

Лекция. Введение. ОБЩАЯ ХАРАКТЕРИСТИКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Учебные вопросы

1. Источники и вид представления экспериментальных данных
2. Цели обработки экспериментальных данных
3. Задачи формирования и обработки экспериментальных данных

1.1. Источники и вид представления экспериментальных данных

Надо измерять все измеримое и делать измеримым то, что пока еще не поддается измерению.

Галилео Галилей

Экспериментальные исследования событий и процессов основаны на наблюдениях, в ходе которых регистрируются различные факты искусственного и естественного происхождения.

Источниками экспериментальных данных являются:

результаты наблюдения за реальными объектами и протекающими в них процессами. Наблюдения могут проводиться в ходе испытаний или в ходе обычной эксплуатации;

результаты моделирования объектов. В первую очередь к ним следует отнести результаты имитационного моделирования;

технические, экономические, научные отчеты и обзоры, публикуемые в различных изданиях, например, сведения о результатах испытаний или о характеристиках однотипных устройств различных производителей;

результаты опросов специалистов и другие источники.

Обработка ЭД, получаемых от различных источников, имеет много общего. Однако организация сбора и интерпретации ЭД специфична для конкретной предметной области. В дальнейшем обработка ЭД будет рассматриваться применительно к результатам наблюдения за функционированием АСОИУ, ее элементов или их моделей.

Вид ЭД определяет форму представления, степень зависимости от времени, характер данных.

Одной из основных форм является *символьная*, которая включает представление данных в виде чисел, двоичных величин или текста. Для задания значений соответствующих величин применяются различные шкалы измерений. Описательные (качественные) признаки измеряются на основе номинальных и порядковых шкал. Номинальные шкалы обеспечивают только группирование объектов по признаку наличия у них некоторых общих свойств, но не позволяют проводить ранжирование объектов. Порядковые шкалы обеспечивают возможность упорядочивания данных по признакам "больше", "меньше", "равно", но при этом не указывается, на сколько одно значение признака больше или меньше другого.

Количественные свойства отображаются числами в относительных или абсолютных шкалах измерений. В относительных шкалах точки начала отсчета и масштаб измерений имеют условный характер. Например, температуру можно измерять в относительных шкалах по Цельсию, Реомюру, Фаренгейту. Исходя из этого, результаты количественного сравнения величин зависят от используемой шкалы, а некоторые операции над количественными признаками недопустимы. Например, температура одного объекта выше температуры другого на три градуса Цельсия, эти три градуса не равны трем градусам шкалы Фаренгейта. Бессмысленно говорить, во сколько раз температура одного объекта выше температуры другого, в частности, нельзя сказать, что температура $+10^{\circ}\text{C}$ в два раза выше, чем $+5^{\circ}\text{C}$. Абсолютная шкала обеспечивает однозначное представление точки отсчета и масштаба. Примерами абсолютных шкал является шкала температур по Кельвину, шкала вероятностей. Эти шкалы позволяют дать однозначные ответы на вопросы о том, на сколько или во сколько раз одна величина больше (меньше) другой. Именно применение относительных и абсолютных шкал дает возможность

проводить количественную обработку ЭД. Но при обработке следует применять только те операции, которые допускаются применяемой шкалой измерений.

Количественные характеристики (параметры) представимы дискретными или непрерывными величинами. Дискретные параметры принимают только отдельные значения, без промежуточных значений между ними. Часто дискретные параметры имеют такое большое количество допустимых значений, что их на практике считают квазинепрерывными. Непрерывные параметры способны принимать любые значения из некоторого допустимого диапазона. В процессе обработки непрерывные величины всегда округляют и представляют ограниченным числом разрядов, т. е. они становятся квазинепрерывными. На практике ЭД отображают с разрядностью, обеспечивающей относительную погрешность не более единиц или десятых долей процента.

Экспериментальные данные могут быть представлены не только в символьной, но и в других формах: *графической* (графики, осциллограммы, штриховые рисунки, цветные изображения и полутоновые рисунки); *аудиоданных*. Такие данные обрабатывают непосредственно или предварительно преобразуют в числовую форму.

В дальнейшем рассматривается только числовая форма представления данных как универсальная и широко распространенная форма представления информации, количественно характеризующая параметры объектов и процессов. А сами параметры считаются непрерывными величинами, если особо не оговорено иное. Обработка данных, представленных в других формах, обладает существенной спецификой и требует отдельного рассмотрения.

