УДК 612.391; 519.72

ПРИМЕНЕНИЕ ПОДХОДОВ ХАРТЛИ И ШЕННОНА К ЗАДАЧАМ ОПРЕДЕЛЕНИЯ КОЛИЧЕСТВА ИНФОРМАЦИИ ТЕХНИЧЕСКИХ СИСТЕМ

А.С. Дулесов, Е.А. Ускова

ГОУ ВПО «Хакасский государственный университет им. Н.Ф. Катанова», г. Абакан, Республика Хакасия

Рецензент В.Е. Подольский

Ключевые слова и фразы: количество информации; мера информации; подход Хартли; подход Шеннона; энтропия.

Аннотация: Рассматриваются вопросы и получены математические выражения для определения количества информации, присущей структуре технических систем. Представлен пример расчета.

Введение. Техническая система представляет собой совокупность взаимосвязанных технических элементов, выделенных из технологической среды и взаимодействующих с окружающей средой для достижения поставленной человеком цели.

Выделяя техническую систему из среды, требуется определить взаимодействие этой системы с окружением независимо от поведения ее элементов.

Одним из характерных признаков функционирования систем является информация. Например, для таких систем как трубопроводные, электрические немаловажное значение приобретают вопросы структурного содержания, надежности, полезности использования и др., которые могут быть решены с позиции определения количества информации.

Постановка задачи. В теории информации для количественной оценки системы выделяют три основных направления: структурное, стохастическое и семантическое [1]. Имея на вооружении данные теоретические направления, покажем далее возможности применения первых двух для определения количества информации, присущей технической системе, когда ее состояние описывается изменением структурного содержания.

Для решения поставленной задачи систему можно отобразить в виде структурной схемы, которая будет являться формальной моделью, отделенной от содержательного наполнения системы. Модель представляет собой плоский неориентированный граф, в котором: вершины – узлы соединения элементов схемы; ребра – элементы схемы. Ребрам приписаны показатели, которые характеризуют свойства, присущие элементам.К свойствам отнесем: отказ или работоспособность элемента сети. Последние измеряются через вероятность отказа q_i или работоспособности p_i элемента i системы, либо через дискретные значения: 0 – элемент в состоянии отказа; 1 – в работоспособном состоянии.

Структурная мера информации (подход Р. Хартли) позволяет рассматривать дискретные состояния системы, используя комбинаторный метод.

Для решения задачи будем полагать, что техническая система состоит из n элементов. Количество информации вычисляется через количество комбинаций, в которых может находиться система вне зависимости от времени. При этом учитываются возможные или реализованные комбинации.

Ранее принято, что каждый элемент i в системе может находиться в одном из двух состояний. Если система включает в себя один элемент, то информация о нем известна, поскольку он либо в работоспособном состоянии, либо – неработоспособном. Количество состояний равно 2. В данных случаях присутствует полная определенность о состоянии системы. При наличии в системе двух элементов, число состояний равно 4, неопределенность в выборе элемента возрастает и, следовательно, увеличивается количество информации о состояниях. С ростом числа элементов в системе увеличивается неопределенность выбора, тем самым больше получаем информации о состоянии системы.

Из этих соображений следует заключение: информация изменяется монотонно от числа возможных состояний системы, которое определяется по выражению

$$N = 2^n. \tag{1}$$

Величина N может быть принята в качестве меры количества информации, но с грубыми допущениями. Дело в том, что применение (1) для определения N как меры количества информации малопригодно: если $N = 1$, информация равна нулю – имеется одно состояние элемента, а информация о нем отсутствует; не выполняется условие линейного сложения количества информации, то есть условие аддитивности.

Вопрос о выборе числа элементов дает нам следующее количество информации

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

определяемое по формуле Хартли и являющееся мерой степени неопределенности состояния системы.

В случае, когда рассматриваются влияния различных факторов, приводящих к изменению структуры системы, можно предложить к применению, опять же согласно Хартли, логарифмическую меру информации

$$I_\Sigma = \sum_{k=1}^{m} \log_2 N_k, \tag{3}$$

где m – количество факторов воздействия; N_k – количество выявленных состояний системы от воздействия k-го фактора.

Таким образом, согласно выражению (3) определяется степень неопределенности состояния системы как сумма $I_\Sigma = I_1 + I_2 + \ldots + I_k$ количества информации, поступающей от системы под воздействием внешних и внутренних факторов. Выражение (3) можно использовать для предварительной оценки состояния системы, поскольку оно учитывает состояние элементов, описываемых числами двоичной системы исчисления.

