Выпуск 22 (39) ОПТИМИЗАЦИЯ 1978 г.
Теория и машинная реализация массовых вычислений

УДК 681.3.06

ИСПОЛЬЗОВАНИЕ ЛОГИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ВЕДЕНИЯ ИНФОРМАЦИОННЫХ МАССИВОВ ЗАДАЧ ОПТИМАЛЬНОГО ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ

А.Б.Венгеров

Современное развитие экономики требует создания и ведения моделей оптимального планирования и управления, отражающих реальную схему разработки плановых заданий в разрезе предприятий, отраслей и народного хозяйства в целом. При достаточно большой размерности модели задача ведения массивов используемой информации превращается в сложную проблему.

Возникающие трудности связаны с необходимостью неоднократного хранения и корректировки больших массивов информации, используемых в задачах оптимального планирования, при пакетной обработке и в диалоговом режиме.

Для облегчения этих трудностей можно использовать схему обработки информации на основе единой информационной модели.

В соответствии с этой схемой, описания структур данных, используемых совместно с программами поиска оптимальных решений, корректировки массивов, пакетной и диалоговой обработки, хранятся вместе с самими данными.

Особенностью этих структур является наличие двух уровней: логического и физического. Такое разделение позволяет достигнуть независимости физических структур хранения информации, учитывающих специфику вычислительных устройств, системы управления данными и пр., от порядка поиска информации по логическим структурам, фиксированного в прикладных программах.

в данной схеме различные задачи "видят" свою информацию через соответствующие части модели - подсхемы. Схема позволяет провести развязку всех задач в концепции "схема-подсхема". При этом различные изменения программ оптимизации, пакетной обработки и диалога не оказывают взаимного влияния друг на друга, что является условием устойчивости прикладных программ и баз данных.

При реализации подобной схемы обработки информации проектирование базы данных можно представить как трехэтапный процесс, где сначала некоторым способом оценивается исследуемый объект, затем создается его логическая модель и, наконец, эта модель воплощается в конкретных вычислительных устройствах.

Для простоты описания объекта, а также для удобства создания логической модели существующие информационные связи представляются в виде функциональных зависимостей.

Если в отношении $R \subseteq D_i \times D_j$ для каждого $d_i \in D_i$ однозначно определяется $d_j \in D_j$, говорят, что D_j функционально зависит от D_i ($D_i \longrightarrow D_j$).

Учитывая рекомендации Кодда [1], задачу построения логической модели можно представить следующим образом. Задано множество $\mathcal{F}$ функциональных зависимостей между элементами (атрибутами) моделируемого объекта. Необходимо синтезировать схему базы данных, представляющую собой множество схем связей, где схема связи $R_i(A_1, \ldots, A_k)$ является k-арным отношением между атрибутами $A_1, \ldots, A_k$.

Рассмотрим следующие свойства функциональных зависимостей [2]:

$X \longrightarrow X$ (рефлексивность); (1)

если $X \longrightarrow Z$, то $X, Y \longrightarrow Z$, где Y - любой атрибут (дополнительность); (2)

если $X \longrightarrow Y$ и $Y, Z \longrightarrow W$, то $X, Z \longrightarrow W$ (псевдотранзитивность). (3)

Для множества $\mathcal{F}$ определим $\mathcal{F}^+$ как замыкание функционального бинарного отношения R на множестве различных совокупностей атрибутов относительно операций (1)-(3).

Функциональная зависимость $f_i \in \mathcal{F}$ называется избыточной, если

$$\mathcal{F}^+ = (\mathcal{F} / f_i)^+. \quad (4)$$

Обозначим через $\mathcal{F}_M$ минимальное покрытие $\mathcal{F}$, т.е. $\mathcal{F}^+ = \mathcal{F}_M^+$ и $\mathcal{F}_M$ не содержит избыточных функциональных зависимостей.

Во время синтезирования логической модели встает необходимость найти $\mathcal{F}_M$ как основу для построения схемы базы данных. Надо заметить, что в общем случае $\mathcal{F}$ может иметь несколько $\mathcal{F}_M^i$ $(i=1,\ldots,p)$, каждому из которых однозначно соответствует логическая модель базы данных.

Для выбора оптимальной относительно некоторых критериев модели необходимо получить и проанализировать все множество вариантов моделей, а следовательно, и все минимальные покрытия.

Введем понятие Θ-контура как неизбыточного множества бисвязных в смысле функциональных зависимостей элементов и докажем следующую теорему.

ТЕОРЕМА I. *Для того чтобы множество $\mathcal{F}$ функциональных зависимостей имело неединственное минимальное покрытие $\mathcal{F}_M^i$, необходимо и достаточно наличие хотя бы одного θ-контура $(\theta \subset \mathcal{F})$.*

ДОКАЗАТЕЛЬСТВО. Обозначим $M_1 \Rightarrow M_2$ как факт выводимости функциональной зависимости $M_1 \rightarrow M_2$ из множества $\mathcal{F}$ после ряда применений операций (I)-(3).

I. Допустим, что последовательным удалением избыточных зависимостей получим из $\mathcal{F}^+$ некоторое $\mathcal{F}_M^i$. Тогда для получения $\mathcal{F}_M^j (j \neq i)$ недостаточно только добавлять к $\mathcal{F}_M^i$ какие-то функциональные зависимости, так как, удалив их, получим $\mathcal{F}_M^i$, а значит, $\mathcal{F}_M^j$ не минимально.

