

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

УДК 004.942

АНАЛИЗ АЛГОРИТМОВ ПРЕОБРАЗОВАНИЯ ИНФОРМАЦИИ В ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ

Г.В. Басалова

Рассмотрен подход к анализу алгоритмов преобразования информации в информационно-вычислительных системах с использованием математического аппарата сетей Петри. Предлагаемый подход позволяет выполнять моделирование и анализ алгоритмов обработки информации в разрабатываемой или уже существующей вычислительной системе. При этом анализ может производиться с помощью вычислительной техники, так как все его этапы строго формализованы.

Ключевые слова: алгоритмы обработки информации, моделирование вычислительных систем, сети Петри.

Одним из основных этапов разработки любых, в том числе и специализированных, вычислительных систем являются изучение и анализ процессов обработки информации в проектируемой системе. В процессе изучения и анализа должны быть уточнены структура вычислительной системы, входные, промежуточные и выходные данные, процедуры обработки информации и последовательность их реализации, основные характеристики специализированной вычислительной системы и ограничения на ее реализацию [1].

При анализе процессов преобразования информации используются информационные модели, предназначенные для отражения информационных связей между объектами. Особенность такого рода моделей заключается в их графическом представлении, но при этом имеется возможность матричного или аналитического способа их изображения. Информационные модели отражают информационные потоки между различными объектами, отношения между ними, содержат идентификаторы объектов, объемные, временные, частотные и другие характеристики как самих объектов, так и входящих и исходящих потоков данных, а также последовательность выполнения расчетов.

В основу описываемого подхода положен принцип последовательного эквивалентного преобразования матричных моделей систем обработки данных в зависимости от этапа анализа системы и необходимости получения требуемых характеристик.

Исходной информацией для моделирования и анализа технологии обработки данных является состав информационных ресурсов (ИР) по каждой задаче, отношения предшествования между ними, области их определения, множество используемых констант, структурированные матрицы смежности и достижимости, определяющие технологию обработки данных.

На основе анализа матриц смежности и достижимости ИР строится информационных граф для каждой задачи. Анализ информационного графа и его подграфов позволяет определить состав процедур, необходимых для решения задачи. Объединения множества процедур и множества ИР по множеству задач определяют полное множество ИР и процедур заданного множества задач обработки данных. Путем проведения специальных операций “выравнивания” и “наложения” графов отдельных задач формируется общий интегрированный граф технологии решения задач обработки данных, матрица смежности которого является результатом логического сложения матриц смежности анализируемых задач.

Интегрированный граф содержит общие и специфические части технологии решения рассматриваемого множества задач. Исходными данными для анализа, систематизации и формирования требований к схеме обработки данных является информация о парных отношениях между наборами информационных ресурсов системы обработки данных, формализуемая в виде матрицы смежности.

Рассмотрим матричные и графовые модели описания исходной информации и процедур ее обработки. Для описания алгоритмов решения задач обработки данных используют два стандартных блока: функциональный блок и блок принятия решений (рис. 1).

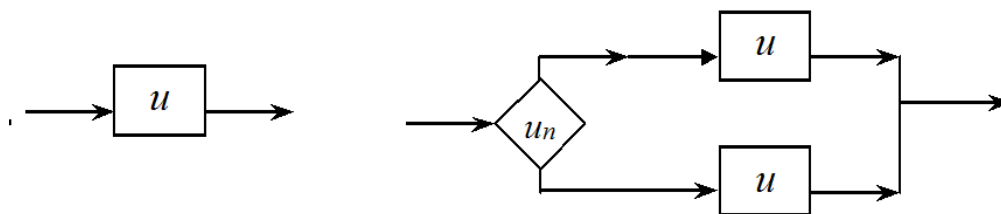


Рис. 1. Блоки для построения моделей обработки информации

Функциональный блок u представляет собой отдельную процедуру обработки данных, под которой понимается некоторое преобразование одних информационных элементов в другие безотносительно к средствам, на которых реализуется это преобразование. Основным требованием при вы-

делении функциональных блоков является одинаковый уровень абстракции и детализации при анализе отдельных задач обработки данных. Под блоком u_n принятия решений понимают процедуру или операцию логического сравнения поступающих входных данных с заданными. В результате операции сравнения управление передается одному (u_1) или другому (u_2) функциональному блоку. С помощью этих блоков описывают циклические участки в алгоритмах обработки данных.