Стохастическая мера информации (подход К. Шеннона) позволяет получать результат определенного выбора среди возможных состояний системы. Получателя информации интересуют ситуации, при которых состояние системы стабильно, она работает надежно и выполняет заданные функции без существенных сбоев в работе. При этом получатель информации заранее знает, что приходящие об элементах системы сообщения, свидетельствующие о высокой вероятности ее работоспособности, близкой к единице, говорят о малой информативности системы. Столь же малоинформативны те состояния системы, у которых вероятности отказа близки к нулю. Появление состояний такого рода маловероятно, и сообщения о таких состояниях поступают получателю крайне редко. В дальнейшем нас будут интересовать только состояния, характеризующие высокий уровень надежности функционирования системы.

Состояния технической системы $X = (x_1, x_2, \ldots, x_i, \ldots, x_N)$ рассматриваются (например, на графе) как возможные исходы с вероятностями присутствия элементов или их групп в работоспособном состоянии и/или в состоянии отказа. Сумма вероятностей всех состояний

$$\sum_{i=1}^{N} p(x_i) = 1. \tag{4}$$

Все множество состояний характеризуется некоторой неопределенностью (по Хартли), зависящей только от мощности множества. Однако всевозможные состояния, в общем случае, не равновероятны. Например, два элемента находящиеся одновременно в работоспособном состоянии, имеют бóльшую вероятность безотказной работы, нежели в состоянии, когда один элемент работает, а другой находится в ремонте (отказе).

Предложим к рассмотрению сложные состояния (события), когда система или подсистема состоит из n элементов. Переход элемента из одного состояния в другое является независимым от состояний других элементов. Поскольку события рассматриваются в двух комбинациях (0 – отказ, 1 – работа), то число состояний определяется по выражению (1). Каждый элемент (или элементы), находящийся в состоянии x_i, выбирается из списка элементов с вероятностями p_i и $q_i = 1 - p_i$. Например, при $n = 2$ по выражению (1) $N = 4$, система будет находиться в состояниях $X = (x_1, x_2, x_3, x_4)$. Каждому событию x_i будут принадлежать следующие комбинации: $(1;1) \in x_1$; $(0;1) \in x_2$; $(1;0) \in x_3$; $(0;0) \in x_4$. Здесь, например, для $(1;0) \in x_3$:

Таблица 1

Выбор показателей состояния

Состояние	Число элементов в состоянии	Вероятность выбора состояния
x_1	n_1	p_1
x_2	n_2	p_2
…	…	…
x_i	n_i	p_i
…	…	…
x_N	n_N	p_N

первый элемент в работе; второй – в нерабочем состоянии. Каждому состоянию x_i приписывается вероятность p_i.

Для обобщенного случая каждое состояние системы будет характеризоваться вероятностями выбора состояния согласно табл. 1.

Поскольку в комбинаторике рассматривается двоичная система счисления, то для простоты дальнейших рассуждений принимается условие: $n_1 = n_2 = \ldots = n_i = \ldots = n_N = n$.

Вероятность того, что состояние x_i содержит n элементов будет равна

$$p = p_1^n q_2^n p_3^n q_4^n \cdots p_i^n q_{i+1}^n \cdots p_{N-1}^n q_N^n$$

или

$$p = p_1^n(1-p_1^n)p_2^n(1-p_2^n)\cdots p_i^n(1-p_i^n)\cdots p_{N/2}^n(1-p_{N/2}^n). \quad (5)$$

При большом числе n элементов в системе образуются (рассматриваются) состояния, в которых частота появления отдельных комбинаций стремится к вероятности выявления этих элементов, то есть $p_i = n_i/n$.

Поскольку ранее принято условие $n_i = n$, то вероятность $p_i = n_i/n = 1$ ($q_i = 1 - p_i = 0$), тогда как все остальные вероятности будут равны нулю (единице). Тем самым, согласно (5) при $p = 1$ ($q = 0$) система обладает абсолютной надежностью (ненадежностью). Реальные технические системы таким свойством не обладают. В них каждое состояние x_i характеризуется конкретными показателями в зависимости от комбинаций (числа состояний) элементов, вероятность появления которых определится по выражению

$$p_i = \prod_{j=1}^{M_i} p_j, \quad M_i = 1, 2, \ldots, \quad (6)$$

при этом

$$p_j = \begin{cases} p_j, & \text{при условии, что элемент находится в рабочем состоянии;} \\ q_j = 1 - p_j, & \text{в состоянии отказа.} \end{cases}$$

Здесь M_i – число состояний, в которых может находиться элемент j в i-м событии (в нашем случае $M_i = 2$). Вероятность совокупности всех событий определим из (5) и (6)

$$p = \prod_{i=1}^{N/2} p_i^{p_i}(1-p_i)^{(1-p_i)}. \quad (7)$$

Выражение (7) показывает, что для абсолютно надежных систем $p = 1$. Тогда система будет иметь только одно состояние $N = 1$ с количеством информации, согласно (2), равным $I = \log_2 N = \log_2 1/p = \log_2 1 = 0$.

