного магнитного потока вызывает изменение индуктивной составляющей суммарного сопротивления катушки и, следовательно, величины тока, протекающего через катушку. Если имеет место дефект, $\Phi_{в.т.}$ уменьшается и Φ_{Σ} изменяется, вызывая изменение тока в катушке. Сравнивая величину тока в катушке измерительной головки, настроенной на бездефектном участки поверхности, с величиной тока в катушке, помещенной на дефектный участок, можно обнаружить наличие дефекта. Метод позволяет обнаружить трещины шириной раскрытия от 0,01мм и длиной от 0,5мм.

Ультразвуковые методы основаны на способности ультразвуковых колебаний распространяться в материале детали в виде направленных пучков и отражаться от границ раздела сред. В качестве границ раздела могут выступать границы детали и дефектов.

Для возбуждения упругих колебаний в контролируемом объекте используются пьезоэлектрические преобразователи. Они представляют из себя пластину из пьезоэлектрического материала (кварц, титанат бария и др.), размещенную в измерительной головке. Под действием приложенного переменного электрического напряжения пластина совершает вынужденные механические колебания. Частота колебаний составляет 1,5 ... 15 МГц. Если измерительную головку приложить к поверхности контролируемой детали, то в детали будут возбуждаться и распространяться упругие волны (рис.1.2). Если на пути распространения ультразвуковых волн отсутствует дефект, то отражение этих волн будет происходить только от передней и задней поверхностей детали. Соответственно регистрирующий прибор зафиксирует два эхо-сигнала (рис.1.2.а).



а - при отсутствии дефекта, б - при наличии дефекта

Рис. 1.2 Схема обнаружения дефекта ультразвуковым методом

Если на пути распространения ультразвуковых волн будет находиться дефект, то регистрирующий прибор зафиксирует три эхо-сигнала, один из которых соответствует дефекту (рис.1.2.б). По величине интервала времени между эхо-сигналами можно судить о глубине расположения дефекта, а по интенсивности пика – о его величине.

Метод позволяет работать с любыми материалами, обнаруживать поверхностные и внутренние дефекты, дефекты пайки и склеивания. В ряде случаев приемник и излучатель могут размещаться в отдельных измерительных головках.

Радиационные методы основаны на использовании проникающих излучений: рентгеновских, гамма лучей, бета лучей. Дефекты материалов определяются путем оценки степени ослабления излучения при его прохождении через деталь. Излучение, прошедшее сплошной материал, имеет меньшую интенсивность по сравнению с лучами, встретившими на своем пути раковины, трещины, рыхлоты. Для регистрации результатов контроля используют рентгеновские пленки и экраны. Радиационные методы весьма сложны в практическом применении и они используются только когда оказываются неэффективными более простые методы.

К физическим методам диагностирования объектов в нерабочем состоянии можно отнести и такой специфический метод, как *метод диагностирования узлов трения по накоплению продуктов износа в масле*.

Из опыта эксплуатации известно, что часть отказов авиационной техники связана с нарушением работы узлов трения из-за недопустимых износов. Процесс разрушения трущихся деталей начинается, как правило, с разрушения поверхностного слоя материала под действием высоких контактных напряжений, что проявляется в виде отрыва частиц материала. Эти частицы уносятся маслом и по их наличию и концентрации можно судить о возникновении и развитии неисправности. Для реализации этого метода авиационная техника и эксплуатирующие организации должны быть оснащены регистрирующими и анализирующими устройствами, позволяющими: обнаруживать и улавливать продукты износа, а также проводить количественную оценку содержания, структуры и химический анализ этих продуктов. Для этой цели используются: магнитные пробки для улавливания частиц; электрические сигнализаторы, срабатывающие при замыкании их электродов частицами; спектральный анализ масла; анализ изменения формы частиц, их поверхности; рассеивание и ослабление пучка света при прохождении через поток масла; анализ изменения вязкости масла. Наиболее часто этот метод применяется для диагностирования узлов трения авиационных ГТД. Более подробно об этом методе можно узнать из литературы [11, 19, 27, 30, 38].

Вторая группа физических методов это физические методы контроля и диагностирования объектов, находящихся в рабочем состоянии. Ниже сделан краткий обзор некоторых из них.

Тепловые методы основаны на анализе теплового (инфракрасного) излучения элементов, деталей или всего объекта [19, 29]. Работоспособное (исправное) изделие имеет определенную картину теплового излучения. Изменение этой картины свидетельствует об изменении режима работы или технического состояния изделия. Увеличение интенсивности излучения может быть следствием перегрева, связанного с дефектом. Тепловые методы разделяют на контактные и бесконтактные (таблица 1.3).

Тепловые методы имеют ряд существенных ограничений. Во-первых, для большинства из них практическое применение (особенно на подвижных деталях) существенно затруднено. Во-вторых, эти методы очень инерционные.

Таблица 1.2 Тепловые методы

Контактные	Бесконтактные
1. Измерение температуры с датчиков (термопары, терморезисторы). Этот метод широко используется в авиации для измерения температуры различных сред, температуры головок цилиндров поршневых двигателей, температуры подшипников. 2. Использование температурно-чувствительных красок и составов, которые меняют цвет при достижении предельной температуры. Например для контроля дисков колес, электрических резисторов. 3. Жидкокристаллические соединения, изменяющие свою окраску в зависимости от температуры.	1. Использование специальных жидкостей, испаряющихся под действием теплового излучения детали с последующей конденсацией пара на специальных тонких мембранах с целью получения визуальной карты теплового поля. 2. Получение тепловой картины объектов контроля на экране электронно-оптических преобразователей инфракрасных излучений . 3. Использование свойств люминофоров изменять интенсивность свечения под воздействием температур.

Ко второй группе физических методов относятся и *виброакустические методы* контроля и диагностики, которые основаны на анализе характеристик шума и вибрации изделий [5, 7, 9, 13, 14, 38].

При работе любой машины возбуждаются вибрация и акустический шум. Вибрация это есть реакция динамической системы на возмущающие силы. Причинами возникновения вибрации могут быть циклические процессы, возникающие при работе машины (вращение роторов, периодические нагрузки и т.п.). В свою очередь, вибрация элементов машины вызывает колебания окружающего воздуха, то есть служит источником акустического шума. В некоторых машинах, например в авиационных ГТД, мощным источником акустического шума является газовая струя, акустическое излучение лопаток компрессора и др.

Известно, что характеристики шума и вибрации зависят от технического состояния машины. В процессе эксплуатации происходит изменение технического состояния узлов и деталей объекта диагностики, а также протекающих в нем рабочих процессов. В результате будут меняться и характеристики вибрации и шума. Причинами изменения виброакустических характеристик могут быть помпажные явления в осевом компрессоре, износ лопаток в проточной части авиационного газотурбинного двигателя, разрушение подшипниковых узлов, износ зубьев зубчатых передач, нарушение соосности валов трансмиссии, износ щеток электродвигателей, износ деталей подвижных соединений и т.п.

Для авиационной техники в качестве характеристик при диагностировании наиболее часто используются общий уровень и спектр вибрации или шума. В настоящее время для авиационных двигателей общий уровень вибрации является параметром, который подлежит обязательной нормировке и контролю бортовыми приборами в полете. Эффективность контроля уровня вибрации авиационных ГТД доказана тем, что около трети всех досрочно снимаемых с крыла двигателей - результат оценки их вибросостояния.

Более глубокое диагностирование возможно при использовании спектральных характеристик [9, 13, 14]. Однако получение этих характеристик и их практическое использование затруднено в связи со сложностью аппаратуры и методик диагностирования. Кроме того, виброакустические методы требуют обработки большого объема статистической информации. Все это несколько сдерживает широкое распространение этих методов.

Упомянутые выше *параметрические методы* основаны на измерении диагностических параметров, которые напрямую связаны с функциональным назначением объекта (например: температура газов за турбиной, давление масла, КПД и т. п.). Эти методы также могут быть, как тестовые, так и функциональные.

Чаще всего параметрические методы применяются к работающим объектам. При этом измеряемые параметры напрямую связаны с функциональным назначением диагностируемого объекта, а в ряде случаев непосредственно характеризует его техническое состояние. Параметрические методы наиболее широко применяются при диагностировании механического, гидравлического и электрического оборудования и поэтому они достаточно полно отражены в литературе [1, 8, 10, 12, 17, 18, 21, 23, 30, 31, 35, 36, 37].

В параметрических методах часто используется понятие *определяющий параметр* [8]. В случае, если техническое состояние объекта диагностируется по небольшому количеству параметров (чаще всего по одному) и эти параметры являются параметрами технического состояния или непосредственно с ними связаны, то такие параметры называются определяющими. Если диагностирование и контроль ведется на основе измерения и преобразования функциональных определяющих параметров, то алгоритм принятия решения о техническом состоянии (решающее правило) может оказаться весьма простым. Приведем некоторые примеры.

Контроль по допуску на определяющий параметр

Пусть имеется объект, на вход которого подается воздействие x . Техническое состояние объекта описывается внутренним параметром z . Выходным или диагностическим параметром является y . Пусть нормативно-технической документацией установлены нормативные значения на параметр технического состояния z_0 , на диагностический параметр y_0 и соответствующие допуски Δz_0 и Δy_0 . Тогда схему контроля по допуску на определяющий параметр можно представить следующим образом (рис.1.3).

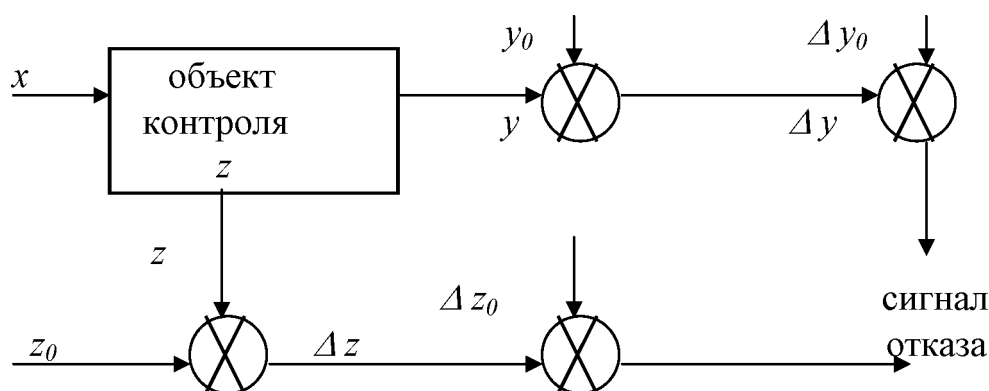


Рис. 1.3 Схема контроля по допуску на определяющий параметр

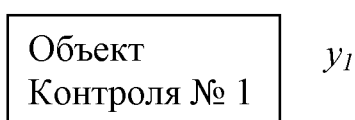
По такой схеме чаще всего контролируются устройства, у которых возможно непосредственное измерение внутренних параметров z (параметров технического состояния), а измеряемые выходные (диагностические) параметры y в явном виде характеризуют техническое состояние объекта. При этом вычисляют разницу между измеренным значением параметра и нормативным значением. Полученную разницу сравнивают с допуском. Сигнал отказа формируется в случае если полученная разница превышает величину допуска, то есть если $|z - z_0| = \Delta z > \Delta z_0$ или $|y - y_0| = \Delta y > \Delta y_0$.

В этой схеме контроля основная проблема связана с установлением величин допусков. Возможно установление нескольких значений допусков.

Применение этого метода возможно и при сложной зависимости между диагностическими параметрами и параметрами технического состояния, но в этом случае необходимо знать диагностическую модель, т.е. математическую зависимость, устанавливающую связь между указанными параметрами. При этом сначала устанавливают допуски на параметры технического состояния, а затем с помощью диагностической модели определяют допуски на измеряемые диагностические параметры.

Контроль двух параллельно работающих элементов.

Метод контроля по допуску на определяющий параметр в упрощенном виде можно распространить для контроля двух параллельно работающих взаимно резервируемых объектов. Сравнению подлежат выходные параметры y_1 и y_2 объектов (рис. 1.4). Разность этих параметров сравнивается с допуском Δy_0 . Если $|\Delta y| - |\Delta y_0| > 0$, то выдается сигнал отказа.



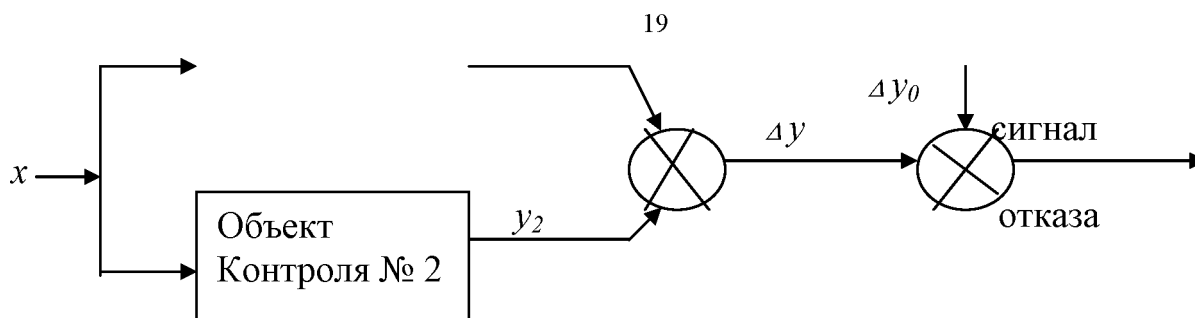


Рис. 1.4 Схема контроля взаимно резервируемых объектов

Этот метод можно расширить, если в качестве второго объекта использовать физическую модель, которая моделирует передаточную функцию контролируемой части объекта. Кроме того, данный подход можно использовать и для проведения текстового диагностирования, если на вход диагностируемого объекта и модели подавать текстовое воздействие.

Контроль одинаковых объектов, работающих параллельно.

На современных воздушных судах широкое распространение имеет многократное резервирование наиболее ответственных агрегатов. Например на Ту-154 датчиком углов крена и тангажа являются три гировертикали, работающие параллельно. Выходной сигнал каждого канала является средним значением сигналов всех трех гировертикалей. Для получения среднего значения используется схема кворум элемента (рис. 1.5)

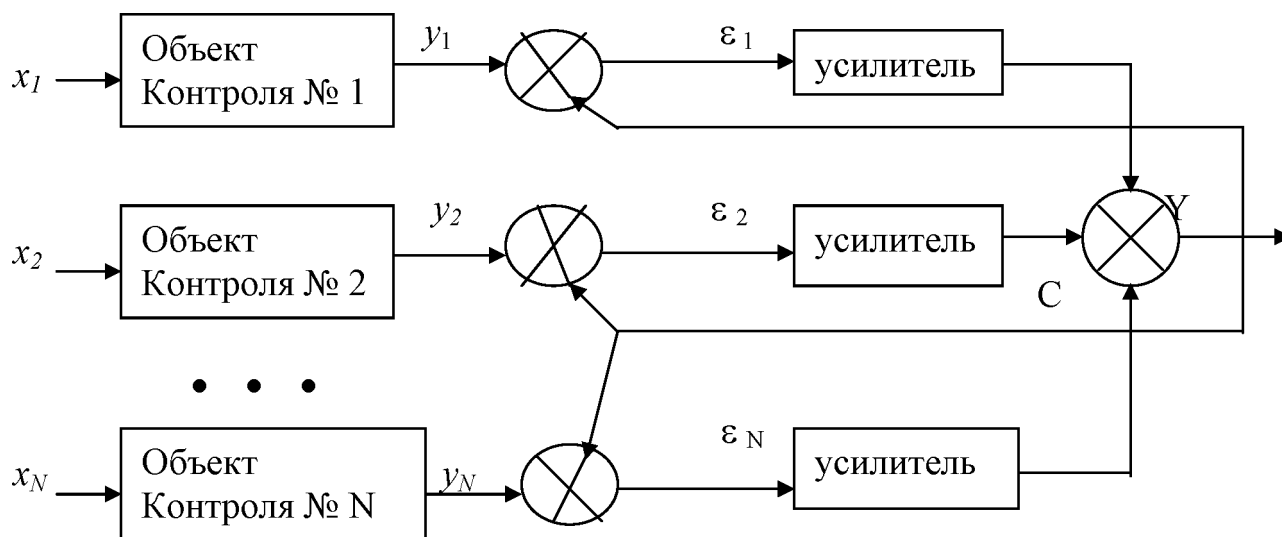


Рис. 1.5 Контроль параллельно работающих объектов на основе схемы кворум элемента

Применение этой схемы также позволяет вести контроль за техническим состоянием параллельно работающих объектов. На входы каждого из N сумматоров, установленных за объектами контроля, подаются выходные сигналы y_i от объектов контроля и результат Y , получаемый на выходе сумматора C . На выходе каждого из N сумматоров получается разность $\varepsilon = y_i - Y$, которая

усиливается соответствующим усилителем. При равных коэффициентах усиления K

$$Y = K \sum_{i=1}^N (y_i - Y) = K \sum_{i=1}^N y_i - KNY,$$

или

$$Y + KNY = K \sum_{i=1}^N y_i.$$

Преобразовывая это выражение относительно Y при условии, что $K \gg 1$, получим следующее выражение

$$Y = \frac{K \sum_{i=1}^N y_i}{(1 + KN)} \approx \frac{1}{N} \sum_{i=1}^N y_i.$$

Из этого соотношения следует, что на выходе сумматора «С» в схеме кворум элемента имеется среднее значение выходных сигналов контролируемых блоков. Контроль за техническим состоянием блоков ведется по значениям ε_i , если одно из значений превышает допуск, то этот канал неисправен. После отказа и выключения i -того канала схема продолжает работать и контролировать состояние $N-1$ каналов.

Логические методы контроля (знаковые).

Если объект контроля обладает линейной функцией преобразования входного сигнала в выходной, то знак приращения определяющего выходного параметра должен совпадать со знаком приращения входного параметра. В этом случае применяется правило знаков:

- если отклонение $+\Delta x$ ($-\Delta x$) вызывает появление отклонения $+\Delta y$ ($-\Delta y$), то объект исправен;

- если отклонение $-\Delta x$ ($+\Delta x$) вызывает появление отклонения $+\Delta y$ ($-\Delta y$), то объект неисправен.

Приведенные выше методы применимы в простейших случаях, когда можно выделять определенные параметры диагностирования. Если количество диагностируемых параметров велико и их связь с параметрами технического состояния не явна, то такие методы не применимы. В этих случаях решение задач диагностики требует создания специальных математических (диагностических) моделей.

2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ В ТЕХНИЧЕСКОЙ ДИАГНОСТИКИ

2.1. Основные понятия

Решение задач технической диагностики требует анализа множества состояний, в которых объекты могут находиться в период эксплуатации, классификации этих состояний, выбора диагностических параметров и признаков, установления связи между конкретными значениями диагностических параметров (признаков) и конкретным техническим состоянием объекта, разработки решающего правила и алгоритма постановки диагноза. Решать эти задачи можно двумя путями: экспериментально или теоретически. Экспериментальный путь наиболее достоверен, поскольку в реальных условиях эксплуатации оценивается связь между диагностическими параметрами (признаками) и параметрами технического состояния. Следует отметить, что чисто экспериментальный путь решение задач диагностики невозможен.

Для решения указанных проблем теоретическим путем необходимо иметь *формализованное описание* связи между параметрами технического состояния и диагностическими параметрами или признаками, то есть необходимо разрабатывать *математические (диагностические) модели* диагностируемых объектов.

Формализованное описание (математическая модель) объекта диагностирования может быть представлено в различной форме: - в аналитической (в виде дифференциальных или алгебраических уравнений); - в табличной форме (в виде таблицы состояний); - в графической (в виде графов причинно-следственных связей); - в форме логических соотношений и т. д. [1, 5, 7, 8, 9, 10, 13, 17, 23, 24, 26, 31].

Математические модели могут быть заданы в явном или неявном виде.

Явная модель объекта диагностирования включает в себя совокупность формализованных описаний всех оговоренных нормативной документацией технических состояний, которое подлежат диагностированию.

Неявная модель объекта технического диагностирования содержит формализованное описание одного технического состояния и правила получения описаний других технических состояний на основе заданного. Причем, чаще всего математическая (диагностическая) модель определяет исправное состояние. Модели для других технических состояний получают на основе модели исправного состояния и правил, связывающих конкретную неисправность с параметрами технического состояния.

В общей постановке математическую модель объекта технической диагностики можно охарактеризовать следующим образом [26]. Пусть имеется объект, у которого техническое состояние описывается внутренними параметрами Z . На вход объекта подаются внешние возмущения X (входные параметры). На выходе объекта наблюдается отклик в виде выходных (диагностических) параметров Y . В самом общем случае соотношение между указанными

параметрами устанавливают вводя понятие *оператор объекта диагностирования* $A(\cdot)$. При этом математическая модель будет иметь следующий вид

$$Y = A(Z, X). \quad (2.1)$$

Вводя индексы i для разных состояний: 0- исправное, i - i -тое неисправное состояние; и индексы j для различных внешних возмущений уравнение (2.1) можно переписать в виде системы уравнений

$$Y_0^j = A(Z_0, X_j); \quad Y_i^j = A(Z_i, X_j), \quad (2.2)$$

где $i = 1 \dots k; j = 1 \dots m$.

Система уравнений (2.2) состоит из $k+1$ уравнений и представляет из себя явную математическую модель объекта диагностирования. Для упрощения записи внутренние параметры Z (параметры технического состояния) включают в оператор и систему уравнений (2.2) при любом внешнем возмущении X записывают в виде

$$Y_0 = A_0(X); \quad Y_i = A_i(X), \quad (2.3)$$

Такой способ задания математической модели является весьма общим. Выбор конкретной формы оператора и есть выбор конкретной математической модели объекта диагностирования. Этот выбор зависит от физических свойств диагностируемого объекта, определяется условиями диагностической задачи и принятыми методами ее решения.

Объекты диагностирования в зависимости от своих свойств делятся на дискретные и непрерывные.

Объект считается непрерывным, если его оператор преобразует совокупность входных сигналов X , определенных на непрерывном (несчетном, континуальном) множестве, в совокупность выходных сигналов Y , принимающих значения на таких же множествах.

Объект считается дискретный, если он имеет оператор, который преобразует совокупность входных величин X , принимающих значение на конечных (счетных) множествах, в совокупность выходных величин Y , принимающих значения на таких же множествах.

Объект называется гибридным, если его оператор осуществляет преобразование непрерывного входного сигнала X в дискретный выходной сигнал Y или наоборот.

Для дискретных объектов разделение технических состояний, в которых он может находиться, на классы не представляет затруднений. Так как значения диагностических параметров и параметров технического состояния счетные, то счетные будут и количество классов технического состояния. Для непрерывных объектов при разделении технических состояний на классы возникают затруднения, связанные с непрерывностью диагностических параметров и бесконечным числом технических состояний.

Практически диагностическую задачу при бесконечном числе технических состояний решить невозможно. Поэтому в любом случае выделяется конечное множество технических состояний, которые требуется диагностировать. Таким образом, при практическом решении диагностической задачи оператор $A(\cdot)$ для

непрерывных объектов заменяется гибридным, реализующим на выходе конечное множество значений диагностических признаков.

Для перехода от непрерывного множества технических состояний, которые есть на самом деле, к дискретному множеству технических состояний, которые можно диагностировать, вводится понятие элементарная проверка объекта [24, 26]. *Элементарная проверка объекта* (ЭПО или ЭП) это процедура подачи на объект отдельного (рабочего или тестового) воздействия и снятия с объекта соответствующего ответа (реакции). Поясним это понятие.

Пусть объект диагностики имеет точки контроля, которые позволяют наблюдать за входными X , внутренними Z и диагностическими Y параметрами. В соответствии с определением каждому входному возмущению X_j из множества входных возмущений X будет соответствовать элементарная проверка π_j из множества проверок Π ($\pi_j \in \Pi$), то есть $X_j \rightarrow \pi_j$. в техническом состоянии s_i из множества S ($s_i \in S$), входное воздействие на объект X соответствует этой проверке. Внутренние параметры Z и оператор A определяются техническим состоянием s_i из множества S ($s_i \in S$). Выходные диагностические параметры Y соответствуют результату j -той проверки, то есть $Y_i^j \rightarrow R_i^j$. Тогда операторное выражение (2.3) в понятиях ЭПО для j -той проверки объекта, находящегося в i -том техническом состоянии запишется следующим образом

$$R_i^j = A_i(\pi_j), \quad (2.4)$$

где $j = 1 \dots n$; $i = 0, 1 \dots k$.

Для исправного состояния это выражение часто записывается в виде

$$R^j = A(\pi_j). \quad (2.5)$$

Выражения (2.4) и (2.5) определяют модель в явном виде. Если имеется зависимость вида (2.5) и правила, позволяющие из множества проверок Π и состояний S получать совокупность выражений вида (2.4), то объект диагностирования задан неявной моделью.

Представленные модели объектов диагностики в операторном виде используются для формального описания диагностических задач в общей постановки, но они непригодны для решения конкретных диагностических задач. Для выполнения диагностирования на практике необходимо иметь математические (диагностические) модели в конкретном виде, то есть необходимо для каждой практической задачи конкретизировать вид оператора $A(\cdot)$.

Наиболее простой формой представления оператора $A(\cdot)$ в случае явной модели является табличная форма. Таблица строится следующим образом. Пусть каждому отказу (неисправности) объекта соответствует s_i неисправное состояние из множества S ($i = 1 \dots n$). Исправное состояние обозначим s_0 . Примем, что столбцы таблицы соответствуют возможным состояниям $s_i \in S$, а строки - проверкам $\pi_j \in \Pi$. В клетку таблицы на пересечении i -того столбца и j -той строки записывается результат R_i^j проверки π_j объекта диагностирования, находящегося в техническом состоянии s_i . Совокупность результатов образуют множество R ($R_i^j \in R$), число элементов этого множества $r = n(k+1)$. Такая

таблица называется *таблицей состояний* объекта диагностирования. Она эквивалентна выражениям (2.4) и (2.5).

Таблица 2.1 Таблица состояний объекта диагностирования

$\pi \setminus s$	s_0	s_1	...	s_i	...	s_k
π_1	R_1	R^1_1		R^1_i		R^1_k
...						
π_j	R^j	R^j_1		R^j_i		R^j_k
...						
π_n	R^n	R^n_1		R^n_i		R^n_k

Анализ этой таблицы позволяет сформулировать свойства множества проверок Π , необходимые для решения задач технического диагностирования.

Первое свойство : множество проверок Π обнаружит любое неисправное состояние объекта диагностирования из множества S , если для каждого состояния $s_i \in S$ найдется хотя бы одна проверка $\pi_j \in \Pi$ такая, что $R^j \neq R^j_i$. Это свойство эквивалентно тому, что столбец таблицы с s_0 отличается от каждого из остальных с s_i . Такая таблица называется *проверяющей или контролирующей*.

Второе свойство: множество проверок Π различают все состояния из множества S , если для каждой пары технических состояний s_i и s_m найдется хотя бы одна проверка π_j такая, что $R^j_i \neq R^j_m$. Это свойство таблицы эквивалентно тому, что все столбцы, определяющие различные состояния, попарно различимы. Такая таблица называется *разделяющей*.

Если таблица обладает только первым свойством, то она позволяет решать только задачи контроля. Если таблица обладает обоим свойствами, то она позволяет решать задачи диагностики.

Исходной информацией для составления таблицы состояний может быть: структурные, функциональные, принципиальные схемы, причинно-следственные связи между параметрами технического состояния, входными и диагностическими параметрами. Одним из наиболее часто используемых способов построение таблицы состояния, является построение этих таблиц на основе функциональных моделей [8, 17, 18, 24, 26].

2.2 Функциональная модель

Если объект диагностики можно представить как несколько взаимосвязанных частей, имеющих самостоятельное функциональное значение, то в качестве математической (диагностической) модели применяется так называемая *функциональная модель*, пользуясь которой можно составить таблицу состояний.

Функциональная модель представляет собой графическое изображение объекта, в котором каждая выделенная часть (функциональный элемент) обозначается в виде прямоугольника со значком Q_i и этот прямоугольник имеет несколько входов x_j , но только один выход y_i (рис.2.1)

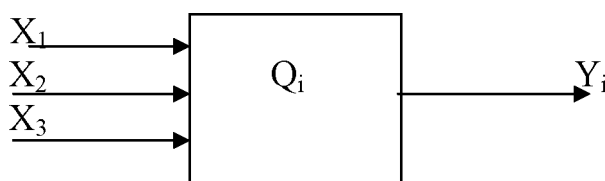


Рис.2.1 Элемент функциональной модели

Количество входов определяется числом внешних воздействий на конкретный элемент. Связи между элементами обозначаются линиями со стрелками, указывающими направление прохождения сигнала. Считается, что в объекте может выходить из строя только один функциональный элемент. При этом i -тое неисправное состояние идентифицируется с отказом i -блока. В этой ситуации запись состояния s представляется в виде вектора $S()$ в котором количество координат соответствует количеству блоков. Значения координат задаются в соответствии со следующими правилами: если блок исправен, то координата равна 1, а если блок неисправен, то координата равна 0. Например, если объект диагностики имеет пять блоков то исправное состояние запишем вектором $S_0(1\ 1\ 1\ 1\ 1)$, при неисправном первом блоке техническое состояние запишется следующим вектором $S_1(0\ 1\ 1\ 1\ 1)$.

Под проверкой π_j понимается операция по оценке выходных сигналов всех блоков объекта при нахождении его в состоянии s_i . При наличии в объекте диагностики n функциональных блоков общее количество технических состояний равно $n + 1$ (n неисправных и одно исправное), а количество проверок равно n .

При формальном описании результатов j -той проверки R_i^j отдельного элемента функциональной схемы при нахождении объекта в i -том состоянии принимаются следующее правила:

1. Результат принимается $= 1$, если блок исправен и на его входы подаются допустимые воздействия.
2. Результат принимается $= 0$, если на входы блока подаются допустимые воздействия, а он неисправен.

3. Результат принимается $= 0$, если блок исправен, но хотя бы на один его вход подаются недопустимое воздействие.

Пользуясь таким представлением можно задать все множество технических состояний объекта, множество проверок, множество результатов и составить таблицу состояний.

Рассмотрим пример построения таблицы состояний объекта диагностики, который можно представить функциональной моделью, состоящей из шести элементов (рис.2.2)

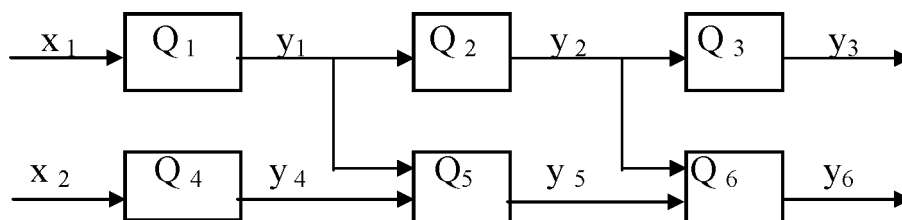


Рис. 2.2 Пример функциональной шести элементной модели

Для такой модели количество технических состояний равно семи. Исправное состояние можно представить шестимерным вектором $S_0(1\ 1\ 1\ 1\ 1\ 1)$, при отказе блока Q_1 техническое состояние представим вектором $S_1(0\ 1\ 1\ 1\ 1\ 1)$ и т.д. Поскольку в объекте шесть блоков, то для проверки функционирования каждого блока необходимо проверить его выходной сигнал. Следовательно, количество необходимых проверок равно шести. Будем считать, что при проведении каждой из проверки π_j ($j = 1 \dots 6$) на вход модели подаются входные воздействия x_1 и x_2 , которые принимают только допустимые значения. В тоже время выходные сигналы блоков y_i ($i = 1 \dots 6$) могут принимать как допустимые, так и недопустимые значения. Допустимое значение выходной сигнал принимает в случае, если блок исправен и на его входы подаются допустимые воздействия. Недопустимое значение выходной сигнал принимает в двух случаях: если блок исправен, но на его входы подаются недопустимые сигналы, или если входные сигналы принимают допустимые значения, но блок неисправен.

С учетом этого и пользуясь изложенными выше тремя правилами результат элементарной проверки блоков объекта, находящегося в i -том техническом состоянии (отказ i -того блока), будем принимать равным 1, если сигнал на выходе проверяемого блока принимает допустимое значение, и равным 0, если сигнал на выходе проверяемого блока принимает недопустимое значение. Например, при отказе блока Q_2 нулевые значения примут результаты R^2_2 , R^2_3 , R^2_6 , поскольку при проведении элементарных проверок π_2 , π_3 , π_6 для блоков Q_2 , Q_3 , Q_6 выходные сигналы примут недопустимые значения. Воспользовавшись приведенными выше рассуждениями, составим таблицу состояний для объекта, функциональная модель которого изображена на рис.2.2.

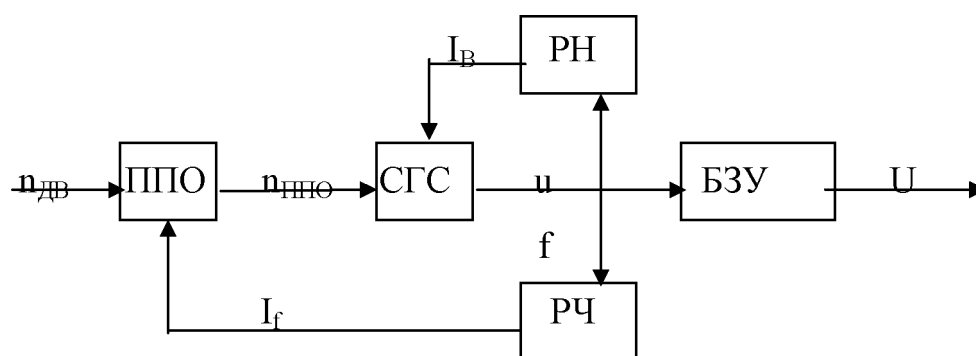
Таблица 2.2 Таблица состояний для шестиэлементного объекта

	S_0	S_1	S_2	S_3	S_4	S_5	S_6
	111111	011111	101111	110111	111011	111101	111110
π_1	1	0	1	1	1	1	1
π_2	1	0	0	1	1	1	1
π_3	1	0	0	0	1	1	1
π_4	1	1	1	1	0	1	1
π_5	1	0	1	1	0	0	1
π_6	1	0	0	1	0	0	0

Полученная таблица является как проверяющей, так и разделяющей, т.е. позволяет решать задачи контроля и диагностики.

Непременным условием применения функциональной модели является наличие у всех элементов (блоков) объекта диагностирования одного выхода. В реальных условиях выходной сигнал некоторых блоков объекта контроля может характеризоваться несколькими параметрами. В этом случае считается, что блок имеет несколько выходных сигналов, каждый из которых характеризует соответствующий параметр. В такой ситуации применяют так называемое «расщепление» выходов. Рассмотрим применение этого приема на примере.

На рис. 2.3 изображена схема системы электроснабжения самолета.



ППО - привод постоянных оборотов; СГС - синхронный генератор самолетный; РН - регулятор напряжения; РЧ - регулятор частоты; БЗУ - блок защиты и управления; $n_{дв}$ - частота вращения коробки приводов двигателя; $n_{гпо}$ - частота вращения генератора; u и f - напряжение и частота на выходе генератора; I_B - ток возбуждения генератора; I_f - управляющее воздействие по частоте на ППО; U - напряжение, подаваемое в самолетную сеть

Рис. 2.3 Схема системы электроснабжения самолета

Выходной сигнал генератора характеризуется двумя параметрами: напряжением и частотой. При составлении функциональной модели проводится «расщепление» выходного сигнала генератора по следующим правилам.

Правило первое. Если выходной сигнал характеризуется несколькими параметрами, то выходной сигнал «расщепляется» на количество сигналов, равное количеству параметров.

Схема системы электроснабжения с «расщепленным» выходным сигналом генератора представлена на рис.2.4.

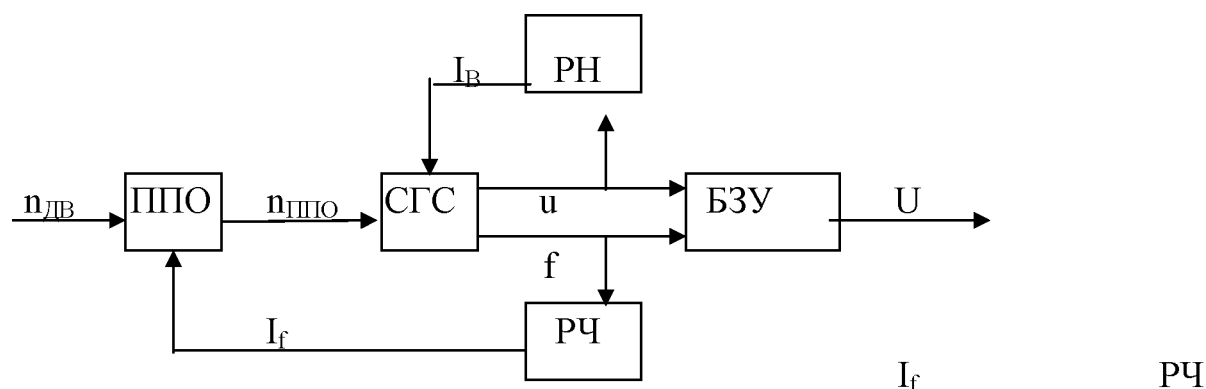


Рис. 2.4 Схема с «расщепленным» выходным сигналом генератора

В соответствии с этой схемой один из блоков (генератор) объекта диагностирования имеет два выхода, что не может быть в функциональной модели. Для приведения в соответствие количества выходов блоков с требованием функциональной модели пользуются вторым правилом.

Правило второе. Если i -тый блок имеет несколько выходов, то на функциональной модели он заменяется несколькими блоками, имеющих только один выход. Число блоков равно числу выходов «расщепляемого» блока.

