



УДК 004.6+57.087.1

Анализ вычислительной сложности методов декомпозиции OLAP-гиперкубов многомерных данных

Ахрем А. А.¹, Носов А. П.^{1,*}, Рахманкулов В. З.¹, Южанин К. В.¹

¹Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия

* nosov@isa.ru

В работе исследуются проблемы редукции (декомпозиции) моделей многомерных данных в виде гиперкубовых OLAP-структур. Проведено исследование вычислительной сложности декомпозиционных методов анализа многомерных гиперкубов полиномиально-логарифмической и полиномиальной степеней сложности. Найдена точная верхняя граница уменьшения сложности декомпозиционных методов анализа исходного OLAP-гиперкуба данных по отношению к недекомпозиционным, и на их основе доказаны критерии эффективного применения редукционных методов анализа гиперкубовых структур по сравнению с традиционными нередукционными методами.

Ключевые слова: OLAP-системы; гиперкуб многомерных данных; методы редукции; вычислительная сложность; вычислительная производительность

Представлена в редакцию: 17.06.2020, исправлена: 15.08.2020.

Введение

В настоящее время становится все более актуальной разработка методов редукции моделей сложных систем, т. е. приведение таких моделей к более простому виду, например, к декомпозиции исходной модели на модели меньшей размерности [1, 2, 3].

Проблемы редукции моделей и анализа вычислительной сложности алгоритмов традиционно интересуют исследователей-математиков и компьютерных специалистов при практической реализации методов моделирования [4, 5, 6, 7, 8, 9, 10]. В частности, проблема снижения размерности признакового пространства в задачах ранжирования и многокритериальной классификации исследовались в работах [4, 5], в которых критерии агрегируют с помощью экспертов, что упрощает порядковую классификацию многокритериальных альтернатив в статическом пространстве оценок. Аналогичный подход представлен в работе [6], когда при решении задач кластеризации исходный набор переменных заменяется новым

набором обобщённых критериев, учитывающих связи переменных. OLAP обработка данных не допускает изменения размерности пространства, однако, поскольку вычислительная сложность OLAP-данных определяется не размерностью гиперкуба, а размером решетки подкубов, которая задает на нем структуру данных, становится возможной декомпозиция на подкубовых структурах.

С увеличением объемов данных падает производительность вычислений кубовых структур. Методы редукции больших кубов данных на подкубы с меньшими объемами позволяют решать проблему снижения производительности вычислений. Решая задачи редукции OLAP-кубов, необходимо исследовать влияние выбранного метода редукции на вычислительную сложность задач и целостность обобщенных результатов.

Аналогичные проблемы возникают при обработке и анализе массивов информации гиперкубовых структур аналитических OLAP-систем, относящихся к классу BigData, или сверхбольших компьютерных систем многомерных данных [4, 5, 6].

В работах авторов [7, 8, 9, 10] по математическим методам анализа многомерных данных аналитических OLAP-систем рассматриваются задачи редукции для случаев, когда агрегирование критериев уже определено решёткой куба, а декомпозиция куба на меньшие по размерности кубы нужна для снижения времени вычисления полной решётки при динамическом изменении данных в кубе. В указанных работах были исследованы методы редукции моделей многомерных данных в виде OLAP-гиперкубов, способствующих уменьшению вычислительной сложности решения задач с большими и сверхбольшими исходными данными в классе задач полиномиальной степени сложности. Доказаны необходимые и достаточные условия эффективного применения методов редукции задач анализа гиперкубов данных по сравнению с традиционными нередукционными методами их решения. Приведены примеры методов декомпозиции кубовых структур, как уменьшающих, так и увеличивающих вычислительную сложность по сравнению с вычислениями по полной модели.

Цель настоящей работы состоит в нахождении условий уменьшения вычислительной сложности решения задач анализа данных редукционными методами, получении точных количественных границ уменьшения сложности декомпозиционных методов из класса полиномиальной степеней сложности, установлении характера зависимости вычислительной производительности от структурных свойств гиперкуба и определении количественных границ вычислительной производительности решения декомпозиционных задач агрегирования данных.