Параметры, вообще говоря, зависят или не зависят от времени. В пособии изучаются оценки параметров, не зависящие от времени.

Результаты наблюдений носят детерминированный или случайный характер. Большинство событий и процессов в автоматизированных системах можно считать случайными. Именно на обработке таких данных сосредоточено внимание в пособии.

Существенной особенностью задач оценивания параметров АСОИУ является наличие цензурирования, под которым понимается отсутствие в результатах наблюдений каких-либо элементов. Цензурирование возникает по ряду причин, например, как следствие разновременности начала и окончания работы различных устройств или из-за того, что время свершения некоторого события выходит за предел периода наблюдения. Обработка таких ЭД требует применения специального математического аппарата.

Таким образом, в данном учебном пособии *в качестве объекта обработки рассматриваются совокупности числовых данных, характеризующих не зависящие от времени случайные значения непрерывных параметров.*

1.2. Цели обработки экспериментальных данных

Искусство измерения является могущественным оружием, созданным человеческим разумом для проникновения в законы природы.

Якоби Б.С. (Мориц Герман)

Основные показатели АСОИУ, например, показатели производительности и надежности, носят вероятностный характер и не могут быть непосредственно измерены [2, 6]. Для их оценки следует применять косвенные способы на основе регистрации соответствующих первичных параметров и последующей обработки накопленных данных с привлечением специальных математических методов. Иначе говоря, ЭД представляют собой лишь наборы возможных случайных значений показателей, зарегистрированных в некоторые моменты времени. Например, продолжительность наработок до отказа некоторой совокупности однотипных устройств можно рассматривать как множество возможных случайных значений показателя "наработка до отказа". Именно по наработкам и необходимо оценить значение этого показателя. Сам показатель как случайная величина характеризуется законом распределения, моментами распределения или другими параметрами, которые и следует определить.

Таким образом, каждое зарегистрированное данное представляет собой частное проявление некоторой закономерности. В обобщенном виде цель обработки ЭД состоит в выявлении присущей этим данным закономерности путем использования сведений о каждом отдельном событии и представлении искомой закономерности в количественном виде.

Основными целями обработки ЭД являются следующие:

оценка значений показателей качества средств, комплексов или системы в целом. На стадиях создания такая оценка проводится в интересах обоснования принимаемых решений по построению объектов, проверки показателей на соответствие требованиям, выявления существенных факторов, влияющих на функционирование объектов, выявления причин несоответствия требованиям. На стадии эксплуатации обработка ЭД проводится также для решения задач управления объектом: изменения режимов работы объекта; изменения порядка обработки информации; обоснования данных для модернизации объекта (изменения конфигурации технических и программных средств); адаптации объекта к условиям функционирования;

сжатие информации о функционировании объекта, ее обобщение для последующего применения в интересах исследования подобных объектов, обоснования данных для создания новых систем;

выявление закономерностей функционирования объекта в конкретных условиях эксплуатации, т. е. установление зависимостей между параметрами объекта, внешней среды и показателями качества объекта. Выявленные закономерности применяют для поиска оптимальных значений параметров при синтезе новых систем, для упрощенного описания объекта в модели суперсистемы;

выявление существенных параметров системы и внешней среды;

изучение типологии объектов (распознавание образов, классификация объектов);

прогнозирование развития объектов в интересах организационного и технологического управления.

Следует помнить, что результаты обработки ЭД не гарантируют достоверного описания неизвестных показателей или закономерностей, их необходимо рассматривать только лишь как более-менее удачную аппроксимацию соответствующих характеристик.

Необходимость сбора и обработки ЭД обусловлена объективными обстоятельствами, так как действительные значения показателей качества сложных объектов обычно существенно отличаются от рассчитанных на стадиях проектирования. Эти различия являются следствием ряда причин:

на стадиях проектирования нет достаточно полных и точных представлений о характеристиках процессов, протекающих в объекте и во внешней среде, поэтому при проектировании приходится вводить существенные допущения и ограничения. На практике часть принятых допущений оказывается не вполне справедливой;

с вводом в действие объекта внешняя среда постепенно меняет свои характеристики, например, меняются стиль и методы работы пользователей, появляются дополнительные потребности в решении задач. Такие изменения заранее предусмотреть невозможно;

условия эксплуатации, квалификация обслуживающего персонала в разных организациях имеют свою специфику;

в ходе эксплуатации изменяются характеристики технических, программных и информационных средств, меняется их взаимное отображение. Эти обстоятельства приводят к изменению характеристик потоков запросов на решение задач и параметров их обслуживания.