С учетом этого правила, предыдущую схему представим следующим образом

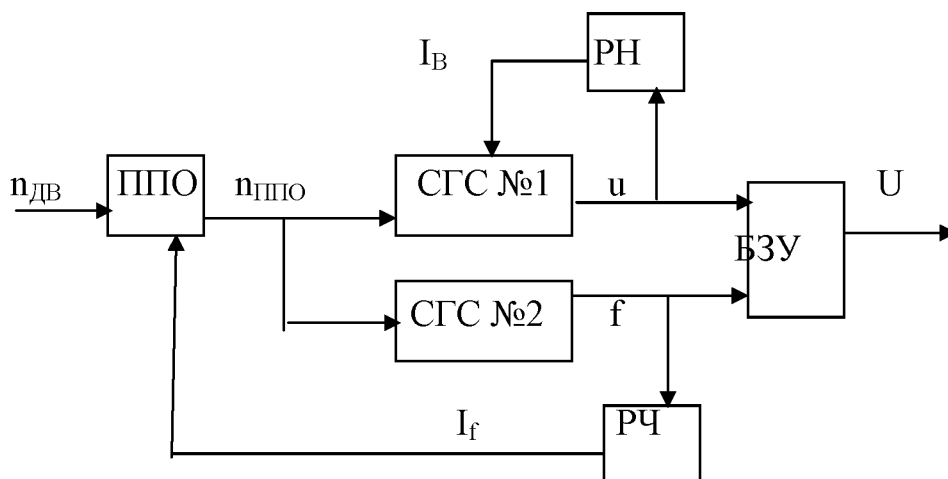


Рис. 2.5 Схема с «расщепленным» выходом генератора

Приведенная схема соответствует требованиям функциональной модели, что позволяет изобразить модель в принятых ранее обозначениях (рис.2.6)



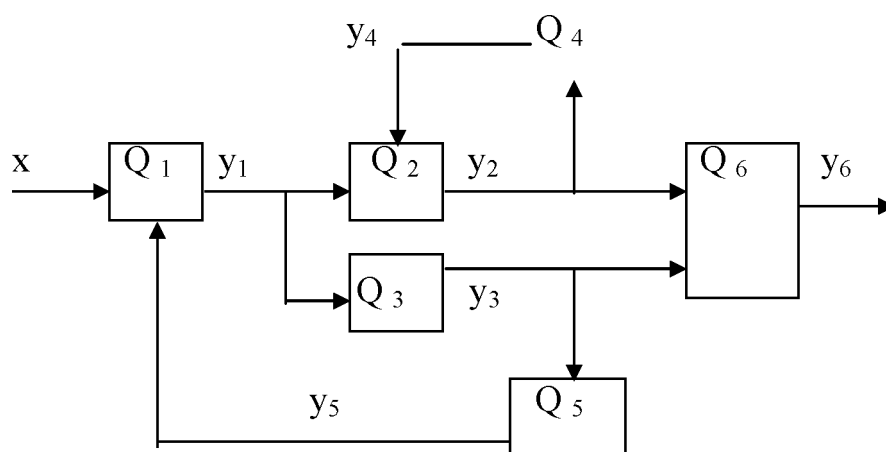


Рис. 2.6 Функциональная модель системы электроснабжения самолета

Пользуясь этой моделью можно составить таблицу состояний. Таблица состояний заполняется исходя из следующих соображений. При исправном состоянии всех блоков сигналы на их выходах будут принимать допустимые значения, то есть результаты всех проверок будут равны 1. При отказе первого блока на выходе всех блоков сигналы будут принимать недопустимые значения, то есть результаты всех проверок будут равны 0. При отказе блока №2 недопустимые значения выходных сигналов будут у блоков №2, №4 и №6, то есть результаты проверок этих блоков будут равны 0. Для остальных блоков результаты проверок будут равны 1. На основе подобных рассуждений заполняется вся таблица.

Таблица 2.3 Таблица состояний системы электроснабжения

	S_0	S_1	S_2	S_3	S_4	S_5	S_6
π_1	1	0	1	0	1	0	1
π_2	1	0	0	0	0	0	1
π_3	1	0	1	0	1	0	1
π_4	1	0	0	0	0	0	1
π_5	1	0	1	0	1	0	1
π_6	1	0	0	0	0	0	0

Представленная таблица является проверяющей, то есть она решает только задачу контроля. Для технических состояний S_1 , S_3 , S_5 , а также для состояний S_2 , S_4 результаты элементарных проверок совпадают. Следовательно, для выбранного набора проверок эти состояния неразличимы между собой.

Для того чтобы исправить это положение необходимо изменить структуру модели и элементарных проверок. Изменение структуры обеспечим путем разрыва обратных связей между блоками Q_4 и Q_2 , а также между блоками Q_5 и Q_1 . Структуру проверок изменим путем введения дополнительных внешних воздействий x_4 и x_5 , которые будут подаваться на входы блоков соответственно

Q_2 и Q_1 вместо выходных сигналов y_4 и y_5 от блоков Q_4 и Q_5 . При таких заменах модель будет выглядеть следующим образом (рис.2.7)

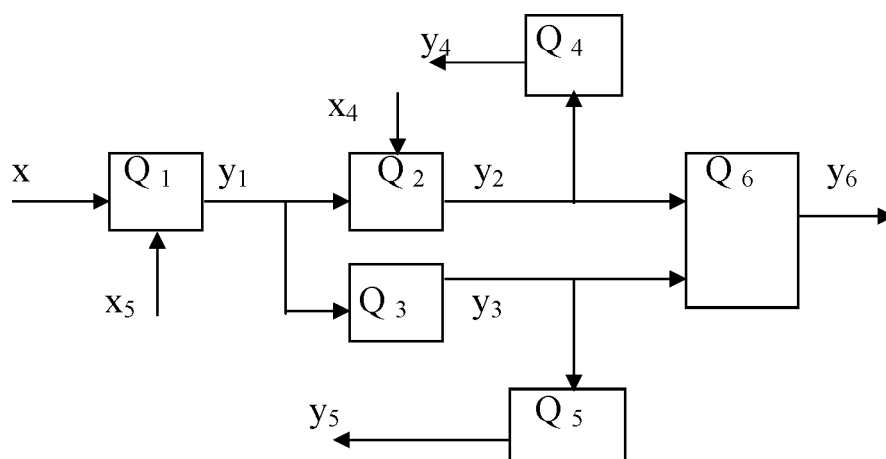


Рис. 2.7 Функциональная модель системы электроснабжения самолета с разорванными обратными связями

При выполнении всех проверок будем считать, входные воздействия x , x_4 и x_5 принимают только допустимые значения. Проведя те же рассуждения, что и при построении таблицы 2.3, получим несколько другую таблицу (таблица 2.4)

Таблица 2.4 Таблица состояний системы электроснабжения с разорванными обратными связями

	S_0	S_1	S_2	S_3	S_4	S_5	S_6
π_1	1	0	1	1	1	1	1
π_2	1	0	0	1	1	1	1
π_3	1	0	1	0	1	1	1
π_4	1	0	0	1	0	1	1
π_5	1	0	1	0	1	0	1
π_6	1	0	0	0	1	1	0

Представленная таблица становится различающей и позволяет решать задачи диагностики. На практике задачу изменения структуры объекта диагностирования путем разрыва обратных связей и подачи на вход проверяемых блоков допустимых воздействий можно решить, например, применяя специальные средства диагностики.

Таким образом, можно сделать следующее важное замечание: если в функциональной диагностической модели (схеме) имеются обратные связи, охватывающие несколько блоков, то в таблице состояний отказы этих блоков

неразличимы. Различие отказов этих блоков возможно только при изменении структуры объекта или при введении дополнительных проверок.

В некоторых случаях при выполнении дополнительных проверок для формирования признаков состояния можно использовать различные уровни выходных параметров проверяемых блоков. Так для выходного напряжения генератора можно сформулировать два признака:

напряжение генератора находится в пределах допуска $y_2^1 \rightarrow u_{\min} \leq y_2 \leq u_{\max}$;
на выходе генератора отсутствует перенапряжение $y_2^2 \rightarrow y_2 \leq u_{\max \text{ доп.}}$

Если на выходе генератора наблюдается большая величина выходного напряжения, то есть выходной параметр y_2^2 принимает недопустимое значение, то это является признаком отказа регулятора напряжения, ведущего к значительному увеличению тока возбуждения генератора и, как следствие, к перенапряжению на выходе генератора.

Напомним, что функциональную модель применяют тогда, когда объект состоит из ярко выраженных функциональных блоков. В тех случаях, когда объект диагноза не имеет явно выраженных блоков, построение функциональной модели может оказаться невозможным.

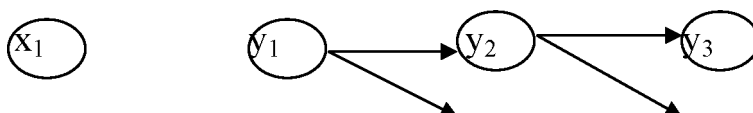
В реальных условиях существуют объекты, для которых аналитические и другие виды зависимостей между параметрами технического состояния и диагностическими параметрами неизвестны, а известно только, что один параметр влияет на другой. То есть имеется информация о том, что существует причинно-следственные связи между параметрами, событиями или явлениями. В таких ситуациях построение таблицы состояний возможно на основе построения и анализа графа причинно-следственных связей [23, 26].

2.3 Модели на основе графов причинно-следственных связей

Графом причинно-следственных связей называется ориентированный граф, вершины которого отображают собой параметры, события или явления, а дуги, соединяющие эти вершины, отражают причинно-следственные связи между соответствующими параметрами, событиями или явлениями. Стрелки на дугах направлены от причины к следствию.

Отметим, что с помощью графа причинно-следственных связей можно анализировать и объекты, представленные в виде функциональной модели. В этом случае граф будет иметь вершины, которые соответствуют входным параметрам модели и выходным параметрам блоков модели. Дуги графов будут соответствовать связям между блоками, а также между входными параметрами и блоками модели.

В качестве примера построим граф причинно-следственных связей для функциональной модели, изображенной на рис. 2.2. В вершинах графа (рис.2.8) в данном случае помещаются входные параметры x_1 и x_2 системы и выходные параметры блоков y_i (где $i = 1 \dots 6$)



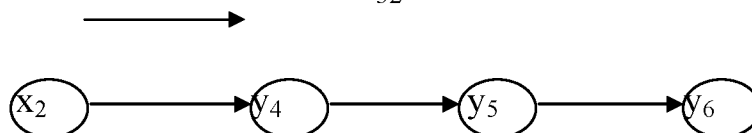


Рис. 2.8 Граф причинно-следственных связей для шестиэлементной модели

Вершины с параметрами x_1 и x_2 называются начальными или корневыми вершинами графа. Вершины с параметрами y_3 и y_6 называются висячими вершинами графа.

Данный пример показывает, что граф более конкретно отражает связь между параметрами, событиями или явлениями, помещёнными в вершины графа. По нему четко видно, что если параметр в вершине графа принимает недопустимое значение, то такое же недопустимое значение принимает параметр, связанный с ним причинно-следственной связью (дугой).

Используя полученный граф можно производить все те же операции, что и по функциональной модели: строить таблицу состояний; выбирать набор проверок (признаков) для проверяющей и различающей совокупностей. Рассмотренный пример является частным, поскольку в вершинах графа находятся измеряемые параметры. Возможности графов намного шире они определяются тремя основными положениями:

1. Путем построения графов причинно-следственных связей можно устанавливать зависимости не только между параметрами, но и между событиями, которые определены логическими высказываниями.

Например: «Значение параметра (предположим T_{Γ}^*) не выходит за пределы заданной для него области допустимых значений». Такие события можно помещать в вершины графа и дугами устанавливать между ними связи. Если истинность этого высказывания (то есть нахождения параметра в норме) условится представлять единицей, то ложность этого события (то есть выход параметра за допуск) можно представлять нулем. Такое описание истинного и ложного событий совпадает с тем описанием, которое было принято раньше для функциональных моделей. Это и является предпосылкой для построения на основе графа причинно-следственных связей таблицы состояний.

2. При построении графа причинно-следственных связей можно в вершины графов проставлять не все входные, внутренние и выходные параметры (события или явления), которые происходят в объекте, а только те из них, которые доступны для измерения или регистрации.

3. Для функциональных моделей принято, что неисправностью объекта является любая неисправность, приводящая к появлению недопустимого значения параметра на выходе одного из блоков. Количество неисправностей системы совпадает с количеством имеющихся блоков. При построении графа причинно-следственных связей количество диагностируемых состояний (дефектов) можно искусственно уменьшить путём задания конкретного перечня неисправностей,

которые требуется обнаружить. Такое задание неисправностей в виде списка равносильно тому, что неисправности, не внесённые в список маловероятны или невозможны. Естественно, что уменьшение числа неисправностей при решении практической диагностической задачи уменьшает число параметров, подлежащих измерению и, как следствие, уменьшают объёмы проверяющей и особенно различающей совокупностей.

В качестве примера, иллюстрирующего особенности применения графов для задач диагностики, рассмотрим построение таблицы состояния с использованием графа причинно-следственных связей для системы смазки редуктора, схема которой изображена на рис. 2.9.

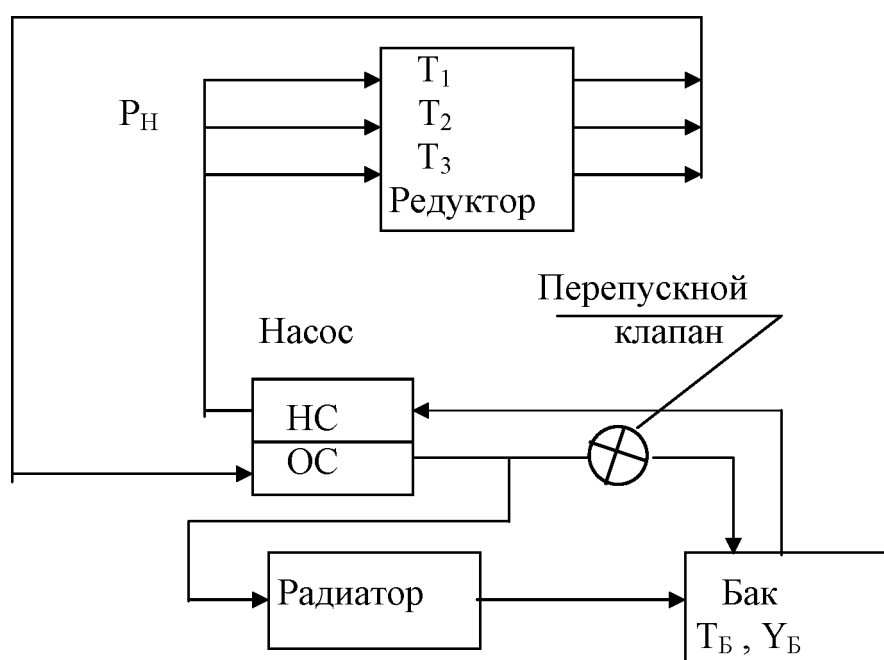


Рис. 2.9 Схема системы смазки редуктора

Будем считать, что выходными параметрами, характеризующими техническое состояние всего редуктора, являются величины температур трех основных подшипников редуктора T_1 , T_2 , T_3 . Смазка редуктора (в том числе и подшипников) осуществляется маслом, циркуляция которого обеспечивается шестеренчатым насосом, который имеет нагнетающую (НС) и откачивающую (ОС) секции и создает в напорной магистрали давление P_H . Стабилизацию температуры масла в баке T_B обеспечивается прокачкой масла через радиатор. Расход масла через радиатор зависит от положения перепускного клапана. Если клапан закрыт (обозначим это событие K_3), то все масло при откачке проходит через радиатор и температура масла в баке падает. Если клапан открыт (обозначим это событие K_0), то происходит увеличение температуры масла в баке. Кроме температуры масла будем контролировать и его уровень в баке Y_B .

Для определения вершин графа составим список недопустимых событий для данной системы:

1. Превышение температуры подшипников предельно допустимого значения H , то есть $T_1 > H$; $T_2 > H$; $T_3 > H$.

2. Давление масла в нагнетающей магистрали ниже нормы H_1 , то есть $P_H < H_1$.
3. Уровень масла в баке ниже минимально допустимого H_2 , то есть $Y_B < H_2$.
4. Температура масла в баке выше максимально допустимой H_3 , то есть $T_B > H_3$.
5. Температура масла в баке ниже минимально допустимой H_4 , то есть $T_B < H_4$.
6. Перепускной клапан постоянно закрыт, то есть имеется событие K_3 .
7. Перепускной клапан постоянно открыт, то есть имеется событие K_O .

Перечисленные события определяют вершины графа. Для установления причинно-следственных связей между этими событиями проведем анализ физики функционирования системы смазки по имеющейся схеме (рис.2.9).

Определим эти связи из следующих соображений. Превышение T_1 , T_2 , или T_3 предельно допустимого значения приводит к отказу редуктора. То есть эти события должны быть заключительными в цепочке причин и следствий (рис.2.10).

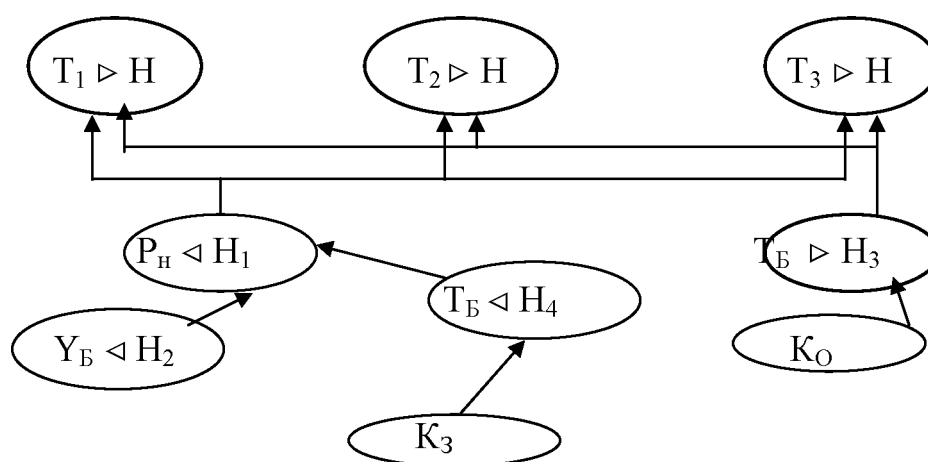


Рис.2.10 Цепочка причинно-следственных связей между неисправностями системы смазки редуктора

Появление повышенной температуры подшипников может произойти по следующим причинам: во первых из-за повышения температуры масла в баке (событие $T_B > H_3$); во вторых из-за падения давления в нагнетающей магистрали (событие $P_H < H_1$). В свою очередь, повышенная температура масла в баке может появиться если перепускной клапан постоянно открыт (произошло событие K_O), то есть масло не охлаждается в радиаторе.

Пониженное давление в напорной магистрали может быть обусловлено двумя причинами: во первых, если уровень масла в баке ниже нормы (событие $Y_B < H_2$); во вторых, если масло очень холодное (событие $T_B < H_4$).

Низкая температура масла в баке появится в том случае, если перепускной клапан будет постоянно закрыт (т.е. произошло событие K_3).

Если цепочку причинно-следственных связей, изображенную на рис.2.10, перевернуть то мы получим граф, у которого в корневых (начальных) вершинах будут помещены события $Y_B < H_2$, K_3 и K_O , а в висячих вершинах будут помещены события $T_1 > H$, $T_2 > H$ и $T_3 > H$. Этот граф (рис.2.11) позволяет

составить таблицу состояний, приняв за неисправности все перечисленные выше недопустимые события.

При таком подходе для этой системы мы определим одно исправное состояние (когда отсутствуют все перечисленные в списке недопустимые события) и девять неисправных состояний (когда происходит только одно из перечисленных в списке недопустимых событий).

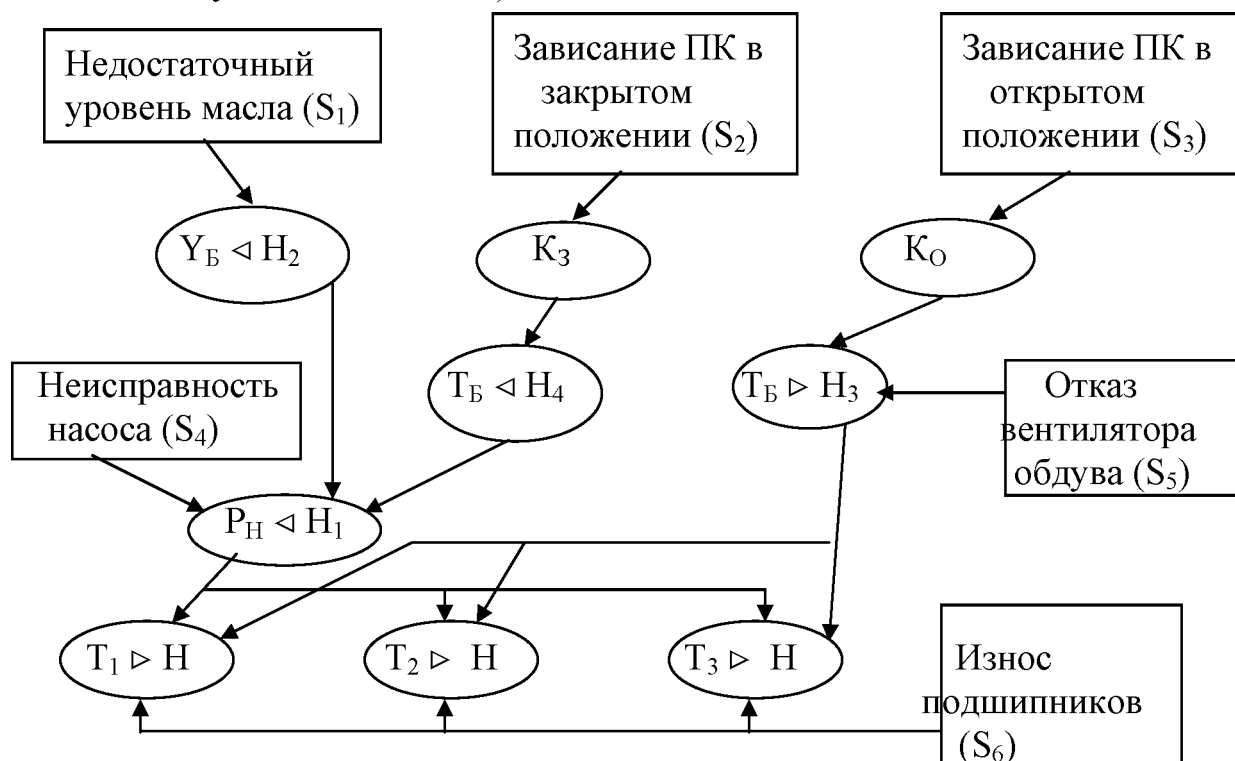


Рис. 2.11 Граф причинно-следственных связей для системы смазки

Под элементарными проверками будем подразумевать оценку появления или не появления указанных в списке событий. Причем, если проверка показывает, что для данного технического состояния проявляется недопустимое событие, то результат проверки принимается равным нулю. И наоборот, если в ходе проверки обнаруживается, что для данного технического состояния проявляется допустимое событие, то результат проверки принимается равным единице. Если теперь при заполнении таблицы в столбец с проверками записывать не события из представленного выше списка, а противоположные им (то есть допустимые события), то полученная таблица будет иметь такой же вид, как и для функциональных моделей. Действительно, для исправного состояния будут наблюдаться только события, противоположные событиям списка, а в соответствующем столбце следует записать единицы. Для отличия допустимых событий от событий, указанных в списке, будем помечать их чертой сверху. Например, $\overline{T_1 > H}$ - температура первого подшипника находится в норме.

С учетом всего сказанного таблица состояний будет содержать десять столбцов и девять строк. Для уменьшения размерности таблицы состояния следует определить список только тех неисправностей, которые мы желаем обнаружить. Например, из статистики неисправностей системы смазки известно,

что в эксплуатации встречаются следующие технические состояния и неисправности:

- S_0 - исправное состояние ;
- S_1 - недостаточен уровень масла в баке ;
- S_2 - перепускной клапан завис в закрытом состоянии ;
- S_3 - перепускной клапан завис в открытом состоянии ;
- S_4 - отказ насоса ;
- S_5 - отказ вентилятора обдува радиатора ;
- S_6 - износ подшипников.

Эти неисправности следует связать с событиями из представленного списка недопустимых событий и наложить их на граф причинно-следственных связей. На рис.2.11 показано, на какие из событий оказывают влияние эксплуатационные неисправности. Пользуясь этим рисунком, составим таблицу состояний, в которую включим только технические состояния, определенные представленным списком неисправностей.

Таблица 2. 5 Таблица состояний для системы смазки

	S_0	S_1	S_2	S_3	S_4	S_5	S_6
$\overline{Y_B \triangleleft H_2}$	1	0	1	1	1	1	1
$\overline{K_3}$	1	1	0	1	1	1	1
$\overline{K_0}$	1	1	1	0	1	1	1
$\overline{T_B \triangleleft H_4}$	1	1	0	1	1	1	1
$\overline{T_B \triangleright H_3}$	1	1	1	0	1	0	1
$\overline{P_H \triangleleft H_1}$	1	0	0	1	0	1	1
$\overline{T_1 \triangleright n}$	1	0	0	0	0	0	0
$\overline{T_2 \triangleright n}$	1	0	0	0	0	0	0
$\overline{T_3 \triangleright n}$	1	0	0	0	0	0	0

В полученной таблице состояния S_i ($i = 0, 1, \dots, 6$) соответствуют состояниям, указанным в списке неисправностей, заданным из опыта эксплуатации. Проверки $\overline{T_1 \triangleright n}$, $\overline{T_2 \triangleright n}$, $\overline{T_3 \triangleright n}$ означают проверку температуры подшипников на предмет не превышения ее предельных значений. Проверка $\overline{P_H \triangleleft H_1}$ это измерение давления в нагнетающей магистрали. Проверки $\overline{T_B \triangleleft H_4}$, $\overline{T_B \triangleright H_3}$, $\overline{Y_B \triangleleft H_2}$ это соответственно измерение температуры и уровня масла в баке.

Из анализа таблицы 2.5 следует, что измерение температур T_1 , T_2 и T_3 дает одинаковую информацию о состоянии масляной системы. Также одинаковую информацию дают проверки событий $\overline{K_3}$ и $\overline{T_B \triangleleft H_4}$. Таким образом, полученная таблица является избыточной.

2.4 Методы оптимизации диагностических тестов

Таблица состояний является удобной формой задания оператора объекта диагностирования. Однако она может содержать избыточное количество проверок, в которых используется большое количество признаков. Поэтому возникает задача выбора минимального количества проверок и признаков, достаточных для решения задач контроля и диагностики.

Пусть в результате анализа объекта диагностирования была составлена функциональная модель и заполнена таблица состояний (таблица 2.6).

Таблица 2.6 Таблица состояний с избыточным числом проверок

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7
π_1	1	0	1	1	1	1	1	1
π_2	1	0	1	1	1	1	0	1
π_3	1	0	0	0	1	1	1	1
π_4	1	0	0	0	0	0	1	1
π_5	1	0	0	0	0	1	1	1
π_6	1	1	0	0	1	0	0	0
π_7	1	0	0	1	1	1	0	0
π_8	1	1	1	1	0	0	0	0
π_9	1	1	1	0	1	0	1	1
π_{10}	1	0	1	1	1	1	1	1
π_{11}	1	1	1	1	0	0	0	0
π_{12}	1	1	1	1	1	1	1	1

Отсутствие одинаковых столбцов в таблице свидетельствует о том, что выбранный набор элементарных проверок (признаков) позволяет различать все восемь состояний, то есть таблица является проверяющей и различающей. Однако этот набор проверок является избыточным и необходимо провести оптимизацию их количества. Оптимизация тестов и выбор минимального количества проверок (признаков) осуществляется в несколько этапов [36].

Первый этап. Выполняется оценка проверок (признаков) на их информативность. На этом этапе отбрасываются те признаки или проверки, которые в строке имеют все нули или все единицы. То есть отбрасываются проверки (признаки), которые не различают состояния, занесенные в таблицу. В нашей таблице такой проверкой является проверка π_{12} .

Второй этап. Просматриваются все проверки (признаки) на предмет их тождественности отображения состояния, то есть просматривается таблица на предмет наличия одинаковых строк. Из тождественных признаков вбираются, как правило, те, которые проще всего измерить. В нашей таблице одинаковые строки соответствуют проверкам π_1 и π_{10} , а также π_8 и π_{11} . Следовательно, из представленного в таблице комплекса проверок следует исключить проверки π_{10} , π_{11} , π_{12} , как неинформативные.

Полученная в результате таблица также является проверяющей и различающей. Однако и эта совокупность проверок все еще остается избыточной. Если бы объект контроля был бы идеально приспособлен для диагностики, то минимальное число проверок J , необходимое для распознавания N состояний, определяется соотношением $J = \log_2 N$. В нашем случае для разделения восьми технических состояний выполняется девять проверок, что явно не соответствует этому соотношению. Поэтому проводится третий этап оптимизации, который может выполняться различными методами [24, 26, 35].

Наиболее часто используют метод определения минимального набора проверок (признаков) с помощью общей различающей логической функции и табличный метод минимизации теста по максимальному числу вхождений проверок в различающую функцию.

Первый метод является математически строгим, позволяет выбрать оптимальный тест, но он достаточно трудоемок. Поэтому рассмотрим более простой и наглядный табличный метод.

Табличный метод минимизации теста по максимальному числу вхождений проверок в различающую функцию

Перепишем таблицу 2.6, исключив из нее неинформативные проверки π_{10} , π_{11} , π_{12} . Полученная в результате исключения этих проверок таблица представлена ниже.

Таблица 2.7 Таблица состояний с избыточным числом проверок

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	W
π_1	1	0	1	1	1	1	1	1	7
π_2	1	0	1	1	1	1	0	1	12

π_3	1	0	0	0	1	1	1	1	15
π_4	1	0	0	0	0	0	1	1	15
π_5	1	0	0	0	0	1	1	1	16
π_6	1	1	0	0	1	0	0	0	15
π_7	1	0	0	1	1	1	0	0	16
π_8	1	1	1	1	0	0	0	0	16
π_9	1	1	1	0	1	0	1	1	12

Пусть в j -той строке результаты проверки π_j примут значение равное единице m_j раз, а значение равное нулю n_j раз. Под числом вхождений проверок (признаков) данной строки w_j понимают произведение количества нулей на количество единиц

$$W_j = m_j \cdot n_j. \quad (2.6)$$

В последнем столбце таблицы 2.7 приведены значения числа вхождений, подсчитанные для соответствующих проверок (строк). Максимальное число $W = 16$ для трех проверок π_5 , π_7 , π_8 . В тест следует выбрать одну из этих проверок. Выбирается тот признак или проверка, который проще измерить. Например, возьмём проверку под номером пять.

Далее таблица перестраивается таким образом, чтобы она разделялась на две части. В левой половине этой таблицы собираются все состояния, у которых результат пятой проверки равен единице (S_0, S_5, S_6, S_7), а в правой половине все состояния, для которых результат равен нулю (S_1, S_2, S_3, S_4) (табл. 2.8).

Таблица 2.8

	1 = 1				1 = 2				
	S_0	S_5	S_6	S_7	S_1	S_2	S_3	S_4	W
π_5	1	1	1	1	0	0	0	0	

π_1	1	1	1	1	0	1	1	1	$0+3=3$
π_2	1	1	0	1	0	1	1	1	$3+3=6$
π_3	1	1	1	1	0	0	0	1	$0+3=3$
π_4	1	0	1	1	0	0	0	0	$3+0=3$
π_6	1	0	0	0	1	0	0	1	$3+4=7$
π_7	1	1	0	0	0	0	1	1	$4+4=8$
π_8	1	0	0	0	1	1	1	0	$3+3=6$
π_9	1	0	1	1	1	1	0	1	$3+3=6$

На втором шаге также считается количество вхождений для каждой проверки (строки), как сумма вхождений проверок, подсчитанных для первой и второй половин таблицы 2.8

$$W = (n_j \cdot m_j)^1 + (n_j \cdot m_j)^2. \quad (2.7)$$

Полученные значения приведены в последнем столбце таблицы 2.8. Максимальное значение числа вхождений имеет проверка под номером семь π_7 . Перестроим таблицу 2.8 по π_7 таким образом, чтобы новая таблица 2.9 делилась на четыре части, и чтобы в каждой из новых частей были собраны состояния, где π_7 равна только единице или только нулю.

Таблица 2.9

	1 = 1		1 = 2		1 = 3		1 = 4		
	S_0	S_5	S_6	S_7	S_3	S_4	S_1	S_2	W
π_5	1	1	1	1	0	0	0	0	

π_7	1	1	0	0	1	1	0	0	
π_1	1	1	1	1	1	1	0	1	1
π_2	1	1	0	1	1	1	0	1	2
π_3	1	1	1	1	0	1	0	0	1
π_4	1	0	1	1	0	0	0	0	1
π_6	1	0	0	0	0	1	1	0	3
π_8	1	0	0	0	1	0	1	1	2
π_9	1	0	1	1	0	1	1	1	2

На третьем шаге количество вхождений для каждой проверки определяется как сумма вхождений, подсчитанных для каждой из четырех частей таблицы

$$W = (n_j \cdot m_j)^1 + (n_j \cdot m_j)^2 + (n_j \cdot m_j)^3 + (n_j \cdot m_j)^4 \quad (2.8)$$

Максимальное число вхождений имеет проверка π_6 . В результате проверки π_5 , π_7 , π_6 не различают только два состояния S_6 и S_7 . Из приведенных таблиц следует, что для их разделения необходимо выполнить проверку π_2 . Таким образом мы получаем минимальный тест для разделения восьми технических состояний, в которых может находиться объект, представленный таблицей 2.6. В этот тест следует ввести проверки π_5 , π_7 , π_6 и π_2 . При этом исходную таблицу 2.6 следует преобразовать к окончательному виду (таблица 2.10)

Таблица 2.10 Таблица состояний с минимальным набором проверок

	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7
π_2	1	0	1	1	1	1	0	1
π_5	1	0	0	0	0	1	1	1

π_6	1	1	0	0	1	0	0	0
π_7	1	0	0	1	1	1	0	0

Из всего вышеизложенного можно построить дерево алгоритма определения технического состояния объекта диагностики, представленного таблицами 2.6 и 2.10 (рис. 2.12).

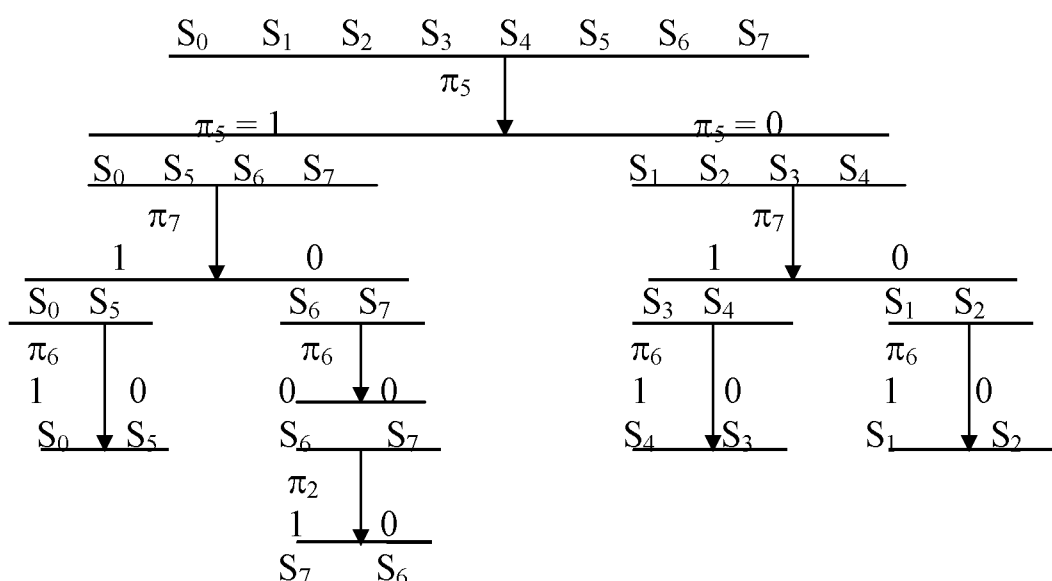


Рис. 2.12 Алгоритм определения состояния объекта

И функциональная модель и граф причинно-следственных связей в конечном итоге определяют математическую модель объекта в виде таблицы состояний. Задание оператора объекта диагностирования в табличной форме достаточно удобно. Однако в ряде случаев (например, когда параметры определены на непрерывном множестве) такое представление оператора невозможно. В таких ситуациях математическая модель может быть представлена в виде аналитических зависимостей между входными возмущениями, параметрами технического состояния и диагностическими параметрами.

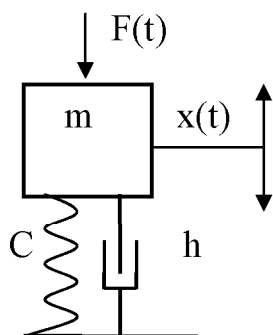
В технической диагностике математические (диагностические) модели объектов, устанавливающие связь между входными возмущениями, параметрами технического состояния и диагностическими параметрами (признаками) в виде аналитических зависимостей (уравнений) называются *аналитическими моделями*. Эти аналитические модели (зависимости) чаще всего могут быть представлены в виде алгебраических или дифференциальных уравнений. Познакомимся с некоторыми подобными моделями.

2.5 Аналитические модели в виде дифференциальных уравнений

Наиболее часто аналитические модели в виде дифференциальных уравнений применяются в вибрационной диагностике. Это связано с тем, что вибрация (упругие механические колебания) механизмов - это их реакция на действие приложенных возмущающих сил, а основным математическим аппаратом для анализа вибрации (упругих механических колебаний) технических устройств является аппарат дифференциальных уравнений [5, 7, 9, 13].

Вибрационное (колебательное) поведение технических систем описывается дифференциальным уравнением (системой уравнений), у которого (как правило) в правой части находится аналитическое выражение, определяющее зависимость возмущающих сил от времени. В левую часть этого уравнения (системы уравнений) помещены производные по времени перемещений точек механизма с коэффициентами, характеризующими инерционные (массовые), упругие (жесткостные) и демпфирующие (рассеивающие) свойства этого механизма. Отсюда следует, что анализ виброакустических процессов для целей диагностики позволяет выявить такие неисправности, которые приводят к изменению действующих на механизм возмущающих сил или к изменению инерционных, жесткостных и демпфирующих свойств этого механизма (например, разбалансировку роторов, нарушение соосности роторов, возникновение трещин, неисправности подшипниковых узлов и т.п.). Поясним это на простейшем примере.

Представим диагностируемый механизм в виде простой одно массовой колебательной системы (рис.2.13)



$F(t)$ - возмущающая внешняя сила;
 m - масса системы;
 C - коэффициент жесткости системы;
 h - коэффициент демпфирования в системе
 $x(t)$ - перемещение центра масс по действием возмущающей силы.

Рис. 2.13 Простейшая одно массовая колебательная система

Колебания центра масс этой системы описывается следующим дифференциальным уравнением

$$m \cdot x''(t) + h \cdot x'(t) + C \cdot x(t) = F(t). \quad (2.9)$$

Общее решение этого уравнения, описывающее зависимость перемещения центра масс от времени, записывается в следующем виде

$$x(t) = x_{CB}(t) + x_B(t), \quad (2.10)$$

где $x_{CB}(t)$ - определяет свободные колебания центра масс, а $x_B(t)$ - вынужденные.