Информационное обеспечение алгоритма обработки данных составляет множество типов ИР, обрабатываемых процедурами алгоритма. В качестве ИР в зависимости от уровня абстракции рассматриваются базы данных, массивы, блоки, записи и т.д. Процедуры обработки данных, входящие в функциональный блок, можно рассматривать как преобразования множества входных либо промежуточных ИР в множество промежуточных либо выходных ИР задачи [2].

Степень абстракции процедур обработки информационных элементов зависит от необходимого уровня детализации описания задач обработки данных при их анализе.

Представим процессы преобразования и обработки информации в вычислительной системе с помощью сети Петри.

Как известно, сеть Петри определяется множеством состояний и переходов, а также входными и выходными функциями переходов. Состояния сети будем обозначать кружками, а переходы черточками (барьерами). Дуги соответствуют функциям, связывающим множества состояний и переходов. Входная функция переходов определяет для каждого перехода множество его входных состояний, а выходная – множество его выходных состояний.

При использовании сетей Петри вводится понятие маркера, под которым понимается метка готовности запуска перехода по каждому из его входных состояний. Наличие маркера будем обозначать точкой в кружочке, соответствующем состоянию. Число точек соответствуют числу маркеров в каждом состоянии.

Переход может сработать, если в каждом его входном состоянии имеется хотя бы один маркер. Размещение маркеров по вершинам-состояниям сети Петри называется ее разметкой.

Таким образом, сеть Петри – это набор $N = (P, T, F, H, M_0)$, где

- P – конечное множество состояний;
- T – конечное множество переходов;
- $F: P \times T \Rightarrow \{0, 1\}$, $H: T \times P \Rightarrow \{0, 1\}$ – функции инцидентности;
- $M_0: P \Rightarrow \{0, 1, 2, \dots\}$ – начальная разметка сети.

Графически сеть Петри представляется в виде ориентированного графа. Вершину-состояние p и вершину-переход t соединяют дугой (p, t) , если $F(p, t)=1$ и дугой (t, p) , если $H(t, p)=1$.

Вершины-состояния помечаются целыми неотрицательными числами или соответствующим числом маркерных точек.

Если все состояния сети обозначить последовательно символами $p_1, p_2, p_3, \dots, p_n$, то разметку всех состояний сети можно представить в виде n -мерного вектора M , координаты которого равны числу маркерных точек в соответствующих состояниях.

Функционирование сети заключается в переходе от одной разметки к другой. Смена разметок происходит в результате срабатывания переходов. Переход t может сработать при разметке M , если

$$M(p) - F(p, t) \geq 0, \forall p \in P.$$

Это означает, что каждое входное состояние перехода t помечено хотя бы одной маркерной точкой.

В результате срабатывания некоторого перехода t , удовлетворяющего условию (1) разметка M заменяется разметкой M' :

$$\forall p \in P: M'(p) = M(p) - F(p, t) + H(t, p),$$

то есть при срабатывании перехода из каждого его входного состояния удаляется и в каждое его выходное состояние добавляется одна маркерная

точка. Это обозначается: $M \xrightarrow{t} M'$.

Пусть $D = \{d_1, d_2, \dots, d_S\}$ – множество ИР специализированной вычислительной системы, где S – число элементов; $U = \{u_1, u_2, \dots, u_L\}$ – множество процедур обработки данных.

Поставим в соответствие каждому элементу $d_i \in D$ вершину-состояние p_i сети. Каждому элементу $u_j \in U$ поставим в соответствие переход t_j сети.

В соответствии со взаимосвязью информационных ресурсов и процедур обработки соединим дугами элементы множеств P и T . Элемент $p_i \in P$ соединяется с элементом $t_j \in T$ дугой (p_i, t_j) , если информационный ресурс d_i является входным элементом процедуры u_j и дугой (t_j, p_i) – если выходным. Так как информационный ресурс может являться входным для нескольких процедур, то для восстановления маркерных точек состояния p_i после срабатывания перехода t_j необходимо также построить дуги (t_j, p_i) для таких t_j и p_i , для которых существует дуга (p_i, t_j) .

Под матрицей смежности ИР M_0 понимают квадратную бинарную матрицу, проиндексированную по обеим осям множеством информационных элементов и содержащую запись 1 в позиции (i, j) $i, j = \overline{1, S}$, если между информационными ресурсами d_i и d_j существует отношение R такое, что для получения информационного элемента d_j непосредственно необходимо обращение к ИР d_i .