Таким образом, фактические значения показателей качества не только отличаются от расчетных, но и меняются с течением времени, имеют свои особенности для одних и тех же типов объектов, эксплуатируемых в различных организациях, претерпевают колебания, зависящие от времени, характера выполняемых на объекте работ.

В зависимости от стадии жизненного цикла АСОИУ задачи обработки ЭД имеют ряд особенностей. На стадии создания имеется принципиальная возможность проведения

активных и пассивных экспериментов. Понятие "активный" подразумевает возможность выбора объема экспериментов, последовательности и значений характеристик воздействий на объект по желанию исследователя. Проведение активных экспериментов позволяет расширить диапазон условий, при которых проводится оценивание качества. А специальным образом подобранные условия проведения исследования и порядок задания внешних воздействий, т. е. рационально обоснованные планы экспериментов, обеспечивают взаимную статистическую независимость результатов испытаний. Эти обстоятельства значительно облегчают обработку ЭД, повышают качество получаемых оценок, позволяют разделить влияние различных факторов при построении модели функционирования объекта. Активные эксперименты в основном проводятся на завершающих стадиях создания в виде испытаний опытных образцов, фрагментов систем и т. п. Вопросы постановки активных экспериментов, методов обработки их результатов изучаются в рамках специальной теории планирования экспериментов.

Реальные условия эксплуатации объекта в конкретной организации могут отличаться от предполагаемых при проведении испытаний. В связи с этим, показатели, установленные в ходе испытаний, несут в себе известную долю абстракции. На стадии создания время испытаний весьма ограничено, что не дает возможности сформировать большой объем ЭД для полноценной оценки искомых показателей.

В *пассивных* экспериментах количество наблюдений, последовательность и значения воздействий определяются реальной обстановкой использования объектов. Иначе говоря, исследователь практически лишен возможности управления качеством и количеством ЭД.

На стадии эксплуатации возможности проведения активных экспериментов значительно ограничены или вообще отсутствуют, что обычно приводит к взаимной зависимости результатов наблюдений. Наличие такой зависимости затрудняет обработку полученных данных, а игнорирование данного обстоятельства приводит к смещению значений получаемых оценок, невозможности разделения влияния различных факторов на показатели функционирования и к другим нежелательным последствиям. Нестационарные условия эксплуатации, влияние на объекты периодических или нерегулярно изменяющихся воздействий (трендов) обуславливают необходимость рассмотрения характеристик не как случайных величин, а как случайных функций. Однако из-за слабой разработанности методов и средств оценки параметров случайных процессов по результатам наблюдения обычно предполагается, что процесс функционирования объекта носит стационарный или кусочно-стационарный характер.

1.3. Задачи формирования и обработки экспериментальных данных

Оставьте трудиться напрасно, стараясь извлечь из одного разума всю мудрость; спрашивайте природу, она хранит все истины и на вопросы ваши будет отвечать вам непременно и удовлетворительно.

чевский. "О важнейших предметах воспитания".

Экспериментальная оценка качества предполагает решение ряда взаимосвязанных задач. К этим задачам следует отнести:

составление перечня регистрируемых параметров, необходимых для вычисления оценок показателей АСОИУ;

выбор моментов и способов регистрации первичных параметров;

выбор или разработку методов преобразования ЭД для оценивания требуемых показателей;

разработку алгоритмов измерения, создание программных, информационных и других средств, реализующих соответствующие процедуры регистрации, хранения, обработки и выдачи данных. Эта группа задач носит прикладной характер и далее не рассматривается.

Составление перечня регистрируемых параметров (определение информационной потребности) включает в себя несколько этапов.

1. Определение состава свойств объекта и внешней среды, подлежащих экспериментальному оцениванию. В этот состав следует включать только необходимые свойства, так как расширение перечня ведет к существенному усложнению задачи, а также к большим накладным расходам ресурсов АСОИУ на оценивание в ущерб выполнению основных функций.