Для того чтобы в ходе решения уравнения (2.9) получить конкретные аналитические зависимости от времени свободных и вынужденных колебаний необходимо определить в конкретном аналитическом виде зависимость от времени возмущающей силы [5, 20]. При изучении вибрационных процессов чаще всего используется представление, как самих колебаний, так и действующих сил в виде ряда Фурье. Поэтому примем, что возмущающая сила, действующая на систему, является гармонической (т. е. $F(t) = F_0 \cdot \sin(\omega t)$, где F_0 - амплитуда возмущающей силы, ω - частота возмущающей силы). С учетом этого выражение (2.9) можно переписать в виде

$$m \cdot x''(t) + h \cdot x'(t) + C \cdot x(t) = F_0 \cdot \sin(\omega t). \quad (2.11)$$

Решая это уравнение, получим, что свободные колебания являются затухающими и определяются выражением

$$x_{CB}(t) = \frac{F_0}{C} \cdot e^{-\frac{h}{2m}t} \cdot \sin(\omega_1 t), \quad (2.12)$$

где $\omega_1 = \sqrt{\omega_0^2 + \frac{h^2}{4m^2}}$ - частота свободных колебаний;

$\omega_0 = \sqrt{\frac{C}{m}}$ - собственная частота колебаний системы.

Необходимо отметить, что в реальных системах $m \gg h$, поэтому свободные колебания механических систем совершаются с частотой практически равной собственной частоте системы (т.е. $\omega_1 \approx \omega_0$). Воспроизвести свободные колебания в реальных условиях можно путем тестового воздействия на объект коротким импульсом возмущающей силы, то есть ударом.

При возбуждении системы (см. рис. 2.13) ударом она будет совершать свободные колебания с частотой, равной собственной частоте системы. Амплитуда колебаний будет зависеть от соотношения между амплитудой силы и жесткостью системы. Скорость затухания колебаний будет определяться величиной соотношения демпфирования и массы в системе. Отсюда следует, что обнаружить неисправность объекта путем анализа его свободных колебаний можно в том случае, если эта неисправность вызовет изменение частоты свободных колебаний или скорости их затухания. Или другими словами, если неисправность повлияет на динамические (жесткостные, демпфирующие, массовые) характеристики системы.

Если поведение системы описывается уравнением (2.11) и на нее постоянно действует гармоническая сила, то через определенный интервал времени свободные колебания затухнут, и система будет совершать только вынужденные колебания. Эти колебания будут также гармоническими, а их аналитическая зависимость определяется выражением

$$x_B(t) = X_0 \sin(\omega t), \quad (2.13)$$

где X_0 - амплитуда вынужденных колебаний; ω - частота вынужденных колебаний.

Частота вынужденных колебаний равна частоте возбуждающей силы, а амплитуда колебаний зависит от амплитуды возбуждающей силы и коэффициента динамичности β :

$$X_0 = F_0 \beta. \quad (2.14)$$

Коэффициент динамичности зависит только от характеристик системы и для нашего случая определяется следующим соотношением

$$\beta = \frac{1}{m \cdot \sqrt{(\omega_0^2 - \omega^2)^2 + 4 \cdot \frac{h^2}{m^2} \cdot \omega^2}}. \quad (2.15)$$

Максимальное значение коэффициент динамичности принимает при равенстве частоты возбуждающей силы собственной частоте системы $\omega = \omega_0$. Такое явление называется *резонансом*, и говорят, что система совершает *резонансные колебания*. Как следует из соотношения (2.15), амплитуда резонансных колебаний зависит от величины возбуждающей силы и демпфирующих свойств системы.

Приведенные соотношения свидетельствуют о том, что диагностика объектов путем анализа вынужденных колебаний возможна в том случае, если неисправности вызывают изменение возбуждающей силы (ее величины или частоты) или изменение динамических свойств самого объекта (массы, жесткости, демпфирования).

Представленные выше зависимости выходного сигнала(колебаний) описывают только одно техническое состояние механизма (исправное), для получения описаний неисправных состояний необходимо установить зависимости между параметрами силы (входным воздействием) и дефектами или зависимости между массой, жесткостью, демпфирования и дефектами, то есть для вибрационной диагностики характерно задание неявных моделей.

Реальные механизмы обладают ∞ числом собственных форм и частот колебаний, для их диагностирования по вибрационным параметрам на практике можно применить следующий приём [13]. Каждой собственной форме и частоте колебаний реального механизма устанавливается соответствие с гипотетической одно массовой системой, и реальный механизм рассматривается как набор одно массовых систем. Любые механизмы с определёнными допущениями можно представить набором одно массовых систем. Сложное входное возмущение можно представить в виде ряда Фурье, что позволяет анализировать реакцию механизма на отдельное гармоническое возмущение.

При таком подходе основная задача это правильно определить основные (определяющие) собственные формы и частоты диагностируемого механизма. Исправное техническое состояние будет описываться характерными только для

исправного объекта амплитудно-частотным составом возмущающей силы и динамическими свойствами (инерционными, жесткостными, демпфирующими) выбранных определяющих форм колебаний. Для задания неисправных технических состояний необходимо установить зависимости амплитудно-частотного состава возмущающей силы и динамических свойств определяющих форм колебаний от конкретных неисправностей механизма.

При более общем подходе реальную механическую (колебательную) систему можно представить как многомерную систему с m входами и n выходами (рис. 2.14)

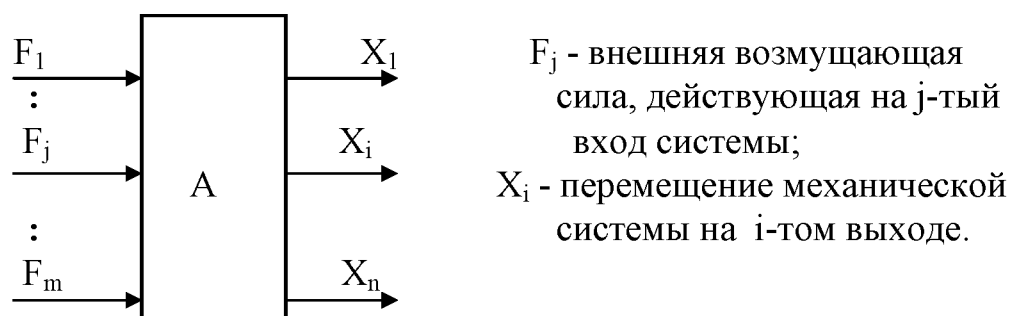


Рис.2.14 Упрощенное представление диагностируемого объекта в виде многомерной колебательной системы

В этом случае оператор A представляет собой систему дифференциальных уравнений с n степенями свободы

$$[M] \cdot [X''] + [H] \cdot [X'] + [C] \cdot [X] = [F] \quad (2.16)$$

Если система имеет n уравнений, то симметричные матрицы $[M]$, $[H]$, $[C]$ имеют размерность $n \times n$ и описывают соответствующие инерционные, демпфирующие и жесткостные свойства системы, а векторы $[X]$ и $[F]$ имеют размерность n и описывают координаты системы и возмущающие силы. Отметим, что и в этом случае диагностическая модель является неявной, что характерно для задач вибрационной диагностики.

Рассмотрим в качестве примера, как изменятся вибрационные характеристики стержня или пластины, которыми можно смоделировать вал или лопатку, при появлении трещины [14]. Представим себе балку на шарнирных опорах или с заземленным концом, у которой появилась трещина (рис. 2.15)

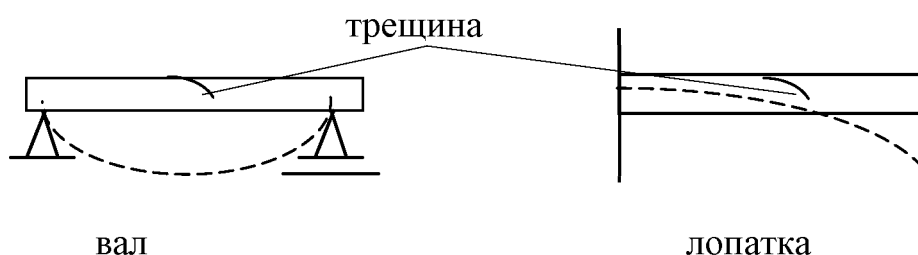


Рис.2.15 Схема балки с трещиной

Будем считать, что балка совершает колебания по первой (определяющей) форме, которая на рисунке изображена пунктирной линией. При колебаниях «берега» трещины будут то сжиматься (при сжатии), то разжиматься (при растяжении). Очевидно, что при сжатии материал ведет себя как сплошной. При растяжении, поскольку «берега» трещины раскрываются, то жесткость балки уменьшается, так как трещина не работает на растяжение. Таким образом, изделие с трещиной имеет различные жесткости в направлениях сжатия и растяжения. Ранее отмечалось, что жесткость определяет динамические свойства (собственную частоту) системы. Следовательно, у изделия с трещиной изменится собственная частота системы, поскольку она определяется соотношением жесткости и массы (см. формулу 2.16).

Так как мы рассматриваем колебания балки по первой (определяющей) форме колебаний, то смоделируем ее одно массовой системой с соответствующими параметрами (собственная частота системы должна быть равна собственной частоте исследуемой балки). Для упрощения рассмотрения не будем учитывать демпфирования в системе. Одно массовая система, моделирующая балку без трещины при принятых допущениях, представлена на рис.2.16. Свободные колебания этой системы описываются дифференциальным уравнением следующего вида: $mx'' + Cx = 0$, которое преобразуется в следующее выражение

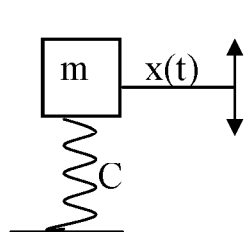


Рис.2.16

$$x'' + \frac{C}{m} \cdot x = 0, \quad (2.17)$$

$$x'' + \omega_0^2 \cdot x = 0$$

где $\omega_0 = \sqrt{C/m}$ собственная частота системы.

Определим начальные условия для момента времени $t = 0$. Будем считать, что для начального момента времени координата центра тяжести массы равна нулю ($x=0$), а его скорость имеет конечное значение V_0 ($x' = V_0$). Уравнение свободных колебаний без демпфирования при наличии в начальный момент скорости определяется следующим уравнением

$$x(t) = \frac{V_0}{\omega_0} \cdot \sin(\omega_0 t). \quad (2.18)$$

Из этого уравнения следует, что при отсутствии трещины свободные колебания совершаются с частотой, равной собственной частоте системы.

Изменим систему, изображенную на рис.2.16 таким образом, чтобы она отражала изменение динамических характеристик при появлении трещины. Наличие трещины учтем добавочной пружиной $C_{\text{доб}}$ (рис.2.17). $C_{\text{доб}}$ - жесткость, которая учитывается только при сжатии (при $x < 0$).

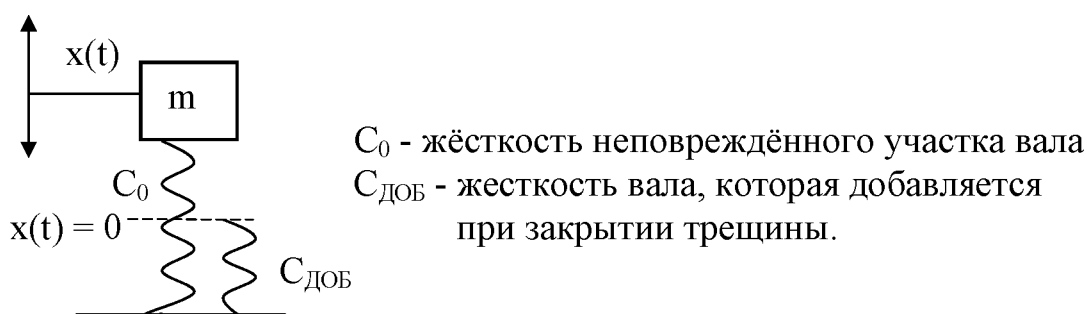


Рис. 2.17 Схема системы, имитирующая трещину

Жёсткость C_0 - работает и на растяжение и на сжатие. Она моделирует жесткость неповреждённого участка вала. В первом приближении можно считать, что отношение добавочной жесткости к основной жесткости равно отношению длины трещины $l_{\text{ТР}}$ к характерному размеру изделия (например, лопатки) $l_{\text{ЛОП}}$ в направлении трещины $\frac{C_{\text{ДОб}}}{C_0} \approx \frac{l_{\text{ТР}}}{l_{\text{ЛОП}}}$. Следовательно, в реальной ситуации

$C_{\text{ДОб}} \ll C_0$. Введем следующие обозначения $\omega_{\text{ОСН}} = \sqrt{\frac{C_0}{m}}$; $\omega_{\text{ДОб}} = \sqrt{\frac{C_{\text{ДОб}}}{m}}$;

$C = C_0 + C_{\text{ДОб}}$; $\omega = \sqrt{\frac{C}{m}}$. В силу $C_{\text{ДОб}} \ll C_0$ имеем $\omega \approx \omega_{\text{ОСН}}$.

С учетом принятых обозначений и допущений свободные колебания системы, изображенной на рис.2.17 будут описываться следующими уравнениями

$$\begin{cases} x'' + \omega_{\text{ОСН}}^2 \cdot x = 0; \text{при } x \geq 0 \\ x'' + \omega^2 \cdot x = 0; \text{при } x \leq 0 \end{cases} \quad (2.19)$$

При задании начальных условий $t = 0$; $x = 0$; $x' = V_0$ можно получить решение этих уравнений. Решение будет периодической функцией с периодом $T \approx \pi/\omega$. Время будет изменяться в интервале от $-\pi/\omega$ до $\pi/\omega_{\text{ОСН}}$. Решение дифференциальных уравнений (2.23) в указанном временном интервале будет иметь следующий вид

$$\begin{cases} x(t) = \frac{V_0}{\omega_{\text{ОСН}}} \cdot \sin(\omega_{\text{ОСН}} \cdot t) \text{ при } 0 \leq t \leq \frac{\pi}{\omega_{\text{ОСН}}} \\ x(t) = \frac{V_0}{\omega} \cdot \sin(\omega \cdot t) \text{ при } -\frac{\pi}{\omega} \leq t \leq 0 \end{cases} \quad (2.20)$$

Функция, заданная уравнениями (2.20) несимметрична относительно начала координат. Справа и слева от начала координат несколько отличаются амплитуды и частоты колебаний. Применим к функции $x(t)$ разложение в ряд Фурье, что

позволит представить ее в виде суммы отдельных гармонических составляющих:

$$x(t) = \frac{a_0}{2} + \sum_{i=1}^{\infty} [a_i \cdot \cos(i\omega_0 t) + b_i \cdot \sin(i\omega_0 t)],$$

где $\omega_0 = \frac{2 \cdot \omega_{OCH} \cdot \omega}{\omega_{OCH} + \omega}$, а коэффициенты ряда Фурье a_i и b_i определяются известными соотношениями

$$a_i = \frac{2}{T} \cdot \int_{-\pi/\omega}^{\pi/\omega_{OCH}} x(t) \cdot \cos(i\omega_0 t) \cdot dt$$

$$b_i = \frac{2}{T} \cdot \int_{-\pi/\omega}^{\pi/\omega_{OCH}} x(t) \cdot \sin(i\omega_0 t) \cdot dt$$

Из анализа последних трех выражений следует, что свободные колебания системы, изображенной на рис. 2.17, не являются гармоническими и содержат несколько составляющих. Например, для значения $i = 1$ будем иметь $a_1 = 0$ $b_1 = V_0 / \omega$. Для значения $i = 2$ соответственно получим следующие значения $a_2 = 0$; $b_2 = \frac{1}{3} \cdot V_0 \cdot \frac{\omega_{ДОБ}^2}{\omega^3}$.

Отсюда отношение амплитуд второй и первой гармонических составляющих будет выражаться следующим уравнением

$$\frac{b_2}{b_1} = \frac{1}{3} \cdot \frac{\omega_{ДОБ}^2}{\omega^2} = \frac{1}{3} \cdot \frac{C_{ДОБ}}{C} = \frac{1}{3} \cdot \frac{l_{ТР}}{l_{ЛОП}}.$$

Обобщая полученные результаты, можно сделать следующие выводы:

1. При появлении трещины свободные колебания системы становятся негармоническими, в сигнале присутствуют высшие гармоники, в частности, вторая гармоника.

2. Отношение амплитуды второй (верхней) гармоники к амплитуде первой (основной) гармоники равно 1/3 отношения размера трещины к характерному размеру изделия.

3. Основная частота свободных колебаний изделия с трещиной меньше частоты свободных колебаний изделия без трещины, поскольку $\omega_0 < \omega_{OCH}$.

В заключение следует отметить, что аналитические модели в виде дифференциальных уравнений применяются не только в вибрационной диагностике. Они используются также, например, в методах диагностирования по температурным полям.

Кроме дифференциальных уравнений аналитические модели могут быть представлены алгебраическими уравнениями [1, 10, 27, 31, 41].

2.6 Аналитические модели на основе алгебраических уравнений

Если связь между входным возмущением, внутренними параметрами объекта диагностирования и выходными диагностическими параметрами представлена в виде алгебраического уравнения или системы алгебраических уравнений, то мы

имеем аналитическую модель алгебраического вида. В зависимости от физического характера объекта диагностирования алгебраические уравнения могут быть линейными или нелинейными. В простейшем случае при описании объекта диагностики подобными моделями задается система линейных алгебраических уравнений, в которых количество уравнений равно числу неизвестных параметров технического состояния или числу диагнозов (то есть числу выбранных для диагноза технических состояний).

Рассмотрим случай представления моделей в виде системы линейных алгебраических уравнений. В общем виде :

$$[A] \cdot [X] + [B] \cdot [Z] + [C] \cdot [Y] = 0. \quad (2.21)$$

В этом уравнении X , Z , Y соответственно вектор входных возмущений действующих на объект, вектор внутренних параметров объекта, вектор выходных диагностических параметров. Матрицы A , B , C - являются матрицами коэффициентов уравнения.

При проведении диагностирования, как правило, располагают значениями диагностических параметров (т.е. известен вектор $[Y]$), по которым необходимо определить значения входных или внутренних параметров (т.е. определить векторы $[X]$ или $[Z]$). С точки зрения математического решения задачи диагностики между векторами входных параметров $[X]$ и вектором внутренних параметров $[Z]$ нет различий. Поэтому векторы $[X]$ и $[Z]$ можно объединить в общий вектор $[R]$, который называется вектором влияющих факторов и тогда уравнение (2.21) примет вид :

$$[Y] = [D] \cdot [R]. \quad (2.22)$$

Вектор $[Y]$ с компонентами y_i называют еще вектором отклика системы на вектор возмущающих факторов $[R]$ с компонентами r_j .

Большинство линейных моделей, приведенных к окончательному виду, разрешены относительно выходных диагностических параметров (откликов) и описываются системой уравнений вида (2.26). Каждое отдельное уравнение в этой системе представляет собой зависимость одного отклика от нескольких влияющих факторов.

В большинстве случаев для задач диагностики имеют значение не абсолютные величины параметров, а отклонение этих параметров от нормы. Поэтому, с учетом линейности системы уравнений, заданных выражением (2.22), его предпочтительнее переписать в виде

$$[\Delta Y] = [D] \cdot [\Delta R], \quad (2.23)$$

где $[\Delta Y]$ - вектор отклонений диагностических параметров (откликов); $[\Delta R]$ - вектор отклонений влияющих факторов.

Линейность модели позволяет представить отклонения в относительном виде, что является особенно удобным :

$$[\delta Y] = [D] \cdot [\delta R]. \quad (2.24)$$

Вектор $[\delta Y]$ есть вектор относительных отклонений откликов (диагностических параметров), который задается своими компонентами δy_i . Вектор $[\delta R]$ есть вектор относительных отклонений влияющих факторов, который задается своими компонентами δr_j .

Матрица $[D]$ называется матрицей коэффициентов влияния. Каждая строка такой матрицы содержит коэффициенты, которые показывают во сколько раз (на сколько %) изменится данный отклик (выходной диагностический параметр) при изменении одного из возмущающих факторов на 1%.

Матрицу $[D]$ удобно представлять в табличном виде. Такая таблица называется таблицей коэффициентов влияния и выглядит следующим образом. В заголовках столбцов указываются относительные отклонения влияющих факторов δr_j . В заголовках строк указываются относительные отклонения откликов (диагностических параметров) δy_i . В клетке, лежащей на пересечении i -той строки и j -того столбца, записывается соответствующий коэффициент влияния d_{ij} , который показывает на сколько процентов изменится i -тый отклик при изменении j -того влияющего фактора. Такая таблица (таблица 2.11) похожа на таблицу состояния, однако в клетках этой таблицы размещены не нули или единицы, а коэффициенты влияния.

Таблица 2.11 Общий вид таблицы коэффициентов влияния

$\delta y_i \setminus \delta r_j$	δr_1		δr_j		δr_m
δy_1	d_{11}		d_{1j}		d_{1m}
δy_2	d_{21}		d_{2j}		d_{2m}
δy_i	d_{i1}		d_{ij}		d_{im}
δy_n	d_{n1}		d_{nj}		d_{nm}

Использование таблицы коэффициентов влияния уже позволяет решать задачи определения технического состояния. Например, зная изменение откликов оценить изменение влияющих факторов, которые непосредственно связаны с параметрами технического состояния.

Отметим, что на практике диагностические задачи, чаще всего, формулируются следующим образом: по известным измеренным значениям диагностических параметров (откликов) y_i определить техническое состояние

(поставить диагноз). Если считать, что влияющий фактор r_j определяет какую либо неисправность, а все отклики (диагностические параметры) могут быть измерены, то предыдущее выражение (2.24) может быть представлено в виде, позволяющим напрямую решать сформулированную выше задачу:

$$[\delta R] = [D]^{-1}[\delta Y] \quad (2.25)$$

Матрица $[D]^{-1}$ является обратной для матрицы коэффициентов влияния $[D]$ и называется диагностической [1]. Диагностическую матрицу также можно представить в виде таблицы, где заголовками столбцов являются отклонения измеряемых параметров δy_i , а заголовками строк - диагнозы технического состояния, определенные отклонениями влияющих факторов δr_j . Данная таблица также похожа на таблицу состояний и ее можно использовать для решения задач диагностики.

Отметим также, что в этом классе моделей могут использоваться алгебраические выражения, в которых несколько откликов связаны с несколькими факторами и они не разрешены относительно откликов, например следующего вида:

$$[C][\delta Y] = [A][\delta R],$$

где $[C]$ и $[A]$ - матрицы коэффициентов влияния соответственно для откликов и для влияющих факторов. Эту систему уравнений можно решать путем последовательной подстановки или в матричном виде:

$$[\delta Y] = [C]^{-1}[A][\delta R].$$

Сделав подстановку $[D] = [C]^{-1}[A]$ получим выражение, соответствующее формуле (2.24). Таким образом, соответствующими преобразованиями и в этом случае можно получить искомую матрицу коэффициентов влияния или диагностическую матрицу.

Мы рассмотрели случай представления модели в виде системы линейных алгебраических уравнений. Однако модель может быть представлена в виде системы нелинейных алгебраических уравнений. Если первоначальные алгебраические уравнения, описывающие объект представлены в нелинейном виде, необходимо произвести их линеаризацию. Линеаризацию проводят в зависимости от вида нелинейности функции. Ниже рассмотрим несколько наиболее часто встречающихся варианта [1].

1. Пусть алгебраическое уравнение, связывающее отклик y с факторами r_j , имеет вид:

$$y = C \cdot f_1(r_1) \cdot f_2(r_2) \cdot \dots \cdot f_m(r_m), \quad (2.26)$$

то есть в виде произведения функций.

При задании функции в таком виде для линеаризации проводится сначала логарифмирование, а затем дифференцирование исходного выражения.

После логарифмирования выражения (2.26) получим следующее уравнение:

$$\lg y = \lg C + \lg f_1(r_1) + \lg f_2(r_2) + \dots + \lg f_m(r_m).$$

После дифференцирования получим соответствующее уравнение:

$$\frac{dy}{y} = 0 + \frac{\partial f_1(r_1)}{\partial r_1} \cdot \frac{dr_1}{f_1(r_1)} + \frac{\partial f_2(r_2)}{\partial r_2} \cdot \frac{dr_2}{f_2(r_2)} + \dots + \frac{\partial f_m(r_m)}{\partial r_m} \cdot \frac{dr_m}{f_m(r_m)}.$$

Каждый член в правой части уравнения умножим и разделим на абсолютное значение соответствующего влияющего фактора. Например, первый член будет иметь вид

$$\frac{\partial f_1(r_1)}{\partial r_1} \cdot \frac{dr_1}{f_1(r_1)} \cdot \frac{r_1}{r_1}.$$

Введем следующие обозначения: $\frac{dr_j}{r_j} = \delta r_j$ - относительное изменение j-того влияющего фактора; $\frac{\partial f_j(r_j)}{f_j(r_j)} \cdot \frac{r_j}{\partial r_j} = d_j$ - коэффициент влияния j-того влияющего фактора на относительное изменение диагностического параметра у. Если произвести соответствующую подстановку, то получим окончательное уравнение в следующем виде:

$$\delta y = d_1 \cdot \delta r_1 + d_2 \cdot \delta r_2 + \dots + d_m \cdot \delta r_m. \quad (2.27)$$

Полученное в результате уравнение является линейным уравнением и связывает между собой относительное изменение измеряемого диагностического параметра (отклика) с относительными изменениями определяющих факторов.

2. Пусть уравнение, связывающие отклик и влияющие факторы, имеет более общий вид:

$$y = f(r_1, r_2, \dots, r_m). \quad (2.28)$$

В этом случае логарифмирование нецелесообразно. Уравнение вида (2.28) следует дифференцировать непосредственно:

$$dy = \frac{\partial f(r_1, r_2, \dots, r_m)}{\partial r_1} dr_1 + \dots + \frac{\partial f(r_1, r_2, \dots, r_m)}{\partial r_m} dr_m.$$

Полученное выражение производной далее следует разделить на исходное уравнение, описываемое формулой (2.28):

$$\frac{dy}{y} = \frac{1}{f(r_1, r_2, \dots, r_m)} \frac{\partial f(r_1, r_2, \dots, r_m)}{\partial r_1} dr_1 + \dots$$

Здесь для упрощения записи представлен только первый член правой части.

Каждый член правой части, как и в первом случае, умножается и делится на абсолютное значение влияющего фактора. Для примера, первый член правой части будет иметь вид:

$$\frac{1}{f(r_1, r_2, \dots, r_m)} \frac{\partial f(r_1, r_2, \dots, r_m)}{\partial r_1} dr_1 \frac{r_1}{r_1}$$

Введем, как и в предыдущем случае, понятие коэффициента влияния и обозначим его d_j : $d_j = \frac{\partial f(r_1, r_2, \dots, r_m)}{f(r_1, r_2, \dots, r_m)} \cdot \frac{r_j}{r_j}$. Тогда, используя понятия, относительное изменение влияющего фактора и относительное изменение отклика, после окончательных преобразований мы получим

$$\delta y = d_1 \cdot \delta r_1 + d_2 \delta r_2 + \dots + d_m \delta r_m. \quad (2.29)$$

Таким образом, мы вновь получили линейное алгебраическое уравнение, разрешенное относительно отклика.

3.В ряде случаев не удастся разрешить уравнение относительно одного отклика. Предположим, что функция имеет вид:

$$\varphi_1(y_1) \cdot \varphi_2(y_2) \cdot \dots \cdot \varphi_n(y_n) = C \cdot f_1(r_1) \cdot \dots \cdot f_n(r_n). \quad (2.30)$$

Поскольку уравнение (2.30) представляет собой произведение нескольких сомножителей, то первоначально следует провести логарифмирование этого выражения:

$$\lg \varphi_1(y_1) + \dots + \lg \varphi_n(y_n) = \lg C + \lg f_1(r_1) + \dots + \lg f_n(r_n)$$

Следующим этапом проведем дифференцирование левой и правой частей. Для упрощения записи оставим только первые члены левой и правой частей:

$$\frac{\partial \varphi_1(y_1)}{\partial y_1} \frac{dy_1}{\varphi_1(y_1)} + \dots = \frac{\partial f_1(r_1)}{\partial r_1} \cdot \frac{dr_1}{f_1(r_1)} + \dots$$

Далее в левой части каждый член умножаем и делим на абсолютное значение соответствующего диагностического параметра, а в правой части каждый член умножаем и делим на абсолютное значение фактора. Введем, как и ранее, понятие коэффициента влияния для каждого влияющего фактора (обозначим d_j)

и, аналогично, понятие коэффициента влияния при соответствующих откликах (обозначим b_i). Используя эти понятия окончательное выражение для относительных изменений откликов и относительных изменений влияющих факторов запишется в виде (для упрощения записи оставим только первые два члена левой и правой частей):

$$b_1 \delta y_1 + b_2 \delta y_2 + \dots = d_1 \delta r_1 + d_2 \delta r_2 + \dots \quad (2.31)$$

Полученное выражение так же определяет линейное алгебраическое уравнение.

4. Наиболее общий случай имеет место тогда, когда и левая и правая части уравнения описываются алгебраическими уравнениями общего вида:

$$\phi(y_1, y_2, \dots, y_n) = f(r_1, r_2, \dots, r_n). \quad (2.32)$$

Поскольку выражение (2.32) нельзя представить в виде произведения сомножителей, то логарифмирование не имеет смысла. На первом этапе лианеризации проводится дифференцирование выражения:

$$\frac{\partial \phi(y_1, y_2, \dots, y_n)}{\partial y_1} dy_1 + \dots = \frac{\partial f(r_1, r_2, \dots, r_n)}{\partial r_1} dr_1 + \dots$$

Здесь для упрощения записи представлены только первые члены левой и правой частей продифференцированного уравнения. В ходе дальнейших преобразований разделим полученное выражение на исходное. Как и ранее, для упрощения записи представим только первые члены левой и правой частей:

$$\frac{\partial \phi(y_1, y_2, \dots, y_n)}{\partial y_1} \frac{dy_1}{\phi(y_1, y_2, \dots, y_n)} + \dots = \frac{\partial f(r_1, r_2, \dots, r_n)}{\partial r_1} \frac{dr_1}{f(r_1, r_2, \dots, r_n)} + \dots$$

Если теперь каждый член левой части поделить и умножить на абсолютное значение соответствующего диагностического параметра, а каждый член правой части поделить и умножить на значение соответствующего фактора, воспользоваться введенными ранее понятиями коэффициентов влияния и провести соответствующие преобразования, то в окончательном виде получим уравнение, соответствующее выражению (2.31).

Приведенные примеры показывают, что если объект диагностики описывается моделью в виде системы алгебраических уравнений, в любой ситуации можно перейти к системе линейных алгебраических уравнений, выраженных относительно каждого параметра или отклика. Путем соответствующих преобразований определить (получить) матрицу (таблицу) коэффициентов влияния или диагностическую матрицу (таблицу).

Рассмотрим пример применения диагностических моделей в виде алгебраических уравнений. Наиболее ярким примером использования подобного

класса моделей являются модели, которые созданы для диагностирования проточной части авиационных газотурбинных двигателей (ГТД) и их систем регулирования по термогазодинамическим параметрам [1, 10, 30, 31, 42].

Для построения диагностических моделей авиационных ГТД при решении задач диагностирования ГВТ и систем регулирования по термогазодинамическим параметрам в качестве исходных используются основные уравнения рабочего процесса, известные из теории ГТД [32, 33]. Проиллюстрируем это на примере двух простейших уравнений: уравнения, связывающего отношения давлений в узлах двигателя, и уравнения баланса мощностей турбины и компрессора.

Уравнение, связывающее отношения давлений в разных сечениях двигателя, имеет следующий вид:

$$\frac{P_c}{P_H} = \frac{P_o}{P_H} \cdot \frac{P_K}{P_o} \cdot \frac{P_o}{P_K} \cdot \frac{P_T}{P_o} \cdot \frac{P_c}{P_T},$$

где P_H - давление в невозмущенном потоке; P_o - давление на входе в компрессор двигателя; P_K - давление за компрессором двигателя; P_T - давление перед турбиной; P_c - давление на срезе сопла.

Обозначив соответствующие отношения давлений через значения степени повышения или понижения давления π и через коэффициенты восстановления давления σ , это уравнение можно переписать в следующем виде:

$$\pi_c = \pi_v \cdot \pi_K \cdot \sigma_{KC} \cdot \frac{1}{\pi_T} \cdot \sigma_c.$$

Это уравнение нелинейное и выражается через произведение сомножителей. Поэтому для его линеаризации необходимо провести сначала логарифмирование затем дифференцирование и далее соответствующие преобразования (см. первый вариант линеаризации). Если рассматривается работа двигателя на земле, то можно считать $\pi_v = \text{const}$. Выполнив необходимые преобразования получим линейное алгебраическое уравнение в относительных отклонениях:

$$\delta\pi_c = \delta\pi_K + \delta\sigma_{KC} + \delta\sigma_c - \delta\pi_T.$$

Данное уравнение устанавливает связь между относительными изменениями параметров, характеризующих техническое состояние компрессора - π_K , турбины - π_T , камеры сгорания - σ_{KC} , сопла - σ_c .

Уравнение баланса мощностей на валу двигателя устанавливает соотношение между работой турбины L_T , работой потребляемой компрессором L_K , а также их коэффициентами полезного действия η_T и η_K . Это уравнение имеет вид:

$$\frac{L_K}{\eta_K} = L_T \cdot \eta_T.$$

После выполнения линеаризации это уравнение в относительных изменениях параметров запишется следующим образом:

$$\delta L_K - \delta \eta_K = \delta L_T + \delta \eta_T .$$

Из теории авиационных ГТД известно, что работу, создаваемую турбиной, можно выразить через газовую постоянную R_T , температуру газа T_T , степень расширения газа на турбине π_T и показатель адиабаты k следующим образом:

$$L_T = \frac{k}{k-1} R_T \cdot T_T \left(1 - \pi_T^{-\frac{k-1}{k}} \right).$$

Выполнив линеаризацию этого уравнения, мы получим соотношение для относительных изменений параметров в следующем виде:

$$\delta L_T = A_T \cdot \delta T_T + A_\pi \cdot \delta \pi_T .$$

Работа, потребляемая компрессором, также может быть выражена через газовую постоянную, угловую скорость вращения U (которая прямо пропорциональна частоте вращения ротора n), степень сжатия компрессора π_K и показатель адиабаты:

$$L_K = \frac{k}{k-1} R_T \cdot U \left(\pi_K^{\frac{k-1}{k}} - 1 \right).$$

После выполнения линеаризации соотношение между относительными изменениями параметров запишется следующим образом:

$$\delta L_K = A_n \cdot \delta n + A_\pi \cdot \delta \pi_K .$$

Воспользовавшись этими соотношениями уравнение баланса мощностей, выраженное в относительных изменениях параметров:

$$A_n \cdot \delta n + A_\pi \cdot \delta \pi_K - \delta \eta_{жс} = A_T \cdot \delta T_T + A_\pi \cdot \delta \pi_T + \delta \eta_T .$$

Последнее выражение устанавливает связь между параметрами, являющимися непосредственными параметрами технического состояния (степень сжатия (расширения), коэффициенты полезного действия), и измеряемыми (диагностическими) параметрами (частота вращения ротора, температура газов). Полученное уравнение является линейным алгебраическим уравнением с соответствующими коэффициентами влияния.

Подобным образом преобразуются и другие уравнения, известные из теории авиационных ГТД. В результате получаем систему алгебраических линейных уравнений, порядок которой зависит от типа двигателя и степени детализации протекающих в нем процессов [1, 10]. Уравнения включают в себя измеряемые параметры, параметры технического состояния и коэффициенты влияния. Наиболее часто в качестве параметров технического состояния (влияющих

факторов) выбираются коэффициенты полезного действия и площади сечений (в первую очередь сопловых аппаратов). В качестве диагностических признаков выбираются частота вращения ротора, температура газа, расход топлива. Коэффициенты влияния определяются расчетным или экспериментальным путем. После определения коэффициентов влияния можно построить матрицу этих коэффициентов (диагностическую матрицу) или соответствующую таблицу. Полученные таблицы непосредственно используются для решения практических задач.

В качестве примера приведем фрагмент матрицы коэффициентов влияния, полученной в работе [10] для двигателя НК-8-2У (таблица 2. 12).

Таблица 2.12

	$\delta\eta_{\text{КНД}}$	$\delta\eta_{\text{КВД}}$	$\delta F_{\text{СА}}$	$\delta\eta_{\text{ТВД}}$	$\delta\eta_{\text{ТНД}}$
$\delta n_{\text{ВД}}$	-0,3	0,4	-0,4	0,6	-0,3
$\delta G_{\text{Т}}$	-1,0	-0,4	0,2	-0,5	-1,2

В этой таблице в качестве диагностических параметров представлены частота вращения ротора высокого давления $n_{\text{ВД}}$ и расход топлива $G_{\text{Т}}$. В качестве параметров технического состояния представлены коэффициенты полезного действия компрессоров низкого и высокого давлений $\eta_{\text{КНД}}$ и $\eta_{\text{КВД}}$, турбин низкого и высокого давлений $\eta_{\text{ТНД}}$ и $\eta_{\text{ТВД}}$, а также площадь соплового аппарата турбины высокого давления $F_{\text{СА}}$.

Одним из простейших методов анализа подобных таблиц является «метод знаков». Знак $-$ у коэффициента влияния свидетельствует о том, что диагностический параметр изменяется в противоположную сторону относительно влияющего фактора, знак $+$ свидетельствует о том, что они изменяются в одну сторону.

Из данной таблицы видно, что если в процессе эксплуатации обнаруживается снижение $\delta n_{\text{ВД}}$, но $\delta G_{\text{Т}}$ увеличивается, то можно с уверенностью сказать, что наблюдается прогар соплового аппарата. В других случаях разделить дефекты турбины и компрессора таким способом затруднительно.

Для решения этой задачи необходимы приемы, позволяющие разделить дефекты компрессора и турбины. Простейшим приемом является увеличение числа диагностических параметров. Если количество параметров будет достаточным, то можно заменить положительные значения коэффициентов влияния на 1, а отрицательные – на 0 и получим таблицу, аналогичную таблице состояний. Однако увеличение числа измеряемых параметров не всегда позволяет решить задачу разделения технических состояний.

Другим возможным вариантом решения задачи диагностики является использование той информации, которая заложена в величине диагностического параметра. Например, данные таблицы 2.12 можно отобразить на плоскости с координатами $\delta n_{\text{ВД}}$ и $\delta G_{\text{Т}}$, то есть представить в двумерном пространстве диагностических признаков (рис.2.18).

