УДК 621.390

ЛОГАРИФМИЧЕСКАЯ МЕРА ИНФОРМАЦИИ СОСТОЯНИЯ ТЕХНИЧЕСКОГО ОБЪЕКТА

Дулесов А.С., Кабаева Е.В.

ФГБОУ ВПО «Хакасский государственный университет им. Н.Ф. Катанова», Абакан, Россия (655017, Хакасия, г. Абакан, пр. Ленина, 90), e-mail: kabaeva_e_v@mail.ru

В работе представлена модель определения логарифмической меры информации. Из структуры технической системы выделяется объект, и рассматриваются его вероятностные состояния отказа и работы. Когда состояния равновероятны, предлагается использовать меру Хартли, а для неравновероятных – меру Шеннона для одного и многих объектов, если они взаимонезависимы. Модель учитывает возможности определения меры информации только для одного объекта. Все состояния объекта разбиты на два класса. Каждый из выделенных классов формируется на основе данных о потоке неравновероятных событий. Для каждого класса состояний объекта определены суммарные и обобщенные вероятности работоспособности и отказа. Данные вероятности нашли применение в полученных математических выражениях для определения меры неопределенности информации. Показано, что полученные формулы идентичны и применимы как при использовании суммарной вероятности, так и обобщенной вероятности.

Ключевые слова: логарифмическая мера информации, энтропия, состояние технического объекта.

LOGARITHMIC MEASURE OF INFORMATION OF THE CONDITION OF TECHNICAL OBJECT

Dulesov A.S., Kabaeva E.V.

Khakass State University n.a. N.F. Katanov, Abakan, Russia (655017, Khakassia, Abakan, Lenin Avenue, 90), e-mail: kabaeva_e_v@mail.ru

The article presents the modifier of logarithmic measure of information model. An object is picked out from the technical system, and its probabilistic states of failure and work are analyzed. When the states are equiprobable it is recommended to use Hartley's measure, and when they are not equiprobable Shanon's measure is preferable for one or more interindependent objects. The model considers the capability of modifying the measure of information only for one object. All states of the object are divided into two classes. Each class is based on data of the flow of the inequiprobable events. The total and generalized probabilities of efficiency and failure are determined for the object's states of each class. The studied probabilities are used in the mathematical formulas for modifying the measure of the uncertainty of information. It is shown that the formulas are identical and may be applied both for the total and generalized probability.

Key words: logarithmic measure of information, entropy, condition of technical object.

Введение. К сложным техническим системам предъявляется ряд требований, среди которых выделяют поддержание высокого уровня надежности (работоспособности). Высоконадежные системы, как правило, подлежат контролю и диагностике с целью своевременного устранения возможных неполадок, появление которых имеет вероятностную природу. В целом систематический контроль позволяет получить общую картину состояния системы. Имея её на руках, можно вырабатывать решения, направленные на сохранение устойчивого поведения системы, сохранение уровня надежности, тем самым решая задачу кибернетики. Кроме этого, отслеживая «движение» системы во времени и пространстве, можно судить о её эволюции или старении, но уже с позиции синергетики.

Естественным процессом в технических системах является старение, которое неразрывно связано с таким понятием, как «неопределенность». Существует множество методологических подходов к анализу процессов и поддержанию работоспособности систем. Один из них основан на использовании теории информации и касается решения задачи получения меры неопределенности информации (энтропии) [4; 5]. В свою очередь значение информационной энтропии служит мерой выбора из возможных альтернатив.

В теории информации нашли свое практическое применение мера Хартли, позволяющая измерять детерминированные процессы конечной длины, и мера Шеннона – процессы любой длительности, анализ которых использует вероятностно-статистические методы. Обе меры входят в структурное и статистическое направления теория информации.

При эксплуатации технического объекта подсистема контроля передает сигналы или сообщения о состоянии системы, из которых формируется набор статистических данных. Их применение и направления теории информации могут быть положены в основу информационного анализа.

Далее предлагаются к рассмотрению некоторые свойства меры информации и несложные математические выкладки для её определения.

