УДК 519.683

В.Г. АКУЛОВСКИЙ

НЕКОТОРЫЕ АСПЕКТЫ ПРЕОБРАЗОВАНИЯ АЛГОРИТМОВ НА ОСНОВЕ ФОРМАЛИЗАЦИИ ИНФОРМАЦИОННЫХ СВЯЗЕЙ

Ключевые слова: *алгебра алгоритмов, регулярные схемы, формализация данных, информационные связи, преобразование алгоритмов.*

Введение. В настоящее время признана принципиально важная роль данных, образующих информационные связи в алгоритмах и программах, [1], однако в современных средствах разработки (например, в таком известном средстве как UML [2]) потоки управления и информационные связи в алгоритмах рассматриваются независимо. В результате не обеспечивается полнота описания алгоритмов и возникает вероятность появления ошибок, вызванных несогласованностью спецификаций этих аспектов алгоритмизации, причем используются графические, т.е. недостаточно формализованные средства описания алгоритмов.

Вместе с тем согласованное рассмотрение потоков управления и информационных связей в алгоритмах позволяет не только устранить (по крайней мере, ослабить) отмеченные недостатки, но и, специфицировав информационные связи в алгоритмах, осуществить их оптимизирующие преобразования. Отметим, что в рамках алгебры алгоритмов эти вопросы ранее не рассматривались.

Сформулируем свойства алгоритмов, на достижение которых направлены полученные результаты. Наиболее качественным (читабельным, легко отлаживаемым и сопровождаемым), очевидно, является алгоритм, содержащий минимальное число коротких связей. Реальные алгоритмы такими свойствами обладают достаточно редко.

В связи с этим цель данной работы — формализация (в рамках алгебраического аппарата) и спецификация информационных связей в алгоритме и доказательство возможности некоторых оптимизирующих преобразований алгоритмов. При этом ограничимся тем, что покажем возможность сокращения длины информационных связей в результате преобразования алгоритмов.

Исходные определения. В качестве формального аппарата используем расширенную алгебру алгоритмов (САА-Р) [3], в которой введено понятие взаимосвязи операторов и данных, операторы определены в виде $(D)A(D')$, что позволило специфицировать входные D и выходные D' данные для каждого оператора. В качестве метода разработки использован метод структурного проектирования программ (МСПП) [4], в рамках которого алгоритмы записываются в виде регулярных схем (РС). В контексте данной работы РС определим следующим образом.

Определение 1. РС будем называть выражение вида

$$(D)A(D')=(D_1)A_1(D'_1)*(D_2)A_2(D'_2)*\ldots*(D_i)A_i(D'_i)*\ldots*(D_n)A_n(D'_n), \quad (1)$$

где $(D)A(D')$ — исходный оператор, любой оператор $(D_i)A_i(D'_i)$, результирующий нижний индекс — порядковый номер оператора. В данном выражении результирующие операторы выполняются последовательно и их суммарная функциональность адекватна функциональности исходного оператора.

Переходя к формализации информационных связей, введем с учетом определения 1 понятие связанных операторов.

Определение 2. Операторы $(D_i)A_i(D'_i)$ и $(D_j)A_j(D'_j)$ (где $i<j$), входящие в РС (1), связаны, если для них выполняется соотношение $D'_i \cap D_j = \varnothing$, т.е. некоторое подмножество выходных данных оператора $(D_i)A_i(D'_i)$ поступает на вход оператора $(D_j)A_j(D'_j)$. Будем говорить, что множество данных ${}_i\vec{D}_j = D'_i \cap D_j$ (${}_i\vec{D}_j \subseteq D'_i$ и ${}_i\vec{D}_j \subseteq D_i$) связывает операторы A_i и A_j (на что указывают используемые индексы) и эти данные будем называть связывающими.

Из определения 2 следует, что данные могут передаваться от любого оператора к одному или нескольким следующим операторам, т.е. слева направо и именно связывающие данные образуют информационные связи в РС. При этом в общем случае некоторый оператор может быть связан со всеми следующими за ним и всеми предшествующими ему операторами.

Используя определения 1 и 2, специфицируем в РС (1) информационные связи. Детализируем входные и выходные данные операторов и дополним систему обозначений. Данные на входе j-го оператора, связывающие его с k-м, обозначим ${}_k\overleftarrow{D}_j$ (где k — «адрес источника», j — «адрес приемника» данных), а на выходе, связывающие его с p-м, — ${}_j\vec{D}_p$ (где j — «адрес источника», p — «адрес приемника» данных).