2. Пусть какими-то удалениями и добавлениями функциональных зависимостей мы получили из $\mathcal{F}_M^i$ некоторое $\mathcal{F}_M^j (i \neq j)$. При этом должно соблюдаться условие

$$(\mathcal{F}_M^i)^+ = (\mathcal{F}_M^j)^+. \qquad (5)$$

3. Для выполнения условия (5) необходимо, чтобы для каждой удаленной из $\mathcal{F}_M^i$ функциональной зависимости $f_m: M_m \rightarrow A_m$, где A_m - одиночный атрибут, M_m - совокупность атрибутов или одиночный атрибут $(m=1,\ldots,K)$, были добавлены одна или несколько функциональных зависимостей, образующих цепочку $M_m \Rightarrow A_m$, поскольку в силу минимальности $\mathcal{F}_m^i$, такой це-

почки нет. Пусть среди этих добавленных зависимостей будет $f_n: M_n \to A_n$. Тогда, в силу ориентированности цепи $M_m \Rightarrow A_m$, будет иметь место цепь

$$M_m \Rightarrow M_n \to A_n \Rightarrow A_m. \qquad (6)$$

4. Для выполнения условия (5) необходимо также, чтобы любая добавленная функциональная зависимость, например f_n, существовала в $\mathcal{F}_M^i$ и исчезла после удаления f_m. То есть цепь $M_n \Rightarrow A_n$ в $\mathcal{F}_M^i$ должна включить f_m, что означает наличие следующей цепи:

$$M_n \Rightarrow M_m \to A_m \Rightarrow A_n. \qquad (7)$$

5. Поскольку $(M_n \neq M_m) \wedge (A_m \neq A_n)$, то из (6) и (7) получаем хотя бы одну пару бисвязных элементов $(M_n, M_m) \vee (A_m, A_n)$, т.е. θ- контур, что и требовалось доказать.

Таким образом, на основании теоремы I задачу анализа и получения множества M минимальных покрытий $\mathcal{F}_M^i$ $(i=1,\dots,M)$ мы свели к получению и анализу θ-контуров.

В силу возможности различных перестановок пар элементов θ-контура получаем множество его состояний S.

Определим степень P θ-контура как количество входящих в него элементов.

Легко видеть, что два θ-контура, имеющих хотя бы один общий элемент, объединяются в один θ-контур. Тогда, фиксируя состояния всех минимальных θ-контуров, принадлежащих множеству $\mathcal{F}$, получим, что количество различных минимальных покрытий множества $\mathcal{F}$ зависит от степеней этих θ-контуров и количества функциональных зависимостей, использующих один элемент какого-либо θ-контура, принадлежащего $\mathcal{F}$.

Обозначим через f_i^n и $f_i^\wedge$ соответственно правый и левый элементы бинарного функционального отношения f_i и введем понятие нормализованной правой части функциональной зависимости $f_i^{\prime n}$ как правой части, полученной добавлением $f_i^\wedge$ к f_i^n на основании свойства (I) функциональных зависимостей.

Докажем следующую теорему.

ТЕОРЕМА 2. Если $(f_i^{\prime n} \supseteq f_j^\wedge) \wedge (f_j^{\prime n} \supseteq f_i^\wedge)$, то $f_i^\wedge \rightleftarrows f_j^\wedge (i,j=1,\dots,K)$, где $f_k^{\prime n} \supset f_k^\wedge$ $(k=1,\dots,K)$.

ДОКАЗАТЕЛЬСТВО. I. Применяя (I) к левым частям f_i и f_j, всегда можно получить нормализованные правые части $f_i^{\prime n}$ и $f_j^{\prime n}$, где $f_k^{\prime\prime} \supset f_k^{\wedge} \quad (k=1,\ldots,K)$.

2. Нормализовав правые части и используя условия теоремы 2, запишем

$$f_i^{\wedge} \longrightarrow f_j^{\wedge} \cup f_i^{\wedge} \cup L_i, \qquad (8)$$

$$f_j^{\wedge} \longrightarrow f_i^{\wedge} \cup f_j^{\wedge} \cup L_j, \qquad (9)$$

где все элементы функциональных зависимостей суть множества имен атрибутов.

Из (8) и (9) получаем $f_i^{\wedge} \longrightarrow f_j^{\wedge}; f_j^{\wedge} \longrightarrow f_i^{\wedge}$, откуда $f_i^{\wedge} \rightleftarrows f_j^{\wedge}$, что и требовалось доказать.

СЛЕДСТВИЕ I. Для того чтобы $f_i^{\wedge} \rightleftarrows f_j^{\wedge}$, необходимо и достаточно, чтобы

$$f_i^{\prime\prime} \cap f_j^{\prime\prime} \supseteq f_i^{\wedge} \cap f_j^{\wedge} \quad (i,j=1,\ldots,K).$$

Таким образом, для отыскания θ-контуров множества функциональных зависимостей можно использовать следующий алгоритм:

I) нормализация правых частей функциональных зависимостей;

2) нахождение непустых объединений правых частей M_δ (блоков);

3) нахождение внутри каждого блока f_i и f_j таких, что $f_i^{\wedge} \cup f_j^{\wedge} \subseteq M_\delta$, и образование из них θ-контуров.

Обычно для удаления избыточных функциональных зависимостей и получения множества M минимальных покрытий выполняется многократный перебор $\mathcal{F}^+$. Используя изложенный подход, можно, вместо этого, рассматривать лишь θ-контуры и пересекающиеся с ними функциональные зависимости, что позволяет резко сократить объем вычислений и упростить анализ логических моделей баз данных.

ЛИТЕРАТУРА

1. GODD E.F. Normalized data base structure: a brief tutorial. IBM Report RJ 935, Nov., 3, 1971.

2. DELOBEL C. CASEY R.G. Decomposition of a data base and the theory of boolean switching functions. - "IBM J. Res. Develop.", v.17, N 5, p.374-386.

Поступила в ред.-изд. отдел
I5.УI.I978 г.