Количество информации, содержащейся в состоянии x_i,

$$I = -\log_2 \prod_{i=1}^{N} p_i^{p_i}(1-p_i)^{(1-p_i)} = -\sum_{i=1}^{N/2}\left[\prod_{j=1}^{M_i} p_j \log_2 p_i + \prod_{j=1}^{M_i} p_j \log_2(1-p_i)\right]. \quad (8)$$

Данное выражение согласуется с формулой К. Шеннона

$$I = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_i p_j \log_2(p_i p_j) \qquad (9)$$

Выражение (8) дает более полное (чем мера Хартли) представление о количестве информации, которую несет система в процессе ее эксплуатации. Поясним данное высказывание на следующем примере. Пусть система состоит из одного единственного элемента (n = 1), который может находиться поочередно в двух состояниях ($N = M = 2$) с вероятностями $p_1 = 0{,}5$ и $q_1 = 1 - p_1 = 0{,}5$. По Хартли будем иметь $I = n\log_2 N = 1\log_2 2 = 1$. Согласно выражению (8)

$$I = -p\log_2 p - (1-p)\log_2(1-p) = -0{,}5\log_2 0{,}5 - 0{,}5\log_2 0{,}5 = -1.$$

Если $p \neq q \ (p + q = 1)$, тогда значение количества информации, полученное по формуле Хартли, будет уступать в точности расчетов форму-ле (8).

В случае, когда количество информации приходится измерять через изменение степени неопределенности, оно по выражению (8) численно совпадет с энтропией исходного множества

$$H = -\sum_{i=2}^{N/2}\left(\prod_{j=1}^{M_i} p_j \log_2 1/p_i + \prod_{j=1}^{M_i} q_i \log_2 1/q_i\right). \qquad (10)$$

Количество информации и энтропия являются логарифмическими мерами и измеряются в одних и тех же единицах. Поскольку в расчетах используется двоичная система счисления, количество информации (энтропия) измеряется в битах. Один бит – количество информации в сообщении об одном из двух равновероятных исходов некоторого фиксированного опыта. В технических системах условие равновероятности двух состояний не соблюдается. Надежность функционирования технических элементов (узлов) системы описывается (в своем большинстве) экспоненциальным распределением вероятности отказов. Тогда в (8) или (10) можно применить натуральный логарифм и единицу измерения Нат.

Согласно выражениям (8) и (10), чем выше надежность функционирования системы, тем меньше энтропия, тем меньше ее неопределенность.

Пример определения меры информации. Имеется два технических элемента (n = 2), которые работают параллельно.

Первый элемент работал безотказно в течении года с вероятностью $p_1 = 0{,}9$, второй – $p_2 = 0{,}8$. Каждый элемент может в любой момент времени находиться в одном из двух состояний: работа или отказ. Количество всех возможных состояний $N = 2^n = 4$, то есть $(1;1) \in x_1;\ (0;1) \in x_2;\ \ (1;0) \in x_3; (0;0) \in x_4$. Каждое состояние имеет вероятность $p_{11} = 0{,}72$; $p_{01} = 0{,}08$; $p_{10} = 0{,}18$ и $p_{00} = 0{,}02$. Их сумма равна единице. Используем выражение (8) для определения количества информации

$$I = -\sum_{i=2}^{2}\left(\prod_{j=1}^{2} p_j \log_2 p_i + \prod_{j=1}^{2} p_j \log_2 q_i\right) =$$

$$= -p_1 p_2 \log_2 p_1 - q_1 p_2 \log_2 q_1 - p_2 q_1 \log_2 p_2 - q_1 q_2 \log_2 q_2 =$$

$$= -(0{,}72\log_2 0{,}9 + 0{,}08\log_2 0{,}1 + 0{,}18\log_2 0{,}8 - 0{,}02\log_2 0{,}2) = 0{,}48 \text{ бит.}$$

Вывод: получены математические выражения, которые позволяют определять количество информации системы в зависимости от ее объема и надежности функционирования. Возможно определение количества информации в индивидуальных событиях, присущих системам при воздействии на них различного рода факторов.

Список литературы

1. Савчук, В.Л. Электронные средства сбора и обработки информации [Электронный ресурс] / В.Л. Савчук. – Режим доступа : http://www.ie. tusur.ru/books/COI/index.htm, свободный (дата обращения 26.02.2009).

Hartley and Shannon's Approaches to Tasks of Information Contents Determination in Technical Systems

A.S. Dulesov, E.A. Uskova

Katanov State University of Khakasia named after N.F. Katanov, Abakan, Republic of Khakasia

Key words and phrases: information contents; information measure; Hartley approach; Shannon approach; entropy.

Abstract: The paper studies the matters of information contents determination in technical systems; mathematical expressions to determine information contents are produced. The example of calculations is presented.

© А.С. Дулесов, Е.А. Ускова, 2009