1. Критерии эффективности редукционных методов решения задач анализа многомерных OLAP-кубов данных

В настоящем разделе статьи будем предполагать, что задан гиперкуб KH_m , имеющий $m > 1$ размерностей. Пусть на KH_m задана решетка L [11, 12], состоящая из n подкубов. В этом случае решетка L определяет гиперкубовую структуру данных, и такой гиперкуб бу-

дем обозначать $K_n H_m$. Пусть на этой структуре анализируются многомерные OLAP-данные из некоторого непустого множества $C(K_n H_m)$. Предположим, что методы решения задач анализа данных из $C(K_n H_m)$ допускают декомпозицию гиперкуба $K_n H_m$ на k ($2 \leq k \leq n$) непересекающихся подкубовых структур L_1, \dots, L_k , в каждую из которых входят n_1, \dots, n_k ($n_1 + \dots + n_k = n$) подкубов исходной решетки L . Допустим, что вычислительная сложность $f(n)$ метода решения задачи из множества $C(K_n H_m)$ принадлежит одному из следующих классов:

– $F^e(n) \triangleq \{f(n, a) = a^n\}$, где $a = \text{const} > 1$ (класс экспоненциальной вычислительной степени сложности);

– $F^p(n) \triangleq \{f(n, p) = n^p\}$, где p — натуральное число (класс полиномиальной вычислительной степени сложности);

– $F^{lp}(n) \triangleq \{f(n; a, p, q) = n^p (\log_a n)^q\}$, где q — целое число (класс полиномиально-логарифмической вычислительной степени сложности);

Классы вычислительной степени сложности определяют значения вычислительной сложности $f = f(n, \cdot)$ нередукционных (недекомпозиционных) методов на всей структуре данных гиперкуба, а вычислительная сложность f^R редукционных (декомпозиционных) методов составляется из вычислительных сложностей метода на подкубовых структурах $f_i = f(n_i, \cdot)$, при этом полагаем

$$f^R(n, \cdot) = \sum_{i=1}^k f(n_i, \cdot). \quad (1)$$

Критерий эффективности методов определяется путем сравнения вычислительных сложностей редукционного и нередукционного методов из одного класса, и эффективнее считается тот из методов, для которого вычислительная сложность меньше.

В работах авторов [8, 9, 10] для классов экспоненциальной и полиномиальной степени сложности были доказаны следующие критерии эффективности на основе уменьшения (увеличения) вычислительной сложности редукционных методов анализа OLAP-кубов многомерных данных по сравнению с нередукционными методами.

I. В работе [8] для методов экспоненциальной степени сложности показано, что при определенных условиях, а именно, при

$$a > 2; \quad n_i \geq 1, \quad f(n_i, a) = a^{n_i}, \quad i = 1, \dots, k; \quad k > 2. \quad (2)$$

вычислительная сложность редукционных методов анализа OLAP-гиперкубов многомерных данных уменьшается по сравнению с традиционными нередукционными методами, т.е. $f^R < f$, при всех $f, f_i \in F^e$, т.е. в этих условиях применение редукционных методов оправдано, так как их эффективность выше.

Отметим, что в случае, когда параметр a удовлетворяет неравенствам $1 < a \leq 2$ возможны ситуации, когда вычислительная сложность декомпозиционных методов анализа

OLAP-гиперкубов уменьшается, увеличивается или равняется вычислительной сложности традиционных недекомпозиционных методов.

Приведем соответствующие примеры.

