EESTI NSV TEADUSTE AKADEEMIA TOIMETISED.
FÜÜSIKA * MATEMAATIKA
ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ФИЗИКА * МАТЕМАТИКА
PROCEEDINGS OF THE ACADEMY OF SCIENCES OF THE ESTONIAN SSR.
PHYSICS * MATHEMATICS

1988, 37, 2

УДК 007.5 : 681.5.01

Х. ТАНИ

ЭНТРОПИЯ СИСТЕМЫ И СОВОКУПНОСТИ

(Представил Х. Абен)

Проблемы оценки структурной сложности систем стали актуальными в связи с развитием методов декомпозиции сложных систем управления. После появления работ Дж. Хартманиса [1] эти методы стали особенно быстро развиваться в рамках теории конечных автоматов, а в настоящее время они приведены как во многих руководствах по проектированию дискретных систем, так и в отдельных работах [2–5].

Основой метода декомпозиции является тот факт, что общее число состояний системы является произведением чисел состояний компонентов, образуемых при декомпозиции. Следовательно, при заданном числе состояний системы компоненты имеют значительно меньше состояний и, стало быть, являются более простыми. Из этого многие авторы сделали вывод, что декомпозицией можно осуществить значительное упрощение систем.

Хотя это верно с точки зрения анализа и синтеза компонентов, естественно, возникает вопрос о том, можно ли, в самом деле, упростить таким путем систему в целом. Ведь для обеспечения выполнения системных функций необходимо образовать дополнительные каналы связи между отдельными компонентами, что ведет к усложнению системы.

Как показал Х. Салум [6, 7], расходы на образование каналов связи настолько большие, что в действительности системы при декомпозиции не упрощаются, а становятся более сложными. Поэтому полезность декомпозиции проявляется только в более наглядной структуре системы, что облегчает анализ и синтез отдельных компонентов.

В данной работе показано, что сложность системы при декомпозиции, в принципе, не может быть ниже сложности исходной системы. В интересах общности работа выполнена в терминах теории информации. Эффективность энтропийных оценок сложности была показана уже С. Ватанабе и другими [8, 9].

Возьмем в качестве модели дискретный, последовательный источник сообщений $G(n)$, имеющий множество выходных символов $A=\{a_1, a_2, \ldots, a_n\}$. В качестве выхода генератора рассмотрим случайные последовательности из n символов

$$\sigma = a_i, a_j, a_k, \ldots, a_n, \ldots, a_q,$$

где $i, j, k, n, q \in \{1, 2, \ldots, n\}$.

Для подсчета энтропии такого источника предполагаем, что распределение вероятностей появления отдельных символов равномерное.

Тогда перед появлением первого символа последовательности вероятность появления i-го символа

$$p_{i1}=\frac{1}{n}, \quad i=1, 2, \ldots, n,$$

а перед появлением j-го символа последовательности

$$p_{ij}=\frac{1}{n-j+1}, \quad j=1, 2, \ldots, n.$$

Следовательно, максимальная энтропия источника также меняется от шага к шагу

$$H_j=-\log_2 p_{ij}=\log_2(n-j+1),$$

а энтропия последовательности из n символов равна сумме условных энтропий по отдельным шагам

$$H=\sum_{j=1}^{n}\log_2(n-j+1). \tag{1}$$

Количество элементов множества A во время генерации очередной последовательности образует убывающую последовательность натуральных чисел, начиная с числа n. Поэтому (1) можно переписать в виде суммы логарифмов натуральных чисел

$$H=\sum_{k=1}^{n}\log_2 k.$$

С другой стороны, количество различных последовательностей источника $G(n)$ равняется $N=n!$, что дает такое же значение энтропии

$$H=-\log_2\frac{1}{n!}=\log_2(n!)=\sum_{k=1}^{n}\log_2 k. \tag{2}$$

Рассмотрим, далее, энтропию совокупности источников $G(n_1)$, $G(n_2)$, ..., $G(n_m)$, множествами выходных символов которых являются попарно непересекающиеся множества A_1, A_2, ..., A_m, содержащие n_1, n_2, ..., n_m элементов соответственно.

