

УДК 65.015.13: 674.01

М.П. Худяков

Худяков Михаил Павлович родился в 1965 г., окончил в 1987 г. Севмашвуз, кандидат технических наук, доцент кафедры автоматизации технологических процессов и технических систем машиностроения Севмашвуза. Имеет более 20 научных работ в области технологии автоматизированного производства.



АНАЛИЗ ЭФФЕКТИВНОСТИ СПОСОБОВ ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИОННЫХ СТРУКТУР В ЗАДАЧАХ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ И СИСТЕМАМИ

Рассмотрены способы представления информации, используемые при проектировании и управлении технологическими процессами. На базе энтропийного подхода получены зависимости для их оценки по степени понижения объема информации.

Ключевые слова: представление информации, информационная база, энтропия.

Проектирование и управление технологическими процессами современных производств осуществляется на основе их информационных моделей. Особенно это характерно для комплексных, распределенных, многопрофильных и многономенклатурных производств, к которым можно отнести лесоперерабатывающие и деревообрабатывающие предприятия. Общий объем информации, генерируемой при проектировании подобных технологических процессов и управлении ими, составляет по укрупненным оценкам от нескольких тысяч до сотен тысяч документов. Управление подобным объемом данных – это существенная проблема, которая на сегодняшний день полностью не решена. В этих условиях поиск способов повышения управляемости информационной базы проектирования и оценка их эффективности является актуальной задачей.

Анализ различных технологических процессов показал, что рост общего объема информации при ее изготовлении описывается кривой 1, интенсивность роста этого показателя – кривой 2, изменение количества актуальной информации – кривой 3 (рис. 1).

Из приведенных на рис. 1 кривых видно, что при сильной неравномерности

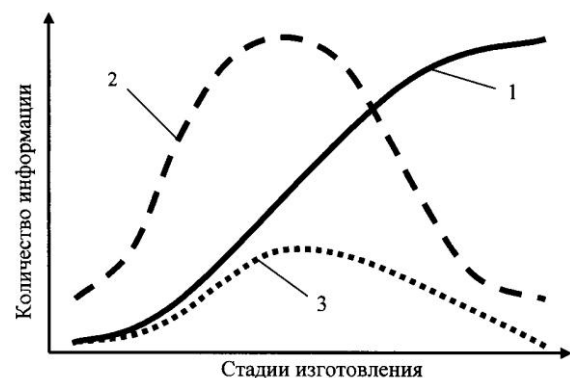


Рис. 1

прироста общей технологической информации количество актуальной в текущий момент времени информации значительно меньше ее общего объема. Следовательно, с одной стороны, задача управляемости информационной базой технологического проектирования состоит в поддержании высокого уровня актуальности технологической информации, с другой – в снижении объема и неравномерности ее прироста. Первая задача решается обычно средствами TDM/EDM/PDM-систем и в данной статье не обсуждается. Рассмотрим методы решения второй задачи.

Любой метод повышения управляемости информационного фонда предусматривает количественный и структурный аспекты. Соответственно критериальные оценки эффективности метода должны предусматривать совместную оценку.

Обозначим за N некоторое общее количество информации, потенциально генерируемое при описании процесса изготовления изделия:

$$N = AS, \quad (1)$$

где A – количество элементов описания реальных объектов (состояния изделия, средства технологического оснащения, операции, процессы и т.д.);

S – число потенциально возможных состояний элементов описания, отражаемых записями p фиксированного размера r (количество полей в записи).

С точки зрения информатики количество возможных состояний некоторого элемента описания (сообщения) определяют по формуле

$$H = k \log_2 m, \quad (2)$$

где H – шенноновская энтропия или мера неопределенности состояния объекта или некоторой ситуации с конечным числом исходов [2];

k – число знаков в сообщении;

m – число знаков в алфавите.

Если сопоставить понятие знака в сообщении и понятие поля записи в описании состояния объекта, понятие числа знаков в алфавите и понятие числа значений поля [1], получим

$$N = A_j p_j \log_2 r_j, \quad (3)$$

где j — количество однотипных элементов описания.

Из формулы (3) следует, что снижение общего объема N информации возможно за счет уменьшения количества элементов описания A , записей p , используемых для записей полей r .