Модель определения меры информации. Техническую систему можно представить в виде структурной схемы с наличием элементов и связей между ними. С позиции оценки надежности в структуру вводят показатели: продолжительность восстановления элемента, частота отказов и др. На их основе определяют вероятность отказа и безотказной работы элемента. Большинство показателей носят вероятностный характер, обусловленный наличием неопределенности поведения системы. Применение меры неопределенности информации может послужить эффективным средством для оценки состояния технической системы, её элементов и структуры. С возможностями применения данной меры в технических системах можно ознакомиться в работах [1; 2]. Изменение состояния влияет на выполнение функций, одна из которых связана с передачей энергии (ресурсов) в системе. Исходя из функциональных особенностей системы возможны, по крайней мере, два варианта оценки состояния (изменения структуры). Первый не связан с учетом потокового процесса, второй учитывает направление потоков в структурной схеме системы. Далее при построении модели примем к рассмотрению первый вариант.

Простейшей моделью количественной оценки структурного содержания системы является подход Хартли, который предложил вычислять количество информации, содержащейся в сообщении. Для нашего случая будем предполагать, что система и каждый из её элементов могут находиться в одном из двух независимых дискретных состояний: работоспособное и неработоспособное. Тогда можно предположить, что от элементов в систему контроля и управления информация поступает в виде сигналов дискретного вида: 0 – элемент системы не работает (находится в неработоспособном состоянии); 1 – элемент работает (находится в работоспособном состоянии). Если предположить, что нас не интересует время пребывания элемента в том или ином состоянии, то общее число всевозможных состояний элементов будет выражаться формулой:

$$N = k^n, \tag{1}$$

где: k = 2 – число возможных состояний элемента или системы; n – количество элементов в рассматриваемой системе.

В формуле (1) общее число состояний (комбинаций) N – это количество сообщений, сформированных из равновероятных и независимых сигналов идущих от элементов. С ростом количества элементов n число комбинаций N увеличивается. Поэтому величину N Хартли использовал в качестве основы для определения меры количества информации. По (1) определяется максимальное количество состояний системы. В реальной обстановке (за определенный промежуток времени, например год) количество состояний всегда будет меньше N. Поскольку нас может интересовать состояние системы на отдельных интервалах времени, то такая мера количества информации, по (1), не удобна для практического использования. Хартли нашел решение, предположив: количество информации I, содержащееся в сообщениях, должно быть функцией от N, то есть $I = f(N)$. Так как число элементов n является показателем степени, то для определения I применяют логарифмическую функцию:

$$I = \log_2 N. \tag{2}$$

Поскольку число состояний k и основание логарифма приняты равными 2, то количество информации при таких условиях принимается за единицу, которая называется «бит» (двоичная единица).

Появление факторов в системе приводит к возникновению того или иного её состояния, которое по отношению к противоположному состоянию независимо. Условие о независимости состояний указывает на то, что общая информация, по (2), будет равна сумме отдельных информаций $I_1 = \log_2 N_1$ и $I_2 = \log_2 N_2$. Здесь N_1 и N_2 – количество состояний, относящихся к работе и отказу элементов системы соответственно. Их общее количество

$$N = N_1 N_2. \tag{3}$$

Например, если рассматривать два элемента в системе, каждый из которых может находиться в каком-либо из двух состояний, то $N = N_1 N_2 = 2 \cdot 2 = 4$, а для 3-х элементов – N = 8.

Прологарифмировав выражение (3), получим:

$$I = \log_k N = \log_k N_1 + \log_k N_2 = I_1 + I_2, \tag{4}$$

которое доказывает свойство аддитивности меры информации Хартли. Данная мера справедлива при условии наличия в системе равновероятных состояний, образующих конечное множество N.

Расширим возможности меры Хартли применительно к задачам поиска дискретного содержания структуры системы, при условии конечности исходного множества N.

Поскольку рассматриваются только два состояния системы – работоспособное и отказ, то они образуют два класса $k = 2$ эквивалентности состояний (см. выражение (4)). Полагаем далее, что каждый из элементов технической системы генерирует состояния, относящиеся только к двум классам эквивалентности.

Если принять во внимание количество элементов n, то согласно (1) полученная мера совпадет с логарифмической мерой Хартли:

$$I = \log_k N = \log_k k^n = n \log_k k = n. \tag{5}$$

Из (5) видно, что количество информации в битах равно количеству элементов системы. Следовательно, для равновероятных и взаимонезависимых друг от друга состояний элементов количество информации может выразиться через формулу:

$$I = n \log_2 k. \tag{6}$$

Таким образом, при наличии в системе n элементов и при равновероятных состояниях $N = k^n$ выражение (6) представляет собой логарифмическую меру информации Хартли.