Считается также (для удобства), что каждый элемент достижим из самого себя, то есть $d_i R d_j$, $i = j$. Наличие такого отношения будет обозначать как $d_i R d_j$, а его отсутствие как $d_i \bar{R} d_j$, чему соответствует запись 0 в позиции (i, j) матрицы M_0 .

Такая формализация позволяет однозначно идентифицировать входные и выходные ИР в алгоритме обработки данных [3].

Матрице смежности M_0 ставится в соответствие граф информационных взаимосвязей $G(D, M_0)$, множеством вершин которого является множество ИР, а дуга (d_i, d_j) соответствует записи 1 в позиции (i, j) в матрице смежности. Взаимосвязь между процедурами обработки, наборами входных и промежуточных данных удобно представлять с помощью таблицы инцидентности, которая представляет собой таблицу вида $T = \|t_{ls}\|$, $l = \overline{1, L}$, $s = \overline{1, S}$, где

$$t_{ls} = \begin{cases} +1, & \text{если } d_s \text{ является входным элементом процедуры } u_l, \\ 0, & \text{если } d_s \text{ не используется процедурой } u_l, \\ -1, & \text{если } d_s \text{ является исходным элементом процедуры } u_l. \end{cases}$$

В таблице T_0 каждая строка отображает процедуру обработки данных, а каждый столбец – использование всеми процедурами рассматриваемого ИР, то есть в строке содержится информация о множестве входных и выходных данных, связанных с анализируемой процедурой.

В s -м столбце таблицы допускается больше одной позиции (l, s) со значением -1, если существуют альтернативные варианты получения соответствующего ИР. В этом случае их число должно совпадать с общим числом процедур, имеющих в пересечении с s -м столбцом значение -1.

Рассмотрим графовое представление двух основных видов взаимосвязи между элементами таблицы T_0 . Выделим два случая. Пусть столбцу информационного ресурса d_s соответствует единственная процедура u_l со значением -1, а входными элементами процедуры u_l являются d_k, d_p, d_m , что обозначим в виде $u_l(d_s) = \{d_k, d_p, d_m\}$ (рис. 2). Этот случай соответствует единственному варианту получения d_s .

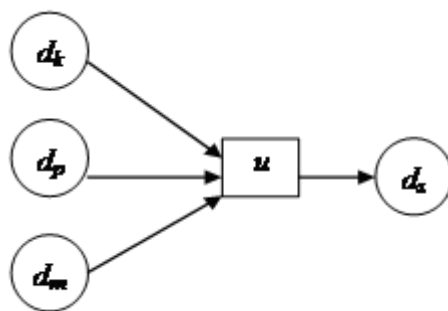


Рис.2. Пример графа для единственного варианта получения информационного ресурса d_s

Пусть столбцу ИР d_s соответствуют процедуры u_l и u_j со значением -1, то есть:

$$u_l(d_s) = \{d_k, d_p\}, \quad u_j(d_s) = \{d_k, d_m\}.$$

Этот случай соответствует наличию двух альтернативных вариантов получения информационного элемента d_s (рис. 3). В первом случае ИР получается с помощью процедуры u_l , входными ИЭ которой являются d_k, d_p во втором – с помощью другой процедуры, имеющей входы d_k, d_m .

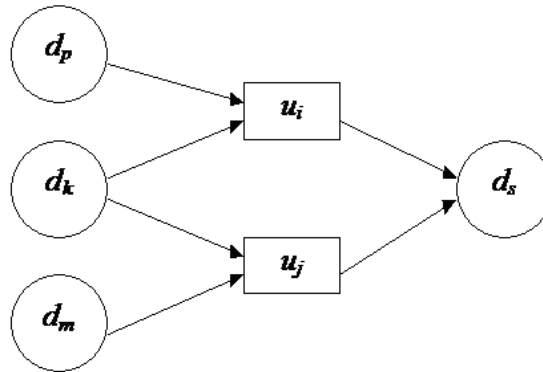


Рис.3. Пример графа при наличии альтернативных вариантов получения информационного ресурса d_s

Для проведения основных процедур анализа задач обработки данных требуется получить структурированную матрицу смежности и построить соответствующие им графы. Для формирования структурированной матрицы смежности необходимо преобразовать исходные данные таким образом, чтобы выявить уровни обработки, последовательность используемых процедур и т.д.