Формула (3) напрямую применима к системам описания лингвистического характера, использующим для этой цели проблемно ориентированные языки. В этом случае в качестве p_j выступает количество необходимых для полного описания j -го объекта записей, а в качестве r_j – количество выраженных лингвистической формулой характеристических признаков j -го объекта описания. Так же применяют формулу (3) и к системам кодирования информации.

Рассмотрим уменьшение количества объектов описания A за счет их типизации. Для этой цели чаще всего применяют различные методы классификации реальных объектов. При их группировке по тем или иным признакам объем описательной информации уменьшается в n раз (степень понижения информационного объема):

$$n = N/N_i = N / \sum (p_i \log_2 r_i + 1) i, \tag{4}$$

где i – количество классификационных групп;

p_i – количество записей в группе;

r_i – количество полей в записях группы.

Для технологий без использования классификаторов $n = 1$.

Для типовой технологии

$$n = A_j p_j \log_2 r_j / (p_j \log_2 r_j + 1) = A_j - A_j / (p_j \log_2 r_j + 1). \tag{5}$$

Для групповых технологий с различными концепциями классификации n соответственно составляет промежуточную величину по отношению к рассмотренным случаям. Чтобы избежать потерь информации в несовпадающих полях и записях для существенно различных объектов описания, объединенных в группы, при составлении группы требуется введение дополнительных полей и записей.

Тогда

$$n = A_j p_j \log_2 r_j / \{ (p_i^{\max} \log_2 r_i^{\max} + 1) i \}, \tag{6}$$

где p_i^{\max} – максимальное количество записей в группе;

r_i^{\max} – максимальное количество полей в записи группы.

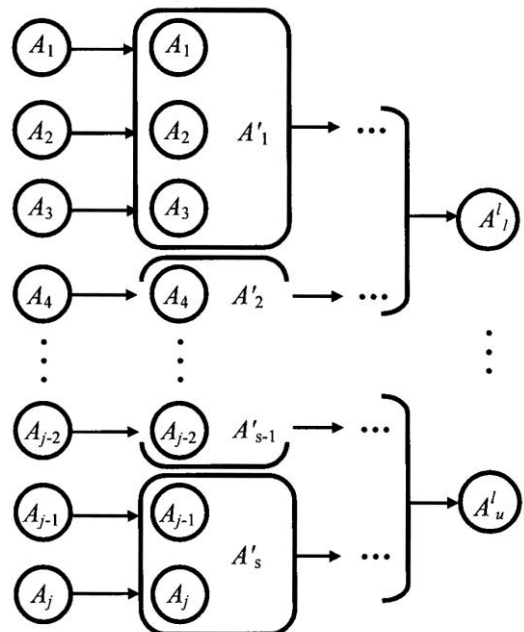
В этом случае максимум n достигается не обязательно при минимуме количества классификационных групп. Таким образом, для заданной совокупности объектов описания существует оптимальный с информационной точки зрения способ классификации с максимальным значением n .

В общем случае процедуру классификации выполняют на нескольких уровнях, что соответствует построению иерархической системы классификации. На каждом уровне степень понижения информационного объема составляет n_q , общая степень понижения информационного объема

$$n = \prod_{q=1}^l n_q, \tag{7}$$

где l – количество уровней иерархии классификационной системы.

Представив произвольную l -уровневую систему классификации в виде графа (рис. 2), получим очевидное равенство



$$N = N_1 N_2 \times \dots \times N_n, \quad (8)$$

из которого следует, что при одинаковой n больший эффект дают системы с наименьшим количеством уровней. Рис. 2

Так как количество уровней классификации зависит от степени понижения размерности информации на начальных уровнях, более информационно эффективны системы с максимальной n на начальных уровнях.