Перепишем РС (1) с учетом введенных обозначений:

$$\begin{gathered}(D)A(D')=(D_1)A_1(D'_1,{}_1\vec{D}_2,{}_1\vec{D}_3,\ldots,{}_1\vec{D}_j,\ldots,{}_1\vec{D}_n)*\\ *(D_2,{}_1\overleftarrow{D}_2)A_2(D'_2,{}_2\vec{D}_3,\ldots,{}_2\vec{D}_j,\ldots,{}_2\vec{D}_n)*\ldots\\ \ldots*(D_j,{}_1\overleftarrow{D}_j,\ldots,{}_l\overleftarrow{D}_j,\ldots,{}_{j-1}\overleftarrow{D}_j)A_j(D'_j,{}_j\vec{D}_{j+1},\ldots,{}_j\vec{D}_p,\ldots,{}_n\vec{D}_n)*\ldots\\ \ldots*(D_n,{}_1\overleftarrow{D}_n,\ldots,{}_j\overleftarrow{D}_n,\ldots,{}_{n-1}\overleftarrow{D}_n)A_n(D'_n).\end{gathered} \quad (2)$$

В построенной РС не только специфицированы данные и связи между операторами, но и все «источники» и «приемники» данных поставлены в однозначное соответствие, т.е. любому ${}_l\vec{D}_p$ соответствует ${}_l\overleftarrow{D}_p$. Поскольку в РС, как правило, не все операторы связаны друг с другом, любое из множеств ${}_l\vec{D}_p$ может быть пустым и соответственно пустым будет и множество ${}_l\overleftarrow{D}_p$.

На основе специфицированных связей решаем задачу оптимизирующего преобразования алгоритмов.

Преобразование алгоритмов. Начнем с определения понятия связей между операторами в РС (2).

Определение 3. Множество $S_j^{\text{л}} = {}_1\overleftarrow{D}_j, {}_2\overleftarrow{D}_j,\ldots,{}_l\overleftarrow{D}_j,\ldots,{}_{j-1}\overleftarrow{D}_j$ назовем множеством левых связей j-го оператора, а множество $S_j^{\text{пр}} = {}_j\vec{D}_{j+1},{}_j\vec{D}_{j+2},\ldots$ $\ldots,{}_j\vec{D}_p,\ldots,{}_j\vec{D}_n$ — множеством его правых связей и будем различать три типа операторов:

• связанные, у которых $S_j^{\text{л}} \neq \varnothing$ и $S_j^{\text{пр}} \neq \varnothing$;

• связанные слева (справа), у которых $S_j^{\text{л}} \neq \varnothing$, а $S_j^{\text{пр}} = \varnothing$ ($S_j^{\text{пр}} \neq \varnothing$, а $S_j^{\text{л}} = \varnothing$);

• не связанные, у которых $S_j^{\text{л}} \neq \varnothing$ и $S_j^{\text{пр}} = \varnothing$.

Множество $S_j^{\text{o}} = {}_1\vec{D}_{j+1}, \ldots, {}_1\vec{D}_n, {}_2\vec{D}_{j+1}, \ldots, {}_2\vec{D}_n, \ldots, {}_{j-1}\vec{D}_{j+1}, \ldots, {}_{j-1}\vec{D}_n$ назовем множеством связей, охватывающих (огибающих) j-й оператор. Количество огибающих связей оператора A_j определим как $K_j^{\text{o}} = |S_j^{\text{o}}|$. Любое из множеств ${}_j\vec{D}_l$ (${}_r\vec{D}_s$), входящих в множества $S_j^{\text{л}}$, $S_j^{\text{пр}}$, S_j^{o} , может быть пустым и в этом случае оно из рассмотрения исключается. Отсюда следует, что множества $S_j^{\text{л}}$, $S_j^{\text{пр}}$, S_j^{o} также могут быть пустыми.

Теперь введем и определим понятие длины информационных связей.

Определение 4. Длину произвольной связи между операторами A_i и A_j определим для случая $i < j$ как ${}_i d_j = j - i$, а для случая $j < i$ — как ${}_i d_j = i - j$.