2. Последовательная декомпозиция сложных свойств до уровня простых, однозначно характеризующих соответствующими показателями. Перечень свойств следует определять с детализацией для АСОИУ в целом и для исследуемых ее компонентов. В результате выполнения данного этапа будет построено дерево свойств исследуемого объекта, корнем которого выступает качество объекта, а листьями – его простые свойства.

3. Выбор показателей, характеризующих каждое свойство или их группы, с учетом специфики реализации конкретных функций объекта. Например, специфическими показателями оцениваются свойства надежности непрерывно и дискретно выполняемых функций.

4. Формирование статистического определения каждого оцениваемого показателя и детализация каждого параметра до уровня таких величин, которые могут быть измерены (зарегистрированы) в ходе наблюдения за работой объекта. **Например, пусть оценивается вероятность своевременной обработки запросов с детализацией по их типам $P_i(t_{ij} < t_{\text{доп}})$, где i – тип запроса, t_{ij} – время обработки j -го запроса i -го типа, $t_{\text{доп}}$ – допустимое время обработки запросов.** Величина t_{ij} формируется путем регистрации времени поступления запроса в систему, выдачи ответа на запрос и вычисления времени обработки. Оцениваемая вероятность $P_i(t_{ij} < t_{\text{доп}})$, при большом количестве запросов эквивалентна частоте m_i/n_i , где m_i – количество своевременно, а n_i – общее количество всех обработанных запросов заданного типа. Тогда регистрации и последующей обработке подлежат следующие сведения: время начала и завершения обслуживания запросов с детализацией по типам; факты завершения обслуживания всех и своевременно обработанных запросов с детализацией по типам. Совокупность величин $t_{\text{доп}}$ и количество типов запросов необходимо рассматривать как набор задаваемых исходных данных.

Детализация параметров должна учитывать структуру объекта, иерархию выполняемых функций, наличие различного рода резервов и другие особенности. Необходимо учитывать принципиальную невозможность регистрации определенных типов событий, в частности ошибок, которые не обнаруживаются средствами контроля, отказов неработающих устройств и т. д. Следовательно, ряд свойств объекта нельзя оценить непосредственно по результатам эксплуатации. Определение соответствующих показателей возможно только косвенным путем на основе других параметров.

5. Объединение первичных регистрируемых параметров, используемых при оценке различных показателей, что и дает информационную потребность. В ходе объединения следует учитывать неоднородность компонент АСОИУ по назначению и режимам работы. Игнорирование неоднородности приводит к смещению оценок.

Выбор моментов и способов регистрации первичных параметров основывается на анализе состава регистрируемых сведений, возможностей аппаратных и программных средств конкретной АСОИУ. Для регистрации первичных параметров нецелесообразно создавать или применять специальную аппаратуру, лучше использовать программные средства. Часть таких средств входит в состав типовых операционных систем, для их применения требуется провести всего лишь настройку. Другие средства подлежат разработке.

Существуют два основных способа регистрации – регистрация системных событий и периодическая регистрация.

Первый способ предусматривает выполнение операций регистрации первичных сведений (проведение измерений) в ходе выполнения управляющих программ операционной системы или программ прикладных задач. Таким образом регистрируются

моменты прихода заявок на решение задач, появление сигналов прерываний от схем контроля и другие подобные события.

Второй способ предполагает включение специальной программы регистрации через определенные промежутки времени. Эта программа производит «моментные снимки» состояния соответствующих элементов АСОИУ. "Фотографированию" подлежат различного рода системные таблицы, отражающие состояние очередей, устройств и т. д.

Достоинством первого способа является высокая точность регистрации измеряемых величин. Однако реализация этого способа сложнее, чем второго, поскольку требуется включение соответствующих программных компонентов во множество других программ. Это усложняет организацию их разработки и увеличивает потребности в ресурсах на создание и выполнение. Кроме того, применение первого способа затрудняет проведение модернизации измерительных средств (приходится перерабатывать большой объем программ).

Второй способ существенно проще в реализации и модификации, так как средства измерений остаются относительно автономными в конструктивном отношении и в отношении включения их в работу. Эти средства в случае необходимости проще модернизировать или отключить, в частности, при возникновении перегрузки системы. Применение данного способа предполагает выбор некоторого спектра частот включения в работу средств регистрации в соответствии с инерционностью процессов, протекающих в исследуемом объекте и во внешней среде. Например, понятно, что интенсивность изменения состояния очередей запросов существенно выше, чем интенсивность отказов устройств. Выбор частоты регистрации проводят с учетом минимизации затрат ресурсов комплекса на проведение измерений при обеспечении заданной точности регистрации параметров.