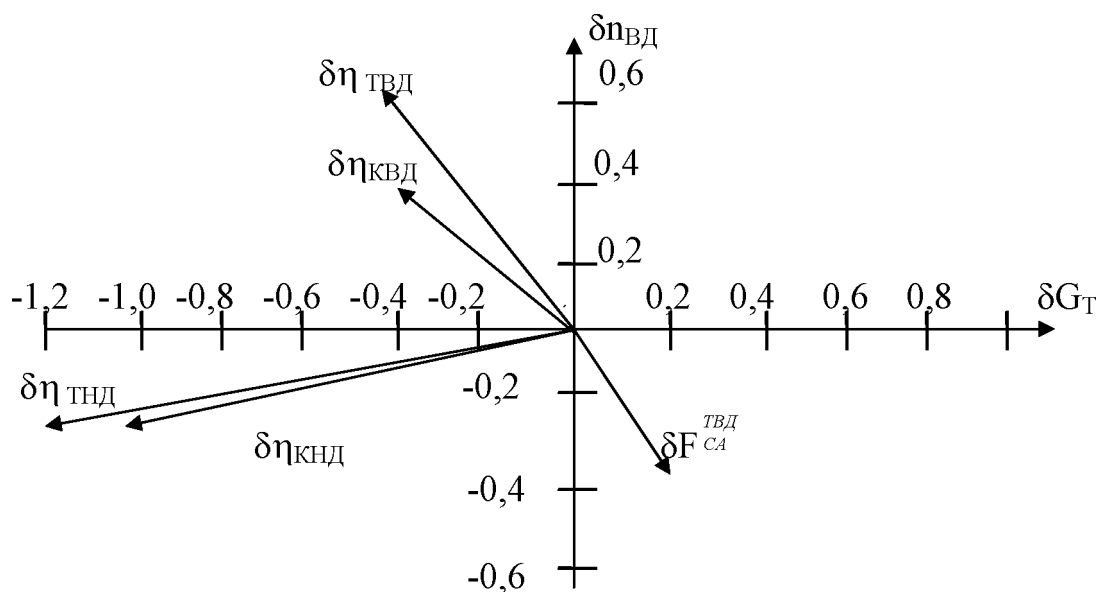


Рис.2.18 Отражение таблицы коэффициентов влияния на плоскости диагностических параметров

В таком случае все технические состояния (диагнозы) представляются в виде точки в пространстве диагностических (измеряемых) параметров. Если для данного примера сформировать диагностические признаки в виде модуля вектора, соединяющего начало координат с точкой диагноза и угла наклона этого вектора к одной из осей, то эти признаки позволяют разделить указанные технические состояния. Существуют и другие методы использования числовых значений коэффициентов влияния или непосредственно диагностических параметров для решения задач распознавания технических состояний (распознавания диагнозов) [1, 2, 9, 14, 29, 31].

В данной главе мы рассмотрели наиболее распространенные виды математических (диагностических) моделей, которые применяются при решении практических задач диагностики технического состояния объектов авиационной техники. Изложенный материал позволяет сделать некоторые выводы.

Математические модели позволяют:

1. Дать формализованное описание связи технического состояния (параметров технического состояния) с выходными (диагностическими) параметрами объекта диагностики.

2. Выделить из бесконечного множества технических состояний объекта диагностики конечный набор диагнозов, которые следует диагностировать, и дать формализованное описание этих технических состояний (диагнозов).

3.Обосновать выбор диагностических (выходных) параметров объекта, для которых следует проводить измерения при выполнении процедур определения технического состояния.

4.Сформировать диагностические признаки, которые должны использоваться в процессе диагностирования.

В заключение отметим, что в ряде случаев создание диагностической модели завершается разработкой таблицы состояний или может быть сведена к подобной таблице. Тогда задача определения технического состояния решается путем непосредственного применения этой таблицы.

В большинстве случаев такой подход не удастся реализовать. Как правило, результатом создания диагностической модели является выбор диагностических параметров или признаков. При этом важное значение имеет эффективность выбранных диагностических параметров или признаков. Эффективность диагностических признаков оценивается так называемой диагностической ценностью признаков.

3. ДИАГНОСТИЧЕСКАЯ ЦЕННОСТЬ ПРИЗНАКОВ

В предыдущей главе было определено, что одним из результатов создания диагностической модели является описание (определение) каждого диагноза набором (комплексом) диагностических параметров или признаков. О различиях между параметрами и признаками уже говорилось в первой главе. Было отмечено, что математическая модель объекта диагностики может быть сформулирована как в понятиях диагностических признаков, так и в понятиях диагностических параметров. Причем, с точки зрения формализованного описания объекта разница между диагностическими признаками и параметрами несущественна. В тоже время, на заключительном этапе принятия решения об отнесении объекта диагностики к одному из диагнозов производится с использованием информации о диагностических признаках.

В зависимости от физической сущности объекта диагностики, особенностей возникающих неисправностей и применяемых методов диагностики диагностические признаки могут быть сформулированы на различных этапах решения задачи (например: на этапе первичного описания объекта, на этапе математического моделирования, на этапе применения методов распознавания). Следует отметить, что ряд методов распознавания для своей реализации требуют только наличие информации о диагностических параметрах. Другие методы при формулировке задачи требуют знание диагностических признаков и их характеристик. Одной из таких характеристик является *разрядность признака* [2, 24, 26]. Рассмотрим это понятие.

Признак называется двухразрядным (простым, бинарным) если в результате обследования объекта диагноза этот признак может быть выражен двумя символами. То есть, если выбранный диагностический признак может принимать только два значения.

Чаще встречается ситуация, когда область определения диагностического параметра непрерывна. При формировании диагностического признака эту область разбивают на m интервалов. Сложным признаком разряда m (m - разрядным признаком) называется результат обследования объекта диагноза, который может быть выражен одним из m символов. M - разрядный признак может принимать m значений. Например, в первой главе на основе диагностического параметра (температуры газов за турбиной авиационного двигателя) был сформирован трехразрядный диагностический признак.

Следует отметить, что для бинарного признака $m = 2$. Кроме того, если имеется одноразрядный признак ($m = 1$), это означает, что результат обследования объекта диагноза выражается одинаково, независимо от технического состояния (диагноза). Такой признак не несет какой-либо диагностической информации и его следует исключить из рассмотрения.

При решении задач технической диагностики важное значение имеет описание объекта в системе диагностических признаков, обладающих большой диагностической ценностью. Количественное определение диагностической ценности признаков производится на основе *теории информации* [2, 6, 40].

Главный принцип: *диагностическая ценность признака* определяется информацией, которая вносится признаком в оценку состояния системы (объекта диагноза).

Центральное место в теории информации занимает понятие *энтропия системы*, которое и используется для оценки количества информации.

3.1 Энтропия системы.

В теории информации *под энтропией системы понимают меру дезорганизованности системы, то есть меру нашего незнания состояния системы.*

Понятие энтропии в теории информации, введенное Винером, Шенноном и Колмогоровым [6, 16, 37], базируется (также как и в термодинамике) на представлении, что в мире порядок борется с хаосом (дезорганизованностью). Если не прикладывать никаких усилий, то объекты мира становятся более дезорганизованными. Или, по-другому, как состояние хаос более вероятен, чем порядок. Поэтому в естественных условиях состояние системы будет становиться все более дезорганизованным, то есть энтропия системы будет возрастать.

Получение информации, в том числе и о состоянии системы, есть операция по наведению порядка. Эта операция направлена на повышение организованности системы. Следовательно, извлечение информации ведет к снижению энтропии системы, но требует определенных затрат. В такой постановке диагностическая ценность признака определяется тем, на сколько уменьшается энтропия системы после определения величины конкретного диагностического признака.

Рассмотрим математическое определение энтропии с позиций теории информации. Первоначально проанализируем факторы, влияющие на энтропию.

1. Допустим имеется система A , которая может находиться в N случайных состояниях A_i ($i = 1 \dots N$). Нахождение системы в одном из этих состояний является случайным событием, каждое из которых может появляться с вероятностями $P(A_1), P(A_2), \dots, P(A_i), \dots, P(A_N)$. Чем больше число состояний, в которых может находиться система, тем больше степень её дезорганизованности. Поскольку степень дезорганизованности определяет энтропию системы, то с увеличением числа состояний энтропия системы будет возрастать. Отсюда следует вывод о том, что математическое выражение для определения величины энтропии должно учитывать количество состояний системы.

2. Будем считать, что события A_i составляют полную группу несовместимых событий, то есть:

$$\sum_{i=1}^N P(A_i) = 1.$$

3. Степень дезорганизованности (то есть энтропии системы) зависит не только от числа состояний n , но и от соотношения между вероятностями $P(A_i)$ нахождения системы в каждом из этих состояний. Поясним на примере. Предположим, что система может находиться в трех состояниях A_1, A_2 и A_3 . Пусть известны вероятности нахождения системы в этих состояниях $P(A_i)$;

$P(A_2)$; $P(A_3)$, причем $P(A_1) = 0,96$; $P(A_2) = 0,02$; $P(A_3) = 0,02$. Очевидно в такой ситуации заранее можно утверждать, что система вероятнее всего будет находиться в состоянии A_1 .

Если $P(A_1) = P(A_2) = P(A_3)$, то дезорганизованность системы будет выше, и следовательно, энтропия больше, чем в первом случае.

Если $P(A_1) = 1$, а $P(A_2) = P(A_3) = 0$, то ни о какой дезорганизованности системы говорить не имеет смысла и в этом случае энтропия системы должна быть равна 0.

С учетом этих требований в теории информации энтропия $H(A)$ системы A , которая может находиться в N состояниях с вероятностями $P(A_i)$ определяется следующим соотношением :

$$H(A) = \sum_{i=1}^N P(A_i) \cdot \log_2 \frac{1}{P(A_i)} = - \sum_{i=1}^N P(A_i) \cdot \log_2 P(A_i), \quad (3.1)$$

где $\log_2$ - двоичный логарифм.

Рассмотри основные свойства энтропии, которые в дальнейшем будут использоваться для определения количества информации.

1. Так как вероятность появления события A_i величина положительная и заключена в интервале от нуля до единицы, то есть $0 \leq P(A_i) \leq 1$, то энтропия системы величина также положительная, то есть $H(A) \geq 0$.

2. Целесообразность использование двоичного логарифма следует из теории связи, поскольку сообщение в одной ячейки памяти может принимать два значения нуль (событие A_1) или единица (событие A_2) с одинаковой вероятностью, то есть $P(A_1) = P(A_2) = 0,5$. Для такой системы

$$H(A) = -P(A_1) \cdot \log_2 P(A_1) - P(A_2) \cdot \log_2 P(A_2) = -\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = 1$$

Эта единица измерения называется двоичной единицей или битом. Таким образом, в качестве единицы измерения энтропии принимается степень неопределенности системы, имеющей два возможных, равновероятных состояния.

3. Можно ввести понятие энтропии для отдельного состояния A_i :

$$H(A_i) = \log_2 \frac{1}{P(A_i)} = -\log_2 P(A_i).$$

Тогда энтропия системы в целом представляет собой среднее значение энтропии отдельных состояний:

$$H(A) = \sum_{i=1}^N P(A_i) \cdot H(A_i).$$

Отсюда следует важный вывод, что в случае введения понятия энтропии i -того состояния энтропию всей системы можно вычислять как среднее значение энтропии отдельных состояний:

$$H(A) = [H(A_i)]_{CP} = [-\log_2 P(A_i)]_{CP}.$$

4. Если вероятность одного из состояний $P(A_k) = 1$, то энтропия всей системы $H(A)$ равна 0. Этот вывод вытекает из следующих соображений. Во первых в сумме, определенной выражением (4.1), будет слагаемое с индексом k равное нулю, то есть:

$$P(A_k) \cdot \log_2 P(A_k) = 1 \cdot 0 = 0.$$

Все остальные слагаемые при $i \neq k$ будут иметь неопределенное значение следующего вида:

$$P(A_i) \cdot \log_2 P(A_i) = 0 \cdot \infty.$$

Однако можно доказать, что эта неопределенность в пределе при стремлении $P(A_i)$ к нулю принимает значение равное нулю:

$$\lim_{P(A_i) \rightarrow 0} \{P(A_i) \cdot \log_2 P(A_i)\} = 0.$$

Таким образом, в соответствии с выражением (4.1) полная энтропия системы равна нулю: $H(A) = 0$. Это условие очевидно и из физических соображений, так как в подобной системе нет никакой неопределенности.

5. Пусть система может находиться в N равновероятных состояниях, то есть:

$$P(A_1) = P(A_2) = P(A_3) = \dots = P(A_i) = \dots = P(A_N) = \frac{1}{N}.$$

Тогда энтропия системы определится следующим образом:

$$H(A) = - \sum_{i=1}^N P(A_i) \cdot \log_2 P(A_i) = - \sum_{i=1}^N \frac{1}{N} \cdot \log_2 \frac{1}{N} = \log_2 N.$$

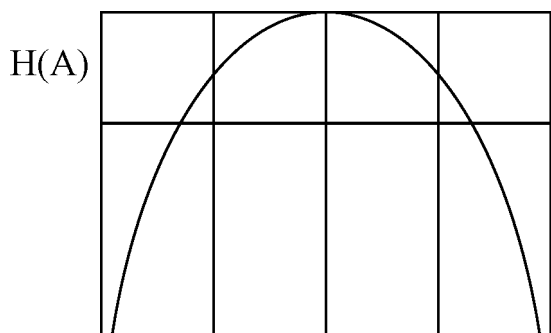
Отсюда следует вывод, что если система может находиться в N равновероятных состояниях, то её энтропия равна логарифму двоичному из числа состояний, то есть: $H(A) = \log_2 N$, где N - число состояний.

6. Если система A может находиться в N состояниях, то энтропия системы максимальна в том случае, когда все состояния равновероятны, то есть

$$H(A) = - \sum_{i=1}^N P(A_i) \cdot \log_2 P(A_i) = \max = \log_2 N,$$

при $P(A_1) = P(A_2) = P(A_3) = \dots = P(A_i) = \dots = P(A_n).$

Поясним последний вывод на следующем примере. Пусть имеется система A , которая может находится в двух состояниях A_1 и A_2 (так называемая бинарная система). Известно $P(A_1)$, а $P(A_2) = 1 - P(A_1)$. Рассмотрим как будет меняться энтропия системы при различных сочетаниях вероятностей $P(A_1)$ и $P(A_2)$.



В первом варианте примем $P(A_1) = 0$. Тогда $P(A_2) = 1$ и, следовательно, $H(A) = 0$ (рис. 3.1).

Во втором варианте примем $P(A_1) = 1$.

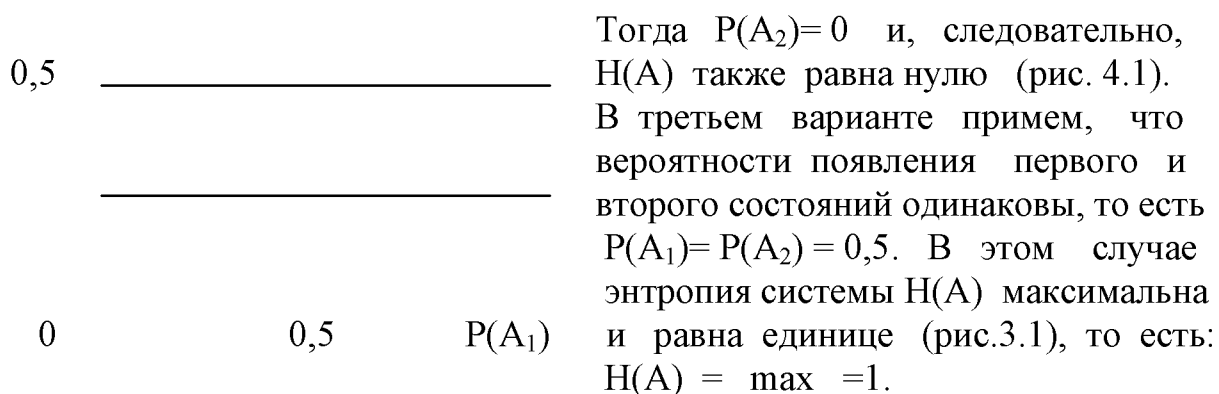


Рис.3.1 Зависимость $H(A)$ бинарной системы от вероятности первого состояния

Рассмотренные понятия самой энтропии и ее свойства позволяют подойти к вопросу определения количества информации [2, 40].

3.2 Определение количества информации.

С точки зрения диагностики важно определять не саму энтропию системы, а то количество информации, которое приобретается о состоянии системы в результате измерения некоторого комплекса (набора, вектора) признаков.

Рассмотрим следующий пример. Пусть объектом диагностики является маслосистема, которая может находиться либо в исправном A_1 , либо в неисправном A_2 состояниях. Для диагностирования системы проводятся измерение температуры t_M или давления P_M масла.

Очевидно, что до выполнения измерений дать ответ о состоянии системы весьма затруднительно. Можно предположить, что исправное и неисправное состояний, являются равновероятными, то есть $P(A_1) = P(A_2) = 0,5$. Таким образом, с учетом изложенного выше, можно определить начальную энтропию системы до проведения измерений $H_{нач}(A)$, которая равна:

$$\begin{aligned} H_{нач}(A) &= -P(A_1) \cdot \log_2 P(A_1) - P(A_2) \cdot \log_2 P(A_2) = \\ &= -\frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{1}{2} \cdot \log_2 \frac{1}{2} = 1 \end{aligned}$$

Пусть для температуры предельное значение составляет 90°C . Из опыта эксплуатации установлено, что при температуре масла $t_M < 90^\circ\text{C}$ в 60% случаев масляная система находится в исправном состоянии, то есть при $t_M < 90^\circ\text{C}$ - $P(A_1) = 0,6$, а $P(A_2) = 0,4$. Следовательно, если в процессе измерения установлено, что температура масла меньше 90°C , то энтропия системы равна:

$$H_t(A) = -0,6 \cdot \log_2 0,6 - 0,4 \cdot \log_2 0,4 = 0,92.$$

Количество информации о состоянии системы A , которое мы получили по результатам измерения температуры $I_A(t_M)$, определим как разность начальной энтропии $H_{HAЧ}(A)$ и значения энтропии, которую имела система после измерения температуры $H_t(A)$:

$$I_A(t_M) = H_{HAЧ}(A) - H_t(A) = 1 - 0,92 = 0,08.$$

Для этой же системы известно, что если давление масла в системе превышает две атмосферы, то вероятность нахождения системы в исправном состоянии равна единице, а в неисправном нулю, то есть при $P_M > 2 \text{ атм.}$ - $P(A_1) = 1$, а $P(A_2) = 0$. Поэтому если в процессе измерения установлено, что давление масла в системе превышает две атмосферы, то энтропия системы равна:

$$H_P(A) = -1 \cdot \log_2 1 - 0 \cdot \log_2 0 = 0.$$

Как и в предыдущем случае, количество информации о состоянии системы A , которое мы получили по результатам измерения давления $I_A(P_M)$, определим как разность начальной энтропии $H_{HAЧ}(A)$ и значения энтропии, которую имела система после измерения давления $H_P(A)$:

$$I_A(P_M) = H_{HAЧ}(A) - H_P(A) = 1 - 0 = 1.$$

В общем случае задача определения количества информации ставится таким образом. Пусть имеется система (объект диагностики) A , которая может находиться в одном из состояний A_i (где $i = 1 \dots N$). В каждом из этих состояний объект может находиться с определенной вероятностью $P(A_i)$. С этим объектом связан комплекс (система, вектор) диагностических признаков K . Этот комплекс (вектор) задается своими компонентами (координатами), число которых равно n . Другими словами объект диагностики описывается в n -мерном пространстве диагностических признаков. В процессе диагностирования, после проведения соответствующих измерений и обработки диагностических параметров, определяются конкретные значения (реализации) диагностических признаков по все n координатам. В свою очередь этот комплекс, состоящий из n реализаций диагностических признаков определяет некоторую j -тую реализацию вектора диагностических признаков K_j , принадлежащую множеству K . Математически эти рассуждения можно записать следующим образом:

$$K = K^* = K_j = \{k_{j1}; k_{j2}; \dots k_{js}; \dots k_{jn}\},$$

где K^* - экспериментально полученная реализация вектора диагностических признаков K ; K_j - экспериментально полученная конкретная j -тая реализация вектора диагностических признаков; k_{js} - экспериментально полученная реализация диагностического признака по s -той координате,

соответствующая j - той реализации вектора диагностических признаков K_j ; $s = 1, \dots, n$; $j = 1, \dots, m$.

В диагностическом смысле задача оценки количества информации формулируется так: - необходимо определить количество информации, которое получается о состоянии системы A , после измерения всех реализаций вектора диагностических признаков, входящих во множество K . В такой постановке количество информации записывается следующим образом:

$$I_A(K) = H(A) - H(A/K), \quad (3.2)$$

где $H(A) = - \sum_{i=1}^N P(A_i) \cdot \log_2 P(A_i)$ - начальная энтропия объекта диагностики, которая определяется по априорным данным; $H(A/K)$ - энтропия системы A после того, как стали известны все реализации вектора диагностических признаков из множества K .

Вычисление начальной энтропии при известных вероятностях $P(A_i)$ не вызывает затруднений. Рассмотрим второй член в правой части выражения (3.2), который называется условной энтропией. Эта условная энтропия вычисляется таким же образом, как и начальная энтропия, только вместо вероятности появления события A_i подставляется условная вероятность нахождения системы в состоянии A_i при наблюдении всех реализации вектора признаков K_j из множества K , то есть

$$H(A/K) = - \sum_{i=1}^N \sum_{j=1}^m P(A_i/K_j) \cdot \log_2 P(A_i/K_j),$$

где $P(A_i/K_j)$ - условная вероятность нахождения объекта диагностирования в состоянии A_i при условии, что вектор диагностических параметров получит реализацию K_j ; $i = 1 \dots N$ - номер технического состояния (диагноза); $j = 1 \dots m$ - номер реализации вектора диагностических признаков.

После подстановки начальной и условной энтропии в выражение (3.2) и соответствующих преобразований, выражение для количества информации, получаемого при обследовании всех диагнозов A_i по всем реализациям вектора диагностических признаков K_j , будет иметь вид:

$$I_A(K) = \sum_{i=1}^N \sum_{j=1}^m P(A_i) \cdot P(K_j/A_i) \cdot \log_2 \frac{P(K_j/A_i)}{P(K_j)}, \quad (3.3)$$

где $P(K_j/A_i)$ - условная вероятность появления j - той реализации вектора признаков K_j при нахождении системы в A_i состоянии;
 $P(K_j)$ - вероятность появления j - той реализации вектора признаков K_j при нахождении объекта диагностики во всех возможных состояниях.

Используя известные из теории вероятностей соотношения между вероятностями совместного появления событий и условными вероятностями появления событий:

$$P(A_i, K_j) = P(A_i) \cdot P\left(\frac{K_j}{A_i}\right) = P(K_j) \cdot P\left(\frac{A_i}{K_j}\right),$$

выражение (4.3) для количества информации можно представить в другом виде:

$$I_A(\mathbf{K}) = \sum_{i=1}^N \sum_{j=1}^m P(K_j) \cdot P\left(\frac{A_i}{K_j}\right) \cdot \log_2 \frac{P\left(\frac{A_i}{K_j}\right)}{P(A_i)} \quad (3.4)$$

или

$$I_A(\mathbf{K}) = \sum_{i=1}^N \sum_{j=1}^m P(A_i, K_j) \cdot \log_2 \frac{P(A_i, K_j)}{P(A_i) \cdot P(K_j)}, \quad (3.5)$$

где $P\left(\frac{A_i}{K_j}\right)$ - вероятность нахождения объекта в диагнозе A_i при условии, что входе процесса диагностирования была получена реализация вектора диагностических признаков K_j ;
 $P(A_i, K_j)$ - вероятность совместного появления следующих событий: нахождение объекта в диагнозе A_i и получении в процессе диагностирования реализации диагностических признаков K_j .

При анализе выражения (3.5) можно сделать вывод, способствующий физическому пониманию расчетных формул, предназначенных для оценки величины получаемой информации. Так, если система признаков выбрана неправильно, то она оказывается не связана с техническим состоянием. Говоря по-другому, все диагнозы объекта и все реализации векторов диагностических признаков являются независимыми событиями, то есть $P(A_i, K_j) = P(A_i) \cdot P(K_j)$. Тогда, как следует из выражения (3.5), $I_A(\mathbf{K}) = 0$. Следовательно, в такой ситуации знание диагностических признаков не дает ни какой информации о техническом состоянии объекта диагностики.

При необходимости определить количество информации, получаемое об отдельном состоянии A_i при обследовании по всем реализациям вектора признаков K_j , в предыдущих выражениях необходимо опустить суммирование по i ($i = 1 \dots N$). Тогда мы получим три выражение для определения $I_{A_i}(\mathbf{K})$:

$$I_{A_i}(\mathbf{K}) = \sum_{j=1}^m P(A_i, K_j) \cdot \log_2 \frac{P(A_i, K_j)}{P(A_i) \cdot P(K_j)}, \quad (3.6)$$

или

$$I_{A_i}(\mathbf{K}) = \sum_{j=1}^m P(A_i) \cdot P\left(\frac{K_j}{A_i}\right) \cdot \log_2 \frac{P\left(\frac{K_j}{A_i}\right)}{P(K_j)}, \quad (3.7)$$

или

$$I_{A_i}(\mathbf{K}) = \sum_{j=1}^m P(K_j) \cdot P\left(\frac{A_i}{K_j}\right) \cdot \log_2 \frac{P\left(\frac{A_i}{K_j}\right)}{P(A_i)}. \quad (3.8)$$

Если необходимо определить количество информации $I_{A_i}(K_j)$, которое получают о состоянии системы A_i , при измерении в процессе диагностирования конкретной реализации вектора диагностических признаков K_j , то в предыдущих выражениях надо вообще опустить суммирование:

$$I_{A_i}(K_j) = P(A_i, K_j) \cdot \log_2 \left[\frac{P(A_i, K_j)}{P(A_i) \cdot P(K_j)} \right] \quad (3.9)$$

или

$$I_{A_i}(K_j) = P(A_i) \cdot P(K_j/A_i) \cdot \log_2 \left[\frac{P(K_j/A_i)}{P(K_j)} \right] \quad (3.10)$$

или

$$I_{A_i}(K_j) = P(K_j) \cdot P(A_i/K_j) \cdot \log_2 \left[\frac{P(A_i/K_j)}{P(A_i)} \right] \quad (3.11)$$

Итак, последние выражения для $I_{A_i}(K_j)$ определяют ожидаемое количество информации о состоянии A_i получаемое при появлении в ходе процесса диагностирования конкретной реализации вектора диагностических признаков K_j .

В общей постановке вектор диагностических признаков может содержать любое число компонент, то есть объект диагностирования может быть определен в пространстве диагностических признаков любой размерности. Следовательно, количество диагностических признаков, являющихся компонентами вектора, может изменяться от одного до бесконечности. Поэтому полученные выражения могут быть применены и к случаю, когда вектор диагностических признаков содержит только одну компоненту, то есть $K_j = k_j$. Тогда, если в выражениях (3.9), (3.10) и (3.11) заменить K_j на k_j , то они будут определять количество информации, получаемое о состоянии A_i при обследовании по конкретному признаку k_j .

3.3 Диагностический вес признака

Последний вывод о том, что данные соотношения справедливы как для случая измерения комплекса признаков, так и для случая измерения отдельного диагностического признака или параметра является весьма существенным. На его основе вводятся ряд важнейших для практики понятий. Одним из таких понятий является *диагностический вес признака* [2, 31].

В технической диагностики понятие *диагностического веса признака* вводится как величина, которая определяется следующим выражением:

$$Z_{A_i}(k_j) = \log_2 \left[\frac{P(A_i/k_j)}{P(A_i)} \right]. \quad (3.12)$$

Данное выражение является эквивалентом количества информации, которое мы получаем о состоянии A_i после измерения признака k_j .

Рассмотрим некоторые свойства диагностического веса признака.

1. Если вероятность появления состояния A_i после измерения признака k_j больше вероятности появления этого состояния без измерения признака (то есть $P(A_i/k_j) > P(A_i)$), тогда $Z_{A_i}(k_j) > 0$.

2. Если условная вероятность появления состояния A_i равна обычной вероятности (то есть $P(A_i/k_j) = P(A_i)$), тогда $Z_{A_i}(k_j) = 0$. Такая ситуация означает, что появление состояния A_i и появление признака k_j являются независимыми событиями.

3. Если условная вероятность появления состояния A_i меньше обычной вероятности (то есть $P(A_i/k_j) < P(A_i)$), тогда $Z_{A_i}(k_j) < 0$. Такая ситуация свидетельствует о том, что появление признака k_j есть отрицание появления состояния A_i .

Учитывая предыдущие выражения (3.9) и (3.10), диагностический вес можно записать и следующим образом:

$$Z_{A_i}(k_j) = \log_2 \left[\frac{P(A_i, k_j)}{P(A_i) \cdot P(k_j)} \right], \quad (3.13)$$

или

$$Z_{A_i}(k_j) = \log_2 \left[\frac{P(k_j/A_i)}{P(k_j)} \right]. \quad (3.14)$$

Определяемые этими соотношениями диагностические веса признаков называются, независимыми (безусловными) диагностическими весами. При определении диагностических весов признаков по этим соотношениям предполагается ситуация, в которой обследование по признаку k_j не зависит от обследования по другим признакам.

Если производится обследование объекта диагностики по двум признакам k_1 и k_2 , последовательно друг за другом, и результаты обследования по k_2 зависят от результатов обслуживания по k_1 , то вводится понятие *условного диагностического веса*.

Пусть в процессе диагностического обследования провели сначала оценку признака k_1 и получили его реализацию k_{1s} . После этого провели оценку признака k_2 и получили его реализацию k_{2j} . Под условным диагностическим весом j - той реализации признака k_2 понимается величина, определяемая следующим выражением:

$$Z_{A_i}(k_{2j}/k_{1s}) = \log_2 \left[\frac{P(k_{2j}/A_i, k_{1s})}{P(k_{2j}/k_{1s})} \right], \quad (3.15)$$

где $P(k_{2j}/A_i, k_{1s})$ является условной вероятностью появлению (реализации) признака k_{2j} при условии, что произошло событие, заключающееся в

совместном появлении диагноза A_i и s - той реализации первого признака (то есть k_{1s});

$P(k_{2j} / k_{1s})$ – условная вероятность появления j - той реализации второго признака k_{2j} при условии, что вначале была получена s - тая реализация первого признака k_{1s} .

Величина, определенное соотношением (3.15), показывает количество информации, получаемое о состоянии A_i по j - той реализации второго признака k_{2j} при условии, что до этого уже была получена s - тая реализация первого признака k_{1s} . При этом диагностический вес признака k_1 определяется по одной из выше приведенных формул (3.12), (3.13), (3.14), так как оценка этого признака производится в начале диагностического процесса и его значение не зависит от результатов оценки других признаков.

При необходимости определить количество информации, которая получается в результате двух последовательных исследований по признакам k_1 и k_2 , вводится понятие суммарного диагностического веса:

$$Z_{A_i}(k_{1s}, k_{2j}) = \log_2 \left[\frac{P(k_{1s} / A_i)}{P(k_{1s})} \right] + \log_2 \left[\frac{P(k_{2j} / A_i, k_{1s})}{P(k_{2j} / k_{1s})} \right]. \quad (3.16)$$

Эта величина показывает количество информации, получаемой о состоянии A_i при последовательном обследовании по признакам k_1 и k_2 и получении при этом соответственно реализаций этих признаков k_{1s} и k_{2j} .

Для практики диагностики бывает важно оценить информативность диагностического признака, имеющего несколько разрядов. Количество информации о диагнозе (состоянии) A_i , получаемое при обследовании по всем реализациям признака k_j , имеющего m разрядов, называется *диагностической ценностью обследования по признаку k_j^m* . Эта диагностическая ценность обследования $C_{A_i}(k_j^m)$ определяется следующим выражением:

$$C_{A_i}(k_j^m) = \sum_{j=1}^m P(k_{js} / A_i) Z_{A_i}(k_{js}), \quad (3.17)$$

где k_{js} - значение признака k_j^m , которое он приобретает при попадании в s - тый разряд; $Z_{A_i}(k_{js})$ – диагностический вес признака k_j разряда s для состояния A_i ; $P(k_{js} / A_i)$ – условная вероятность появления k_j признака в разряде s для состояния A_i .

Диагностическая ценность обследования по m – разрядному признаку можно использовать для:

1. Оценки эффективности применения для диагностирования, выбранного диагностического m - разрядного признака;
2. Для оптимизации числа разрядов признака.

Проиллюстрируем последний вывод на примере [2].

Пусть имеется некоторый диагностический параметр y , который может изменяться от нуля до единицы, то есть $0 \leq y \leq 1$. Обследование большого числа объектов показало, что все значения y в данном диапазоне равновероятны, но для объектов с диагнозом A_1 они лежат в пределах от 0,5 до 0,75. Требуется определить оптимальное количество интервалов, при котором обследование по данному признаку давало бы максимум информации, то есть признак будет обладать максимальной диагностической ценностью.

Рассмотрим несколько вариантов разбиения интервала $0 \leq y \leq 1$.

1. Разобьем этот интервал на два разряда: $0 - 0,5$ и $0,5 - 1$ и сформируем диагностический признак k_1 . Примем, что этот признак принимает значение k_{11} при попадании параметра y в первый интервал, то есть при $0 \leq y < 0,5$, или значение k_{12} при попадании параметра y во второй интервал, то есть при $0,5 \leq y \leq 1$. С учетом изложенного выше вероятности появления этих значений признаков (без относительно к диагнозу) равны: $P(k_{11}) = P(k_{12}) = 0,5$. В тоже время, вероятности появления этих признаков при нахождении объекта в состоянии A_1 различны, а именно $P(k_{11}/A_1) = 0$ и $P(k_{12}/A_1) = 1$.

В соответствии с формулой (4.17) диагностическую ценность двух разрядного признака k_1 вычислим следующим образом:

$$\begin{aligned} C_{A_1}(k^2_1) &= P(k_{11}/A_1) \cdot \log_2 \frac{P(k_{11}/A_1)}{P(k_{11})} + P(k_{12}/A_1) \cdot \log_2 \frac{P(k_{12}/A_1)}{P(k_{12})} = \\ &= 0 \cdot \log_2 0 + 1 \cdot \log_2 \frac{1}{0,5} = 1 \end{aligned}$$

2. Разобьем интервал изменения диагностического параметра y на четыре разряда: $0 \dots 0,25$; $0,25 \dots 0,5$; $0,5 \dots 0,75$; $0,75 \dots 1$. Будем считать, что при попадании параметра в указанные разряды, признак k_1 принимает соответственно значения k_{11} , k_{12} , k_{13} , k_{14} . Как и в предыдущем случае вероятности появления этих значений без относительно к диагнозу равны между собой, то есть $P(k_{11}) = P(k_{12}) = P(k_{13}) = P(k_{14}) = 0,25$. Условные вероятности появления указанных значений признака k_1 при нахождении объекта в диагнозе A_1 равны: $P(k_{11}/A_1) = 0$, $P(k_{12}/A_1) = 0$ и $P(k_{14}/A_1) = 0$, а $P(k_{13}/A_1) = 1$. Воспользуемся соотношением (4.17) для определения диагностической ценности обследования по четырех разрядному признаку k_1 :

$$\begin{aligned} C_{A_1}(k^4_1) &= P(k_{11}/A_1) \cdot \log_2 \frac{P(k_{11}/A_1)}{P(k_{11})} + P(k_{12}/A_1) \cdot \log_2 \frac{P(k_{12}/A_1)}{P(k_{12})} + \\ &+ P(k_{13}/A_1) \cdot \log_2 \frac{P(k_{13}/A_1)}{P(k_{13})} + P(k_{14}/A_1) \cdot \log_2 \frac{P(k_{14}/A_1)}{P(k_{14})} \end{aligned}$$

Очевидно, что в этом выражении из четырех слагаемых только одно с $P(k_{13}/A_1) = 1$ не будет равно нулю. Таким образом:

$$C_{A_1}(k^4_1) = P(k_{13}/A_1) \cdot \log_2 \frac{P(k_{13}/A_1)}{P(k_{13})} = 1 \cdot \log_2 \frac{1}{0,25} = 2.$$

3. Если произвести увеличение разрядности признака до восьми и выполнить аналогичные рассуждения, то вычисленное значение диагностической ценности обследования по полученному восьми разрядному признаку также будет равно двум: $C_{Ai}(k_1^8) = 2$.

Рассмотренный пример позволяет сделать очень важный вывод.

С увеличением числа разрядов диагностическая ценность обследования по заданному признаку будет возрастать и достигнет максимума. При дальнейшем увеличении разрядности величина диагностической ценности не будет меняться, а трудоемкость обследования будет увеличиваться.

Подводя итог данного раздела, отметим, что основные положения, изложенные в нем, являются весьма важными при разработке методов и средств диагностики и их следует использовать при формировании диагностических признаков и выборе их разрядности.

На конечном этапе, диагностическая задача сводится к необходимости применения формализованных методов распознавания (классификации) диагнозов, основанных на математической обработке числовых данных, полученных в процессе измерения диагностических признаков или параметров. Эти методы в технической диагностики получили название «методы распознавания (разделения, классификации) диагнозов». Ниже рассмотрены некоторые подобные методы, наиболее часто используемые на практике.

4. МЕТОДЫ РАСПОЗНАВАНИЯ В ТЕХНИЧЕСКОЙ ДИАГНОСТИКЕ

4.1 Основные понятия

В общем случае задача распознавания диагнозов формулируется следующим образом [2, 3, 26].

Технический объект (система) может находиться в n состояниях, то есть для этого объекта может наблюдаться совокупность диагнозов $D = (D_1, D_2, \dots, D_i, \dots, D_n)$, где $i = 1 \dots n$.

Каждое техническое состояние (диагноз D_i) описывается комплексом признаков $K = (K_1, K_2, \dots, K_j, \dots, K_L)$, где $j = 1 \dots L$. Каждый j -тый признак обладает m_j разрядами.

Требуется построить решающее правило, с помощью которого объект, предъявленный для диагностики технического состояния, по полученному комплексу m_j - разрядных признаков мог быть отнесен к одному из возможных диагнозов D_i .

Для наглядности описания задач распознавания диагнозов удобно дать им геометрическую интерпретацию. При этом учтем, что в зависимости от свойств объекта диагностики, законов изменения параметров технического состояния, законов изменения диагностических параметров и, следовательно, диагностических признаков существует два подхода к решению задач распознавания диагнозов: детерминистский и вероятностный.

Детерминистский подход имеет место в том случае, если каждый диагноз может быть описан конкретным набором диагностических признаков, которые принимают строго определенные значения. При этом этот набор значений признаков имеет место только для данного диагноза. С детерминистскими методами мы уже сталкивались на примере составления таблицы состояния.

Дадим геометрическую интерпретацию задач распознавания диагнозов для детерминистского подхода. Пусть в результате оценки технического состояния определены конкретные значения диагностических признаков, то есть получено, что объект диагностирования характеризуется комплексом признаков $K = (K_1, K_2, \dots, K_j, \dots, K_L)$. Пусть признак K_j определяет значение j -той координаты некоторого вектора в L -мерном пространстве. Тогда набор признаков $(K_1, K_2, \dots, K_j, \dots, K_L)$ определяет вектор, соединяющий начало координат и точку с соответствующими координатами в L -мерном пространстве. При таком представлении каждому диагнозу будет соответствовать строго определенная точка в L -мерном пространстве признаков. Предыдущие рассуждения будут справедливы и для случая, если в качестве координат L -мерного пространства использовать понятие «диагностический параметр». Поэтому описание диагнозов в L -мерном пространстве диагностических признаков или диагностических параметров эквивалентны.

