ИНФОРМАЦИОННЫЙ АНАЛИЗ ИЕРАРХИЧЕСКИХ СИСТЕМ


С.Ш. Кажикенова,
доцент
Карагандинского
государственного
технического
университета,
канд. техн. наук,
член-кор. МАИН

В.П. Малышев,
заведующий
лабораторией
энтропийно-
информационного
анализа Химико-
металлургического
института
им. Ж. Абишева,
д-р техн. наук,
профессор,
академик МАИН,
лауреат
Государственной
премии Республики
Казахстан

Самоорганизующиеся иерархические системы относятся к классу многоуровневых и многоцелевых систем. Использование меры определенности и неопределенности информации позволяет анализировать общие механизмы энтропийно-информационных закономерностей технологических переделов, являющихся фундаментальной основой всех самопроизвольно протекающих процессов накопления информации, приводящих к самоорганизации технологических систем. Поэтому очень важно найти адекватные математические модели оптимального решения и постановки задач для анализа химико-металлургических процессов с количественной оценкой их технологического совершенства.

Для информационного анализа качества технологических продуктов и процессов их получения количественные оценки ценности информации могут производиться только после предварительной договоренности о том, что в каждом конкретном случае имеет ценность для рассматриваемых металлургических процессов. Алгоритмы вычисления информационной емкости системы, предложенные Шенноном, позволяют выявить соотношение количества детерминированной информации и количества стохастической информации, которую нельзя заранее предугадать, и тем самым дать возможность определить качественную и количественную оценку определенной технологической схемы. При общей характеристике энтропийно-информационного анализа любых объектов широко используется статистическая формула Шеннона для выражения неопределенности любой системы [1]:

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i, \qquad (1)$$

где p_i — вероятность обнаружения какого-либо однородного элемента системы в их множестве N;

$\sum_{i=1}^{N} p_i = 1$, $p_i \geq 0$, i = 1,2,...,N.

Рассмотрим применение данной формулы для количественной оценки неопределенности качества продукта или технологического передела через неопределенность главного элемента системы. В качестве вероятности обнаружения главного элемента технологической системы можно принять его содержание в продукте, выраженное в долях единицы. До опубликования созданной К. Шенноном теории Р. Хартли предложил определять количество информации по формуле [2]:

$$H_{max} = \log_2 N, \qquad (2)$$

где H_{max} — количество информации; N — число элементов системы.

Для энтропийно-информационного анализа технологического передела необходимо выбрать единую меру статистических и детерминистических начал в любом целом. Наиболее полно эта мера выражается в информации, которая может быть выражена в различных отношениях: свободная и связанная, субъективная и объективная, реальная и потенциальная и т. д. [3]. Столь же правомерно использование энтропии как меры неупорядоченности, которая также охватывает весь спектр состояний системы. Информация, как мера определенности, отражает функцию структурного начала в технологической системе, а энтропия, как мера неопределенности, — ее бесструктурного дополнения [4].

Теорема 1. Если $\overline{I(d)}$, $\overline{I(h)}$ — относительные значения информации $I(d)$, энтропии $I(h)$ и на основании закона сохранения суммы энтропии и информации выполнено условие:

$$\overline{I(d)} + \overline{I(h)} = 1, \qquad (3)$$

то $\overline{I(d)}$ есть решение уравнения:

$$\overline{I(d)}^n + \overline{I(d)} - 1 = 0,$$

где $n \in Z$, $n \geq 0$.

Доказательство. В равенстве (3) $\overline{I(d)} = \dfrac{I(d)}{H}$, $\overline{I(h)} = \dfrac{I(h)}{H}$ есть относительные значения информации и энтропии.

Их гармония возможна тогда и только тогда, когда пропорциональны их относительные изменения [5]:

$$\frac{d\overline{I(h)}}{\overline{I(h)}} = n\frac{d\overline{I(d)}}{\overline{I(d)}}. \qquad (4)$$

$$\ln \overline{I(h)} = n\ln \overline{I(d)} + \ln c \Rightarrow \ln\overline{I(h)} = \ln\overline{I(d)}^n \cdot c \Rightarrow$$
$$\Rightarrow \overline{I(h)} = c \cdot \overline{I(d)}^n. \qquad (5)$$

Зададим начальные условия для определения произвольной постоянной:

$$n = 1 \Rightarrow \overline{I(h)} = \overline{I(d)}. \qquad (6)$$

Так как $n \in Z$, $n \geq 0$, где Z — множество целых чисел, то $c = 1$. Таким образом:

$$\overline{I(h)} = \overline{I(d)}^n \Rightarrow \overline{I(d)}^n + \overline{I(d)} - 1 = 0. \qquad (7)$$

Что и требовалось доказать.