Энтропия отдельных источников

$$H_j=\sum\log_2 i, \quad j=1, 2, \ldots, m.$$

Образуя из данной совокупности систему источников путем композиции в виде

$$A=\bigcup_{i=1}^{m}A_i,$$

имеющей количество элементов

$$n=\sum_{j=1}^{m}n_j,$$

получим источник, энтропия которого определяется выражением (2).

Теперь легко убедиться, что суммарная энтропия совокупности меньше, чем энтропия системы, т. е.

$$\sum_{j=1}^{m}H_j<H,$$

так как

$$\sum_{j=1}^{m}\sum_{k=1}^{n_j}\log_2 k<\sum_{i=1}^{n}\log_2 i, \tag{3}$$

или, выражая то же самое в числе комбинаций,

$$K_1 = \prod_{j=1}^{m} (n_j!) < n!. \tag{4}$$

Выражения типа (3) и (4) допускают два вывода:

1. суммарная энтропия компонентов (а следовательно, и сложность) меньше, чем сложность системы;

2. ввиду сильной нелинейности энтропию нельзя использовать в качестве меры сложности систем, состоящих из компонентов.

Оказывается, однако, что оба эти вывода неверны. Для доказательства проследим задачу в обратном порядке.

Пусть имеется источник $G(n)$ с множеством выходных символов A из n различных элементов. Разобьем множество A на попарно непересекаемые подмножества $A_1, A_2, \ldots, A_m$, называемые компонентами источника $G(n)$, количество элементов которых равняется $n_1, n_2, \ldots, n_m$ соответственно. В терминах теории автоматов этому соответствует разбиение

$$\pi = \{B_1, B_2, \ldots, B_m\},$$

где B_j — блоки, $j = 1, 2, \ldots, m$.

Разбиение π может, однако, быть выполнено очень многими способами (даже при фиксированных значениях $n_1, n_2, \ldots, n_m$) в смысле состава блоков B_j. Численно это определяется количеством комбинаций

$$K_2 = C_n^{n_1} \times C_{n-n_1}^{n_2} \times \ldots \times C^{n_i}_{\left(n - \sum_{j=1}^{i-1} n_j\right)} \times \ldots \times C_{n_m}^{n_m} =$$

$$= \prod_{j=1}^{m} C^{n_j}_{\left(n - \sum_{i=1}^{j-1} n_i\right)}. \tag{5}$$

Это является источником дополнительной энтропии, что содержательно означает свободу выбора декомпозиции. Поэтому общее число комбинаций при декомпозиции значительно выше

$$K = K_1 K_2. \tag{6}$$

Теперь можно сформулировать следующую теорему.

Т е о р е м а. *Энтропия последовательного источника сообщений не изменится при разбиении его на компоненты.*

Д о к а з а т е л ь с т в о. Для доказательства достаточно показать, что $K = n!$. В самом деле, подставляя (4) и (5) в (6), получим:

$$K = K_1 K_2 = \prod_{i=1}^{m} (n_i!) \prod_{i=1}^{m} C^{n_i}_{\left(n - \sum_{j=1}^{i-1} n_j\right)} =$$

$$= \frac{n_1! \cdot n_2! \ldots n_m! \cdot n! \, (n - n_1)! \, (n - n_1 - n_2)! \ldots (n - n_1 - \ldots - n_m)!}{n_1! \cdot n_2! \ldots n_m! \, (n - n_1)! \, (n - n_1 - n_2)! \ldots (n - n_1 - \ldots - n_m)!} = n!$$

Выражая (6) в логарифмической форме, получим

$$\log_2 K = \log K_1 + \log_2 K_2,$$

или

$$H = H_1 + H(\pi). \tag{7}$$

В (7) энтропия H_1 является суммарной энтропией компонентов

$$H_1 = \sum_{i=1}^{m} H_i,$$

а энтропия $H_2(\pi)$ появилась из-за того, что разбиение можно выполнить K_2 разными способами. Поскольку она не определяется ни одним из компонентов и поэтому является как бы невидимой, будем называть ее скрытой энтропией. Ее величина определяется распределением количества элементов в разбиении π.

Теперь можно написать баланс энтропии при декомпозиции источника из n элементов на m компоненты, содержащие n_1, n_2, ..., n_m выходных символов соответственно

$$\sum_{i=1}^{m} H_i + H(\pi) = H(n!). \qquad (8)$$

Таким образом, при различных декомпозициях происходит только перераспределение энтропий между компонентами и изменение значения скрытой энтропии, а общая энтропия является постоянной величиной, определяющейся значением n.