Отметим, что поскольку практически значения параметров получаются с учетом погрешностей измерений, то для каждого диагноза, как в пространстве

параметров, так и в пространстве признаков мы будем иметь не точку, а некоторую область. Но главной особенностью детерминистской постановки задачи является то, что области различных диагнозов не пересекаются. Например, на рис.4.1 изображен случай (сплошные линии), когда объект диагностики может находиться в двух диагнозах D_1 и D_2 , а его техническое состояние описывается двумя признаками (параметрами) K_1 и K_2 .

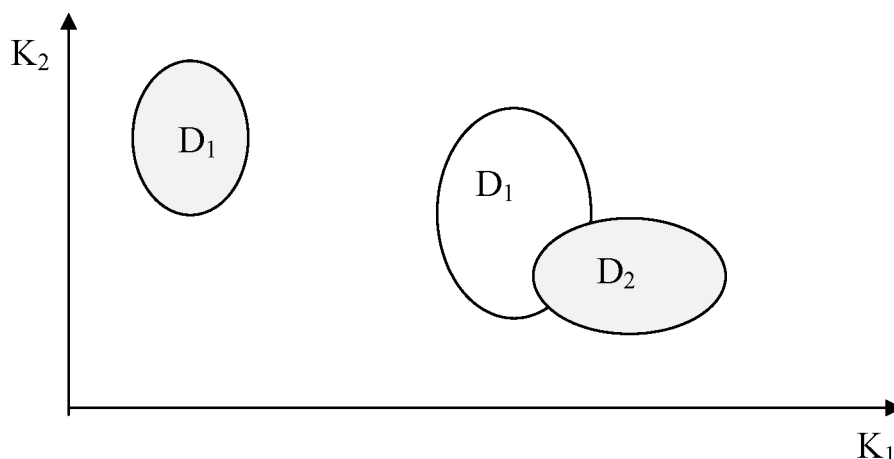


Рис.4.1 Представление областей диагнозов в пространстве признаков

Таким образом, при детерминистском подходе задача разработки решающего правила сводится к разделению пространства признаков (или параметров) на области диагнозов.

При детерминистском подходе для описания постановки диагностической задачи можно воспользоваться понятиями теории вероятностей. В случае перехода к понятиям теории вероятностей трактовка будет следующей. Если мы имеем набор признаков, принявших определенное значение, то вероятность диагноза, в область которого попала точка, определенная этим набором признаков, равна единице.

Такая трактовка предполагает, что и каждый признак принимает конкретное значение только при одном диагнозе. При других диагнозах он принимает другое значение.

К сожалению, детерминистская постановка диагностической задачи зачастую невозможна, так как на диагностические параметры (а, следовательно, и на сформированные на их основе диагностические признаки) воздействует огромное количество случайных факторов (изготовление элементов систем в пределах допусков, внешние условия работы, точность измерения и т.п.). В результате диагностические признаки (параметры) имеют существенный разброс значений даже для случая, когда объекты находятся в одном и том же диагнозе. Поскольку диагностический признак обладает существенным разбросом, то он становится величиной, которая только с некоторой вероятностью определяет появление конкретного диагноза.

В такой ситуации используется *вероятностный подход* к решению задач технической диагностики. При использовании вероятностных методов

постановка задач распознавания диагнозов формулируется следующим образом.

1. Имеется технический объект, который может находиться в одном из n диагнозов $D = (D_1, D_2, \dots, D_i, \dots, D_n)$. Появление конкретного диагноза D_i в данный момент времени случайно и определяется соответствующей вероятностью $P(D_i)$.

2. Известна совокупность признаков $K = (K_1, K_2, \dots, K_j, \dots, K_L)$, каждый из которых является случайной величиной и возможность его появления характеризуется вероятностью $P(K_j)$.

3. Требуется сформулировать правило, с помощью которого можно по заданному комплексу признаков отнести объект к одному из диагнозов. То есть определить вероятность нахождения объекта в i -том диагнозе (появления события D_i) при условии, что экспериментально получен конкретный набор значений признаков $K^* = (K_1^*, K_2^*, \dots, K_j^*, \dots, K_L^*)$, где K_j^* - конкретная реализация случайной величины K_j .

В геометрической интерпретации ситуация, соответствующая вероятностному подходу означает, что области диагнозов пересекаются, как это показано на рис.4.1 пунктирной линией для диагноза D_1 .

Вероятностные методы распознавания являются более общими. Поэтому ниже подробно рассмотрены наиболее широко используемые методы, относящиеся к этому классу.

4.2 Метод Байеса.

Метод Байеса является одним из наиболее простых и мощных методов. Этот метод основан на вычислении условной вероятности появления такого события как диагноз D_i при появлении конкретной реализации комплекса признаков K^* [2, 3, 27].

Рассмотрим первоначально основные положения этого метода на простейшем случае, когда имеется диагноз D_i и один бинарный признак K_j , встречающийся при появлении этого диагноза.

Определим некоторые понятия:

1. $P(D_i)$ - априорная (до опытная) вероятность появления диагноза D_i . Эту вероятность определяют по статистическим данным на начальном этапе применения метода исходя из следующих соображений. Если при обследовании N объектов диагноза установлено, что из них N_i имеют диагноз D_i , то вероятность появления этого диагноза определяется соотношением

$$P(D_i) = \frac{N_i}{N}.$$

2. $P(K_j / D_i)$ - априорная условная вероятность появления признака K_j у объектов имеющих техническое состояние (диагноз) D_i . Эта вероятность так же определяется на начальном этапе по имеющимся статистическим данным. Если из N обследованных объектов N_i находилось в диагнозе D_i , а из них N_{ij} объектов имели признак K_j , то условная вероятность появления признака K_j у объектов с диагнозом D_i вычисляется следующим образом

$$P(K_j/D_i) = \frac{N_{ij}}{N_i}.$$

3. $P(K_j)$ - априорная вероятность появления признака K_j у всех объектов независимо от их состояния. То есть, если из N объектов независимо от их технического состояния у N_j был обнаружен признак K_j , то эта вероятность определяется следующим соотношением

$$P(K_j) = \frac{N_j}{N}.$$

Напомним некоторые положения теории вероятностей. Пусть мы имеем два события A и B . Известны вероятности появления этих событий $P(A)$ и $P(B)$, а также условная вероятность появления события A при уже состоявшемся событии B $P(A/B)$ и условная вероятность появления события B при уже состоявшемся событии A $P(B/A)$. Тогда вероятность одновременного появления событий A и B $P(A,B)$ определяется следующей формулой

$$P(A,B) = P(A) P(B/A) = P(B) P(A/B).$$

Воспользовавшись этой формулой и данными выше понятиями можно записать вероятность одновременного появления диагноза D_i и признака K_j следующим образом

$$P(D_i, K_j) = P(D_i) P(K_j/D_i) = P(K_j) P(D_i/K_j).$$

В этом выражении величина $P(D_i/K_j)$ - это условная вероятность существования диагноза D_i при обнаружении признака K_j , то есть это та величина, которая ищется при вероятностном подходе к решению задачи распознавания диагнозов. После соответствующих преобразований из последнего выражения получим формулу Байеса

$$P(D_i/K_j) = P(D_i) P(K_j/D_i) / P(K_j). \quad (4.1)$$

Формула (4.1) получена для случая, когда при постановке диагноза используется один простой признак.

Для принятия решения о диагнозе при использовании набора (комплекса) признаков применяется *обобщенная формула Байеса*, которую можно получить из следующих соображений. Если диагностирование проводится по комплексу признаков, то в результате обследования мы получаем конкретную реализацию каждого j -того признака K_j^* и, следовательно, конкретную реализацию комплекса признаков K^* в целом. В этом случае формула Байеса предстанет в виде:

$$P(D_i/K^*) = P(D_i) \frac{P(K^*/D_i)}{P(K^*)}, \quad (4.2)$$

где $P(D_i / K^*)$ - условная вероятность нахождения объекта диагностики в диагнозе D_i при условии, что в ходе обследования была получена реализация K^* комплекса признаков K ; $P(K^*)$ - вероятность появления конкретной реализации K^* комплекса признаков K у всех диагностируемых объектов, независимо от их технического состояния; $P(K^*/D_i)$ - условная вероятность появления конкретной реализации K^* комплекса диагностических признаков K для объектов, находящихся в диагнозе D_i .

Преобразуем последнее выражения с учетом следующих соображений.

Примем, что система может находиться только в одном из n технических состояний, тогда

$$\sum_{i=1}^n P(D_i) = 1.$$

Будем считать, что отдельные диагностические признаки, входящие в состав комплекса признаков, независимые. Такое допущение вполне справедливо для реальных условий при большом числе влияющих факторов. Тогда условную вероятность $P(K^*/D_i)$ в соответствии с известными положениями теории вероятностей можно представить как произведение

$$\begin{aligned} P(K^*/D_i) &= P(K^*_1/D_i)P(K^*_2/D_i)...P(K^*_j/D_i)...P(K^*_L/D_i) = \\ &= \prod_{j=1}^L P(K^*_j/D_i) \end{aligned}$$

где $P(K^*_j / D_i)$ - условная вероятность появления конкретной реализации K^*_j j -того признака при нахождении объекта диагностики в диагнозе D_i ; $j = 1... L$.

Вероятность же появления конкретной реализации комплекса признаков при нахождении объекта во всех диагнозах $P(K^*)$ можно представить следующим образом

$$P(K^*) = \sum_{i=1}^n P(D_i)P(K^*/D_i) = \sum_{i=1}^n P(D_i) \prod_{j=1}^L P(K^*_j/D_i).$$

С учетом последних соотношений уравнение (4.2) перепишем в окончательном виде

$$P(D_i / K^*) = \frac{P(D_i) \prod_{j=1}^L P(K^*_j / D_i)}{\sum_{i=1}^n P(D_i) \prod_{j=1}^L P(K^*_j / D_i)}. \quad (4.3)$$

Полученное уравнение называется *обобщенной формулой Байеса*.

Сделаем некоторые замечания к полученным соотношениям.

1. Поскольку проверяемый объект обязательно будет находиться в одном из диагнозов D_i , то с учетом соотношения (4.3) можно записать:

$$\sum_{i=1}^n P(D_i / K^*) = 1.$$

2. Если реализация некоторого комплекса признаков K^* встречается только для одного диагноза D_s , а для остальных диагнозов не встречается, то такая реализация комплекса признаков называется детерминирующей для диагноза D_s . Для этой реализации комплекса признаков справедливо соотношение

$$P(K^* / D_i) = \begin{cases} = 0 & \text{при } i \neq s \\ \neq 0 & \text{при } i = s \end{cases}.$$

Тогда из обобщенной формулы Байеса следует, что

$$P(D_i / K^*) = \begin{cases} = 1 & \text{при } i = s \\ = 0 & \text{при } i \neq s \end{cases}.$$

Анализ этих замечаний свидетельствует о том, что детерминистский подход является частным случаем вероятности.

Для практического использования метода Байеса на первоначальном этапе необходимо произвести расчет априорных вероятностей появления i -тых диагнозов и условных априорных вероятностей появления m -ного разряда j -того признака при нахождении объекта диагностики в D_i -том диагнозе. Расчет этих вероятностей производится на основе статистического материала, полученного из эксплуатации. Результаты этих расчетов и результаты окончательных расчетов вероятностей появления диагнозов при полученном комплексе признаков удобно представлять в табличном виде.

Рассмотрим порядок диагностирования с помощью метода Байеса.

На первоначальном этапе на основе собранного статистического материала определяют:

- набор диагнозов D_i , подлежащих распознаванию;
- априорные вероятности появления этих диагнозов $P(D_i)$.
- априорные вероятности появления m -ного разряда j -того признака K_{jm} при нахождении объекта в диагнозе D_i , то есть $P(K_{jm} / D_i)$.

Для удобства работы эти данные представляются в виде таблицы, которая называется диагностической матрицей в методе Байеса [2].

Таблица 4.1 Диагностическая матрица в методе Байеса

$P(K_{jm}/D_i) \setminus D_i$		D_1	D_2	...	D_i	...
K_1	$P(K_{11}/D_i)$					
	$P(K_{12}/D_i)$					
	$P(K_{13}/D_i)$					
K_2	$P(K_{21}/D_i)$					
	$P(K_{22}/D_i)$					
...	...	...	...	...	...	...
K_j	$P(K_{jm}/D_i)$					
...	...	...	...	...	...	...
K_l	$P(K_{lm}/D_i)$					
$P(D_i)$						

Составление этой таблицы является важнейшим моментом при диагностировании с использованием метода Байеса. Поскольку эксплуатация ведется непрерывно, то исходные данные постоянно пополняются и элементы диагностической матрицы должны постоянно уточняться.

Пользуясь данными, представленным в диагностической матрице, для каждого из диагнозов D_i с использованием обобщенной формулы Байеса (4.3) рассчитываются после опытные (апостериорные) вероятности появления диагноза D_i при условии, что входе измерений получена конкретная реализация комплексах признаков K_s^* , то есть $P(D_i/K_s^*)$. Для удобства пользования результаты расчета заносятся в таблицу следующего вида:

Таблица 4.2

D_i	$P(D_i/K_s^*)$	$P(D_i/K_1^*)$	$P(D_i/K_2^*)$	...	$P(D_i/K_s^*)$
D_1		$P(D_1/K_1^*)$	$P(D_1/K_2^*)$	...	$P(D_1/K_s^*)$
D_2		$P(D_2/K_1^*)$	$P(D_2/K_2^*)$	...	$P(D_2/K_s^*)$
...		...	...	...	...
D_n		$P(D_n/K_1^*)$	$P(D_n/K_2^*)$		$P(D_n/K_s^*)$

Решающим правилом является выбор диагноза имеющего максимальную рассчитанную вероятность при данном наборе признаков, то есть при использовании метода Байеса объект с комплексом признаков K_s^* относится к диагнозу с наибольшей рассчитанной (апостериорные) вероятностью $P(D_i/K_s^*)$.

Следует отметить, что метод Байеса обладает рядом недостатков:

1. Для реализации данного метода необходим большой объем

экспериментальных начальных данных, которые можно получить только из эксплуатации.

2. Методу характерны большие погрешности при распознавании редких диагнозов из-за грубости оценок вероятности появления этих диагнозов.

3. Отсутствует однозначный критерий для выбора порогового значения вероятностью $P(D_i/K_s^*)$, по которому принимается решение по диагнозу.

Несмотря на имеющиеся недостатки метод Байеса достаточно эффективен и прост в реализации, в том числе и при использовании вычислительной техники.

Поясним метод Байеса на примере. Пусть при эксплуатации авиационного ГТД контролируется время выбега ротора при останове - t и вибрация корпуса двигателя V . В качестве диагностических признаков принимаются: K_1 - время выбега ротора меньше требуемого по техническим условиям; K_2 - повышенная вибрация. Для данного типа ГТД появление этих признаков связано со следующими диагнозами: D_1 - увеличенный зазор по бандажным полкам рабочих лопаток турбины; D_2 - износ беговых дорожек подшипника ротора. Исправное состояние будем обозначать D_3 .

В ходе эксплуатации было установлено, что для двигателей, находящихся в исправном состоянии, признак K_1 не встречается, а признак K_2 встречается у 10% двигателей. То есть $P(K_1/D_3) = 0$, а $P(K_2/D_3) = 0,1$.

Известно, что 85% двигателей отрабатывают свой ресурс без дефектов, для 10% двигателей наблюдается диагноз D_1 , а для 5% - диагноз D_2 . Поэтому принимаем $P(D_1) = 0,1$; $P(D_2) = 0,05$; $P(D_3) = 0,85$.

Далее, в ходе сбора статистики определено, что признак K_1 встречается у 15% двигателей, находящихся в диагнозе D_1 , и у 55% двигателей, находящихся в диагнозе D_2 . Следовательно $P(K_1/D_1) = 0,15$ и $P(K_1/D_2) = 0,55$. Признак K_2 встречается у 10% двигателей, находящихся в диагнозе D_1 , и у 50% двигателей, находящихся в диагнозе D_2 , то есть $P(K_2/D_1) = 0,1$ и $P(K_2/D_2) = 0,50$.

Составим диагностическую матрицу. При этом будем считать, что признаки K_1 и K_2 бинарные и обозначим: $\underline{K}_1$ - отсутствие первого признака и $\underline{K}_2$ - отсутствие второго признака. Тогда вероятности не появления первого и второго признаков соответственно определяются соотношениями

$$P(\underline{K}_1/D_i) = 1 - P(K_1/D_i) \text{ и } P(\underline{K}_2/D_i) = 1 - P(K_2/D_i).$$

Таблица 4.3 Диагностическая матрица

D_i	D_1	D_2	D_3
-------	-------	-------	-------

$P(K_j / D_i)$			
$P(K_1 / D_1)$	0,15	0,55	0
$P(K_1 / D_2)$	0,85	0,45	1
$P(K_2 / D_1)$	0,1	0,5	0,1
$P(K_2 / D_2)$	0,9	0,5	0,9
$P(D_i)$	0,1	0,05	0,85

Пользуясь формулой Байеса, определим вероятность появления каждого из диагнозов при различных сочетаниях признаков. Например, в соответствии с формулой (4.3), вероятность нахождения двигателя в диагнозе D_1 при наблюдении обоих признаков определится из соотношения

$$P(D_1 / K_1, K_2) = \frac{P(D_1)P(K_1/D_1)P(K_2/D_1)}{\{P(D_1)P(K_1/D_1)P(K_2/D_1) + P(D_2)P(K_1/D_2)P(K_2/D_2) + P(D_3)P(K_1/D_3)P(K_2/D_3)\}}.$$

Результаты расчетов сведем в таблицу 4.4, подобную таблице 4.2

Таблица 4.4

	$P(D_i / K_1, K_2)$	$P(D_i / \underline{K}_1, K_2)$	$P(D_i / K_1, \underline{K}_2)$	$P(D_i / \underline{K}_1, \underline{K}_2)$
D_1	0,0984	0,0811	0,4954	0,0897
D_2	0,9016	0,1074	0,5046	0,0132
D_3	0	0,8115	0	0,8971

В соответствии с полученными результатами можно сделать следующее заключение:

1. Если при диагностировании были обнаружены оба признака (повышенная вибрация и малое время выбега ротора), то можно достаточно уверенно предположить, что у подшипника ротора появился износ беговых дорожек.

2. Если при диагностировании оба признака отсутствуют, то наиболее вероятно исправное состояние двигателя.

3. Если при диагностировании обнаружен только первый признак (увеличение времени выбега), то уверенно можно сказать, что двигатель неисправен, но для разделения неисправных состояний необходимы дополнительные исследования.

4. Если при диагностировании обнаружен только второй признак (повышенная вибрация), то с большой степенью вероятности можно предположить, что двигатель находится в исправном состоянии.

4.3 Методы принятия статистических решений.

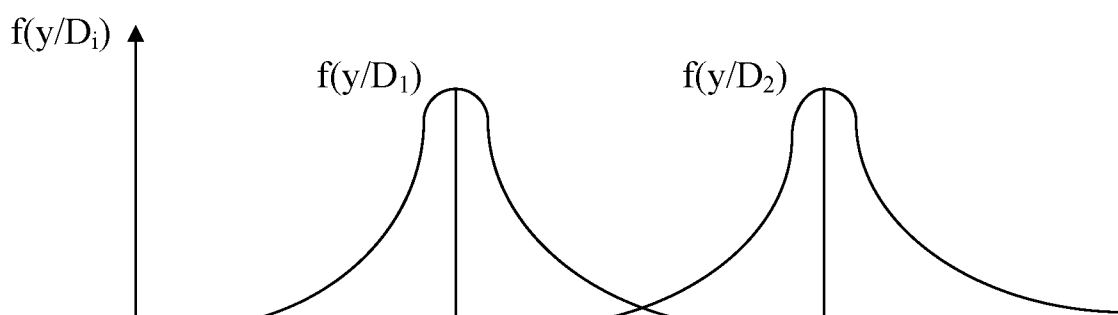
Эти методы также относятся к вероятностным методам. Они отличаются от Байесовского подхода в первую очередь правилами принятия решения. Напомним, что в методе Байеса предпочтение отдается диагнозу, имеющему максимальную величину условной вероятности появления при полученной в ходе обследования объекта диагностики реализации комплекса диагностических признаков. В этих же методах решающее правило определяется с наложением некоторого условия оптимальности, которое позволяет сократить необходимый для решения диагностической задачи объем статистической информации [2, 3, 27].

Кроме того, эти методы эффективны в случаях, когда имеются затруднения в формировании диагностического признака, поскольку не определено пороговое значение параметра, по которому принимается решение о нахождении объекта в том или ином диагнозе, то есть отсутствует критерий, необходимый для принятия решения.

Первоначально рассмотрим простейший случай, когда при решении диагностической задачи объект описывается только одним диагностическим параметром (одномерный случай). При таком рассмотрении задача распознавания диагнозов формулируется следующим образом.

Пусть диагностируется объект, который может находиться в исправном (диагноз D_1) или в неисправном (диагноз D_2) состояниях. Диагностика объекта производится с использованием выходного диагностического параметра «у», который является случайной величиной и для каждого из этих состояний описывается своим законом распределения. Обозначим распределение плотности вероятности этого параметра для исправного состояния $f(y/D_1)$, а для неисправного - $f(y/D_2)$. Примем, что при переходе объекта в неисправное состояние величина параметра возрастает. Ставится задача выбрать некоторое пороговое значение (порог) этого параметра y_0 . После выбора этого порога решающее правило формулируется таким образом: при превышении порогового значения диагностическим параметром объект следует относить к неисправному состоянию, и наоборот при значениях диагностического параметра ниже порога будем относить объект к исправному состоянию. В символьном виде это условие запишется следующим образом:

$$\text{если } y > y_0 \text{ то } y \in D_2; \quad \text{если } y < y_0 \text{ то } y \in D_1. \quad (4.4)$$



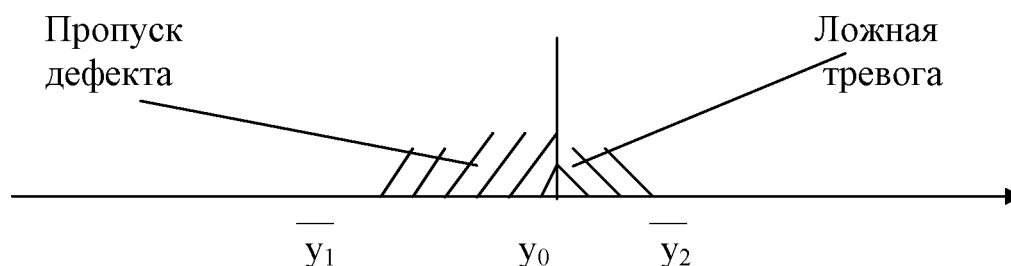


Рис 4. 2 Плотности вероятности параметра y для исправного и неисправного состояний

Существенным в этой постановке является то, что непрерывные распределения диагностического параметра « y » для исправного и неисправного состояний пересекаются (рис.4.2). Поскольку распределения $f(y/D_1)$ и $f(y/D_2)$ пересекаются, следовательно принципиально невозможно выбрать y_0 , которое не давало бы ошибочных решений. Поэтому выбор порогового значения параметра y_0 осуществляется на основе заданного условия оптимальности, которое базируется на понятиях ошибки типа «ложная тревога» и «пропуск дефекта». Определим эти понятия.

Ложная тревога. После выбора порога y_0 может возникнуть ситуация, когда $y > y_0$ и, в соответствии с условием (4.4), принимается, что объект относится к неисправному состоянию (к диагнозу D_2). На самом деле объект исправен, то есть принадлежит диагнозу D_1 . Такое событие вполне возможно из-за того, что непрерывные распределения $f(y/D_1)$ и $f(y/D_2)$ пересекаются (рис.4.2). Такое событие называется ложная тревога и обозначается H_{21} , где индекс «2» обозначает тот диагноз, который мы приняли, а индекс «1» – тот диагноз, который есть на самом деле.

Пропуск дефекта. Пропуск дефекта это событие, заключающееся в том, что неисправный объект принимается за исправный. Оно возникает тогда (рис.4.2), когда измеренная величина диагностического параметра меньше выбранного порога ($y < y_0$), но принадлежит распределению $f(y/D_2)$. Будем обозначать это событие H_{12} . Как и предыдущем случае первый индекс соответствует принятому диагнозу, а второй - действительному.

Поскольку события «ложная тревога» и «пропуск дефекта» случайные, то для их характеристики следует определить вероятность их появления при некотором пороговом значении диагностического параметра y_0 .

Вероятность появления сложного события «ложная тревога» $P(H_{21})$ это есть вероятность одновременного появления двух простых событий: нахождения объекта диагностики в диагнозе D_1 и появление значения диагностического параметра больше порогового $y > y_0$ для исправного состояния (то есть из распределения $f(y/D_1)$). В соответствии с положениями теории вероятностей

$$P(H_{21}) = P(D_1)P(y > y_0/D_1) = P_1 \int_{y_0}^{\infty} f(y/D_1) dy, \quad (4.5)$$

где $P(D_1) = P_1$ - априорная вероятность нахождения объекта в диагнозе D_1 ;

$P(y > y_0/D_1) = \int_{y_0}^{\infty} f(y/D_1) dy$ - вероятность события $y > y_0$ для исправного изделия.

Последняя вероятность соответствует площади под кривой $f(y/D_1)$ при $y > y_0$ (рис.4.2).

Вероятность появления сложного события «пропуск дефекта» $P(H_{12})$ это есть вероятность одновременного появления двух простых событий: нахождение объекта диагностики в диагнозе D_2 и появления значения диагностического параметра меньше порогового $y < y_0$ для неисправного состояния (то есть из распределения $f(y/D_2)$). Аналогично предыдущему эта вероятность определим следующим соотношением

$$P(H_{12}) = P(D_2)P(y < y_0/D_2) = P_2 \int_{-\infty}^{y_0} f(y/D_2) dy, \quad (4.6)$$

где $P(D_2) = P_2$ - априорная вероятность нахождения объекта в диагнозе D_2 ;

$P(y < y_0/D_2) = \int_{-\infty}^{y_0} f(y/D_2) dy$ - вероятность события $y < y_0$ для неисправного изделия.

Последняя вероятность соответствует площади под кривой $f(y/D_2)$ при $y < y_0$ (рис.4.2).

Полная вероятность принятия ошибочного решения складывается из вероятностей событий «ложная тревога» и «пропуск дефекта». Учитывая, что в различных ситуациях затраты на ликвидацию ошибок по ложной тревоге и пропуску дефекта различны, то им устанавливают «цены» (весовые коэффициенты): C_{21} – цена ложной тревоги и C_{12} – цена пропуска дефекта. Разумеется, цены ошибок имеют условное значение, но они должны учесть предполагаемые последствия ложной тревоги и пропуска дефекта. С учетом заданных цен и определяют суммарную ошибку:

$$R = C_{21}P_1 \int_{y_0}^{\infty} f(y/D_1) dy + C_{12}P_2 \int_{-\infty}^{y_0} f(y/D_2) dy. \quad (4.7)$$

Выражение, определенное соотношением (4.7), называется функцией риска.

Основная идея методов принятия статистических решений заключается в следующем. В зависимости от наличия и объема исходных статистических данных, по которым определяются цены ошибок C_{21} и C_{12} , априорные вероятности диагнозов P_1 и P_2 , плотности вероятностей $f(y/D_1)$ и $f(y/D_2)$, задается условие минимизация функции риска R и далее определяется пороговое значение y_0 для этого условия.

Метод минимального риска.

Этот метод применим тогда, когда имеется полная информация о ценах C_{12} и C_{21} , о вероятностях нахождения объектов в исправном D_1 и неисправном D_2 диагнозах, о функциях $f(y/D_1)$ и $f(y/D_2)$. Задача ставится таким образом, чтобы пороговое значение y_0 соответствовало минимальному значению функции риска

R.

Для нахождения экстремума функции риска продифференцируем ее по y_0 и приравняем полученное выражение к нулю:

$$\frac{dR}{dy_0} = -C_{21} \cdot P_1 \cdot f(y_0/D_1) + C_{12} \cdot P_2 \cdot f(y_0/D_2) = 0.$$

После преобразования получим:

$$C_{12} \cdot P_2 \cdot f(y_0/D_2) = C_{21} \cdot P_1 \cdot f(y_0/D_1).$$

Откуда отношение плотностей вероятностей, называемое отношением правдоподобия, для исправного и неисправного состояний при граничном значении диагностического параметра равно:

$$\frac{f(y_0/D_1)}{f(y_0/D_2)} = \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}. \quad (4.8)$$

Пользуясь выражением (3.8), при известных ценах C_{12} и C_{21} , вероятностях P_1 и P_2 и функциях $f(y/D_1)$ и $f(y/D_2)$ можно вычислить пороговое значение диагностического параметра y_0 . В дальнейшем это пороговое значение следует использовать для проверки выполнения условий соотношения (4.4) и отнесения объекта к одному из диагнозов.

Отметим, что для классификации объектов необязательно вычислять пороговое значение параметра. Ее можно проводить путем определения величины отношения правдоподобия для текущего значения диагностического параметра и сравнения этой величины со значением, рассчитанным по формуле (4.8).

Действительно, численное значение отношения $\frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}$ равно величине отношения правдоподобия $\frac{f(y_0/D_1)}{f(y_0/D_2)}$, вычисленному при подстановки в него порогового значения диагностического параметра y_0 . Из рис.4.2 следует, что величина отношения правдоподобия, вычисленного для значений диагностического параметра y меньше порогового y_0 , будет больше величины отношения правдоподобия, вычисленного для порогового значения параметра. То есть при $y < y_0$ выполняется условие $\frac{f(y/D_1)}{f(y/D_2)} > \frac{f(y_0/D_1)}{f(y_0/D_2)}$. Для значений диагностического параметра больше порогового это соотношение будет обратным, т. е. при $y > y_0$ выполняется условие $\frac{f(y/D_1)}{f(y/D_2)} < \frac{f(y_0/D_1)}{f(y_0/D_2)}$. С учетом этих соотношений классификацию диагностируемых объектов можно проводить следующим образом.

После экспериментального определения величины диагностического параметра y^* вычисляют отношение правдоподобия $\frac{f(y^*/D_1)}{f(y^*/D_2)}$. Полученное значение сравнивают с величиной $\frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}$. Решающее правило будет выглядеть следующим образом:

$$\begin{aligned}
&\text{если } \frac{f(y^*/D_1)}{f(y^*/D_2)} > \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}, \quad \text{то } y^* \in D_1; \\
&\text{если } \frac{f(y^*/D_1)}{f(y^*/D_2)} < \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}, \quad \text{то } y^* \in D_2.
\end{aligned}
\tag{4.9}$$

Сделаем два замечания к полученным результатам.

1. Для одно-модальных законов распределения диагностического параметра пороговое значение параметра y_0 расположено на числовой оси между средними значениями y_1 и y_2 , то есть $\bar{y}_1 < y_0 < \bar{y}_2$ (рис.4.2). Это вытекает из следующего. Поскольку ищется минимум функции риска, то вторая производная этой функции в точке y_0 должна быть больше нуля, то есть

$$dR'' = -C_{21} \cdot P_1 \cdot f'(y_0/D_1) + C_{12} \cdot P_2 \cdot f'(y_0/D_2) > 0.$$

После переноса в правую часть, будем иметь следующее неравенство:

$$C_{12} \cdot P_2 \cdot f'(y_0/D_2) > C_{21} \cdot P_1 \cdot f'(y_0/D_1).$$

Проведя преобразования, в окончательном виде получим следующее соотношение:

$$\frac{C_{12} \cdot P_2}{C_{21} \cdot P_1} > \frac{f'(y_0/D_1)}{f'(y_0/D_2)}.$$

$$\tag{4.10}$$

Величина отношения $\frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}$ всегда положительная. В тоже время значение производной $f'(y/D_1)$ при $y > \bar{y}_1$ всегда меньше нуля, а производной $f'(y/D_2)$ при $y < \bar{y}_2$ всегда больше нуля. Поэтому соотношение (4.10) строго выполняется при условии $\bar{y}_1 < y_0 < \bar{y}_2$.

2. Часто распределения вероятностей диагностического параметра y в исправном и неисправном состояниях подчиняются нормальному закону. При этом определены параметры законов распределений (средние значения и среднеквадратические отклонения) соответственно для исправного ($D_1 \rightarrow \bar{y}_1$ и σ_1) и для неисправного ($D_2 \rightarrow \bar{y}_2$ и σ_2) состояний. Подставив эти параметры законов в известную формулу, получим уравнения (4.11), описывающие распределения плотностей вероятностей диагностического параметра y соответственно для исправного и неисправного состояний:

$$\begin{cases} f(y/D_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} \cdot e^{-\frac{(y-\bar{y}_1)^2}{2\sigma_1^2}} \\ f(y/D_2) = \frac{1}{\sigma_2 \sqrt{2\pi}} \cdot e^{-\frac{(y-\bar{y}_2)^2}{2\sigma_2^2}} \end{cases} \quad (4.11)$$

Отношение правдоподобия для этого случая будет выглядеть следующим образом:

$$\frac{f(y_o/D_1)}{f(y_o/D_2)} = \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1} = \frac{\sigma_2 \cdot e^{-\frac{(y_o-\bar{y}_1)^2}{2\sigma_1^2}}}{\sigma_1 \cdot e^{-\frac{(y_o-\bar{y}_2)^2}{2\sigma_2^2}}}.$$

Выполним последовательно преобразования этого выражения

$$\frac{\sigma_2}{\sigma_1} \cdot e^{-\left[\frac{(y_o-\bar{y}_1)^2}{2\sigma_1^2} - \frac{(y_o-\bar{y}_2)^2}{2\sigma_2^2}\right]} = \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1},$$

и далее, в окончательном виде:

$$e^{-\left[\frac{(y_o-\bar{y}_1)^2}{2\sigma_1^2} - \frac{(y_o-\bar{y}_2)^2}{2\sigma_2^2}\right]} = \frac{\sigma_1}{\sigma_2} \cdot \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}.$$

После логарифмирования последнего выражения получим

$$-\frac{1}{2} \left[\frac{(y_o - \bar{y}_1)^2}{\sigma_1^2} - \frac{(y_o - \bar{y}_2)^2}{\sigma_2^2} \right] = \ln \frac{\sigma_1}{\sigma_2} \cdot \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}. \quad (4.12)$$

Последнее выражение (4.12) представляет собой квадратное уравнение относительно y_o . Решив это уравнение, мы получим два корня: y_{01} и y_{02} . Из полученных корней, в качестве порогового значения диагностического параметра y_o , с учетом сделанного выше первого замечания, выбирается положительное значение, удовлетворяющее условию $(\bar{y}_1 < y_o < \bar{y}_2)$.

Еще раз отметим, что этот метод применяется в том случае, если имеется полная информация о всех функциях и величинах входящих в уравнение риска. Такая ситуация не всегда имеет место.

Метод минимального числа ошибочных решений.

Этот метод применяется тогда, когда стоимости ошибки ложной тревоги C_{21} и пропуска дефектов C_{12} точно не известны, но из практических соображений можно принять, что они примерно одинаковы, то есть $C_{21} \approx C_{12}$. С учетом этого допущения определяют полную вероятность принятия ошибочного решения как сумму вероятностей ошибок ложной тревоги и пропуска дефекта в виде

$$P_{ош} = P_1 \int_{y_o}^{\infty} f(y/D_1) \cdot dy + P_2 \int_{-\infty}^{y_o} f(y/D_2) \cdot dy. \quad (4.13)$$

Выражение (4.13) определяет так называемую функцию суммарной ошибки. В методе минимального числа ошибочных решений в качестве условия оптимальности при определении порогового значения диагностического параметра y_0 принимается минимум функции суммарной ошибки.

Для нахождения значения диагностического параметра, при котором функция суммарной ошибки имеет минимум, продифференцируем выражение (4.13) и приравняем его нулю:

$$\frac{dP_{\text{ош}}}{dy_0} = -P_1 \cdot f(y/D_1) + P_2 \cdot f(y/D_2) = 0.$$

Из последнего формулы отношение правдоподобия для порогового значения диагностического параметра y_0 определяется следующим образом:

$$\frac{f(y_0/D_1)}{f(y_0/D_2)} = \frac{P_2}{P_1}. \quad (4.14)$$

Отметим, что и в методе минимального числа ошибочных решений для классификации объектов необязательно вычислять пороговое значение параметра y_0 . Как и в методе минимального риска классификацию можно проводить путем определения величины отношения правдоподобия для текущего значения диагностического параметра и сравнения этой величины со значением, рассчитанным по формуле (4.14). Для метода минимального числа ошибочных решений решающее правило будет выглядеть следующим образом:

$$\begin{aligned} \text{если } \frac{f(y^*/D_1)}{f(y^*/D_2)} &> \frac{P_2}{P_1}, \text{ то } y^* \in D_1; \\ \text{если } \frac{f(y^*/D_1)}{f(y^*/D_2)} &< \frac{P_2}{P_1}, \text{ то } y^* \in D_2. \end{aligned} \quad (4.15)$$

Если дважды продифференцировать выражение (4.13) для функции суммарной ошибки и провести рассуждения, аналогично как и в методе минимального риска, то можно показать, что и для метода минимального числа ошибочных решений в случае одно-модальных распределений плотностей вероятности выполняется условие $\bar{y}_1 < y_0 < \bar{y}_2$.

Если распределения диагностического параметра в исправном и неисправном состояниях заданы нормальным законом, то, произведя такие же преобразования, как и в методе минимального риска, мы получим квадратичное уравнение относительно y_0 . Решив это уравнение, определим пороговое значение y_0 .

Напомним, что применение этого метода предполагает равенство цен ошибок ложной тревоги и пропуска дефекта. Если при решении практических задачи методом минимального числа ошибочных решений оказывается, что $C_{12} \gg C_{21}$ (как это часто бывает на самом деле), то полученное пороговое значение параметра y_0 в результате даст большое значение ошибки пропуска дефекта. Мы получим так

называемое “неосторожное решение”.

Метод максимального правдоподобия.