Классификационные системы не являются единственным механизмом сокращения количества элементов описания. Один из альтернативных механизмов – системы искусственного интеллекта, чаще всего реализуемые в форме экспертных систем (ЭС) как совокупность баз данных и знаний, дополненных интерпретатором. Для ЭС продукционного типа число решений для «дерева» произвольного вида

$$N = 2^z \{1 + \sum [(y_v - b_v)/(k_v + s_v)]_{v=1, z}\} = 2^z \{1 + \sum [(y_v - b_v)/2^v]_{v=1, z}\}, \quad (9)$$

где z – суммарное количество шагов вывода;

v – номер шага;

k_v – количество логических связей v -го уровня типа «если x , то y »;

s_v – количество логических связей v -го уровня типа «иначе»;

$k_v = s_v = 1$;

y_v – количество связей v -го уровня типа «или»;

b_v – количество связей v -го уровня типа «и».

Степень понижения информационного объема на v -м шаге вывода

$$n_v = 2 + (y_v - b_v)/n_{v-1}, \quad (10)$$

где n_{v-1} – степень понижения информационного объема на $(v-1)$ -м шаге вывода.

Количество элементов массива данных, необходимых для вывода, для дерева решений произвольного вида определим по формуле

$$N = n_0(1 + n_1(1 + n_2(1 + \dots + (n_v)\dots)). \quad (11)$$

Из формулы (11) видно, что с точки зрения размера информационного массива более выгодно задавать структуру дерева решений с максимальной степенью понижения в шагах с большими номерами.

Дерево решений в ЭС рассматриваемого типа может быть приведено к эквивалентному бинарному дереву решений путем добавления вершин и ребер. Для него количество элементов массива данных, необходимых для вывода, определим по формуле

$$N' = N + y - b, \quad (12)$$

где N – количество вершин исходного дерева;

y – общее на дерево количество связей типа «или»;

b – общее на дерево количество связей типа «и».

Следовательно, при прочих равных условиях ЭС с преобладанием дизъюнктивных операций более эффективна в смысле объема хранимой информации, чем ЭС с конъюнкцией или ЭС на бинарных деревьях.

Для ЭС фреймового типа соотношения (9) – (12) сохраняются, если вместо количества шагов подставить количество уровней вложенности исходного и всех ссылочных фреймов, а выражение $(y_v - b_v)$ заменить на $(f_v + d_v - 2)$, где f_v – количество атрибутов исходного фрейма, d_v – количество атрибутов всех ссылочных фреймов.

Для ЭС в виде простой семантической сети

$$N = ae, \quad (13)$$

где a – общее число понятий сети;

e – общее число отношений между понятиями.

В семантической сети, вершины (понятия) которой представлены фреймами или продукциями, в формуле (13) следует вместо a подставить выражения, соответствующие зависимостям (9) – (12).

Необходимо отметить, что представленная на рис. 2 система классификации отражает хотя и наиболее распространенный, но частный случай, когда группы не включают общих записей и полей. При использовании общих записей и полей зависимость (6) преобразуют к виду

$$n = A p_j \log_2 r_j \{ (p_i^{\max} \log_2 r_i^{\max} + 1 - p_i^{\text{com}} \log_2 r_i^{\text{com}}) i \}, \quad (14)$$

где p_i^{com} – количество общих записей в группе;

r_i^{com} – количество общих полей в записи группы.

Очевидно, что при некотором усложнении структуры взаимосвязей между элементами описания общие поля и записи позволяют еще в большей степени понизить объем информации. Методы, основанные на разделении общих данных, широко применяют в объектных моделях.

Для объектно-реляционных моделей вполне применима зависимость (13), где для a выполняется (14). Аналогично для других комбинированных моделей в силу мультипликативности получены формулы для определения размерности их описания.

Таким образом, получены формулы для оценки различных методов повышения управляемости информационного фонда в технологических процессах и производственных системах, которые могут быть использованы при анализе, синтезе и управлении большими информационными массивами.

СПИСОК ЛИТЕРАТУРЫ

1. *Дмитриев В.* Прикладная теория информации / В. Дмитриев. – М.: Высш. шк., 1989.

2. *Шеннон К.* Математическая теория связи / К. Шеннон // Работы по теории информации и кибернетике. – М.: Иностран. лит-ра, 1963.

Севмашвтуз

Поступила 16.02.04

M.P. Khudyakov

Analysis of Presentation Technique Efficiency for Informational Structures in Technological Process and System Control

Information presentation modes used for designing and control over technological processes are analyzed. Dependences for their assessment according to the degree of information volume reduction are obtained based on the entropy approach.