Вначале осуществляют переупорядочение ИР в матрице смежности по уровням их обработки. С этой целью используют матрицу достижимости ИР. Под матрицей достижимости B_0 понимается квадратная бинарная матрица, проиндексированная одинаковым образом по обеим осям множеством информационных ресурсов $D = \{d_s\}, s = \overline{1, S}$. Запись 1 или 0 в каждой позиции (i, j) матрицы достижимости соответствует наличию или отсутствию для всех упорядоченных пар ИР (d_i, d_j) отношения достижимости R , обладающего свойством транзитивности.

Информационный ресурс d_j достижим из информационного ресурса d_i ($d_i R d_j$), если на графе информационных взаимосвязей можно указать направленный путь от вершины d_i к вершине d_j , либо $d_i = d_j$, то есть если для получения информационного ресурса d_j используется ИР d_i .

Матрица достижимости определяется с использованием матрицы смежности T_0 и свойства транзитивности отношения достижимости. При этом заданной матрице достижимости может соответствовать некоторое

множество матриц смежности T_0 , любая из которых имеет одну и ту же матрицу достижимости B_0 . Следовательно, имеется множество графов информационных взаимосвязей, любой из которых содержит необходимую информацию для построения матрицы достижимости.

Процедура выделения уровней обработки с использованием матрицы достижимости B_0 состоит в следующем. Информационный ресурс $d_i \in D$, принадлежит множеству элементов верхнего уровня матрицы достижимости, если $R(d_i) \cap A(d_i) = R(d_i)$, где $R(d_i)$ и $A(d_i)$ – соответственно множество предшествования и достижимости информационного элемента d_i .

С использованием матрицы достижимости производится итеративное разделение множества ИР на подмножества в соответствии с уровнями их обработки.

Упорядоченному множеству ИР соответствует структурированный граф информационных взаимосвязей, ИР которого разделены на различные уровни.

Обозначим структурированную матрицу смежности T_{0C} . В этой матрице информационные ресурсы разделены на подмножества в соответствии с уровнями их обработки. Информационные ресурсы, столбцы которых в матрице T_{0C} не содержат единиц (нулевые столбцы) являются входными элементами задач обработки данных, а информационные ресурсы, соответствующие высшему уровню обработки L_1 , являются выходными ресурсами. Остальные ресурсы являются промежуточными.

Для полного анализа необходимо получение взаимосвязи между информационными ресурсами и процедурами в алгоритме обработки данных. Поэтому на втором этапе анализа, используя информацию, содержащуюся в структурированной матрице смежности и таблице инцидентности задач обработки данных, строится структурированная матрица смежности технологии задач обработки данных [4]. Она отражает существующую взаимосвязь между информационными ресурсами и процедурами в общем алгоритме обработки данных. В этой матрице процедуры упорядочены по уровням их использования, а информационные ресурсы – по уровням их обработки.

Таким образом, рассмотренный метод позволяет выполнять моделирование и анализ алгоритмов обработки информации в разрабатываемой или уже существующей специализированной вычислительной системе. При этом анализ может проводиться с помощью вычислительной техники, так как все его этапы строго формализованы.

Список литературы

1. Гудман С., Хидетниemi С. Введение в разработку и анализ алгоритмов / пер. с англ. М.: Мир, 1981. 366 с.

2. Типизация разработки модульных систем обработки данных / Мамиконов А.Г., Кульба В.В., Косяченко С.А. М.: Наука, 1989. 165 с.

3. Мамиконов А.Г., Кульба В.В. Синтез оптимальных модульных систем обработки данных. М.: Наука, 1986. 275 с.

4. Методы и модели анализа данных: OPLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. СПб.: БХВ-Петербург, 2004.

Басалова Галина Валерьевна, канд. техн. наук, доц., galina_basalova@mail.ru, Россия, Тула, Тульский государственный университет

ANALYSIS OF TRANSFORMATION ALGORITHMS IN INFORMATION-COMPUTING SYSTEMS

G.V. Basalova

The article considers approaches to the analysis of data processing algorithms in computing systems using mathematical apparatus of Petri nets. The proposed approach allows performing simulations and analysis of information processing algorithms in a developed or existing computer system. The analysis can be performed with the help of computer technology, because it is strictly formalized.

Key words: data processing algorithms, modeling of computer systems, Petri nets.

Basalova Galina Valerievna, candidate of technical sciences, docent, galina_basalova@mail.ru, Russia, Tula, Tula State University