Этот метод также опирается на формулу (4.7), определяющую функцию риска. Идея этого метода вытекает из следующих соображений. В практических задачах вероятность неисправного состояния мала, то есть $P_1 \gg P_2$. В тоже время цена пропуска дефекта значительно больше цены ложной тревоги, то есть $C_{21} \ll C_{12}$. Таким образом, если неизвестны вероятности состояний и цены ошибок, то можно считать

$$C_{12} \cdot P_2 \approx C_{21} \cdot P_1.$$

Тогда выражение (3.7) можно представить в виде

$$R_{MAX. nm.} = \int_{y_0}^{\infty} f(y/D_1) dy + \int_{-\infty}^{y_0} f(y/D_2) dy. \quad (4.16)$$

В данном методе условием оптимизации является нахождение порогового значения y_0 , удовлетворяющему минимуму выражения (4.16). Как и в предыдущих случаях для нахождения минимума проведем дифференцирование этого выражения и после преобразований получим отношение правдоподобия, соответствующее пороговому значению диагностического параметра y_0 в виде

$$\frac{f(y_0/D_1)}{f(y_0/D_2)} = 1. \quad (4.17)$$

Для этого метода справедливы все те же замечания, что и для предыдущих. А именно:

1. Классификацию объектов диагностики можно проводить с использованием порогового отношения правдоподобия, рассчитанного по формуле (4.17). Для метода максимального правдоподобия решающее правило будет выглядеть следующим образом:

$$\text{если } \frac{f(y^*/D_1)}{f(y^*/D_2)} > 1, \text{ то } y^* \in D_1; \quad (4.18)$$

$$\text{если } \frac{f(y^*/D_1)}{f(y^*/D_2)} < 1, \text{ то } y^* \in D_2.$$

2. В случае одно-модальных распределений диагностического параметра для исправного D_1 и неисправного D_2 состояний выполняется условие $\bar{y}_1 < y_0 < \bar{y}_2$.

3. Если распределения диагностического параметра в исправном и неисправном состояниях заданы нормальным законом, то произведя необходимые преобразования, мы получим квадратичное уравнение относительно y_0 . Решая это уравнение определим пороговое значение y_0 .

Решение по методу максимального правдоподобия оказывается близким к решению по методу минимального риска.

Метод минимакса.

Метод предназначен для ситуации, когда отсутствуют предварительные статистические сведения о вероятностях появления диагнозов D_1 и D_2 , то есть о

P_1 и P_2 . Решение ведется для «наихудшего случая». Рассматривается ситуация, когда P_1 и P_2 имеют наименее благоприятные значения, приводящие к максимальному значению функции риска R . Значение порога y_0 ищут такое, которое минимизирует функцию риска R при этих неблагоприятных значениях P_1 и P_2 . Отсюда и название метода - *метод минимакса*.

В такой постановки функция риска становится функцией двух параметров: y_0 и P_1 (так как $P_2 = 1 - P_1$):

$$R(y_0, P_1) = C_{21} P_1 \int_{y_0}^{\infty} f(y/D_1) dy + C_{12} (1 - P_1) \int_{-\infty}^{y_0} f(y/D_2) dy. \quad (4.19)$$

Поскольку функция записана как функция двух параметров P_1 и y_0 , будем проводить минимизацию функции риска по двум этим параметрам. Для нахождения экстремумов по P_1 и y_0 необходимо взять частные производные по этим аргументам и приравнять их нулю. Продифференцировав выражение (4.19) по y_0 и приравняв нулю, получим:

$$\frac{\partial R}{\partial y_0} = -C_{21} P_1 f(y_0/D_1) + C_{12} f(y_0/D_2) - C_{12} P_1 f(y_0/D_2) = 0.$$

После соответствующих преобразований и подстановки $y = y_0$ приходим к отношению правдоподобия (назовем его условием 1) в виде:

$$\frac{f(y_0/D_1)}{f(y_0/D_2)} = \frac{C_{12} \cdot (1 - P_1)}{C_{21} \cdot P_1}. \quad (4.20)$$

Проведя дифференцирование выражения (3.19) по аргументу P_1 и, приравняв производную нулю, получаем следующее выражение

$$\frac{\partial R}{\partial P_1} = C_{21} \int_{y_0}^{\infty} f(y/D_1) dy - C_{12} \int_{-\infty}^{y_0} f(y/D_2) dy = 0.$$

После необходимых преобразований получаем еще одно соотношение (назовем его условием 2):

$$C_{21} \int_{y_0}^{\infty} f(y/D_1) dy = C_{12} \int_{-\infty}^{y_0} f(y/D_2) dy. \quad (4.21)$$

Далее требуется определить y_0 и P_1 такие, чтобы они удовлетворяли условиям 1 и 2, то есть необходимо определить корни системы уравнений, состоящей из двух уравнений, заданных условиями 1 и 2.

Данную систему уравнений решают с помощью метода приближений Ньютона, который связывает между собой исходное (предыдущее) $y_{0(n-1)}$ и последующее y_{0n} приближения для порогового значения диагностического параметра. Для выполнения этих приближений вводят функцию вида

$$\phi(y_0) = C_{21} \int_{y_0}^{\infty} f(y/D_1) dy - C_{12} \int_{-\infty}^{y_0} f(y/D_2) dy$$

Это выражение, которое определяет $\varphi(y_0)$, удобнее переписать с использованием интегральных функций распределения диагностического параметра в исправном $F(y/D_1)$ и неисправном состояниях $F(y/D_2)$. При такой замене последнее выражение примет вид:

$$\varphi(y_0) = C_{21}[1 - F(y_0/D_1)] - C_{12}[F(y_0/D_2)]. \quad (4.22)$$

Уравнение, связывающее между собой исходное (предыдущее) $y_{0(n-1)}$ и последующее y_{0n} приближения порогового значения, по методу Ньютона записывается следующим образом:

$$y_{0(n)} = y_{0(n-1)} - \frac{\varphi(y_{0(n-1)})}{\varphi'(y_{0(n-1)})}. \quad (4.23)$$

Производную функции $\varphi(y_0)$ можно определить с учетом выражения (4.22):

$$\varphi'(y_0) = -C_{21}f(y_0/D_1) - C_{12}f(y_0/D_2). \quad (4.24)$$

На базе полученных соотношений и проводится определение порогового значения диагностического параметра y_0 путем последовательного вычисления приближений. В качестве первого приближения y_0 принимается величина

$y_{01} = \frac{\bar{y}_1 + \bar{y}_2}{2}$. Второе приближение вычисляется с использованием соотношений (4.22), (4.23), (4.24) по выражению $y_{02} = y_{01} - \frac{\varphi(y_{01})}{\varphi'(y_{01})}$. Третье - по выражению $y_{03} = y_{02} - \frac{\varphi(y_{02})}{\varphi'(y_{02})}$ и так далее.

Расчет ведется до тех пор, пока отличия между последующими приближениями станет меньше выбранной погрешности. После определения y_0 его значение подставляется в формулу соответствующую условию 1, и из этой формулы определяется вероятность P_1 . После необходимых преобразований формулу для вычисления P_1 получим в следующем виде:

$$P_1 = \frac{C_{12}}{C_{12} + C_{21} \left[\frac{f(y_0/D_1)}{f(y_0/D_2)} \right]}.$$

Далее вычисляют вероятность $P_2 = 1 - P_1$. После определения порогового значения диагностического параметра, вероятностей исправного и неисправного состояний становится возможным определить величины ошибок ложной тревоги и пропуска дефекта, а также величины суммарной ошибки и функции риска.

Выше был рассмотрен случай, когда неисправность приводила к увеличению диагностического параметра. Если неисправность приводит к уменьшению параметра, то соотношение между плотностями распределения будет выглядеть следующим образом (рис.4.3).

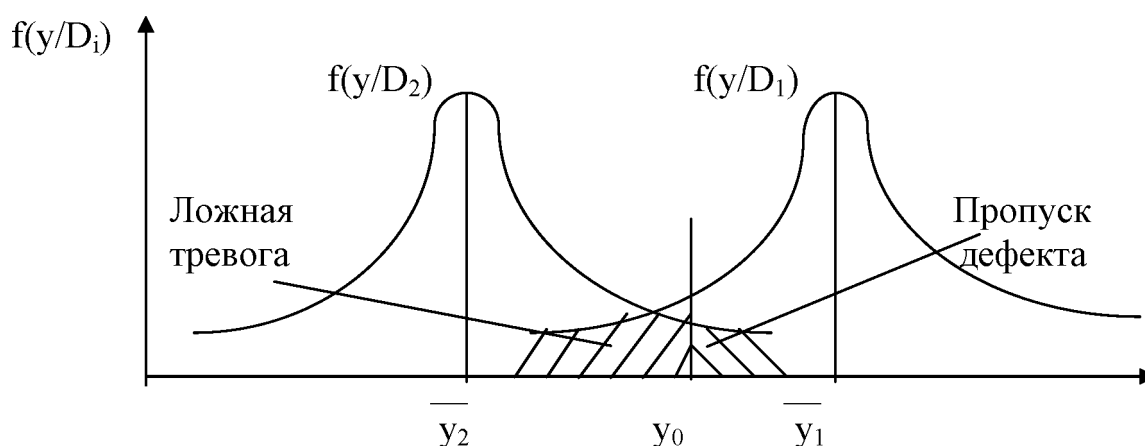


Рис 4. 3 Плотности вероятности «у» для исправного и неисправного состояний в случае, когда при неисправности «у» уменьшается

При этом условие классификации (4.4) измениться, и будет выглядеть следующим образом

$$\text{если } y > y_0 \text{ то } y \in D_1; \quad \text{если } y < y_0 \text{ то } y \in D_2. \quad (4.25)$$

Площади под кривыми плотностей распределения, относящиеся соответственно к ошибкам ложной тревоги и пропуска дефекта, поменяются местами (рис.4.3). А выражение для функции риска будет иметь следующий вид:

$$R = C_{21}P_1 \int_{-\infty}^{y_0} f(y/D_1)dy + C_{12}P_2 \int_{y_0}^{\infty} f(y/D_2)dy. \quad (4.26)$$

Отличие последнего выражения от выражения (4.7) заключается в изменении пределов интегрирования. Как и в предыдущих случаях, для нахождения минимума функции риска продифференцируем выражение (4.26) и полученную производную приравняем нулю:

$$\frac{dR}{dy_0} = C_{21}P_1 f(y_0/D_1) - C_{12}P_2 f(y_0/D_2) = 0.$$

После небольших преобразований получим отношение правдоподобия:

$$\frac{f(y_0/D_1)}{f(y_0/D_2)} = \frac{C_{21}P_1}{C_{12}P_2},$$

которое совпадает с отношением правдоподобия, определенным выражением (4.8). Если повторить все выкладки, которые были проведены для метода минимального риска при $\bar{y}_1 < \bar{y}_2$, то мы получим аналогичные результаты и для рассматриваемого случая. То есть при $\bar{y}_2 < \bar{y}_1$ справедливы следующие замечания:

Для классификации объектов диагностики с помощью отношения правдоподобия можно использовать решающее правило, определенное

выражениями (4.9).

2. Для одно модальных законов распределения имеет место соотношение $\bar{y}_2 < y_0 < \bar{y}_1$.

3. Если распределение диагностического параметра в исправном и неисправном состояниях подчиняется нормальному закону, то для определения численного значения порога y_0 следует использовать квадратное уравнение (4.12).

Аналогичным образом можно показать, что результаты, полученные для методов минимального числа ошибочных решений и максимального правдоподобия при условии $\bar{y}_1 < \bar{y}_2$, будут справедливы и при условии $\bar{y}_2 < \bar{y}_1$.

В методе минимакса для случая $\bar{y}_2 < \bar{y}_1$, из за перестановки пределов интегрирования в функции риск произойдут некоторые изменения. Для этого случая функция $\phi(y_0)$ будет выглядеть следующим образом:

$$\phi(y_0) = C_{21} \int_{-\infty}^{y_0} f(y/D_1) dy - C_{12} \int_{y_0}^{-\infty} f(y/D_2) dy,$$

а при представлении через интегральный закон эта функция будет иметь вид:

$$\phi(y_0) = C_{21} F(y_0/D_1) - C_{12} [1 - F(y_0/D_2)].$$

Производная же этой функции определится следующим выражением:

$$\phi'(y_0) = C_{21} f(y_0/D_1) + C_{12} f(y_0/D_2).$$

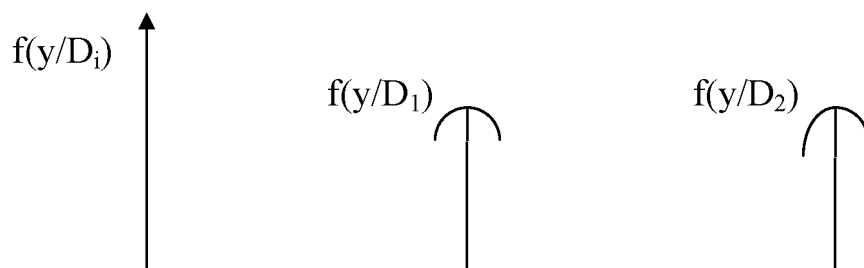
Во всем остальном формулы и порядок определения порогового значения диагностического параметра по методу минимакса не изменятся.

Отметим, что на практике при определении порогового значения диагностического параметра с помощью методов принятия статистических решений, как правило, используются несколько методов. В дальнейшем сравнивают результаты, и выбирается то решение, которое более соответствует условиям поставленной задачи.

В заключение сделаем три замечания, касающиеся некоторых особенностей применения методов принятия статистических решений.

1. Решение при наличии зоны неопределенности.

В некоторых случае, когда требуется высокая надежность распознавания (например, если цена пропуска дефекта C_{12} очень высока), имеет смысл ввести так называемую зону неопределенности. При попадании диагностического признака в эту зону принимается решение об отказе диагностирования. Графически этот случай иллюстрируется на рисунке 4.4.



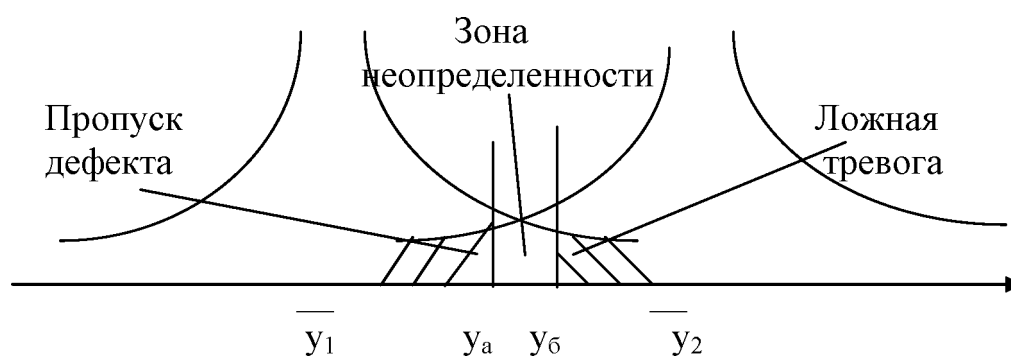


Рис 4. 4 Принятие статистических решений при наличии зоны неопределенности

В этом случае решающее правило будет выглядеть следующим образом :

если $y \leq y_a$ то $y \in D_1$; если $y \geq y_b$ то $y \in D_2$;

если $y_a < y < y_b$ то отказ от распознавания.

Ситуация отказа от распознавания является событием, которое свидетельствует о том, что имеющейся информации недостаточно для принятия решения.

При введении зоны неопределенности изменится выражение для функции риска . В этом выражении появятся слагаемые, характеризующие вероятность такого события, как отказ от распознавания. То есть в функции риска должны появиться дополнительные слагаемые, определяющие вероятность выполнения условия $y_a < y < y_b$ при нахождении объекта диагностики в исправном и неисправном состояниях. Пользуясь известными соотношения теории вероятностей [2, 4] эти дополнительные слагаемые представим в виде:

$P_1 \int_{y_a}^{y_b} f(y/D_1) dy$ - вероятность выполнения условия $y_a < y < y_b$ при

нахождении объекта в исправном состоянии;

$P_2 \int_{y_a}^{y_b} f(y/D_2) dy$ - вероятность выполнения условия $y_a < y < y_b$ при нахождении

объекта в неисправном состоянии.

Полностью функция риска при наличии зоны неопределенности запишется следующим образом:

$$R = C_{21} P_1 \int_{y_b}^{\infty} f(y/D_1) dy + C_{12} P_2 \int_{-\infty}^{y_a} f(y/D_2) dy + \\ + C_0 \left[P_1 \int_{y_a}^{y_b} f(y/D_1) dy + P_2 \int_{y_a}^{y_b} f(y/D_2) dy \right] ,$$

где C_0 - цена отказа от распознавания.

Из последнего выражения следует, что функция риска становится функцией двух переменных y_a и y_b . Для их определения необходимо найти частные производные функции риска по этим переменным и приравнять их нулю. Выполнив соответствующие преобразования можно получить два отношения

правдоподобия, которые имеют следующий вид:

$$\frac{f(y_a / D_1)}{f(y_a / D_2)} = \frac{(C_{12} - C_0) \cdot P_2}{C_0 \cdot P_1} \quad \text{и} \quad \frac{f(y_6 / D_1)}{f(y_6 / D_2)} = \frac{C_0 \cdot P_2}{(C_{21} - C_0) \cdot P_1}.$$

Решая эти соотношения, определяют численные значения для y_a и y_6 .

Проведя соответствующие рассуждения и преобразования аналогичные соотношения можно получить для методов минимального числа ошибочных решений и максимального правдоподобия.

2. Решение в случае нескольких технических состояний.

Все рассмотренные выше результаты относились к двум состояниям (диагнозам) D_1 и D_2 . Методы принятия статистических решений можно распространить и на любое число состояний. Наиболее простым приемом в этом случае является следующий подход:

- 1) Выбирается один из наиболее вероятных диагнозов D_i , и его принимают за первый диагноз D_1 .
- 2) Все остальные диагнозы объединяют в один, который условно определяется как диагноз D_2 .

В дальнейшем решение проводится теми же способами, что и для двух состояний. Последовательно принимая за диагноз D_1 каждый из известных диагнозов можно определить решающие правила для всей совокупности диагнозов.

3. Решение для случая, когда техническое состояние объекта определяется несколькими параметрами.

Рассмотрим случай, когда объект может находиться в двух диагнозах D_1 и D_2 , а его техническое состояние характеризуется комплексом диагностических параметров, который можно представить в виде вектора: $\bar{y} = \{y_1, y_2, \dots, y_n\}$. Пусть известны, полученные из статистических материалов, многомерные распределения вектора диагностических параметров в исправном и неисправном состояниях:

$$\left. \begin{aligned} f(\bar{y}/D_1) &= f(y_1, y_2, y_3, \dots, y_n/D_1) \\ f(\bar{y}/D_2) &= f(y_1, y_2, y_3, \dots, y_n/D_2) \end{aligned} \right\}.$$

При многомерном описании объекта n -ым количеством диагностических параметров каждому объекту можно поставить в соответствие некоторый вектор, координаты которого определены соответствующими диагностическими параметрами. Причем каждое техническое состояние будет отражаться в этом n -мерном пространстве параметров некоторой областью. Примем, что исправному состоянию D_1 соответствует область S_1 , а неисправному D_2 - область S_2 .

При такой постановке задачи решающее правило формулируется следующим

образом:

если вектор параметров $\bar{y}$ попадает в область S_1 , то объект находится в исправном состоянии D_1 ;

если вектор параметров $\bar{y}$ попадает в область S_2 , то объект находится в неисправном состоянии D_2 .

С учетом изложенного выше функция риска примет следующий вид:

$$R = C_{21}P_1 \int_{S_2} f(\bar{y}/D_1) dy + C_{12}P_2 \int_{S_1} f(\bar{y}/D_2) dy.$$

Таким образом, при описании объекта n -ным количеством диагностических параметров функция риска записывается через n -мерные интегралы.

После соответствующих преобразований придем опять к соотношению правдоподобия, которое в этом случае будет иметь вид:

$$\frac{f(\bar{y}_o/D_1)}{f(\bar{y}_o/D_2)} = \frac{C_{12} \cdot P_2}{C_{21} \cdot P_1}.$$

Отметим, что как правило, параметры считаются независимыми. Поэтому из теории вероятностей следует:

$$f(\bar{y}/D_1) = f(y_1/D_1) \cdot f(y_2/D_1) \cdot \dots \cdot f(y_n/D_1),$$

$$f(\bar{y}/D_2) = f(y_1/D_2) \cdot f(y_2/D_2) \cdot \dots \cdot f(y_n/D_2),$$

что позволяет упростить отношение правдоподобия.

Следует помнить, что методы принятия статистических решений рекомендуется использовать тогда, когда размерность задачи не превышает 2 – 3 (то есть когда объект описывается не более чем двумя или тремя параметрами) и когда число распознаваемых диагнозов не превышает трех. Для решения задачи диагностирования в случае описания объекта большим числом параметров более приемлемы так называемые методы распознавания в пространстве признаков.

4.4 Методы распознавания в пространстве параметров

Методы распознавания (разделения) диагнозов в пространстве параметров [2, 3] основываются на том, что каждый объект описывается набором (комплексом)

диагностических параметров, которые принимаются в качестве компонентов n -мерного вектора. Поэтому каждому объекту можно поставить в соответствие точку в пространстве параметров, в которой находится конец этого вектора. Считается, что точки, отображающие одинаковые технические состояния (диагнозы), группируются в одной области параметров, то есть образуют “компактное множество”. Следовательно, каждому диагнозу соответствует некоторая область в пространстве параметров. В такой трактовке задача постановки диагноза сводится к выработке формализованного метода, позволяющего определить в какой области пространства параметров находится конец вектора, компоненты которого соответствуют комплексу диагностических параметров, измеренного у контролируемого объекта.

Прежде чем переходить непосредственно к самим методам распознавания рассмотрим некоторые *свойства пространственных параметров*.

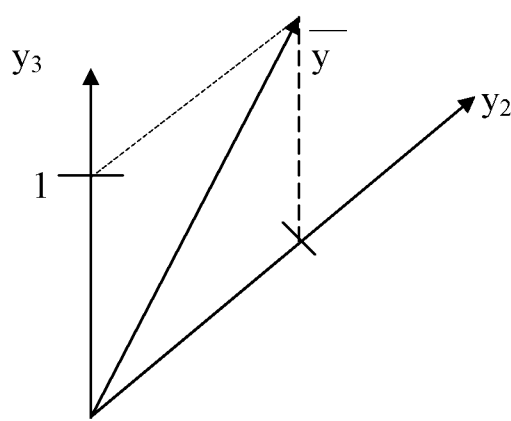
1. Как отмечалось ранее (см. гл. 1) каждый объект (его техническое состояние) может характеризоваться комплексом (набором) диагностических признаков или параметров, которые в равной степени могут приниматься в качестве компонент n -мерного вектора. В зависимости от того, что используется для описания объекта (параметры или признаки) компоненты этого вектора могут быть дискретные или непрерывные.

Дискретные компоненты встречаются тогда, когда техническое состояние объекта описывается диагностическими признаками или их разрядами.

В качестве *непрерывных компонент* могут выступать непосредственно непрерывные диагностические параметры (давление P , температура t , расход топлива G_T и другие).

2. Если для описания технического состояния объекта используются простые бинарные (двух разрядные) признаки, то компоненты вектора могут выражаться двоичными числами, то есть сочетанием 0 и 1. Например, в трехмерном пространстве техническое состояние объекта может быть определено следующим вектором $\bar{y} = \{0, 1, 1\}$.

На рис.4.5 изображено трехмерное пространство признаков и вектор с указанными координатами. При таком описании объектов диагностики любое из технических состояний (диагнозов) будет находиться в вершине единичного куба. Таким образом, если число признаков n , то каждый i -тый диагноз D_i будет являться вершиной единичного n -мерного куба.



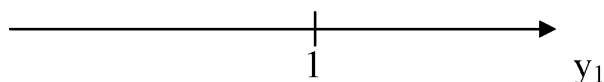


Рис. 4.5 Трехмерное пространство признаков

3. Областью диагноза D_i называется множество точек пространства признаков, которое соответствует техническим объектам, обладающим состоянием D_i . Запись $\bar{y} \in D_i$ имеет двойной смысл: во первых это означает, что объект находится в состоянии D_i ; во вторых это означает, что точка описываемая вектором $\bar{y}$ попадает в область D_i .

При использовании понятий пространства диагностических параметров (признаков) и определении каждого состояния в виде области диагноза, задача диагностики может решаться путем отыскания уравнения разделяющей функции, или разделяющей плоскости. Уравнение разделяющей функции выводится через так называемые дискриминантные функции.

Дискриминантные функции.

Пусть в n -мерном пространстве параметров содержатся точки, принадлежащие различным диагнозам $D_1, D_2, \dots, D_i, \dots, D_m$. Для каждой области пространства параметров (диагноза) D_i определим скалярную функцию $f_i(\bar{y}) = f_i(y_1, y_2, \dots, y_k, \dots, y_n)$. Функция $f_i(\bar{y})$ называется дискриминантной функцией диагноза D_i , если она отвечает условию:

$$f_i(\bar{y}) > f_j(\bar{y}) \text{ при } \bar{y} \in D_i, \text{ где } i \neq j, \text{ а } j = 1, 2, \dots, m.$$

То есть, дискриминантная функция $f_i(\bar{y})$ принимает для точек области D_i (диагноза D_i) n -мерного пространства параметров наибольшее значение по сравнению с дискриминантными функциями других диагнозов. Линейная дискриминантная функция диагноза D_i будет иметь вид:

$$f_i(\bar{y}) = \lambda_{i1} \cdot y_1 + \lambda_{i2} \cdot y_2 + \dots + \lambda_{in} \cdot y_n + \lambda_{i(n+1)}, \quad (4.27)$$

где $\lambda_{i1}, \dots, \lambda_{in}, \lambda_{i(n+1)}$ - весовые коэффициенты линейной дискриминантной функции $f_i(\bar{y})$.

Поясним это положение на следующем примере. Пусть мы имеем два диагноза D_1 и D_2 . Техническое состояние объектов диагностики характеризуется двумя параметрами y_1 и y_2 . Пусть для этих диагнозов заданы некоторые функции, которые определяются следующими выражениями:

$$\begin{aligned} f_1(y_1, y_2) &= \lambda_{11} \cdot y_1 + \lambda_{12} \cdot y_2 + \lambda_{13}, \\ f_2(y_1, y_2) &= \lambda_{21} \cdot y_1 + \lambda_{22} \cdot y_2 + \lambda_{23}. \end{aligned} \quad (4.28)$$

Функция $f_1(y_1, y_2)$ называется дискриминантной функцией диагноза D_1 , если она

всегда больше $f_2(y_1, y_2)$ при условии, что объект относится к диагнозу D_1 , то есть если $\bar{y}=\{y_1, y_2\} \in D_1$. Функция $f_2(y_1, y_2)$ называется дискриминантной функцией диагноза D_2 , если она всегда больше $f_1(y_1, y_2)$ при условии, что объект относится к диагнозу D_2 , то есть если $\bar{y}=\{y_1, y_2\} \in D_2$.

Разделяющая функция.

Допустим, что области диагнозов D_1 и D_2 имеют общую границу, как это изображено на рис.4.6.

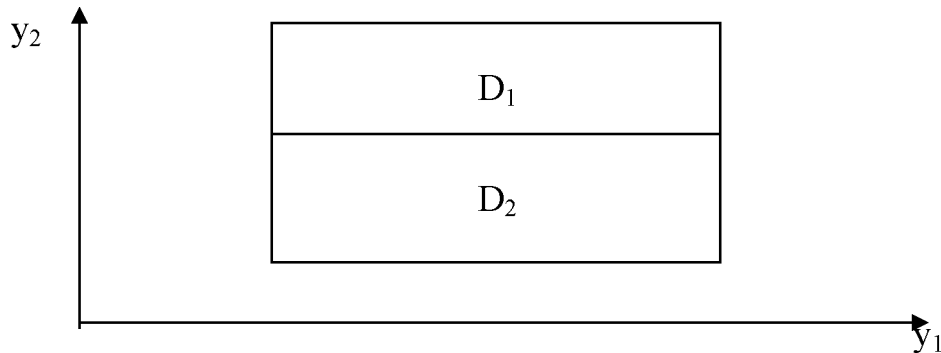


Рис.4.6 Области диагнозов с общей границей

Существование общей границы у областей диагнозов определяет равенство на этой границе дискриминантных функций, то есть на общей границе областей $f_1(y_1, y_2) = f_2(y_1, y_2)$. При $\bar{y}=\{y_1, y_2\} \in D_1$ выполняется условие $f_1(y_1, y_2) > f_2(y_1, y_2)$, а при $\bar{y}=\{y_1, y_2\} \in D_2$ - следующее условие $f_1(y_1, y_2) < f_2(y_1, y_2)$.

На основе этих свойств дискриминантной функции вводится понятие разделяющей функции, которая в нашем случае определится таким образом:

$$f(\bar{y}) = f(y_1, y_2) = f_2(y_1, y_2) - f_1(y_1, y_2). \quad (4.29)$$

Последнее выражение определяет линейную разделяющую функцию для двух диагнозов в двух мерном пространстве диагностических параметров, которая представляет собой прямую линию. Подставив выражения (3.28) в формулу (3.29) получим:

$$f(\bar{y}) = (\lambda_{21} - \lambda_{11}) \cdot y_1 + (\lambda_{22} - \lambda_{12}) \cdot y_1 + (\lambda_{23} - \lambda_{13}).$$

Последнее выражение после несложных преобразований можно записать в виде

$$f(\bar{y}) = \lambda_1 \cdot y_1 + \lambda_2 \cdot y_1 + \lambda_3,$$

где λ_i – весовые коэффициенты разделяющей функции.

При количестве параметров больше двух линейная разделяющая функция описывает разделяющую плоскость. В случае двух диагнозов при n -мерном пространстве параметров выражение для линейной разделяющей функции будет иметь следующий вид

$$f(\bar{y}) = \lambda_1 \cdot y_1 + \lambda_2 \cdot y_2 + \dots + \lambda_i \cdot y_i + \dots + \lambda_n \cdot y_n + \lambda_{(n+1)}, \quad (4.30)$$

а весовые коэффициенты этой функции определяются соотношением $\lambda_i = (\lambda_{2i} -$

λ_{li}).

С использованием понятия разделяющей функции решающее правило будет выглядеть следующим образом:

$$\begin{aligned} f(y_1, y_2) > 0 &\Rightarrow y \in D_1, \\ f(y_1, y_2) < 0 &\Rightarrow y \in D_2. \end{aligned} \quad (4.31)$$

В ряде случаев для повышения надежности распознавания вводят зону неопределенности или так называемый «порог чувствительности» ε (рис. 4.7).

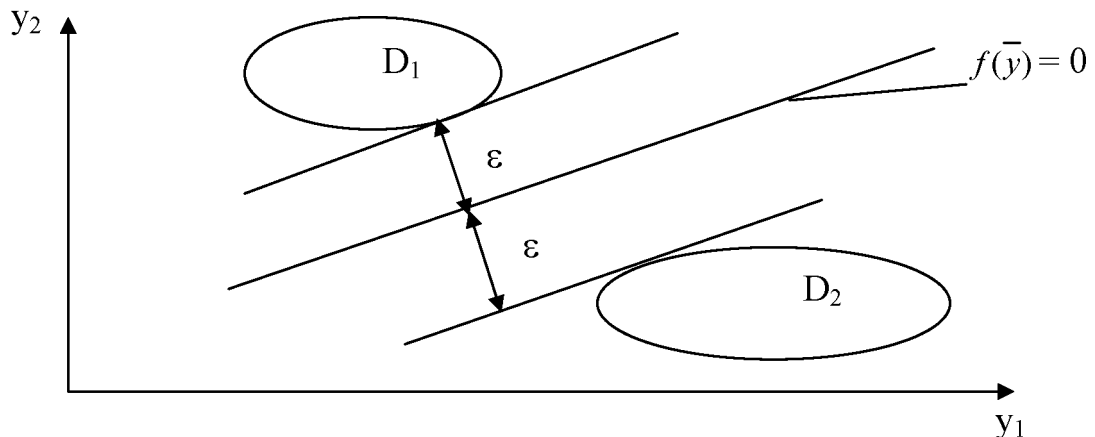


Рис. 4.7 Области диагнозов с зоной неопределенности

Тогда решающее правило будет выглядеть следующим образом, если

$$f(y_1, y_2) \geq \varepsilon \Rightarrow D_2, \text{ если } f(y_1, y_2) \leq -\varepsilon \Rightarrow D_1,$$

где ε - пороговое значение разделяющей функции.

Если выполняется условие $-\varepsilon < f(y_1, y_2) < \varepsilon$, то принимается решение об отказе от диагноза.

Все сказанное выше позволяет сделать заключение о том, что в общем случае для двух диагнозов D_i и D_j разделяющая функция определяется через дискриминантные функции этих диагнозов, то есть в виде:

$$f(\bar{y}) = f_i(\bar{y}) - f_j(\bar{y}).$$

Методы распознавания в пространстве параметров, которые строятся на основе линейных разделяющих функций, называются *линейными методами распознавания (разделения)*. В свою очередь, диагнозы, которые возможно разделить с помощью этих методов, называются *линейно разделимыми*.

Как следует из выражения (4.30) знание весовых коэффициентов полностью определяет разделяющую функцию. Комплекс весовых коэффициентов можно представить как вектор с числом компонент $n+1$. Поэтому если вектор весовых коэффициентов $\bar{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{n+1})$ определен, то решать задачу распознавания диагнозов можно путем вычисления значений разделяющей функции при тех значениях измеренных параметров, которые получены из опыта. После расчета значения разделяющей функции следует воспользоваться решающим правилом,

заданным выражением (4.31).

Таким образом, основной задачей в такой постановке является поиск вектора весовых коэффициентов. Прежде чем переходить к поиску этого вектора проанализируем его свойства. Анализ проведем на основе следующих соображений.

Выше отмечалось, что вектор весовых коэффициентов задается $n+1$ компонентами, то есть описывается $n+1$ -мерным пространством. Дополним n -мерный вектор диагностических параметров еще одним параметром (компонентом) $y_{n+1} \equiv 1$, то есть дополненный вектор диагностических параметров запишем в следующем виде:

$$\bar{y}_* = \{y_1, y_2, \dots, y_k, \dots, y_n, y_{n+1}\}.$$

При таком дополнении разделяющую функцию $f(\bar{y}_*)$ можно представить в виде скалярного произведения вектора весовых коэффициентов $\bar{\lambda}$ и дополненного вектора диагностических параметров $\bar{y}_*$:

$$f(\bar{y}_*) = \bar{\lambda} \cdot \bar{y}_*. \quad (4.32)$$

Это скалярное произведение описывает уравнение плоскости в $n+1$ -мерном пространстве (гиперплоскость) и в развернутом виде записывается следующим образом

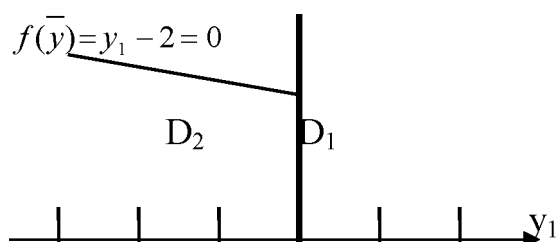
$$f(\bar{y}_*) = \lambda_1 \cdot y_1 + \lambda_2 \cdot y_2 + \dots + \lambda_k \cdot y_k + \dots + \lambda_n \cdot y_n + \lambda_{n+1} \cdot y_{n+1} = 0. \quad (4.33)$$

Последние два выражения позволяют сформулировать два свойства вектора весовых коэффициентов и разделяющей гиперплоскости, которые будут использоваться при их определении:

1. Вектор весовых коэффициентов и гиперплоскость перпендикулярны друг другу.
2. В дополненном пространстве признаков разделяющая гиперплоскость всегда проходит через начало координат.

Поясним на простейшем примере указанные свойства.

Пусть объект диагностики может находиться в одном из двух диагнозов D_1 или D_2 . Техническое состояние объекта характеризуется диагностическим параметром y_1 (рис. 4.8).



Разделяющая функция описывается уравнением $f(\bar{y}) = y_1 - 2 = 0$, а решающее правило будет выглядеть следующим образом:
если $y_1 < 2 \Rightarrow D_2$,

-1 0 1 2 3 4

если $y_1 > 2 \Rightarrow D_1$.

Рис.4.8 Разделяющая функция в основном пространстве признаков

Введем дополнительный параметр $y_2 \equiv 1$. Тогда уравнение разделяющей функции $f(\bar{y})$ определится следующим выражением $f(\bar{y}) = y_1 - 2y_2 = 0$ (рис.4.9), а выражение для вектора весовых коэффициентов будет выглядеть следующим образом: $\bar{\lambda} = \{1, -2\}$.

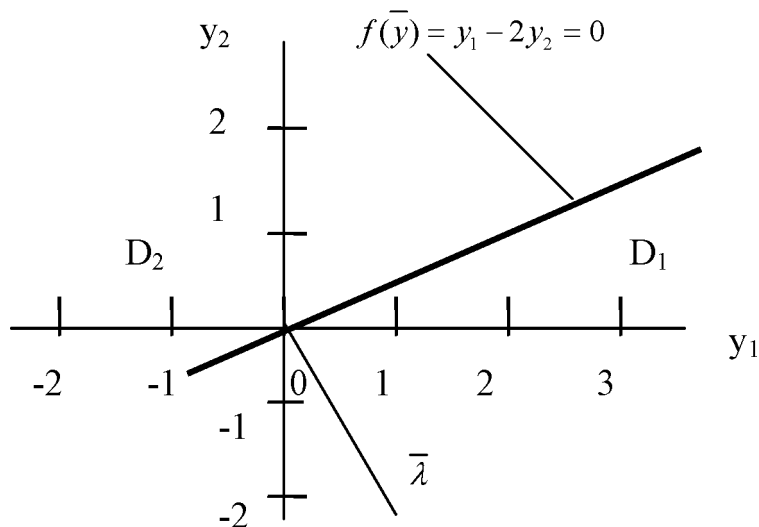


Рис.4.9 Разделяющая функция в дополненном пространстве признаков

Полученные соотношения и рис.4.9 свидетельствуют о том, что разделяющая функция в дополненном пространстве диагностических признаков проходит через начало координат, а вектор весовых коэффициентов перпендикулярен к ней.

Нахождение вектора весовых коэффициентов разделяющей функции.

Рассмотрим случай, когда объект диагностики может находиться в двух диагнозах D_1 и D_2 , а его техническое состояние характеризуется двумя диагностическими параметрами y_1 и y_2 . Решение поставленной задачи в двухмерном варианте позволяет использовать наглядную графическую иллюстрацию, а, в то же время, перенос полученных результатов на многомерный

случай не вызывает затруднений.

Решение рассматриваемой задачи возможно при наличии некоторой априорной информации о значениях диагностических параметров объектов, находящихся в обоих технических состояниях. Пусть известны значения диагностических параметров y_1 и y_2 для числа M_1 технических объектов (образцов), находящихся в диагнозе D_1 , и для числа M_2 образцов, находящихся в диагнозе D_2 . Эти образцы изображены на рис. 4.10 в виде точек, куда попадают концы соответствующих векторов. Совокупность векторов попадающих в область диагноза D_1 называется обучающей выборкой диагноза D_1 . Совокупность векторов попадающих в область диагноза D_2 называется обучающей выборкой диагноза D_2 .