Математическое описание процесса развития любой системы задается формулой [6]:

$$\frac{\partial^2 I}{\partial M^2} > 0,\ \frac{\partial^2 I}{\partial N^2} > 0,$$

где M — масса; N — число элементов технологической системы.

Положительная вторая производная свидетельствует об ускоренном развитии системы. Суть этого ускорения в том, что при переходе на более высокий структурный уровень технологического процесса вступает в действие закон или принцип прогрессивного увеличения разнообразия [7]. В математическом понимании принцип увеличения разнообразия состоит в том, что с переходом на более высокие структурные уровни число элементов, образующих данный структурный уровень, имеющий различные признаки, увеличивается по закону:

$$N_n = N_0^{k^n}, \qquad (8)$$

где n — порядковый номер, рассматриваемого уровня, $n \in Z$, $n \geq 0$; N_n — число элементов n-го уровня; k — длина кода элементов на каждом уровне; N_0 — число элементов уровня, принятого за начало отсчета $n = 0$.

Теорема 2. Пусть N_n — число элементов n-го уровня иерархической системы, $n \in Z$, $n \geq 0$. I_0 есть емкость информации нулевого уровня. Тогда емкость информации n-го уровня в расчете на один элемент выражается формулой:

$$I_n = k^n I_0, \qquad (9)$$

где k — длина кода элементов на каждом уровне иерархической системы.

Доказательство.

$$I_n = \log N_n = \log N_0^{k^n} = k^n \log N_0 = k^n I_0.$$

Что и требовалось доказать.

Теорема 3. Информационная емкость иерархической системы и n-го уровня определяются равенствами:

$$I_{\Sigma_n} = \sum_{i=0}^{n} \frac{H_{i(\max)}}{(i+1)} = \log N \sum_{i=0}^{n} \frac{\prod_{m=0}^{i} k_m}{(i+1)!},$$

$$I_n = \frac{H_{n(\max)}}{(n+1)} = \frac{\prod_{m=0}^{n} k_m \log N}{(n+1)!}, \qquad (10)$$

где $H_{n(\max)}$ — максимально возможная энтропия системы.

Доказательство. Согласно закону сохранения, количество детерминированной информации рассчитывается как разность между $H_{n(\max)}$ и H_n. Тогда:

$$I_n(d) = H_{n(\max)} - H_n => H_{n(\max)} = H_{n(\max)} - H_n + I_n(h) =>$$

$$=> I_n(h) = H_n . \qquad (11)$$

Максимум информации находится по формуле Хартли, которая применительно к уровневой выразится следующим образом:

$$H_{n(\max)} = \log N_n . \qquad (12)$$

где N_n определено формулой (9).

Информационная емкость технологической системы зависит от информационных свойств системы и определяется формулой [5]:

$$I_n = \frac{\Delta I_n}{(n+1)!} , \qquad (13)$$

где ΔI_n — максимальное приращение информации.

Так как $\Delta I_n = H_{n(\max)}$, то:

$$I_{\Sigma_n} = \sum_{i=0}^{n} \frac{H_{i(\max)}}{(i+1)!} = \log N \sum_{i=0}^{n} \frac{\prod_{m=0}^{i} k_m}{(i+1)!} ,$$

$$I_n = \frac{\prod_{m=0}^{n} k_m \log N}{(n+1)!} .$$

Что и требовалось доказать.

Из формулы (11) следует, что информационная емкость любого уровня технологической системы, за исключением нулевого, всегда меньше максимально возможного и не может превзойти некоторое определенное для каждого уровня значение.

Теорема 4. Информационная емкость технологической системы определяется по ее стохастической части.

Доказательство. Доказательство проведем от противного. Допустим, что информационная емкость технологической системы определяется по ее детерминированной части $I_n = I_n(d)$. Тогда из равенства (12) получим:

$$I_n(d) = H_{n(\max)} - H_n = H_{n(\max)} - I_n(h) =>$$

$$=> I_n(h) = H_{n(\max)} - I_n = H_{n(\max)} - \frac{H_{n(\max)}}{(n+1)!} =$$

$$= H_{n(\max)} \left[1 - \frac{1}{(n+1)!} \right]. \qquad (14)$$

Если $n = 0$, то $I_0(h) = 0$. Это означает, что этот уровень оказывается полностью детерминированным, и мы приходим к противоречию. Если же информационная емкость технологической системы определяется по ее стохастической части $I_n = I_n(h)$, то:

$$I_n(h) = \frac{H_{n(\max)}}{(n+1)!} .$$

Если $n = 0$, то $I_0(d) = 0$ и детерминация полностью отсутствует.