В (6) состояния (работа и отказ) слиты воедино. Однако противоположные состояния желательно разделять, поскольку при их слиянии утрачивается смысл оценки уровня надежности через меру информации. Кроме этого, вероятности нахождения элемента в каждом из состояний неравнозначны. Поскольку важнейшей задачей является сохранение высокого уровня надежности элемента или системы, то на практике вероятность работоспособного состояния всегда будет выше противоположной вероятности. Во избежание слияния информаций (противоположных по своей сути при оценке надежности системы) нужно каждый класс эквивалентности рассматривать в отдельности.

Неравновероятность наличия состояний для элемента или системы вынуждает использовать формулу Шеннона. Последняя является мерой неопределенности наличия или вероятностного присутствия состояния элемента в том или ином классе эквивалентности. Рассмотрим получение формулы на следующем примере.

Например, оценивая надежность технической системы, состояния её элементов рассматриваются на длительных интервалах времени (год и более). На выделенном участке времени состояния перемежаются, следуя друг за другом, образуя поток событий. В этом потоке каждое из событий характеризуется видом (отказ или работа), временем появления и окончания, а также иными показателями. Эти состояния фиксируются в органе управления, одной из задач которого является сохранение высокого уровня работоспособности системы. Решая эту задачу (в нашем случае через определение количества информации), имеющийся поток событий классифицируют, относя события к конкретному i-му элементу либо к самой системе. Так, для одного из элементов, имея поток событий, можно определить вероятности появления каждого из них: p_i и q_i – вероятность нахождения i-го элемента в работоспособном и неработоспособном состоянии. Вероятности появления событий одного вида образуют суммарную вероятность, $p_i + q_i = 1$. Тогда количество информации при неравновероятных и взаимонезависимых событиях, содержащееся в одном элементе, определяется по формуле Шеннона:

$$I = -(p\log_2 p + q\log_2 q)\,. \tag{7}$$

Если рассматривать элементы системы как независимо функционирующие, тогда по формуле Шеннона можно определить информацию как

$$I = -[\sum_{i=1}^{n}(p_i \log_2 p_i) + \sum_{i=1}^{n}(q_i \log_2 q_i)]. \tag{8}$$

Вероятности в (8), стоящие перед логарифмом, усредняют значение самого логарифма. Если не разделять состояния по виду, то данное выражение можно переписать как

$$I_1 = -\sum_{i=1}^{n}(p_i \log_2 p_i) = -p_{cp}\log_2(\prod_{i=1}^{n} p_i) \tag{9}$$

при условии $\sum_{i=1}^{n} p_i = 1$. В (9) $p_{cp} = \dfrac{\sum_{i=1}^{n}(p_i \log_2 p_i)}{\sum_{i=1}^{n}\log_2 p_i}$ – среднее значение вероятности появления событий всех n элементов.

Однако (8) малопригодно из-за наличия взаимосвязей между элементами, и по этой причине состояния элементов будут определять состояния самой системы. Если ставится задача определить количество информации, содержащейся в системе, то она потребует выполнения некоторых условий: 1) следует иметь совокупные данные о состояниях системы за длительный промежуток времени; 2) иметь данные по каждому из элементов. Например, исходя из второго условия, решение задачи можно получить, опираясь на результаты, представленные в работе [3]. Далее будем рассматривать возможности определения меры информации лишь для одного элемента, исключив из рассмотрения систему.

Далее рассмотрим возможности в определении меры информации лишь для одного элемента (объекта). В данном случае будет справедливым использование выражения (7) при условии $p + q = 1$. Тогда максимум информации достигается при $p = q$ и будет равен 1.

В выражениях для определения меры информации справедливость применения логарифма по основанию 2 объясняется разбиением всего множества состояний элемента на два класса эквивалентности: работоспособные состояния и их вероятности отнесем к первому классу k_1, неработоспособные – ко второму классу k_2. Оба класса эквивалентности включают в себя целое число состояний $m = G + L$, где G – число состояний работоспособности, L – неработоспособности элемента системы. В первом классе присутствует множество G состояний с суммарной вероятностью $p = \sum_{g=1}^{G} p_g$, во втором – L с суммарной вероятностью $q = \sum_{l=1}^{L} q_l$. Тем самым каждый класс разбит на отдельные неравновероятные состояния.