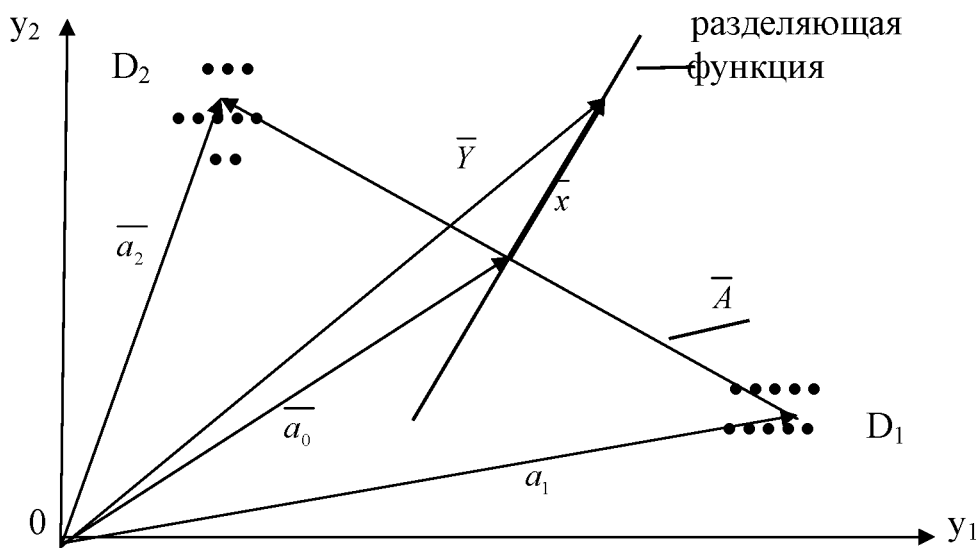


Рис. 4. 10 Построение разделяющей функции в двух мерном пространстве параметров

Введем понятие «средний вектор» или «эталон» для каждого из диагнозов D_1 и D_2 . В векторной форме для диагноза D_1 эталон определится следующим выражением:

$$\bar{a}_1 = \frac{1}{M_1} \sum_{s=1}^{M_1} \bar{a}_{s1},$$

где $\bar{a}_{s1}$ - s - тый вектор, конец которого попадает в область диагноза D_1 .

Очевидно, что в скалярной форме нахождение эталона сводится к определению средних значений по каждой координате.

Соответственно для диагноза D_2 в векторной форме эталон определяется следующим выражением:

$$\bar{a}_2 = \frac{1}{M_2} \sum_{s=1}^{M_2} \bar{a}_{s2},$$

где $\bar{a}_{s2}$ - s - тый вектор, конец которого попадает в область диагноза D_2 .

Определим вектор $\overline{A}$, соединяющий концы эталонов. Из рис. 4.10 видно, что этот вектор можно определить как разность эталонов:

$$\overline{A} = \overline{a_2} - \overline{a_1}. \quad (3.34)$$

Без доказательств примем, что разделяющая функция проходит через середину этого вектора и она перпендикулярна к нему. Это соответствует физическому смыслу разделяющей функции, поскольку она должна быть равно удалена от обеих областей диагнозов и от D_1 и от D_2 . Определим среднюю точку на векторе $\overline{A}$ и соединим ее с началом координат. Вектор $\overline{a_0}$, соединяющий начало координат с этой точкой (рис.4.10), определится следующим соотношением:

$$\overline{a_1} + \frac{\overline{A}}{2} = \overline{a_0},$$

или после соответствующих подстановок и преобразований - следующим соотношением:

$$\overline{a_0} = \overline{a_1} + \frac{\overline{a_2} - \overline{a_1}}{2} = \frac{\overline{a_1} + \overline{a_2}}{2}. \quad (4.35)$$

Рассмотрим вектор $\overline{Y}$, конец которого лежит на прямой, принятой за разделяющую функцию (рис.4.10). Поскольку конец вектора $\overline{Y}$ лежит на разделяющей функции, то, если подставить значение его координат в уравнение разделяющей функции, уравнение разделяющей функции примет значение равное нулю. Этот вектор можно представить как сумму вектора $\overline{a_0}$ и вектора $\overline{x}$, лежащего на разделяющей функции, то есть:

$$\overline{Y} = \overline{a_0} + \overline{x} \quad \text{или} \quad \overline{x} = \overline{Y} - \overline{a_0}. \quad (4.36)$$

Поскольку вектор $\overline{x}$ лежит на разделяющей функции, то его скалярное произведение с вектором $\overline{A}$ равно нулю: $\overline{A} \cdot \overline{x} = 0$. Произведя соответствующую подстановку, используя выражения (4.34) и (4.36) получим:

$$(\overline{a_2} - \overline{a_1}) \cdot (\overline{Y} - \overline{a_0}) = 0.$$

Раскрывая скобки и произведя замену $\overline{a_0}$ через выражение (4.35) мы придем к следующему соотношению:

$$(\overline{a_2} - \overline{a_1}) \cdot \overline{Y} - \overline{a_0} \cdot (\overline{a_2} - \overline{a_1}) = (\overline{a_2} - \overline{a_1}) \cdot \overline{Y} - \frac{\overline{a_2} + \overline{a_1}}{2} \cdot (\overline{a_2} - \overline{a_1}) = 0. \quad (4.37)$$

Для выполнения непосредственных вычислений необходимо перейти от векторной формы записи уравнения к скалярной, для чего вместо векторов в формулу (4.37) нужно поставить их координаты. Для нашего двух мерного случая вектора, представленные в формуле (4.37), определим через их координаты в следующем виде:

$$\bar{Y} = \{y_1, y_2\}, \quad \bar{a}_1 = \{a_{11}, a_{12}\}, \quad \bar{a}_2 = \{a_{21}, a_{22}\}.$$

Подставив координаты векторов в формулу (4.37) получим следующее выражение:

$$(a_{21} - a_{11}) \cdot y_1 + (a_{22} - a_{12}) \cdot y_2 + \frac{1}{2}(a_1^2 - a_2^2) = 0, \quad (4.38)$$

где a_1^2 и a_2^2 квадраты модулей соответствующих векторов.

Сопоставляя полученное выражение с выражением (4.33) для разделяющей функции в дополненном пространстве параметров, получим выражения для вычисления компонент весового вектора в рассматриваемом нами примере:

$$\lambda_1 = (a_{21} - a_{11}), \quad \lambda_2 = (a_{22} - a_{12}), \quad \lambda_3 = \frac{1}{2}(a_1^2 - a_2^2).$$

Полученные результаты несложно перенести на n -мерное пространство параметров. Для этого представим рассматриваемые вектора $\bar{Y}$, $\bar{a}_1$, $\bar{a}_2$ через их координаты (компоненты) в n -мерном пространстве параметров:

$$\bar{Y} = \{y_1, y_2, y_3, \dots, y_n\}, \quad \bar{a}_1 = \{a_{11}, a_{12}, a_{13}, \dots, a_{1n}\},$$

$$\bar{a}_2 = \{a_{21}, a_{22}, a_{23}, \dots, a_{2n}\}.$$

Проведя рассуждения аналогичные двух мерному случаю, получим выражение для разделяющей функции (гиперплоскости), справедливое для n -мерного случая:

$$(a_{21} - a_{11}) \cdot y_1 + (a_{22} - a_{12}) \cdot y_2 + \dots + (a_{2n} - a_{1n}) \cdot y_n + \frac{1}{2}(a_1^2 - a_2^2) = 0. \quad (4.39)$$

Данное уравнение соответствует уравнению разделяющей плоскости, записанному в общем виде. Из этого уравнения определим выражения для компонент весового вектора:

$$\lambda_1 = a_{21} - a_{11}, \quad \lambda_2 = a_{22} - a_{12}, \quad \dots, \quad \lambda_n = a_{2n} - a_{1n},$$

$$\lambda_{n+1} = \frac{1}{2}(a_1^2 - a_2^2).$$

(4.40)

Знание компонент весового вектора позволяет проводить процедуру диагностирования.

*Алгоритм определения координат весового вектора
и постановки диагноза на его основе.*

Постановка задачи. Пусть имеются диагнозы, D_1 и D_2 каждый, из которых описывается набором векторов.

1. Определяют эталоны (средние вектора) для каждого диагноза:

$$\bar{a}_1 = \frac{1}{M_1} \sum_{s=1}^{M_1} \bar{a}_{s1} \quad \text{и} \quad \bar{a}_2 = \frac{1}{M_2} \sum_{s=1}^{M_2} \bar{a}_{s2},$$

а точнее их компоненты:

$$a_{11}, a_{12}, \dots, a_{1n}, a_{21}, a_{22}, \dots, a_{2n}.$$

2. Определяют модули эталонных векторов:

$$a_1^2 = a_{11}^2 + a_{12}^2 + \dots + a_{1n}^2 \quad \text{и} \quad a_2^2 = a_{21}^2 + a_{22}^2 + \dots + a_{2n}^2.$$

2. Определяют компоненты весового вектора $\bar{\lambda}$ (весовые коэффициенты разделяющей функции):

$$\lambda_1 = a_{21} - a_{11}, \lambda_2 = a_{22} - a_{12}, \dots, \lambda_n = a_{2n} - a_{1n}, \lambda_{n+1} = \frac{1}{2}(a_1^2 - a_2^2).$$

4. Проводят диагностирование. Предъявленный объект представляется вектором своих параметров $\bar{Y} = \{y_1, y_2, y_3, \dots, y_n\}$. Далее определяют скалярное произведение $\bar{Y} \cdot \bar{\lambda}$. Если $\bar{Y} \cdot \bar{\lambda} > 0$ то объект относят к D_2 ; если $\bar{Y} \cdot \bar{\lambda} < 0$, то объект относят к D_1 .

Применение разделяющих функций или плоскостей можно реализовать просто в том случае, если эти функции линейны и количество диагнозов не превышает двух.

Для решения задач в случае большего количества диагнозов или при нелинейной постановке чаще используются метрические методы распознавания в пространстве параметров.

Метрические методы распознавания в пространстве параметров.

Эти методы также основаны на представлении (описании) диагностируемого объекта вектором в n -мерном пространстве параметров. Базовым понятием в этих методах является понятие *метрика пространства* диагностических параметров. Поэтому рассмотрим первоначально это понятие. Начнем с наиболее распространенной метрики - евклидовой.

Евклидова метрика. Пусть имеются две точки в пространстве параметров a и

у. Точка а описывается в пространстве координатами $a_1, a_2, \dots, a_j, \dots, a_n$, а точка у описывается в пространстве координатами $y_1, y_2, \dots, y_j, \dots, y_n$. Если расстояние между этими точками определяется соотношением

$$l(a, y) = \sqrt{(a_1 - y_1)^2 + (a_2 - y_2)^2 + \dots + (a_j - y_j)^2 + \dots + (a_n - y_n)^2}, \quad (4.41)$$

то такое пространство называется евклидовым. Величина, определенная по формуле (4.41), называется *евклидовым расстоянием* или *евклидовой метрикой*. В евклидовом пространстве его метрика обладает рядом свойств. Приведем эти свойства:

$$\begin{aligned} l(a, y) &\geq 0, \quad l(a, y) = l(y, a), \quad l(a, a) = 0, \\ l(a, c) &\leq l(a, b) + l(b, c). \end{aligned} \quad (4.42)$$

В ряде диагностических задач целесообразнее использовать так называемые *обобщенное расстояние порядка k* или *обобщенную метрику порядка k*. Обобщенная метрика порядка k для n - мерного пространства параметров определяется следующим соотношением:

$$l_k(a, y) = \left(\sum_{j=1}^n |a_j - y_j|^k \right)^{\frac{1}{k}},$$

где k может принимать значения от 1 до ∞ . При $k = 2$ последнее выражение определяет евклидову метрику.

Наиболее часто применяется, евклидова метрика, которая записывается без корня:

$$L^2 = l_2(a, y) = \sum_{j=1}^n (a_j - y_j)^2. \quad (4.43)$$

Рассмотренные выше метрики применимы только в том случае, если векторы диагностических параметров (признаков) имеют одинаковую размерность, так как при вычислении метрик используется операция сложения разностей различных параметров. Такое пространство параметров называют однородным или изотропным.

Если диагностические параметры (признаки) имеют разные единицы измерения, то пространство параметров неоднородно. При использовании такого пространства параметров производится нормализация координат путем введения весовых коэффициентов:

$$L^2 = \sum_{j=1}^n \lambda_j^2 \cdot (a_j - y_j)^2, \quad (4.44)$$

где λ_j - весовой коэффициент по j - той координате.

Чаще всего в качестве нормализующего весового коэффициента по j - той координате используют величина обратная среднеквадратическому отклонению диагностического параметра y_j , то есть $\lambda_j = \frac{1}{\sigma_j}$. В этом случае весовой

коэффициент λ_j имеет определенный физический смысл. Чем меньше рассеивание диагностического параметра или признака, тем больше его вклад в метрику, то есть тем больше его значимость для диагностики технического состояния объекта.

Диагностирование с помощью метрик называется *диагностированием по расстоянию между диагнозами в пространстве параметров*. Диагностирование по расстоянию различают на :

- методы диагностирования *по расстоянию до эталона*;
- методы диагностирования *по расстоянию до множества*.

Диагностирование по расстоянию до эталона.

В этом методе отнесение предъявленного для распознавания объекта к одному из диагнозов совершается по наименьшему расстоянию до эталона, характеризующего этот диагноз. В качестве эталона для некоторого диагноза принимается «типичный» для данного диагноза объект. Наиболее естественно определить «типичный» объект как некий средний для данного диагноза объект.

Пусть имеется диагноз D_i . Известно, что M_i объектов попадают в область диагноза D_i и каждый из этих объектов характеризуется своим вектором параметров a_{is} (где s изменяется от 1 до M_i). С учетом сказанного выше в качестве эталона примем объект, который описывается средним вектором параметров $\overline{a_i}$ (где $\overline{a_i} = \frac{1}{M_i} \sum_{s=1}^{M_i} \overline{a_{is}}$). Компоненты этого вектора по каждой из координат определяются как средние значения соответствующих координат всех M_i объектов, попадающих в диагноз D_i . Например, по j -той координате компонента a_i^j для среднего вектора $\overline{a_i}$ определится следующим образом:

$$a_i^j = \frac{1}{M_i} \sum_{s=1}^{M_i} a_{is}^j, \quad (4.45)$$

где a_{is}^j - значение j -той координаты (j -того параметра) s -того объекта, входящего в диагноз D_i .

После определения эталона можно проводить процедуру диагностирования.

Алгоритм диагностирования по расстоянию до эталона.

1. Для каждого i -того диагноза по априорным данным определяют эталоны, которые задаются средними векторами $\overline{a_i}$. Координаты средних векторов рассчитывают по формуле (4.45).

2. Для не изотропных пространств параметров определяют среднее квадратичные значения σ_{ij} для i -того диагноза по j -той координате (j -тому диагностическому параметру).

3. Проводят расчет расстояний (метрики) между вектором, характеризующим предъявленный к диагностированию объект, и векторами эталонов.

4. Объект относят к тому диагнозу, с которым имеет минимальное расстояние или минимальную метрику : $L_i = \min$.

Необходимо отметить, что надежность распознавания диагноза тем выше, чем меньше расстояние L_i по сравнению с другими. Сопоставление этих расстояний можно провести следующим образом. Введем величину, обратную этому расстоянию, то есть $1/L_i$. Сопоставление расстояний L_i ведется по следующему отношению:

$$\varepsilon_i = \frac{1/L_i}{\sum_{m=1}^M \frac{1}{L_m}} ,$$

где M - число известных диагнозов, i и m - текущие номера диагнозов. Вычисленная по этому отношению величина эквивалентная вероятности появления i -ого диагноза, так как $\sum_{i=1}^M \varepsilon_i = 1$.

Диагностирование по расстоянию до множества.

В этом методе оценивается расстояние не до одного эталона, а до всех точек, попадающих в данную область диагноза D_i . При этом в качестве оценки расстояния до диагноза чаще всего используются один из двух достаточно простых критериев.

В качестве первого критерия используется минимальное расстояние от конца вектора диагностических параметров, характеризующего диагностируемый объект, до конца вектора диагностических параметров, характеризующего объект, принадлежащий диагнозу D_i . В этом случае алгоритм распознавания строится следующим образом.

Алгоритм диагностирования по минимальному расстоянию.

1. Определяется расстояние $l(\bar{Y}, \bar{a}_{is})$ от предъявленного объекта, характеризуемого вектором $\bar{Y}$, до всех объектов данного диагноза D_i , характеризуемых собственными векторами диагностических признаков $\bar{a}_{is}$:

$$l(\bar{Y}, \bar{a}_{is}) = L_{is}^2 = \sum_{j=1}^n \lambda_{ij}^2 \cdot (a_{is}^j - y_j)^2 . \quad (4.46)$$

2. Для диагноза D_i определяют минимальное расстояние $l(\bar{Y}, \bar{a}_{is})_{\min}$.

3. Подобную процедуру выполняют для все известных диагнозов.

4. Проводят сравнение выявленных минимальных расстояний по все диагнозам. Объект относят к тому диагнозу, с которым он имеет наименьшее расстояние из этих минимальных расстояний.

В качестве второго критерия используют среднее расстояние от конца вектора, характеризующего диагностируемый объект, до области диагноза D_i . В этом случае алгоритм распознавания строится следующим образом.

Алгоритм диагностирования по среднему расстоянию.

1. Определяется расстояние $l(\bar{Y}, \overline{a_{is}})$ от предъявленного объекта, характеризуемого вектором $\bar{Y}$, до всех объектов данного диагноза D_i , характеризующих собственными векторами диагностических признаков $\overline{a_{is}}$ в соответствии с формулой (4.46).

2. Для каждого диагноза D_i определяют среднее значение расстояний до диагностируемого объекта:

$$l_{CP}(\bar{Y}, \overline{a_{is}}) = l(\bar{Y}, D_i) = \frac{1}{M_s} \sum_{s=1}^{M_s} l(\bar{Y}, \overline{a_{is}}).$$

3. Объект относится к тому техническому состоянию (к тому диагнозу), с которым он имеет минимальное среднее расстояние.

Напомним, что применение методов распознавания предполагает обладание некоторым количеством априорной информации. Так использование методов разделения в пространстве параметров требует, чтобы каждый из диагнозов был предварительно описан первоначальной выборкой, в которую входили бы не менее десяти объектов.

5. ОСНОВЫ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЯ АВИАЦИОННОЙ ТЕХНИКИ.

5.1 Основные положения

В предыдущих главах были изложены основы технической диагностики, то есть науки об определении состояния технических систем в настоящий момент времени. Между тем выше отмечалось, что при оценке состояния авиационной техники необходимо решать три типа задач, а именно задачи генезиса, диагноза и прогноза.

Важность решения задач прогноза связано с тем, что использование методов прогнозирования технического состояния для авиационной техники направлено на оценку остаточных ресурсов, на обоснование режимов технической эксплуатации, на выбор периодичности и объемов технического обслуживания и ремонта, а также на выявление потенциальных отказов и обеспечение принятия решения до возникновения фактического отказа.

К сожалению, в настоящее время задачи прогнозирования в практическом плане разработаны слабо, но учитывая их важность ниже будут рассмотрены основные положения прогнозирования состояния авиационной техники [2, 8, 14, 27, 31].

Задачу прогнозирования технического состояния авиационной техники будем рассматривать как задачу прогнозирования потенциальных неисправностей или отказов. В теоретическом плане для решения прогностических задач необходимо:

- сформулировать задачу прогноза конкретной неисправности (или отказа) и определить связанные с этой неисправностью диагностические параметры;
- накопить необходимые исходные данные о изменении выбранных диагностических параметров до настоящего времени;
- разработать (выбрать) модель развития технического состояния по данной неисправности (модель изменения диагностических параметров при развитии дефекта);
- выбрать метод прогнозирования и определить параметры модели прогноза технического состояния;
- произвести расчет и по результатам расчета принять решение.

При такой постановке задачи прогнозирование технического состояния авиационной техники базируется на выявлении тенденции изменения (тенденции развития) диагностических признаков и параметров. Исходными данными для прогнозирования являются полученные в процессе эксплуатации значения диагностических признаков и параметров, зарегистрированные при различных значениях наработки изделия. Важнейшей характеристикой исходных данных является минимальная длина интервала предыстории, на котором проявляются закономерности изменения прогнозируемого явления. Свойства этой априорной информации в значительной степени определяют выбор той или

иной модели для формализованного описания развития неисправности и изменения диагностических параметров.

Формирование прогнозной модели является довольно сложной самостоятельной задачей. При этом без учета физической сущности прогнозируемого явления не обойтись. Определение законов изменения физических процессов основано на одном из фундаментальных постулатов, сформулированном К. Шенном [40]. Суть постулата в том, что основные закономерности, наблюдаемые в прошлом, будут сохраняться и будущем.

Сложность и разнообразие прогнозируемых физических явлений привело к созданию большого количества методов прогнозирования. Выбор и подгонка того или иного метода прогнозирования является довольно сложной задачей. На правильный выбор метода в первую очередь влияют количество априорной информации о прогнозируемом явлении, ее достоверность, адекватность прогнозируемому явлению.

Итак, в процессе прогнозирования можно выделить четыре основных этапа: сбор и подготовка исходных данных, идентификация прогнозируемого явления, выбор метода прогнозирования, собственно прогнозирование.

Из всего сказанного выше следует, что качество исходных данных имеет огромное значение на окончательные результаты прогнозирования. Рассмотрим некоторые особенности изменения диагностических параметров, которые могут оказать влияние на процедуру прогнозирования.

Если зарегистрировать изменение диагностических параметров по времени в процессе наработки изделия, то мы получим реализацию некоторой непрерывной функции. По виду таких функций можно судить об изменении технического состояния диагностируемого объекта. При сравнении записей одноименных диагностических параметров для однотипных объектов можно обнаружить, что эти записи отличаются друг от друга. В зависимости от свойств зарегистрированных в эксплуатации исходных данных могут возникнуть различные ситуации.

На рис. 5.1,а показан случай, когда закономерность изменения параметра по наработке достаточно очевидна. Для этого случая можно указать на два основных свойства реализаций этих параметров.

1. Для реализации параметра по одному объекту характерно практическое отсутствие случайного разброса (флуктуаций).
2. Реализации параметра по различным объектам имеют идентичный вид и практически не пересекаются (не перемешиваются).

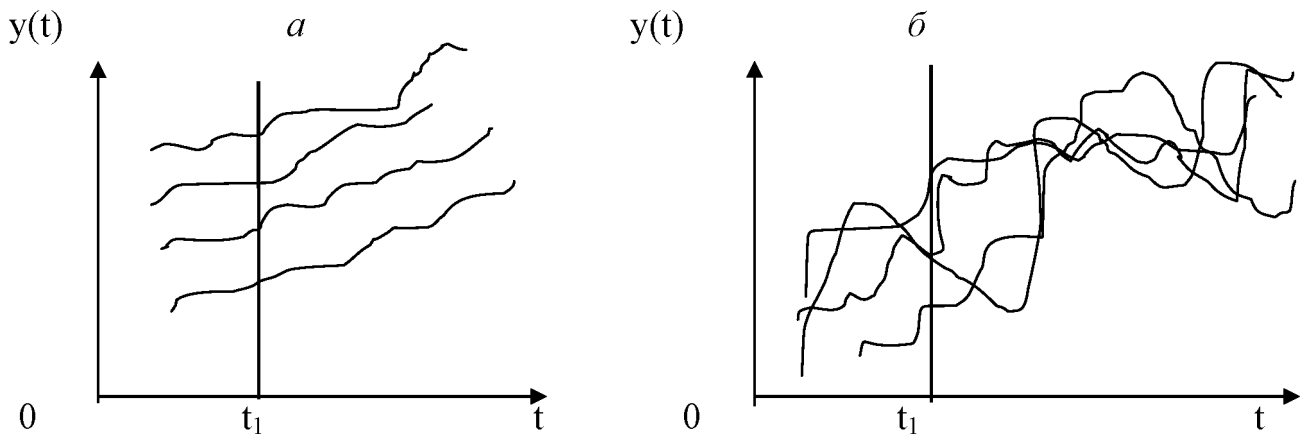


Рис.5.1 Слабое (а) и сильное (б) перемешивание реализаций диагностических параметров

На рис.5.1,б показан случай, когда в зависимости параметра по наработке закономерности явно не прослеживаются. И в этой ситуации также можно выделить два основных свойства этих реализаций.

1. Даже для реализации параметра по одному объекту наблюдается существенный разброс значений (сильная флуктуация параметра).
2. Реализации одноименных параметров с различных экземпляров однотипных объектов сильно пересекаются (сильно перемешаны).

В зависимости от того, какому из этих случаев соответствует зафиксированная реализация диагностического параметра, будут различаться и методы прогнозирования. В то же время, не смотря на существенное различие зависимостей, изображенных на рис. 5.1,а и 5.1,б они обладают рядом общих свойств, которые заключаются в следующем.

Проведем сечение в момент времени t_1 . Для этого момента времени совокупность значений диагностического параметра, взятые с разных реализаций (то есть зарегистрированные для различных объектов) будут представлять собой совокупность значений случайной величины. В целом же, набор реализаций любого диагностического параметра, полученный по парку эксплуатируемых объектов, будет представлять собой совокупность (ансамбль) реализаций так называемой случайной многомерной функции.

Для случайной многомерной функции характерно то, что ее значение случайным образом зависит и от времени и от номера реализации. Следовательно, для описания поведения диагностического параметра по всему парку эксплуатируемых объектов (по всей совокупности реализаций) необходимо использовать аппарат теории вероятностей. Причем в каждый момент времени диагностический параметр будет характеризоваться соответствующим законом распределения (например, плотностью вероятностей $f(y)$) и моментными характеристиками (например, средним значением m_y , дисперсией D_y или средним квадратическим отклонением σ_y и так далее). Отсюда следует, что прогнозирование технического состояния для всего парка эксплуатируемых объектов, возможно, осуществить путем отслеживания изменения по наработке

закона распределения или моментных характеристик диагностического параметра, полученных по всей совокупности реализаций этого параметра.

В качестве примера рассмотрим случай, когда распределение диагностического параметра в каждый момент времени описывается нормальным законом. Для задания этого закона достаточно определить среднее m_y и среднее квадратическое отклонением σ_y этого параметра. Рассчитав по известным зависимостям [2, 4, 20] эти моментные характеристики для различных моментов времени мы определим поведение исследуемого диагностического параметра во всем интервале времени эксплуатации и для всего парка эксплуатируемых объектов.

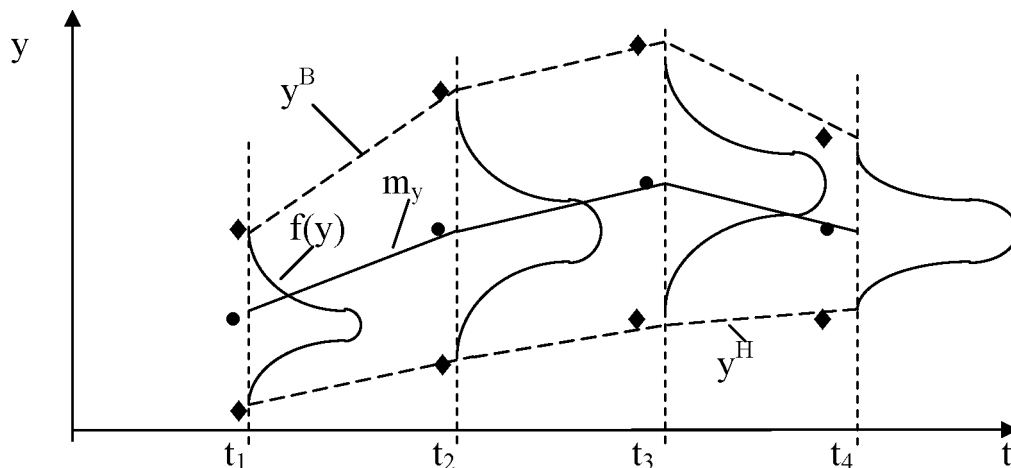


Рис.5.2 Изменение во времени закона распределения, среднего значения и доверительного интервала

На рис. 5.2 представлено изменение во времени среднего значения m_y параметра y и его доверительного интервала, в который попадает диагностический параметр. Если принять доверительную вероятность $P = 0,997$, то доверительный интервал равен $\pm 3\sigma_y$. Поэтому верхняя граница статистического диапазона изменения диагностического параметра определялась по следующему выражению:

$$y^B = m_y + 3\sigma_y.$$

Соответственно нижняя граница - по следующему выражению:

$$y^H = m_y - 3\sigma_y.$$

Для наглядности на рис.5.2 изображены и функции плотности распределения параметра $f(y)$ для выбранных моментов времени.

Подводя итог проведенным выше рассуждениям еще раз отметим, что при описании поведения диагностического параметра по всему парку эксплуатируемых объектов рассмотренный подход справедлив для обоих случаев, изображенных на рис.5.1. Следовательно, при прогнозировании технического состояния всего парка эксплуатируемых объектов необходимо использовать вероятностно-статистические характеристики диагностического параметра, определенные по всей совокупности реализаций этого параметра.

На практике часто ставится задача провести так называемое индивидуальное прогнозирование. То есть провести прогнозирование для одного эксплуатируемого объекта по единственной реализации диагностического параметра, зарегистрированной на этом объекте. При такой постановке диагностической задачи ее решение будет зависеть от того, к какому виду относятся зарегистрированные реализации диагностического параметра: к виду, изображенному на рис.5.1,а или на рис.5.1,б.

Если реализации диагностического параметра имеют слабое перемешивание, то индивидуальное прогнозирование проводят на основе прогнозирования изменений значений непосредственно самого параметра.

Если реализации диагностического параметра имеют сильное перемешивание, то индивидуальное прогнозирование можно проводить путем оценки изменений вероятностно-статистических характеристик диагностического параметра, вычисленных по одной реализации, зарегистрированной с диагностируемого объекта.

Поэтому прежде чем начать процесс прогнозирования, прежде чем выбрать модель и метод прогноза следует оценить степень перемешивания реализаций. Провести подобную оценку можно несколькими способами: непосредственно по внешнему виду ансамбля реализаций; исходя из физической сущности процесса, характеризуемого анализируемым диагностическим параметром; определив величину нормированного коэффициента корреляции диагностического параметра для моментов времени t_1 и t_2 .

Величина нормированного коэффициента корреляции $r_y(t_1, t_2)$ характеризует степень тесноты связи между совокупностями значений реализаций диагностического параметра для двух моментов времени t_1 и t_2 и поэтому является объективной оценкой степени перемешивания этих реализаций. Этот коэффициент оценивается следующим соотношением:

$$r_y(t_1, t_2) = \frac{1}{\sigma_{y1} \cdot \sigma_{y2}} \cdot \frac{\sum_{i=1}^n (y_i(t_1) - m_{y1})(y_i(t_2) - m_{y2})}{n}, \quad (5.1)$$

где n - число реализаций параметра y ; $y_i(t_1)$ и $y_i(t_2)$ - значение параметра с i -той реализации для моментов времени t_1 и t_2 ; m_{y1} , m_{y2} , σ_{y1} , σ_{y2} - соответственно средние и средние квадратические отклонение параметра для моментов времени t_1 и t_2 .

Считается, что если $r_y(t_1, t_2) \geq 0,6$, то связь между значениями диагностического параметра в моменты времени t_1 и t_2 достаточно тесная, степень перемешивания реализаций слабая и можно вести индивидуальное прогнозирование непосредственно по величине самого параметра. Если $0,3 \leq r_y(t_1, t_2) < 0,6$, то перемешивание реализаций умеренное. Прогнозирование имеет смысл вести только с использованием вероятностно-статистических характеристик диагностического параметра. При $r_y(t_1, t_2) < 0,3$, наблюдается

сильное перемешивание и слабая связь, что свидетельствует о не целесообразности проведения прогнозирования.

Вслед за оценкой степени перемешанности реализаций диагностического параметра выбирают модель и метод прогноза. На этом этапе решается ряд важных вопросов.

Во первых, определяется какое будет проводиться прогнозирование: индивидуальное или для группы эксплуатируемых объектов.

Во вторых, определяется по какой характеристике будет проводиться прогнозирование. Например, по зависимости самого диагностического параметра от наработки или по зависимости от наработки вероятностно- статистических характеристик этого параметра.

В третьих, определяется вид аналитической зависимости, с помощью которой следует аппроксимировать выбранную для прогнозирования характеристику.

В четвертых, на базе имеющихся статистических данных проводится расчет параметров выбранной аналитической зависимости.

На заключительном этапе по полученным аппроксимирующим зависимостям проводят расчет на заданный интервал наработки прогнозируемой характеристики и на основе полученных расчетных значений принимают решение о техническом состоянии эксплуатируемого объекта или объектов в будущем.

Как отмечалось выше, выбор прогнозируемой характеристики определяется физической сущностью протекающих в диагностируемом объекте процессов и зависит от степени перемешанности реализаций диагностического параметра. Выбор вида аналитического выражения, аппроксимирующего зависимость прогнозируемой характеристики от наработки, также существенно зависит от физической сущности процессов, протекающих в объекте.

Накопленный опыт эксплуатации свидетельствует о том, что в практических задачах технической диагностики авиационной техники наиболее распространенными являются следующие модели изменения прогнозируемых характеристик по наработке:

1. Модель в виде полинома первой степени (линейная зависимость):

$$Z = f_1(t) = a_0 + a_1 t. \quad (5.2)$$

2. Модель в виде полинома второй степени (квадратичная зависимость):

$$Z = f_2(t) = a_0 + a_1 t + a_2 t^2, \quad (5.3)$$

где Z - выбранная в качестве прогнозируемой характеристика (сам параметр или его вероятностно-статистическая характеристика);

t - время наработки; a_i ($i = 0, 1, 2$) - параметры (коэффициенты) прогнозной модели.

3. Модель в виде экспоненциальной зависимости:

$$Z = f_3(t) = e^{\alpha t}, \quad (5.4)$$

где α - параметр экспоненциальной прогнозной модели.

Расчет параметров выбранной прогнозной модели осуществляется формализованными методами. Наиболее распространенным методом является *метод наименьших квадратов* [4, 20]. В основу этого метода положено требование минимизации квадратов отклонений значений прогнозируемой характеристики, полученных из эксплуатации, от значений прогнозируемой характеристики, рассчитанных по модели:

$$u = \sum_{i=1}^n (Z_{\text{э}}(t_i) - Z(t_i))^2 = \min, \quad (5.5)$$

где $Z(t_i)$ - расчетное значение прогнозируемой характеристики в момент времени t_i ; $Z_{\text{э}}(t_i)$ - значение прогнозируемой характеристики, определенное по данным эксплуатации на момент наработки t_i .

Для примера рассмотрим модели в виде многочленов. При выборе линейной зависимости вида (5.2) требование (5.5) будет выглядеть следующим образом:

$$u = \sum_{i=1}^n (Z_{\text{э}}(t_i) - a_0 - a_1 t_i)^2 = \min. \quad (5.6)$$

Нахождение коэффициентов a_0 и a_1 , соответствующих минимуму функции (5.6) требует определения частных производных и приравнивания их нулю, то есть

$$\begin{aligned} \frac{\partial u}{\partial a_0} &= -2 \sum_{i=1}^n (Z_{\text{э}}(t_i) - a_0 - a_1 t_i) = 0, \\ \frac{\partial u}{\partial a_1} &= -2 \sum_{i=1}^n (Z_{\text{э}}(t_i) - a_0 - a_1 t_i) t_i = 0. \end{aligned}$$

После раскрытия скобок, суммирования и соответствующих преобразований получим систему из двух уравнений:

$$a_0 n + a_1 \sum_{i=1}^n t_i = \sum_{i=1}^n Z_{\text{э}}(t_i) \quad \text{и} \quad a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n t_i^2 = \sum_{i=1}^n Z_{\text{э}}(t_i) \cdot t_i.$$

Решив эту систему мы определим параметры аппроксимирующей зависимости (5.2) a_0 и a_1 .

Для квадратичной зависимости (5.3) требование (5.5) будет выглядеть несколько по иному:

$$u = \sum_{i=1}^n (Z_{\text{э}}(t_i) - a_0 - a_1 t_i - a_2 t_i^2)^2 = \min.$$

Проведя аналогичные преобразования, что и в случае линейной зависимости, можно получить систему из трех уравнений:

$$\begin{aligned}
a_0 n + a_1 \sum_{i=1}^n t_i + a_2 \sum_{i=1}^n t_i^2 &= \sum_{i=1}^n Z_{\vartheta}(t_i), \\
a_0 \sum_{i=1}^n t_i + a_1 \sum_{i=1}^n t_i^2 + a_2 \sum_{i=1}^n t_i^3 &= \sum_{i=1}^n Z_{\vartheta}(t_i) \cdot t_i, \\
a_0 \sum_{i=1}^n t_i^2 + a_1 \sum_{i=1}^n t_i^3 + a_2 \sum_{i=1}^n t_i^4 &= \sum_{i=1}^n Z_{\vartheta}(t_i) \cdot t_i^2.
\end{aligned}$$

Решив эту систему уравнений, мы определим параметры прогнозной модели вида (5.3).

После определения параметров моделей (5.2), (5.3), (5.4) можно вести собственно прогнозирование. Однако необходимо сделать следующее замечание. Полученные по моделям значения прогнозируемых характеристик являются наиболее вероятными (средними) значениями. Поскольку диагностические параметры обладают свойством случайности, то, очевидно, что полученные в эксплуатации значения прогнозируемых характеристик будут отличаться от расчетных. Поэтому в ходе прогнозирования следует оценить разброс и доверительный интервал для используемых характеристик.

Разброс прогнозируемой характеристики оценивается путем расчета ее среднего квадратического отклонения. Если ведется индивидуальное прогнозирование и в качестве прогнозируемой характеристики выбран сам параметр, то рассчитывают его среднее квадратическое отклонение по той реализации, которая получена с диагностируемого экземпляра объекта. Если ведется прогнозирование для группы объектов, а в качестве прогнозируемой характеристики опять же выбран сам параметр, то его среднее квадратическое отклонение рассчитывается по всему ансамблю реализаций параметра.

При использовании для прогнозирования вероятностно-статистических характеристик диагностического параметра выполняется оценка средних квадратических отклонений этих характеристик. Известно, что среднее квадратическое отклонение вероятностно-статистической характеристики случайной величины связана со средним квадратическим значением самой случайной величины [2, 20]. Поэтому при использовании для прогнозирования вероятностно-статистических характеристик диагностического параметра первоначально выполняют расчет среднего квадратического отклонения самого диагностического параметра, а затем выполняют перерасчет для вычисления среднего квадратического отклонения прогнозируемой характеристики.

Расчет доверительного интервала зависит от того, какому закону распределения подчиняется прогнозируемая характеристика. Если в качестве прогнозируемой характеристики выбран сам диагностический параметр, то обычно за основу принимается нормальный закон. Если принято среднее значение параметра, то расчет доверительного интервала ведется с применением закона распределения Стюдента. Если прогнозируется дисперсия диагностического параметра, то используется закон распределения хи-квадрат.

В завершении этой главы приведем примерный порядок анализа экспериментальных данных и выполнения расчетов, которому рекомендуется придерживаться при прогнозировании состояния авиационной техники.