Что и требовалось доказать.

Теорема 5. Предельная степень детерминации и неустранимой стохастичности технологической системы определяются по формулам:

$$d_{\Sigma_n} = \frac{I_{\Sigma_n}(d)}{H_{\Sigma_{n(\max)}}}, \quad h_{\Sigma_n} = \frac{I_{\Sigma_n}(h)}{H_{\Sigma_{n(\max)}}} ,$$

где $I_{\Sigma_n}(d)$, $I_{\Sigma_n}(h)$ — системные детерминированная и стохастическая составляющие, $H_{\Sigma_{n(\max)}}$ — системная максимальная информация.

Доказательство. На основании равенства (12) получим:

$$I_n(d) = \log N \prod_{m=0}^{n} k_m \left[1 - \frac{1}{(n+1)!} \right] ,$$

$$I_{\Sigma_n}(d) = \log N \sum_{i=0}^{n} \prod_{m=0}^{i} k_m \left[1 - \frac{1}{(i+1)!} \right] \qquad (15)$$

Максимальная информация n-го уровня и суммарное значение максимальной информации с учетом формул (9) и (13) определяется как:

$$H_{n(\max)} = \log N \prod_{m=0}^{n} k_m ,$$

$$H_{\Sigma_{n(\max)}} = \log N \sum_{i=0}^{n} \prod_{m=0}^{i} k_m . \qquad (16)$$

Для установления предельной степени детерминации и неустранимой стохастичности технологического передела вычислим пределы:

$$d_{\Sigma_n} = \lim_{h \to 0} \frac{I_{\Sigma_n}(d) + I_{\Sigma_n}(h)}{H_{\Sigma_{n(\max)}}} = \frac{I_{\Sigma_n}(d)}{H_{\Sigma_{n(\max)}}} , \qquad (17)$$

$$h_{\Sigma_n} = \frac{I_{\Sigma_n}(h)}{H_{\Sigma_{n(\max)}}} . \qquad (18)$$

Что и требовалось доказать.

Для предельных характеристик технологической системы степень детерминации равна коэффициенту избыточности $R = 1 - H_r / H_{max}$, используемому в теории информации; отношение равно h / d коэффициенту стохастичности $G = H_r / I_r = (1 - R) / R$, используемому в теории информации; I_r — реализованная информация, H_r — энтропия системы в рассматриваемый момент, R — коэффициент избыточности.

Смысл избыточной информации связан с познанием технологической системы, при котором извлекаемая информация всегда меньше объективно содержащейся в ней в виде детерминированных соотношений. Коэффициент стохастичности может изменяться от нуля до бесконечности и более детально отражает стохастические и детерминированные свойства технологической системы. Величина, равная $1 / G = Q$, дает более наглядное представление о возможностях непредсказуемого развития технологической системы, поэтому по аналогии с коэффициентом стохастичности ее можно назвать коэффициентом детерминированности.

При подстановке равенства (8) в (10), (15)–(16) получим формулы для определения всех видов информации иерархической системы:

$$I_n(h) = \frac{k_n \log N}{(n+1)!}, \quad I_{\Sigma_n}(h) = \log N \sum_{i=0}^{n} \frac{k^i}{(i+1)!}, \quad (19)$$

$$I_n(d) = k^n \left[1 - \frac{1}{(n+1)!}\right] \log N,$$

$$I_{\Sigma_n}(d) = \log N \sum_{i=0}^{n} k^i \left[1 - \frac{1}{(i+1)!}\right], \quad (20)$$

$$H_{n(\max)} = k^n \log N, \; H_{\Sigma_{n(\max)}} = \log N \sum_{i=0}^{n} k^i. \quad (21)$$

Из формул для детерминированной составляющей и максимальной информации технологической системы следует, что они не имеют конечных пределов при $n \to \infty$ и являются неограниченными функциями. Относительно стохастической части (19) мы имеем сходящийся по признаку Даламбера числовой ряд.

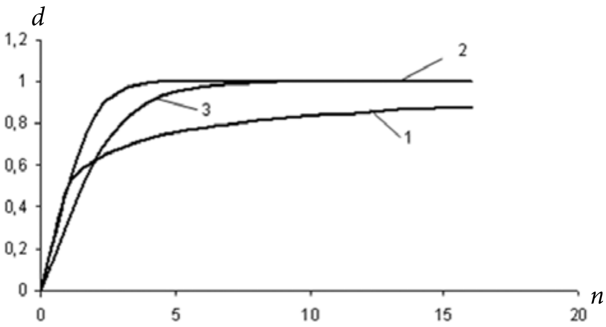
Сопоставим расчетные данные новой модели, рассчитанные по формулам (19)–(21), с расчетными данными теоремы 1 и представим сравнительные показатели по степени детерминации иерархической системы в таблице. Различие новой и модели, построенной по теореме 1, проиллюстрируем графически в координатах n, d. Мы видим, что при переходе на более высокий структурный уровень вступает в действие закон или принцип прогрессивного увеличения разнообразия [5]. Так как распределение вероятностей по этим уровням не влияет на качество продукции, то при расчетах достаточно ограничиться только междууровневыми корреляциями.