Пример

Пусть имеется последовательный генератор, содержащий 18 выходных символов. Его энтропия

$$H = \log_2(18!) = 52{,}5076 \text{ битов.}$$

При разбиении его на компоненты, содержащие $n_1=5$, $n_2=3$, $n_3=4$ и $n_4=6$ элементов соответственно, суммарная энтропия компонентов равняется

$$H_1 = \log_2(5!) + \log_2(3!) + \log_2(4!) + \log_2(6!) = 23{,}5688 \text{ битов.}$$

Из баланса энтропии (8) можно определить значение скрытой энтропии

$$H(\pi) = H - H_1 = 52{,}5076 - 23{,}5688 = 28{,}9388 \text{ битов.}$$

Для контроля вычислим значение скрытой энтропии из (5): $H(\pi) = \log_2(C_{18}^5 \cdot C_{13}^3 \cdot C_{10}^4 \cdot C_6^6) = 28{,}9388$ битов, что совпадает со значением, полученным из баланса (8).

Рассмотрим, далее, как изменяется значение скрытой энтропии при изменении разбиения. Легко определить крайние точки.

Во-первых, при 1-разбиении [2]

$$H(\pi) = \log_2(C_n^n) = 0.$$

Другая крайность имеется при 0-разбиении: скрытая энтропия

$$H(\pi) = \log_2(C_n^1\, C_{n-1}^1 \times \ldots \times C_1^1) = H(n!),$$

а суммарная энтропия компонентов равна нулю, так как

$$H_1 = \sum_{i=1}^{m} \log_2(1) = 0.$$

При других разбиениях, очевидно, значение скрытой энтропии имеет некоторое промежуточное значение. Например, при разбиении источника, множество выходных символов которого состоит из n элементов, на два компонента, содержащие n_1 и $n-n_1$ элементов, значение скрытой энтропии имеет максимум при $n_1=n/2$, так как значение биномного коэффициента $C_n^{n_1}$ при этом максимальное.

Так как значение скрытой энтропии

$$H_S = \sum_{i=1}^{m} \log_2 \left(C^i_{\left(n - \sum_{1}^{i-1} n_j\right)} \right) \qquad (9)$$

определяется разбиением π, то возникает вопрос: можно ли слагаемые выражения (9) привязать к соответствующим компонентам разбиения? Оказывается, что так просто сделать это нельзя, так как значения отдельных слагаемых в (9) зависят от порядка записи компонентов. Это является естественным, так как свобода выбора компонента в ходе выполнения декомпозиции уменьшается. Так, при выборе первого компонента имеется большой выбор — из n элементов можно выбрать n_1 элементов, а при выборе второго компонента — из $n-n_1$ элементов и т. д. При образовании последнего компонента уже нет никакого выбора — в состав него входят элементы, которые остались после выбора всех предыдущих компонентов.

Поэтому для привязки к каждому компоненту доли скрытой энтропии необходимо установить между компонентами отношение порядка

$$A_i \prec A_k \prec A_l \prec \ldots \prec A_q,$$

$$i, k, l, q \in 1, 2, \ldots, m.$$

Тогда можно соответственно упорядочить и разбиение

$$\pi(B_i, B_k, B_e, \ldots, B_q).$$

Вычисляя скрытую энтропию в том же порядке

$$H(\pi) = \log_2 (C_n^{n_i}) + \log_2 (C_{n-n_i}^{n_2}) + \ldots + \log_2 (C_{nq}^{nq}),$$

получим

$$H(\pi) = H_i + H_k + H_l + \ldots + H_q.$$

Слагаемые этого выражения соответствуют скрытой энтропии упорядоченных компонентов. Для рассмотренного примера имеем

$$H(\pi) = \log_2 (8568) + \log_2 (286) + \log_2 (210) + \log_2 (1).$$

Взяв это же разбиение в другом порядке, например, $\pi(4, 3, 5, 6)$, имеем

$$H(\pi) = \log_2 (3060) + \log_2 (364) + \log_2 (462) + \log_2 (1).$$

Итак, в случае упорядоченных компонентов, их энтропию можно рассматривать как сумму

$$H_i^0 = H_{1i} + H(\pi)_i = \log_2 (n_i!) + \log_2 (C_{n-R}^{n_l}),$$

где R — сумма количеств элементов предшествующих компонентов.