Исходя из определений 3 и 4, общую (суммарную) длину левых (правых) связей оператора A_k запишем $L_k^{\text{л}} = \sum_{p=1}^{k-1} {}_p d_k$ ($L_k^{\text{пр}} = \sum_{p=k+1}^{n} {}_k d_p$), а общую длину огибающих k-й оператор связей — в виде $L_k^{\text{o}} = \sum_{l=1}^{k-1} \sum_{p=k+1}^{n} {}_l d_r$. Во всех случаях любой из элементов ${}_i d_j$, входящих в приведенные выражения (в соответствии с определением 3), может быть равен нулю.

Учитывая, что кроме рассмотренных связей существуют и другие их виды, которые в [5] названы обратными, приведем определения локально независимых операторов.

Определение 5. Локально независимыми назовем любые два не связанных (см. определение 2) оператора $(D)A(D') * (\hat{D})B(\hat{D}')$, если для них выполняются следующие ограничения: $D \cap \hat{D} = \varnothing$, $D \cap \hat{D}' = \varnothing$, $D' \cap \hat{D}' = \varnothing$, т.е. эти операторы не имеют никаких информационных связей между собой. В результате для них выполняется тождественное соотношение $(D)A(D') * (\hat{D})B(\hat{D}') = (\hat{D})B(\hat{D}') * (D)A(D')$, т.е. операторы могут перемещаться в рамках РС.

Далее, обобщая определения 5, введем понятие области независимости.

Определение 6. Областью независимости слева от оператора $(D_i)A_i(D_i')$, входящего в РС (1), назовем непрерывную последовательность операторов $(D_m)A_m(D_m') * \ldots * (D_{i-1})A_{i-1}(D_{i-1}')$, где оператор не связан с данным (см. определение 2) и предшествует ему (см. определение 1) и для каждого выполняются соотношения из определения 5. В рассматриваемом (более общем) случае эти соотношения представляют

$$\left(\bigcup_{p=m}^{i-1} D_P\right) \cap D_i = \varnothing, \quad \left(\bigcup_{p=m}^{i-1} D_P\right) \cap D_i' = \varnothing, \quad \left(\bigcup_{p=m}^{i-1} D_P'\right) \cap D_i' = \varnothing, \tag{3}$$

где $m \geq i-1$. Область независимости справа от оператора $(D_i)A_i(D_i')$ определяется аналогично и формально записывается в виде соотношений

$$D_i \cap \left(\bigcup_{p=i+1}^{r} D_P\right) = \varnothing, \quad D_i \cap \left(\bigcup_{p=i+1}^{r} D_P'\right) = \varnothing, \quad D_i' \cap \left(\bigcup_{p=i+1}^{r} D_P'\right) = \varnothing, \tag{4}$$

где $r \geq i+1$. Будем говорить, что оператор $(D_i)A_i(D_i')$ обладает областью независимости, если для него выполняются соотношения (3) и/или (4).

Исходя из определений 5 и 6, будем утверждать, что каждый оператор в своей области независимости не имеет никаких информационных связей и может быть перемещен в рамках этой области.

Докажем, что перемещение операторов в рамках их зоны независимости может обеспечить сокращение длины информационных связей в РС.

Теорема. В любой РС, содержащей операторы, обладающие областями независимости, длина информационных связей сократится в результате перемещения этих операторов в рамках их областей независимости при выполнении следующих условий. Если некоторый оператор A_i в результате перемещения и перенумерации операторов становится j-м (A_j), то в зависимости от его типа (см. определение 3) длина информационных связей в РС сократится при условии выполнения соотношений:

1) для не связанных операторов при перемещении в произвольном направлении $L_j^{\text{о}} < \hat{L}_i^{\text{о}}$, при этом перемещение 1- и n-го (последнего) операторов к сокращению связей не приведет;

2) для связанных слева (справа) операторов при перемещения их влево (вправо) $L_j^{\text{о}} < \hat{L}_i^{\text{о}} + \Delta_j^{\text{л}}$ ($L_j^{\text{о}} < \hat{L}_i^{\text{о}} + \Delta_j^{\text{пр}}$), а для 1- и n-го соответственно $L_j^{\text{о}} < \Delta_1^{\text{пр}}$ и $L_j^{\text{о}} < \Delta_n^{\text{л}}$;

3) для связанных операторов при перемещения влево (вправо) $L_j^{\text{о}} + \overline{\Delta}_j^{\text{пр}} < \hat{L}_i^{\text{о}} + \Delta_j^{\text{л}}$ ($L_j^{\text{о}} + \overline{\Delta}_j^{\text{л}} < \hat{L}_i^{\text{о}} + \Delta_j^{\text{пр}}$).