Выделив 2 эквивалентных класса, где каждому из них принадлежит собственное множество неравновероятных состояний, информацию согласно (7) можно определить по выражению:

$$I^* = -[(\sum_{g=1}^{G} p_g)\log_2(\sum_{g=1}^{G} p_g) + (\sum_{l=1}^{L} q_l)\log_2(\sum_{l=1}^{L} q_l)] \tag{10}$$

$$\text{при условии } \sum_{g=1}^{G} p_g + \sum_{l=1}^{L} q_l = p + q = 1, \tag{11}$$

где p_g и q_l – вероятность g-го работоспособного и l-го неработоспособного состояний соответственно, ($m = G + L$) – полное количество состояний элемента. Выражения (7) и (10) совпадают и могут быть применены тогда, когда получены данные путем отслеживания потока событий либо ранее обобщенные статистические данные. При равенстве вероятностей состояний элемента – $p_g = q_l$, (например $p_g = q_l = 0{,}125$ и $G = L = 4$) по выражению (10) и при соблюдении условия (11) получим максимальное значение информации $I^* = 1$, тогда как по (8) – $I = 3$. Тем самым при равенстве вероятностей первая величина означает максимальное значение информации, содержащейся в одном элементе, вторая – в 3-х независимо функционирующих элементах. В последнем случае использование (8) неправомерно.

Часто в практике расчетов уровня надежности аналитик опирается на наличие статистических данных. При этом он может взять уже готовые обобщенные значения либо, накапливая опыт эксплуатации, рассматривает поток событий, получая тем самым серию вероятностей, и на основе теоремы сложения вероятностей находит суммарное значение $p = p_1 + ... + p_g + ... + p_G$ и соответственно суммарное q. Подставив эти значения в (7), можно определить количество информации, содержащееся в одном элементе.

Таким образом, выражение (10) представляет собой логарифмическую меру неопределенности информации, содержащуюся в одном элементе, с учетом разделения на работоспособное и неработоспособное состояния.

Отметим еще одну особенность в получении меры неопределенности. Она связана с тем, что формула Шеннона согласуется с формулами Хартли (3)-(6) для равновероятных событий. Если рассматривать поток неравновероятных состояний (событий), то принимая во внимание (3), обобщенные вероятности каждого из классов находятся по Шеннону как

$$P = p_1 \cdot p_2 \cdot ... \cdot p_g \cdot ... \cdot p_G = \prod_{g=1}^{G} p_g \text{ и } Q = \prod_{l=1}^{L} q_l. \qquad (12)$$

При их определении работает теорема умножения вероятностей, поскольку предполагается, что уровень надежности элемента можно представить в виде последовательно протекающих независимых событий[1]. Имея обобщенные вероятности по (12), можно сделать вывод о том, что мера информации для каждого из классов эквивалентности обладает свойством аддитивности. Тогда меру информации можно определить по формуле:

$$I_1 = -p_{cp} \log_2 P - q_{cp} \log_2 Q = -[p_{cp} \log_2 (\prod_{g=1}^{G} p_g) + q_{cp} \log_2 (\prod_{l=1}^{L} q_l)]. \qquad (13)$$

В (13) величины p_{cp} и q_{cp} усредняют значение информации.

Если в данном выражении будут известны p_{cp} и q_{cp}, то оно будет соответствовать формуле (10). По сути выражения для определения усредненных значений должны

[1] Следует отметить, что события могут быть и взаимозависимыми.

учитывать факт того, что в каждом эквивалентном классе события не являются однородными по характеру появления и содержанию причин, их породивших. Следовательно, основание логарифма при определении информации для одного класса событий должно отличаться от уже принятого основания, равного 2.

В теории о логарифмах известно выражение $\log_c a \cdot \log_a b = \log_c b$, которое в нашем случае (например, для k_1 класса) выглядит как

$$\log_2 P \cdot \log_P p = \log_2 p\,. \tag{14}$$

Из выражения (14) вытекает следующее:

$$\log_P p = \frac{\log_2 p}{\log_2 P}. \tag{15}$$

Отношение в (15) может быть рассмотрено как плотность информации. Тогда (например, для класса k_1) можно записать соотношение:

$$-p\log_2 p = -p\log_P p\log_2 P, \tag{16}$$

где $p_{cp} = p\log_P p$.