Учитывая достоинства и недостатки указанных способов, их следует применять совместно: одна часть сведений регистрируется на основе системных событий, а другая – на основе моментных снимков.

Информационная потребность и совокупность способов регистрации первичных параметров составляют модель наблюдаемости АСОИУ.

Результатами выполнения процедур регистрации являются совокупности данных, характеризующие системные события или состояния ресурсов системы в определенный момент времени.

Выбор или разработка методов преобразования ЭД для оценивания требуемых показателей предполагает формализацию описания процедур первичной, предварительной и основной обработки результатов регистрации.

Первичная обработка данных направлена на преобразование зарегистрированных величин к виду, удобному для последующего хранения и обработки. При этом не требуется применения сложного математического аппарата. В ходе первичной обработки данные подвергаются "сжатию" (например, результат регистрации заносится в соответствующий классификационный разряд статистического ряда) и записываются в специальные массивы, хранящиеся в основной или внешней памяти ЭВМ.

Основная особенность алгоритмов регистрации и первичной обработки состоит в необходимости их реализации в реальном масштабе времени с жесткими ограничениями на время включения и выполнения. При этом они не должны вносить заметных искажений в выполнение основных программ АСОИУ.

Предварительная обработка данных связана с их обобщением, сортировкой по системным событиям и периодам наблюдения.

Основная обработка зарегистрированных данных направлена на определение тех показателей и функций, которые вытекают из целей экспериментального исследования. Реализация соответствующих процедур предусматривает широкое использование сложного математического аппарата с большим объемом вычислений.

Предварительная и основная обработка выполняются в фоновом режиме, поэтому они не предъявляют жестких требований к ресурсам АСОИУ. Эта обработка вообще

может проводиться на других вычислительных средствах вне реального масштаба времени.

Выполнение процедур регистрации и обработки неизбежно связано с внесением ошибок. Источниками ошибок являются:

инструментальные ошибки. Связанны с погрешностью процедур регистрации (например, с задержками включения в работу) и возможными сбоями в их выполнении. Эти ошибки в основном носят случайный характер;

методические ошибки. Определяются принятыми допущениями и ограничениями используемых методов первоначальной и предварительной обработки (представление непрерывных величин дискретными значениями, обобщение измерений в виде классификационного статистического ряда и т. д.). Методические ошибки имеют как случайный, так и систематический характер.

В дальнейшем будем считать, что по каждой оцениваемой характеристике формируется совокупность наблюдений $X=(x_1, x_2, \dots, x_n)$, состоящая из множества результатов регистрации. Отметим ряд основных свойств данной совокупности:

конкретные значения являются результатом проявления одной и той же закономерности;

совокупность состоит из варьирующихся величин, отличающихся своими значениями. Свойство вариации вызывает необходимость обработки всего множества наблюдений (если бы элементы совокупности были тождественны друг другу, то не было бы потребности обрабатывать всю совокупность);

результаты регистрации носят случайный характер и содержат в себе неустранимые ошибки.

Результаты обработки ЭД носят частный характер, и без должного обоснования по ним нельзя делать обобщающие выводы относительно аналогичных объектов, функционирующих в других условиях.

Пособие содержит описание методов и моделей обработки экспериментальных данных, применяемых в целях оценивания параметров (показателей) автоматизированных систем и систем телекоммуникаций. Основное внимание уделяется изложению сущности, вопросам применения указанных методов и моделей в инженерной практике. Теоретические аспекты методов обработки ЭД не рассматриваются, для их изучения можно рекомендовать такие фундаментальные работы, как [7, 8].

Изложение теоретического материала сопровождается примерами, иллюстрирующими сущность проводимых преобразований ЭД. Для проведения расчетов применялись типовые пакеты прикладных программ со стандартной разрядностью внутреннего представления вещественных чисел (например, MS Excel использует 8 – 9 разрядов в представлении вещественных чисел). В тексте пособия результаты вычислений представлены с округлением, существенно меньшим количеством разрядов, чем использовалось при вычислениях (это вносит некоторые искажения в представление результатов вычислений)