Во всех приведенных случаях $L_j^{\text{о}}$ — длина связей, огибающих оператор A_j (в новом месте расположения), а $\hat{L}_j^{\text{о}}$ — длина этих связей в исходном месте расположения, после удаления оператора A_i, $\Delta_j^{\text{пр}}$ ($\Delta_j^{\text{л}}$) — величина, на которую сократится общая длина правых (левых) связей, $\overline{\Delta}_j^{\text{пр}}$ ($\overline{\Delta}_j^{\text{л}}$) — прирост общей длины правых (левых) связей в результате перемещения оператора в его новое место расположения.

Доказательство. При перемещении операторов в их области независимости на изменение длины связей в РС влияют три (и только три) фактора: сокращение общей длины огибающих связей в месте удаления оператора и ее увеличение в месте вставки оператора, сокращение левых (правых) связей при перемещении оператора влево (вправо), удлинение левых (правых) связей при перемещении оператора вправо (влево). Никаким другим изменениям информационные связи не подвергаются.

Первый фактор. Разобьем его на два этапа.

Этап 1. Во всех случаях, когда некоторый оператор A_i извлекается из последовательности операторов, образующих его область независимости, а операторы, входящие в РС, перенумеровываются, длина каждой отдельной, огибающей этот оператор связи, сократится на единицу. Это следует из определения 4 и того факта, что все индексы в диапазоне от $i+1$ до n уменьшатся на единицу. Суммарную длину огибающих связей, полученную после извлечения i-го оператора, запишем $\hat{L}_i^{\text{о}} = L_i^{\text{о}} - K_i^{\text{о}}$ (где $L_i^{\text{о}}$ — исходная длина этих связей, а $K_i^{\text{о}}$ — их количество).

Этап 2. Во всех случаях, когда извлеченный ранее оператор A_i вставляется в последовательность операторов, образующих его область независимости, и после перенумерации получает номер j (A_j), увеличится на 1 каждая отдельная исходная связь, которая после вставки оператора A_j будет его охватывать. Это следует из определения 4 и того факта, что все индексы в диапазоне от $j+1$ до $n-1$ увеличатся на 1. Суммарную длину огибающих связей, полученную после вставки j-го оператора, запишем $L_j^{\text{о}} = \tilde{L}_j^{\text{о}} + K_j^{\text{о}}$ (где $\tilde{L}_j^{\text{о}}$ — исходная длина связей, а $K_j^{\text{о}}$ — их количество).

Влияние первого фактора на общую длину связей в РС (увеличение или сокращение) зависит от количества огибающих связей в исходном и результирующих местах размещения оператора, т.е. от выбора места, в которое перемещаем оператор A_i.

Второй фактор. В результате перемещения оператора влево (вправо) длина его левых (правых) связей со всеми предшествующими (последующими) операторами сократится на величину перемещения, что докажем, основываясь на определении 4.

При перемещении оператора A_i влево на j-е место (вправо на p-е место), где $j < i$ ($p > i$) — длина некоторой левой (правой) связи между операторами A_j (A_p) и A_k (A_l), где $k < j$ ($1 > p$), будет определяться соотношением ${}_k d_j = {}_k d_i - {}_i d_j$ (${}_p d_l = {}_i d_l - {}_i d_p$). Общая длина левых (правых) связей оператора A_j (A_p) в этом случае будет определяться соотношением $L_j^{\text{л}} = L_i^{\text{л}} - \sum_{m=i}^{j-1} {}_m d_j$ ($L_p^{\text{пр}} = L_i^{\text{пр}} - \sum_{m=i}^{p-1} {}_m d_p$) и она, очевидно, меньше левых (правых) связей исходного оператора. Поэтому величину, на которую сократится суммарная длина левых (правых) связей, запишем $\Delta^{\text{л}} = L_i^{\text{л}} - L_j^{\text{л}}$ ($\Delta^{\text{пр}} = L_i^{\text{пр}} - L_p^{\text{пр}}$).