Тогда (13) с учетом средних величин можно записать в виде:

$$\begin{aligned} I_1 = -p\log_P p \cdot \log_2 P - q\log_Q q \cdot \log_2 Q = -[(\sum_{g=1}^{G} p_g)\log_{\prod p}(\sum_{g=1}^{G} p_g) \cdot \log_2(\prod_{g=1}^{G} p_g) + \\ +(\sum_{l=1}^{L} q_l)\log_{\prod q}(\sum_{l=1}^{L} q_l) \cdot \log_2(\prod_{l=1}^{L} q_l)]. \end{aligned} \tag{17}$$

Примем условие: G = L = 4; p_g = q_l = 0,125. Тогда по выражению (17) и при соблюдении условия (11) максимальное значение меры информации $I_1 = -0{,}5\log_{0{,}000244} 0{,}5 \cdot \log_2 0{,}000244 -$ $-0{,}5\log_{0{,}000244} 0{,}5 \cdot \log_2 0{,}000244 = -1{,}0,$ что подтверждает соответствие выражению (10).

Заключение. Из рассмотренного выше следует, что структура технической системы, состоящая из элементов и связей между ними, подлежит информационному анализу и оценке с позиции надежности. Каждый из элементов может находиться в одном из двух состояний: работа или отказ. Количество состояний, если они равновероятны, определяет значение меры Хартли согласно (6) и разделяется на два класса эквивалентности: класс работоспособных и класс неработоспособных состояний элемента системы. Если события неравновероятны, то мера информации для одного элемента может быть определена по формуле (7). Когда элементы взаимонезависимы, то по формуле Шеннона (8) и (9) можно определить меру информации для системы в целом.

Рассматривая состояния лишь для одного элемента или объекта, каждый из выделенных классов формируется на основе данных о потоке неравновероятных событий. Для каждого из классов эквивалентности можно определить суммарные и обобщенные вероятности работоспособности и отказа элемента. Эти показатели применимы для определения меры неопределенности информации элемента по полученным выражениям (10) и (17) с разделением на класс работоспособных и неработоспособных состояний. Показано, что (10) и (17) идентичны и применимы: первое выражение – при наличии суммарной вероятности, второе – при обобщенной вероятности.

На основе использования вышеупомянутых формул можно для однотипных элементов определить меру неопределенности и на основе полученных величин выделить менее надежный из них.

Список литературы

1. Вильчинская О.О., Гатауллин И.Н., Головинов С.О. и др. Определение количества информации в структуре технической системы // Информационные технологии: приоритетные направления развития. Кн. 5 : монография. – Новосибирск : ЦРНС – Изд-во «Сибпринт», 2010. – 261 с.
2. Дулесов А.С., Семенова М.Ю., Хрусталев В.И. Свойства энтропии технической системы // Фундаментальные исследования. – 2011. – № 8 (часть 3). – С. 631-636.
3. Дулесов А.С., Ускова Е.А. Применение подходов Хартли и Шеннона к задачам определения количества информации технических систем // Вопросы современной науки и практики. Университет им. В.И. Вернадского. – 2009. – № 2 (16). – С. 46-50.
4. Дулесов А.С., Ускова Е.А. Применение формулы Хартли для оценки структурных связей элементов в задаче обеспечения надежного функционирования технических систем // Вопросы современной науки и практики. Университет им. В.И. Вернадского. – 2009. – № 6 (20). – С. 37-41.
5. Кузнецов Н.А. Информационное взаимодействие в технических и живых системах // Информационные процессы. – 2001. – Т. 1. – № 1. – С. 1-9.

Рецензенты:

Нагрузова Любовь Петровна, д.т.н., профессор кафедры «Строительство» Хакасского технического института – филиала ФГАОУ ВПО «Сибирский федеральный университет», г. Абакан.

Булакина Елена Николаевна, д.т.н., профессор кафедры «Автомобили и автомобильное хозяйство» Хакасского технического института – филиала ФГАОУ ВПО «Сибирский федеральный университет», г. Абакан.