5.2 Порядок обработки результатов измерений при прогнозировании.

1. Проводится упорядочение статистических данных, полученных с эксплуатации. Это упорядочение осуществляется с помощью таблицы, которая содержит набор горизонтальных строк с равномерными интервалами значений диагностического параметра и набором вертикальных колонок с равномерными интервалами наработки (табл. 5.1).

Таблица 5.1

$\Delta Z \backslash t_i$	Δt_1	Δt_2	Δt_3	...
ΔZ_1				
ΔZ_2				
...				

Вид такой таблицы позволяет качественно оценить наличие закономерности в зависимости диагностического параметра от наработки. Если все клетки таблицы равномерно заполнены, то закономерности нет. Если в противоположных углах таблицы наблюдаются нули, то закономерность есть.

2. Проводят вычисления нормированного коэффициента корреляции по формуле (5.1) для различных моментов времени. По величине этого коэффициента оценивают степень перемешанности реализаций диагностического параметра. Кроме того, рассчитав и построив зависимость этого коэффициента от разности $\Delta t = t_2 - t_1$ (то есть $r_y = f(\Delta t)$), определяют, для какого максимального промежутка времени наблюдается значимая связь между значениями диагностического параметра, что позволяет оценить максимально допустимый интервал прогнозирования.

3. На основании поставленных задач по прогнозированию и с учетом полученных результатов по оценке перемешивания реализаций диагностического параметра определяются с типом прогнозирования (индивидуальное или групповое), а также решают вопрос по выбору прогнозируемой характеристики (сам диагностический параметр или его вероятностно-статистическая характеристика).

4. По имеющимся эксплуатационным данным строят зависимость прогнозируемой характеристики от наработки. Если в качестве прогнозируемой характеристики выбирается сам диагностический параметр, то строят непосредственно зависимость этого параметра от наработки по данным снятым с объекта, для которого проводится прогнозирование. Если в качестве прогнозируемой характеристики выбрана какая либо вероятностно-статистическая характеристика диагностического параметра, то сначала проводят

расчет этой характеристики для различных моментов наработки, а затем строят указанную зависимость.

5. Анализируя вид полученной зависимости, а также используя данные о физических процессах, протекающих в диагностируемом объекте, определяют вид прогнозной модели, с помощью которой следует аппроксимировать выбранную для прогнозирования характеристику. По имеющимся статистическим данным вычисляют параметры модели и среднее квадратическое отклонение прогнозируемой характеристики.

6. По полученной модели выполняют расчет значения прогнозируемой характеристики на заданный интервал наработки и оценивают доверительный интервал этой характеристики. Пользуясь полученными расчетными значениями, принимают решение о будущем техническом состоянии эксплуатируемого объекта.

В заключение отметим, что в процедуре прогнозирования имеются ряд этапов, которые достаточно хорошо формализованы, и есть этапы, правильность выполнения которых существенно зависит от опыта исполнителя. Но в любом случае огромное значение для прогнозирования, как и для диагностирования в целом, имеют объективность, достоверность и точность данных, получаемых из эксплуатации. Поэтому все большую роль в решении задач оценки технического состояния эксплуатируемой авиационной техники начинают играть бортовые устройства регистрации, на которые записываются параметры полета и определяющие параметры наиболее важных систем воздушного судна.

6. БОРТОВЫЕ УСТРОЙСТВА РЕГИСТРАЦИИ.

6.1 Назначение бортовых устройств регистрации.

Ранее неоднократно говорилось о требованиях, предъявляемых к диагностической информации. Особую ценность представляет информация, зарегистрированная с помощью бортовых устройств регистрации параметров полета [12, 18, 25, 36]. Бортовые устройства регистрации (БУР) предназначены для записи и хранения ряда основных параметров полета (высоты, скорости, углов атаки, перегрузок) и определяющих параметров наиболее важных технических систем воздушного судна (ВС). Анализ записанной информации позволяет решать задачи повышения безопасности полетов и надежности авиационной техники, а также совершенствования процессов ее эксплуатации. Общая структура использования записей БУР представлена на рис.6.1.



Рис.6.1 Задачи, решаемые с помощью БУР

БУР являются составной частью единого комплекса регистрации и обработки полетной информации, который состоит из двух частей. Одна из них, включающая бортовые устройства записи и хранения, устанавливается на борту ВС. Другая часть, состоящая из средств обработки, анализа и документирования зарегистрированных БУР параметров, входит в состав устройств наземного

технического обслуживания. В целом результаты обработки информации, получаемой с БУР используются для:

- технического диагностирования наиболее важных систем ВС;
- анализа причин конкретных летных происшествий и предпосылок к ним;
- оценки качества пилотирования (работы) экипажа;
- сбора статистического материала, что направлено на совершенствование авиатехники для повышения ее надежности, на совершенствование системы технической эксплуатации, на совершенствование системы летной эксплуатации.

Конструктивно на борту ВС устанавливаются два типа БУР: эксплуатационный и аварийный. Отличие между ними заключается в том, что аварийный БУР защищен. Он помещается в специальный контейнер, который, как правило, состоит из трех оболочек: теплопоглощающей, ударозащитной и теплоизоляционной. Информацию, записанную на аварийный БУР, анализируют только в экстренных случаях, например при расследовании авиационных происшествий. Эксплуатационный БУР не имеет такой защиты и информацию с него анализируют периодически, например после каждого полета.

В связи с важностью информации, регистрируемой БУРами, к ним предъявляется ряд специфических требований. Основными требованиями, предъявляемыми к БУР являются :

- обеспечения сохранности информации при пожарах, появлении ударных перегрузок до 100...200 g, при попадании носителя информации в топливные и масляные смеси;
- малые габаритные размеры;
- достаточный объем (количество) записываемой и сохраняемой информации.

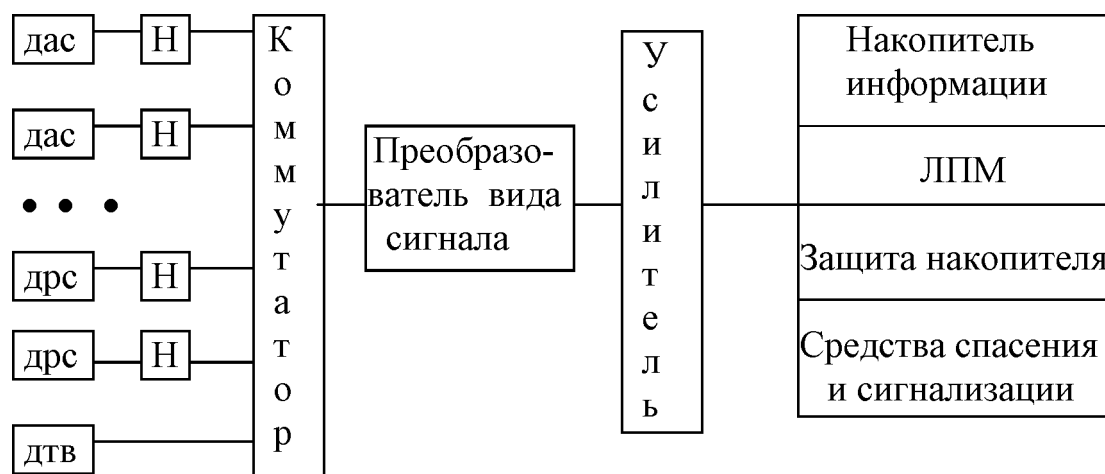
Выполнение первой группы требований обеспечивается применением термостойких, механически прочных и влагонепроницаемых защитных контейнеров для накопителей информации, а также применение при их изготовлении материалов с высокими прочностными, термическими и коррозионными характеристиками. В ряде случаев БУРы оборудуются устройством, которое катапультирует накопитель информации при возникновении аварийной ситуации. При этом накопитель снабжается сигнализатором места нахождения, например радиомаяком. Кроме того, аварийные накопители имеют яркую окраску защитного контейнера.

Минимальные масса и габаритные размеры достигаются путем оптимизации количества регистрируемых параметров и специальных методов уплотнения записи регистрируемых параметров.

В настоящее время применяются БУРы, в которых запись регистрируемых параметров осуществляется тремя способами. Первый способ - это механическая запись параметров с помощью чернил, царапанья перьями или иглами, наколов иглами на носителе информации. Второй способ - это фотооптическая запись с помощью лучей света на фотопленке. Третий способ - это магнитная запись на магнитной пленке или магнитной проволоке. Кроме того, в накопителях может применяться различная форма записи информации, а

именно: аналоговая или цифровая. В зависимости от способа записи информации БУРы делятся на механические, оптические и магнитные.

Несмотря на применение различных способов записи и отличия в конструктивном исполнении структурные схемы применяемых в настоящее время бортовых устройств регистрации имеют много общего (рис 6.2).



дас – датчики аналоговых сигналов; дрс - датчики разовых сигналов;
дтв - датчик текущего времени; Н - нормализаторы; ЛПМ - лентопротяжный механизм.

Рис.6.2 Примерная структурная схема БУР

В качестве первичных преобразователей информации используются два типа датчиков. С помощью датчиков аналоговых сигналов регистрируются параметры полета и определяющие параметры систем ВС, например, высота и скорость полета, давление и температура жидкостей и газов и т.п. С помощью датчики разовых сигналов регистрируют так называемые разовые команды, сигнализирующие о выходе какого либо параметра за пределы допустимых значений (появление опасной температуры, возникновение пожара, отказ двигателя и др.), о включении различных устройств ВС или о нахождении различных механизмов в крайних положениях (выпуск шасси или механизации, включение реверса тяги и др.).

С помощью нормализаторов электрические сигналы с выходов датчиков приводятся к одному уровню, то есть масштабируются. Например, сигналы с датчиков нормализуются к диапазону 0...1 вольт независимо от индивидуальных диапазонов и характеристик датчиков.

Коммутатор обеспечивает регистрацию многих параметров на небольшое количество дорожек записи, имеющихся на ленте. Таким способом обеспечивается уплотнение информации, поскольку на одну дорожку пишется последовательно несколько сигналов.

Преобразователь вида сигнала обеспечивает кодирование сигналов. Например, осуществляет преобразование электрических сигналов, характеризующих значения измеряемых величин в дискретный код.

После усиления в усилителе сигнал поступает в записывающее устройство (накопитель информации), где и регистрируется на носителе информации.

В зависимости от способа записи и конструктивных особенностей для ряда БУРов часть указанных блоков может отсутствовать. Например, для регистраторов с механической записью отсутствуют коммутатор и общий усилитель. Датчики измеряемых параметров и отдельные усилительные устройства обеспечивают преобразование этих параметров в механические перемещения записывающих перьев.

Ниже приведены основные характеристики и технические данные ряда бортовых устройств регистрации, применяемых в настоящее время в эксплуатации.

6.2 Характеристики и устройство БУР.

Выше отмечалось, что по способу записи бортовые регистраторы делятся на механические, оптические и магнитные. Наиболее простыми являются механические. Поэтому рассмотрение характеристик БУР начнем именно с них.

Механические БУРы.

Барограф АД-2. Предназначен для записи барометрической высоты полета (в диапазоне до 7000 метров) и отметок времени. Запись осуществляется пером с пастой на бумажную ленту, которая закреплена на барабане. Барабан приводится в вращение часовым механизмом и делает один оборот за шесть часов. Важным моментом при подготовке к работе барографа является настройка нулевой линии, соответствующей атмосферному давлению в момент подготовки с учетом температурной поправки.

Самописец КЗ-63. Самописец обеспечивает запись барометрической высоты полета H_6 , приборной скорости полета $V_{пр}$, вертикальной перегрузки n_y и отметок времени. Запись параметров и отметок времени осуществляется путем царапанья специальными иглами по эмульсионному слою киноплёнки, которая до этого обработана фиксатором без проявления.

Перемещение ленты осуществляется лентопротяжным механизмом на одной из двух скоростей: малой и большой. Малая скорость составляет 4,2...5,2 мм/мин, а большая - 4,2...5,2 мм/с. Лента размещается на катушке, на которой имеется индикатор критического остатка ленты. Запас ленты составляет 10 метров, что хватает на 20...25 часов полета.

Оптические БУРы.

Система автоматической регистрации параметров полета САРПП-12. Эта система обеспечивает запись с помощью световых лучей на фотопленке шириной 35 мм шести аналоговых и девяти разовых сигналов. САРПП -12 выпускается в нескольких вариантах для установки на самолеты и вертолеты. Структурная схема системы изображена на рис. 6.3.

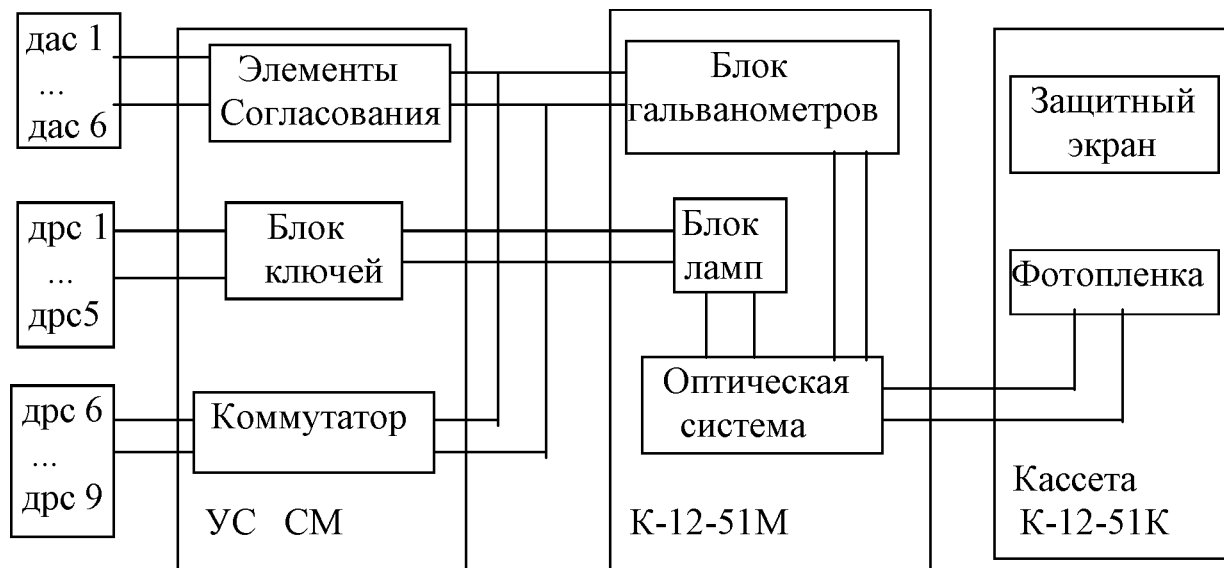


Рис. 6.3 Структурная схема САРПП-12

В систему входят:

- шесть датчиков аналоговых сигналов дас 1,..., дас 6;
- девять датчиков разовых сигналов дрс 1,..., дрс 9, которые регистрируют разовые команды типа «Давление масла ниже нормы», «шасси убрано» и др;
- устройство согласующее, служащее для согласования характеристик датчиков аналоговых сигналов с характеристиками оптических гальванометров осциллографа К-12-51М, а также для расположения в нем элементов управления;
- светолучевой оптический осциллограф К-12-51М, который является непосредственным регистратором параметров;
- блок кассеты К-12-51К с носителем информации фотопленкой, длина которой составляет 12 метров.

Аналоговые датчики дас 1... дас 6, служащие для регистрации параметров полета и систем ВС, через индивидуальные согласующие элементы подключены каждый к своему оптическому гальванометру. Зеркала этих гальванометров освещаются одной лампой. Оптическая система направляет световые лучи от зеркалец на фотопленку кассеты, при этом сигналы от датчиков записываются на соответствующие отдельные каналы.

Пять разовых сигналов от датчиков дрс1,..., дрс5 записываются на фотопленку отдельными световыми лучами (на отдельные каналы), которые создаются одной из пяти ламп блока ламп и оптической системой осциллографа

К-12-51М. Каждая из пяти ламп блока включается блоком ключей при появлении разового сигнала на соответствующем датчике дрс1,..., дрс5.

Остальные четыре сигнала с датчиков дрс6,..., дрс9 записываются путем коммутации на те же каналы, что и сигналы от датчиков аналоговых сигналов дас1,..., дас4. На последнем двенадцатом канале записываются отметки времени.

Магнитные БУРы.

В качестве бортовых устройств регистрации на современных ВС наиболее широкое применение нашли малогабаритные самописцы, регистрирующие измеряемые параметры на магнитную ленту в течении одного или нескольких полетов. К ним относятся следующие магнитные системы регистрации параметров МСРП – 12, МСРП – 64, МСРП – 256. Число, следующее за буквенным обозначением системы, указывает количество каналов, на которые регистрируются параметры. В регистраторах типа МСРП осуществляется преобразование физических величин измеряемых параметров (скорость и высота полета, давление и температура в системах, частоты вращения роторов и т.д.) в электрические сигналы с последующим их кодированием в электроимпульсный код и фиксированием его на непрерывно движущейся магнитной ленте. Основные технические характеристики магнитных регистраторов приведены в таблице 6. 1

Таблица 6.1 Основные данные МСРП

Характеристика	МСРП – 12–16	МСРП – 64–2	МСРП- 256
ЛА, на которых установлена аппаратура	АН-26, АН-24, АН-12, Ми-6, Ми-26	Ту-154, Ил-62, Як-42, Ил-76, Ту-134	Ил-86
Количество каналов регистрации	12	64	256
Количество регистрируемых параметров	12 аналоговых 12 разовых команд	48 аналоговых 32 разовых команд	228 аналоговых 128 разовых команд
Общее время записи, ч	30	25...50	25
Время сохранения записи	75 минут полета	25...50 часов полета	12,5...25 часов полета
Средства наземной обработки	ДУМС Луч-74 Луч-84	Луч-74 Луч-84	Анализ-86
Масса аппаратуры, кг	32	40	105 (128)

Магнитная система регистрации параметров МСРП - 12. Эта система обеспечивает одновременную запись 12 аналоговых и 12 разовых сигналов на ферромагнитную ленту шириной 6,25 мм. Запись производится на одну дорожку ленты при ее движении в одном направлении и на вторую дорожку при ее движении в обратном направлении в течении всего времени полета. При этом на ленте всегда остается записанной информация о последних 75 минутах полета, а более ранние записи стираются. Каждый из датчиков аналоговых сигналов (параметров) подключается к записывающему устройству 12 раз в секунду. Каждый из датчиков разовых сигналов - через пять секунд. Скорость протяжки магнитной ленты составляет 96 мм/с, плотность записи - два импульса на мм. Запас ленты - 250 м. На рис.6.4 приведена структурная схема МСРП-12.

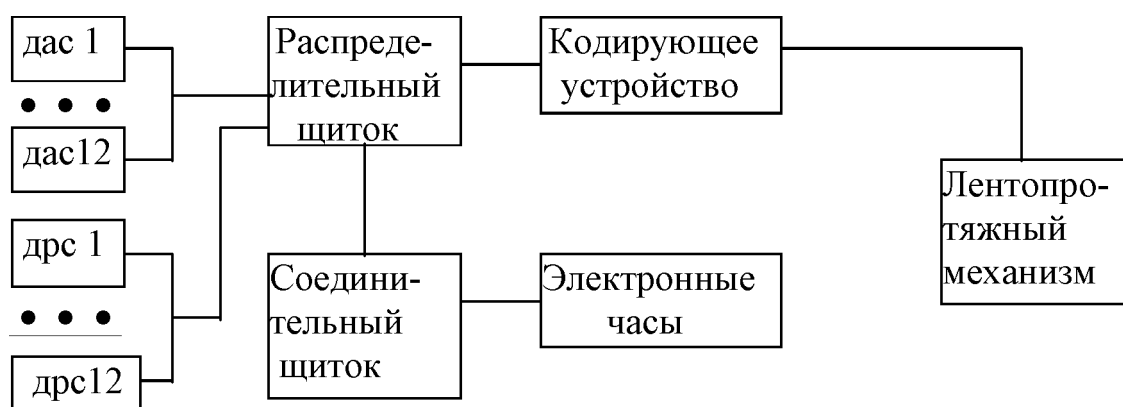


Рис. 6.4 Структурная схема МСРП - 12

Двенадцать датчиков аналоговых сигналов через распределительный щиток подключены к соответствующим входам кодирующего устройства. Последнее устройство преобразует напряжение каждого датчика во времяимпульсный код. Импульсы напряжения с выхода кодирующего устройства подводятся к магнитной головке записи в лентопротяжном механизме и записываются на магнитную ленту.

Через каждые пять секунд на время 0,5 секунды распределительный щиток отключает датчики аналоговых сигналов и вместо них подключает к кодирующему устройству датчики разовых сигналов. С выходов этих датчиков напряжения разовых сигналов кодируются и записываются на магнитную ленту. Через соединительный щиток на кодирующее устройство поступают импульсы времени от электронных часов, которые также записываются на магнитную ленту.

Лентопротяжный механизм с двумя кассетами и магнитными головками записи и стирания располагаются в защитном шаровом контейнере.

Питание системы осуществляется от основной самолетной сети постоянного тока. При отказе этой сети происходит автоматическое переключение питания МСРП - 12 от аварийного источника постоянного тока.

Магнитная система регистрации параметров МСРП - 64. Система предназначена для регистрации 59 аналоговых сигналов, 32 разовых сигналов, текущего времени и служебной информации. К служебной информации относятся сведения о номере самолета, дате и номере рейса. Номенклатура и число регистрируемых параметров различно для различных типов ВС. Так, например, система МСРП-64 самолета ТУ-154 Б регистрирует 48 аналоговых параметра и 56 разовых команд. Причем увеличение количества разовых команд на 24 связано с использованием для этой цели шести каналов, предназначенных для записи аналоговых сигналов, с применением уплотнителя разовых сигналов.

Носителем информации служит магнитная лента шириной 19,5 мм, размещенная на двух кассетах лентопротяжного механизма. Длина ленты 250м. При скорости движения ленты 2,67 мм/с длительность записи составляет примерно 20 часов при движении ленты сначала в одном, а затем в другом направлении.

Структурная схема МСРП-64 показана на рис.6.5. Датчики аналоговых

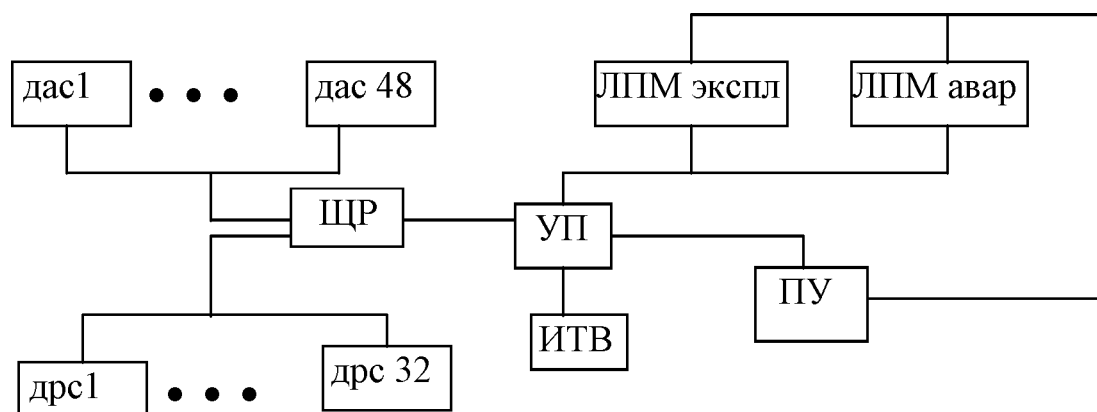


Рис. 6.5 Структурная схема системы МСРП-64

сигналов дас1,..., дас 48 и датчики разовых сигналов дрс1,..., дрс32 подключены к распределительному щитку (ЩР) и их электрические сигналы поступают на вход соответствующих каналов устройства преобразования (УП). В устройстве преобразования осуществляется преобразование электрических сигналов от датчиков в цифровой код. Аналоговые сигналы подключены к УП через коммутатор, и их кодирование производится поочередно, в строгой последовательности с частотой один или два герца.

К устройству преобразования подключены также индикатор текущего времени (ИТВ) и пульт управления (ПУ). Индикатор текущего времени предназначен для индикации астрономического времени и преобразования его в

цифровой код. Пульт управления служит для принудительного включения и контроля работоспособности системы, в том числе и ее лентопротяжных механизмов (ЛПМ), а также для кодирования опознавательных данных: номера рейса, номера ВС и даты полета.

Аварийный и эксплуатационный бортовые накопители с соответствующими лентопротяжными механизмами (ЛПМ экспл и ЛПМ авар) предназначены для записи и хранения измерительной и служебной информации на магнитную ленту. Для сохранения информации в случае летного происшествия ЛПМ авар помещен в контейнер, состоящий из трех оболочек: теплопоглощающей, ударозащитной и теплоизоляционной. Контейнер эксплуатационного ЛПМ не имеет ударной и тепловой защиты и снабжен легкоъемной крышкой.

Питание системы МСРП-64 осуществляется от основной сети постоянного тока ВС или, при ее отказе, от бортовых аккумуляторов. Включение системы в работу осуществляется принудительно с пульта управления или автоматически от сигнализатора скорости движения самолета.

Дальнейшим развитием системы МСРП-64 является система МСРП-256. Система МСРП-256 построена на тех же принципах, но имеет существенно увеличенное число каналов.

6.3 Краткие сведения о наземных средствах обработки полетной информации

С целью определения численных значений зарегистрированных в полете параметров, а также определения моментов времени, в которые возникали внештатные ситуации или срабатывали различные системы ВС, производится обработка записей бортовых устройств регистрации с помощью наземных средств.

В механических и оптических регистраторах запись осуществляется сразу в наглядной графической форме. Поэтому обработка записей заключается в их расшифровывании оператором, которое выполняется с помощью масштабной линейки и тарифовочных графиков для каждого параметра. При расшифровке записей сделанных на прозрачном носителе (например, фотопленке) применяются увеличители с экраном (типа «Микрофот»). Никаких средств автоматической расшифровки механических и оптических записей в настоящее время в эксплуатации не применяется. Поэтому процесс их расшифровки трудоемкий и требует высокой квалификации оператора.

Для обработки записей, сделанных на магнитной ленте, используются специальные устройства, осуществляющие декодирование и расшифровку записей.

Декодирующее устройство магнитной системы - ДУМС. Декодирующее устройство типа ДУМС предназначено для воспроизведения магнитных записей, сделанных на МСРП-12, декодирование и регистрацию их в непрерывной форме на фотобумагу или фотопленку шириной 120 мм. При этом одновременно может

быть декодирована и записана информация не более чем четырех информационных каналов плюс отметки времени.

В комплект ДУМС входят: лентопротяжный механизм с воспроизводящими головками, устройство преобразования кодированного сигнала в непрерывный аналоговый или разовый сигнал, светолучевой магнитоэлектрический осциллограф К-12-22, пульт управления. Устройство преобразования осуществляет декодирование записанного на магнитную ленту кода в электрическое напряжения, значение которого пропорционально значениям регистрируемых параметров. Далее эти напряжения подаются на осциллограф.

Наземное декодирующее устройство НДУ-8. Это устройство предназначено для декодирования магнитных записей, сделанных на МСРП-64. В состав этой системы входят:

- блок воспроизведения с лентопротяжным механизмом и регулируемым усилителем;
- блок цифровой индикации служебных параметров и астрономического времени;
- блок преобразования, куда входят восемь преобразователей «код-напряжение» для аналоговых сигналов, преобразователь кода разовых сигналов, преобразователь кода времени;
- блок управления со цифровым индикатором для индикации разовых сигналов или цифровых значений аналоговых параметров.

Результаты декодирования регистрируются с помощью светолучевого осциллографа К20-22 на фотобумаге. Одновременно можно записать восемь аналоговых сигналов, которые изображаются в виде графических временных зависимостей. Появлению или отключению разового сигнала соответствуют появление или пропадание на фотобумаге следа от светового луча.

На цифровой индикатор выводится служебная информация и астрономическое время. Также на цифровой индикатор могут поочередно выводиться значения аналоговых сигналов.

Для дешифрирования и одновременного автоматизированного анализа записей с БУР типов МСРП-64 и МСРП-256 применяются специализированные созданные на базе ЭВМ системы, типа «Луч-74», «Луч-84», «Анализ-86» и их модификации. Эти системы представляют собой мини-ЭВМ, оборудованную дополнительными устройствами ввода и вывода информации, обеспечивающими ввод информации с магнитных лент различных бортовых накопителей типа МСРП и вывод информации в виде текста, таблиц и графиков на обычную или специальную бумагу.

С помощью этих специализированных систем решают следующие задачи.

1. Выполняют экспресс-анализ полетной информации, который обеспечивает обнаружение событий, влияющих на безопасность полета, и документирование их.
2. Проводят автоматизированную обработку полетной информации с выдачей результатов обработки в виде текста и графиков в координатах величина параметра - время.

3. Осуществляют индикацию опознавательных данных и служебной информации на цифровых индикаторах и на экране дисплея.
4. Выполняют решение различных технических задач с программированием их решения на алгоритмических языках.

Обобщенная структурная схема специализированной системы декодирования и автоматизированного анализа записей бортовых устройств регистрации типа МСРП представлен на рис.6.6.

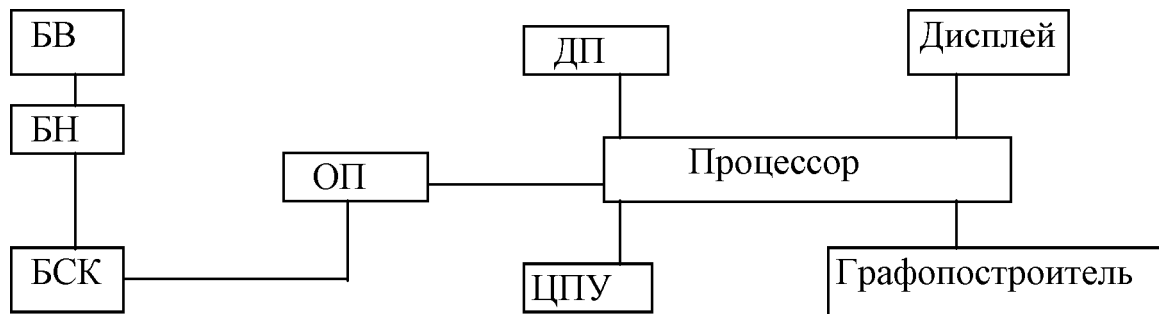


Рис. 6.6 Обобщенная структурная схема системы обработки полетной информации

По функциональному назначению составные части системы объединяются в следующие две группы. Первая группа это аппаратура ввода информации с магнитной ленты, куда входят блок воспроизведения (БВ), блок нормализации (БН) и блок сопряжения и кодирования (БСК). Вторая группа это собственно управляющий вычислительный комплекс, куда входят процессор, оперативная память (ОП), долговременная (внешняя) память (ДП), а также аппаратура вывода информации - дисплей, графопостроитель и цифропечатающее устройство (ЦПУ).

В процессе обработки полетной информации на лентопротяжный механизм блока воспроизведения устанавливается кассета с магнитной лентой, снятая с регистраторов типа МСРП. С блока воспроизведения, после нормализации в БН, информация кодируется в БСК и подается в оперативную память вычислительного комплекса, где осуществляется ее обработка, хранение, вывод на внешние устройства в виде, удобном для восприятия оператором, а также документирование.

Для выполнения обработки и анализа полетной информации специализированные система снабжаются комплексом программного обеспечения. В этот комплекс, как правило, входят следующие программы:

1. Аппроксимация тарировочных данных датчиков аналоговых сигналов;
2. Автоматизированная обработка информации с выводом ее в виде графиков на графопостроитель или в табличном виде на ЦПУ;

3. Экспресс-анализ полетной информации, состоящий из двух частей (модулей) - контроль техники пилотирования и контроль работоспособности основных систем ВС;
4. Контроль выполнения летно-эксплуатационных ограничений;
5. Распечатка каталога сообщений о появлении событий, связанных со срабатыванием систем ВС, в том числе, и при появлении нештатных ситуаций.

В заключение отметим, что повышение требований к современным ВС в части обеспечения безопасности полетов предполагает как совершенствование систем регистрации полетных данных, то есть БУР, так и систем обработки полетной информации. Это, в первую очередь, связан с необходимостью применения для повышения уровня безопасности полетов более совершенных методов контроля и диагностики, обеспечивающих обнаружение широкого класса неисправностей на самой ранней стадии их развития.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Ахметзянов А.М., Дубравский Н.Г., Тунаков А.П. Диагностика состояния ВРД по термогазодинамическим параметрам. - М.: Машиностроение, 1983.
2. Бюргер И.А. Техническая диагностика. - М.: Машиностроение, 1978.
3. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. - М.: Наука, 1974.
4. Вентцель Е.С. Теория вероятностей. - М.: Наука, 1969.
5. Вибрация в технике: Справочник. В 6 - ти том. /Ред. совет: В.Н. Челомей (пред.). - М.: Машиностроение, 1979. - Т. 5 /Под ред. М.Д. Генкина, 1981.
6. Винер Н. Кибернетика, или управление и связь в животном и машине. - М.: Наука, 1983.
7. Генкин М.Д., Соколова А.Г. Виброакустическая диагностика машин и механизмов. - М.: Машиностроение, 1987.
8. Диагностирование и прогнозирование технического состояния авиационного оборудования: Учеб. пособие для вузов / В.Г. Воробьев, В.В. Глухов, Ю.В. Козлов и др.; Под ред. И.М. Синдеева. - М.: Транспорт, 1984.
9. Дорошко С.М. Контроль и диагностирование технического состояния газотурбинных двигателей по вибрационным параметрам. - М.: Транспорт, 1984..
10. Епишев Н.И. Техническое диагностирование ГТД по параметрам, регистрируемым при работе двигателя. - Куйбышев : КуАИ, 1985 .
11. Ермаков Г.И. Диагностирование технического состояния авиационных двигателей путем анализа работавшего масла. - М.: ЦНТИ ГА, 1985. - 42 с.
12. Ильинский В.М. Системы контроля авиационных силовых установок. - М.: Транспорт, 1980.
13. Карасев В.А., Максимов В.П., Сидоренко М.К. Вибрационная диагностика газотурбинных двигателей. - М.: Машиностроение, 1978.
14. Карасев В.А., Ройтман А.Б. Доводка эксплуатируемых машин. Вибродиагностические методы. - М.: Машиностроение, 1986.
15. Коллакот Р.А. Диагностирование механического оборудования. Пер. с англ. - Л.: Судостроение, 1980.
16. Колмогоров А.И. Интерполирование и экстраполирование стационарных случайных последовательностей. //Известия АН СССР. Математика. 1941. № 5. С.3 - 14.
17. Константинов В.Д., Глухов В.В. Куликов Г.А. Методы контроля технического состояния авиационного оборудования. - М.: МИИГА, 1986.
18. Константинов В.Д., Куликов Г.А. Средства контроля технического состояния авиационного оборудования. - М.: МИИГА, 1987.
19. Контроль технической исправности самолетов и вертолетов: Справочник / Под ред. В.Г. Александрова. - М.: Транспорт, 1976.
20. Корн Г.А., Корн Т.М. Справочник по математике для научных работников и инженеров, - М.: Наука, 1973.

21. Лихачев В.Я., Васин А.С., Гликман Б.Ф. Техническая диагностика пневмогидравлических систем ЖРД. - М.: Машиностроение, 1983.
22. Макаров Р.А. Средства технической диагностики машин. - М.: Машиностроение, 1981.
23. Маркович Х.П. Использование граф-модели для решения задач технической диагностики // Кибернетика и диагностика. - Рига: Зинатне, 1968. Вып. 2.
24. Мозгалеvский А.В., Гаскаров Д.В. Техническая диагностика (непрерывные объекты). - М.: Высшая школа, 1975.
25. Нападов А.П., Игонин Н.Н. Обработка и анализ полетной информации. - Самара: КуАИ, 1988.
26. Основы технической диагностики. Модели объектов, методы и алгоритмы диагноза./ В.В. Карибский, П.П. Пархоменко, Е.С. Согомонян и др.; Под ред. П.П. Пархоменко. - М.: Энергия, 1976.
27. Пивоваров В.А. Диагностика летательных аппаратов и авиадвигателей (основы теории и прикладные вопросы). - М.: МГТУ ГА, 1995.
28. Приборы для неразрушающего контроля материалов и изделий: Справочник. В 2-х кн. /Под ред. В.В. Ключева. - М.: Машиностроение, 1986. Кн. 1 - 488 с.; Кн. 2
29. Симбирский Д.Ф. Температурная диагностика двигателей. - Киев: Техника, 1976.
30. Сиротин Н.Н. Конструкция и эксплуатация, повреждаемость и работоспособность газотурбинных двигателей. – М.: РИА «ИМ – Информ», 2002.
31. Сироткин Н.Н., Коровкин Ю. М. Техническая диагностика авиационных газотурбинных двигателей. - М.: Машиностроение, 1979.
32. Теория авиационных воздушно-реактивных двигателей. / Под ред. С.М. Шляхтенко. - М.: Машиностроение, 1974.
33. Теория двухконтурных турбореактивных двигателей. / Под ред. С.М. Шляхтенко, В.А. Сосунова. - М.: Машиностроение, 1979.
34. Техническая эксплуатация летательных аппаратов: Учеб. для вузов / Н.Н. Смирнов, Н.И. Владимиров, Ж.С. Черненко и др.; Под ред. Н.Н. Смирнова. - М.: Транспорт, 1990.
35. Техническая диагностика гидравлических приводов. / Под ред. Т.М. Башты . - М.: Машиностроение 1989.
36. Техническая эксплуатация авиационного оборудования: Учебник для вузов /В.Г. Воробьев, В.Д. Константинов, В.Г. Денисов и др.; Под ред. В.Г. Воробьева. - М.: Транспорт, 1990.
37. Технические средства диагностирования: Справочник / В.В. Ключев, П.П. Пархоменко, В.Е. Абрамчук и др.; Под общ. ред. В.В. Ключева. - М.: Машиностроение , 1989.
38. Технология эксплуатации, диагностики и ремонта газотурбинных двигателей: Учеб. пособие./Ю.С. Елисеев, В.В. Крымов, К.А. Малиновский, В.Г. Попов. – М.: Высш. шк.; 2002
39. Третьяков О.Н. Методы и средства эксплуатационного сопровождения ресурса авиационных ГТД. – М.: Изд-во МАТИ, 1999.

40. Шеннон К. Работы по теории информации и кибернетике. - М.: Изд-во иностранной литературы, 1963.
41. Ямпольский В.И., Белоконь Н.И., Пилипосян Б.Н. Контроль и диагностирование гражданской авиационной техники. - М.: Транспорт, 1990.
42. Ярославцев Н.Л., Попов В.Г. Техническая диагностика. – М.: РИО МАТИ им. К.Э. Циолковского, 1995.