Третий фактор. При перемещении оператора A_i влево на j-е место (вправо на p-е место), где $j < i$ ($p > i$) — длина связи между операторами A_j (A_p) и A_r (A_s), где $r > i$ ($s < i$), будет определяться соотношением ${}_j d_r = {}_i d_r + {}_j d_i$ (${}_p d_s = {}_i d_s + {}_p d_i$), т.е. в обоих случаях связи удлиняются на величину перемещения. Общая длина левых (правых) связей оператора A_j (A_p) в этом случае будет определяться соотношением $L_j^{\text{л}} = L_i^{\text{л}} - \sum_{m=i}^{j-1} {}_m d_j$ ($L_p^{\text{пр}} = L_i^{\text{пр}} + \sum_{m=i}^{p-1} {}_m d_p$) и она, очевидно, больше длины левых (правых) связей исходного оператора. Отсюда прирост общей длины левых (правых) связей запишем $\overline{\Delta}^{\text{л}} = L_j^{\text{л}} - L_i^{\text{л}}$ ($\overline{\Delta}^{\text{пр}} = L_p^{\text{пр}} - L_i^{\text{пр}}$).

Первый и последний операторы РС согласно определениям 3, 4 и выражению (2) обладают такими свойствами. У первого A_1 (последнего A_n) оператора РС множества $S_1^{\text{л}}$ и $S_1^{\text{о}}$ ($S_n^{\text{пр}}$ и $S_n^{\text{о}}$) пусты и соответственно $L_1^{\text{л}} = 0$, $L_1^{\text{о}} = 0$ ($L_n^{\text{пр}} = 0$ и $L_n^{\text{о}} = 0$).

Подводя итог изложенному, сделаем следующие выводы. Первый фактор влияет на длину связей во всех случаях, но в случае независимых операторов он единственный, что и отражено в соотношении 1, причем перемещение первого и последнего операторов в связи с приведенными свойствами $L_1^{\text{о}} = 0$ и $L_n^{\text{о}} = 0$ (т.е. ни одна связь их не охватывает) к сокращению длины связей не приведет.

Наряду с первым фактором для связанных слева (справа) операторов в соотношении 2 учтено влияние второго фактора, а для связанных — влияние второго и третьего в соотношении 3. При этом первый и последний операторы не могут быть связанными в соответствии со свойствами $L_1^{\text{л}} = 0$ и $L_n^{\text{пр}} = 0$ и в качестве таковых не рассматриваются. В случае связанных слева (справа) операторов из свойства $L_1^{\text{о}} = 0$ ($L_n^{\text{о}} = 0$) следует свойство $\hat{L}_1^{\text{о}} = 0$ ($\hat{L}_n^{\text{о}} = 0$), что и ведет к трансформации соотношения 2 к соответствующим частным случаям.

Поскольку в приведенных соотношениях влияние всех факторов учтено для каждого типа операторов, то, очевидно, выбор нового места расположения операторов (если оно существует) с учетом этих соотношений неизбежно приведет к сокращению суммарной длины информационных связей в РС. Теорема, таким образом, доказана.

Прежде чем завершить данную работу, сделаем следующее замечание. Хотя возможность перемещения операторов ограничена областями независимости и, таким образом, исключено их взаимное влияние, на такое перемещение следует наложить еще одно достаточно сильное ограничение. Оно связано с необходимостью обеспечения требуемой алгоритмом последовательности действий, например последовательности ввода–вывода данных, выполняемых операторами. Однако пока это ограничение не формализовано, контроль за его удовлетворением оставляем за разработчиком.

Заключение. Поставленная в работе цель достигнута, поскольку показана возможность реализации (в рамках алгебраического аппарата) оптимизирующих преобразования алгоритмов, в частности сокращения длины информационных связей на основе их формализации. Дальнейшая формализация информационных связей целесообразна для расширения возможностей преобразования алгоритмов, в частности рассмотренных последовательных алгоритмов в параллельные (квазипараллельные).

СПИСОК ЛИТЕРАТУРЫ

1. Турский В. Методология программирования. — М.: Мир, 1981. — 264 с.
2. Леоненков А.В. Самоучитель UML. — СПб.: БХВ-Петербург, 2001. — 304 с.
3. Акуловский В.Г. Формализация взаимосвязей операторов и данных в рамках расширенной алгебры алгоритмов // Кибернетика и системный анализ. — 2008. — № 6. — С. 170–182.
4. Ющенко Е.Л., Цейтлин Г.Е., Грицай В.П., Терзян Т.К. Многоуровневое структурное проектирование программ: Теоретические основы, инструментарий. — М.: Финансы и статистика, 1989. — 208 с.
5. Акуловский В.Г. Некоторые подходы к контролю и преобразованию алгоритмов на основе анализа специфицируемых данных // Проблеми програмування. — 2008. — № 4. — С. 84–93.

Поступила 07.07.2009