Величину H_i^0 будем называть полной энтропией компонента i.

Объединив в упорядоченной системе два компонента с множествами выходных символов A_i и A_j, можно суммировать полные энтропии

$$H^0(A_i \cup A_j) = H_i^0 + H_j^0. \tag{10}$$

Таким образом, закон сложения энтропий при объединении компонентов действителен только для полных энтропий.

Резюмируя полученные результаты, можно ввести определения совокупности и системы источников сообщений.

1° Совокупностью источников сообщений будем называть некоторое количество несвязанных источников $G_1, G_2, \ldots, G_m$ с множествами выходных символов $A_1, A_2, \ldots, A_m$, содержащих $n_1, n_2, \ldots, n_m$ элементов соответственно.

Свойства совокупности:

1. Для каждого члена можно определить энтропию

$$H_i = \log_2 (n_i!), \quad i = 1, 2, \ldots, m.$$

2. Энтропию объединения двух членов совокупности нельзя определить как сумму энтропий членов, так как

$$H(n_i)+H(n_j)<H(n_i+n_j).$$

3. Элементы множеств выходных символов компонентов должны быть различимыми только в пределах данного множества.

2° Системой источников сообщений будем называть некоторое количество упорядоченных источников G_1, G_2, ... , G_m, для которого определены:

— общее количество элементов $n=\Sigma\, n_i$;

— разбиение $\pi(n_k, n_i, n_j, \ldots, n_m, \ldots, n_q)$;

— энтропии компонентов и скрытая энтропия.

Свойства системы:

1. Существование баланса энтропии (8).
2. Для каждого компонента может быть определена полная энтропия.
3. Имеет место закон сложения полных энтропий (10).

Для иллюстрации различий между совокупностью и системой можно сказать, что при декомпозиции источника в совокупность источников исходный источник не восстановим. Для восстановления исходного источника требуется, чтобы результатом декомпозиции была система, а не совокупность.

ЛИТЕРАТУРА

1. *Hartmanis, J.* Algebraic Structure Theory of Sequencial Machines. New York, Prentice-Hall, Inc., Englewood Cliffs, 1966.
2. *Кеэваллик А. Э.* Автоматика и вычислительная техника, № 1, 17—24 (1974).
3. *Lee, S. C.* In: Modern Switching Theory and Digital Designs. New York, Prentice-Hall, Inc., Englewood Cliffs, 1978, 236—297.
4. *Баранов С. И.* Синтез микропрограммных автоматов. Л., «Энергия», 1979, 52—86.
5. *Фридман А., Менон Р.* Теория и проектирование переключительных схем. М., «Мир», 1978, 290—374.
6. *Салум Х. Л.* Кибернетика, № 4, 8—12 (1986).
7. *Салум Х. Л.* Электрические станции, № 9, 43—45 (1986).
8. *Watanabe, S.* IBM J. Res. Develop., **4**, № 1, 56—82 (1960).
9. *Лаусмаа Т. М.* Изв. АН ЭССР. Физ. Матем., **31**, № 4, 390—398 (1982).

Институт кибернетики
Академии наук Эстонской ССР

Поступила в редакцию
16/IV 1987

H. TANI

SÜSTEEMI JA KOMPONENTIDE KOGUMI ENTROOPIA

Artiklis on näidatud, et diskreetse süsteemi keerukust ei ole võimalik vähendada süsteemi väiksemateks komponentideks dekomponeerimise teel, sest komponentide summaarse keerukuse vähenemisega kaasneb nendevaheliste sidemete keerukus. On esitatud varjatud entroopia ja täisentroopia mõisted ning näidatud täisentroopiate liidetavust.

H. TANI

ENTROPY OF THE SYSTEM AND OF THE COLLECTION

Using the concept of entropical complexity it is shown that the complexity of a discrete system cannot be diminished by decomposition of it into components because the diminished complexity of components is compensated by increasing complexity of links between them. Additionally, the definitions of hidden and full entropy are given and the additivity of full entropies is shown.