Сравнение степеней детерминации и стохастичности по расчетам теоремы 1 и теоремы 5 для $k = 2$, $N_0 = 2$

$n \in Z$	$\overline{I_n(d)}$	$\overline{I_n(h)}$	d_n	h_n	d_{Σ_n}
0	0	1	0	1	0
1	0,5000	0,5000	0,5000	0,5000	0,3333
2	0,6180	0,3820	0,8333	0,1667	0,6190
3	0,6823	0,3177	0,9583	0,0417	0,8000
4	0,7245	0,2755	0,9917	0,0083	0,8989
5	0,7549	0,2451	0,9986	0,0014	0,9496
6	0,7781	0,2219	0,9998	0,0002	0,9749
7	0,7965	0,2035	1,0	0	0,9875
8	0,8111	0,1889	1,0	0	0,9937
9	0,8245	0,1755	1,0	0	0,9969
10	0,8354	0,1646	1,0	0	0,9984
11	0,8446	0,1554	1,0	0	0,9992
12	0,8528	0,1472	1,0	0	0,9996
13	0,8599	0,1401	1,0	0	0,9998
14	0,8662	0,1338	1,0	0	0,9999
15	0,8722	0,1278	1,0	0	1,0
16	0,8774	0,1226	1,0	0	1,0

Из данных таблицы видно, что уровневая детерминация d_n и гармонизированная детерминация $\overline{I_n(d)}$ совпадают только для первых двух уровней. В дальнейшем уровневая детерминация резко возрастает, приближаясь на седьмом уровне к единице. Системная детерминация занимает промежуточное положение. Очевидно, $\overline{I_n(d)}$ ближе к системной детерминации, для которой значение детерминации меньше за счет вклада нижних уровней, отличающихся большей стохастичностью. Системная детерминация, как видно, существенно зависит от длины кода элемента k. На втором уровне технологической схемы получается значение, практически совпадающее с отношением золотого сечения.

Отсюда следует, что при произвольной элементной базе должны отличаться особой распространенностью трехуровневые системы с бинарным принципом организации. Однако все три модели требуют идентификации при сравнении с практическими данными, что требует самостоятельного анализа.



Зависимость степени детерминации от уровня:
n — номер уровня; *d* — степень детерминации; 1 — зависимость по теореме 1; 2, 3 — уровневая и системная детерминации по новой модели

Рассмотренный подход, на наш взгляд, полностью соответствует основным требованиям системного энтропийно-информационного анализа, так как обеспечивает при моделировании иерархической системы технологических процессов целостность ее рассмотрения за счет общетеоретических и методических концепций, позволяющих целиком удерживать в поле зрения всю систему в целом для решения задачи на всех уровнях. Кроме того, на основе учета основных элементов в системе и связей между ними обеспечивает полноту и всесторонность рассмотрения. Предложенный алгоритм упрощения при моделировании позволяет адекватно отразить реальный технологический передел и учесть определяющие факторы в иерархической системе.

Доказанные в данной работе теоремы показывают неразрывную связь детерминированной и стохастической составляющих, из которых первая является доминирующей и обеспечивающей устойчивость, а вторая определяет наиболее тонкие изменения и оптимальную информационную емкость технологических систем. В связи с этим заключаем, что энтропийно-информационный подход к изучению технологических систем является объективно необходимым.

Литература:

1. Шеннон К.Э. Математическая теория связи / Работы по теории информации и кибернетике. — М.: ИЛ, 1963. — С. 243.

2. Хартли Р. Передача информации / Теория информации и ее приложения. — М.: ИЛ, 1959. — С. 5.

3. Налимов В.В. Вероятностная модель языка. — М.: Наука, 1979. — 69 с.

4. Волькенштейн М.В. Энтропия и информация. — М.: Наука, 1986. — 192с.

5. Малышев В.П. Вероятностно-детерминированное отображение. — Алматы: Гылым, 1994. — 376 с.

6. Седов Е.А. Эволюция и информация. — М.: Наука, 1976. — 232 с.

7. Сороко Э.М. Структурная гармония систем. — Минск: Наука и техника, 1984. — 264 с.