Пример 1. Пусть $a = 3/2$, $n = 5$, $n_1 = 4$, $n_2 = 1$. В этом случае имеем:

$$a^5 = \left(\frac{3}{2}\right)^5 = \frac{243}{32}; \quad a^{n_1} = a^4 = \frac{81}{16} = \frac{162}{32}; \quad a^{n_2} = a^1 = \frac{3}{2} = \frac{48}{32}. \quad (3)$$

Принимая во внимание (3) получаем

$$f(5, 3/2) = \frac{243}{32} > \frac{162}{32} + \frac{48}{32} = \frac{210}{32} = f(4, 3/2) + f(1, 3/2) = f^R(5, 3/2), \quad (4)$$

т.е. вычислительная сложность традиционного метода решения задач анализа OLAP-кубов многомерных данных больше вычислительной сложности редукционного метода. Применение редукционного метода оправдано.

Пример 2. Пусть $a = 3/2$, $n = 3$, $n_1 = 2$, $n_2 = 1$. В этом случае будем иметь:

$$f(3, 3/2) = \left(\frac{3}{2}\right)^3 = \frac{27}{8}; \quad f(2, 3/2) = \frac{9}{4} = \frac{18}{8}; \quad f(1, 3/2) = a^1 = \frac{3}{2} = \frac{12}{8}. \quad (5)$$

Используя равенства (5) получаем

$$f(3, 3/2) = \frac{27}{8} < \frac{18}{8} + \frac{12}{8} = \frac{30}{8} = f(2, 3/2) + f(1, 3/2) = f^R(3, 3/2), \quad (6)$$

т.е. в этом случае вычислительная сложность традиционного метода анализа гиперкуба OLAP-данных меньше вычислительной сложности редукционного метода. Применение редукционного метода менее эффективно, и для решения задач анализа лучше использовать нередукционный метод.

Пример 3. Пусть $a = 2$, $n = 2$, $n_1 = n_2 = 1$. В данном случае получаем следующую оценку соотношений:

$$f(2, 2) = 2^2 = 4 = 2 + 2 = f(1, 2) + f(1, 2) = f^R(2, 2), \quad (7)$$

т.е. в рассматриваемом случае вычислительные сложности традиционного и декомпозиционного методов анализа OLAP-гиперкуба данных совпадают.

II. В работе [9] исследуются задачи анализа OLAP-гиперкубов данных с применением методов полиномиальной степени сложности. Показано, что при $p \geq 2$ вычислительная сложность редукционных методов уменьшается по сравнению с вычислительной сложностью нередукционных методов, т.е. $f^R < f$, при всех $f, f_i \in F^p$, и, следовательно, анализ данных, основанный на использовании подкубовых структур имеет преимущество над методами, в которых такая декомпозиция не учитывается.

Отметим, что при $p = 1$ ни один из методов не имеет преимуществ над другими, так как $f^R = f$ ввиду простого равенства $n^1 = n_1^1 + \dots + n_k^1$, справедливого при любом $k \geq 2$ в силу предположения о характере разбиения на подкубовые структуры (1).

Таким образом, критерий эффективности $f^R < f$ определяет преимущество редукционных методов анализа OLAP-гиперкубов данных над нередукционными методами при всех $p \geq 2$ в классе методов полиномиальной степени сложности.

2. Критерий эффективности редукционных методов в одном подклассе полиномиально-логарифмической степени сложности

В работе [10] для методов анализа OLAP-гиперкубов данных из класса полиномиально-логарифмической степени сложности $F^{lp}(n) \triangleq \{f(n; a, p, q) = n^p(\log_a n)^q\}$ показано, что при $p > 1$ и $q > 1$ выполняется критерий эффективности $f^R < f$, устанавливающий преимущество в данном классе редукционных методов над нередукционными.

Рассмотрим ситуацию, когда задача анализа OLAP-гиперкуба данных решается методами полиномиально-логарифмической степени сложности из рассматриваемого нами класса F^{lp} , где вычислительная сложность метода определяется формулой

$$f(n, a) = n(\log_a n)^{-1}, \quad a > 1, \quad (8)$$

т.е. подслучай [10], когда $p = 1$ и $q = -1$.

В этом случае справедлива следующая теорема.

Теорема 1. Для методов из подкласса полиномиально-логарифмической степени сложности F^{lp} вида (8) выполняется критерий эффективности:

$$f^R > f, \quad (9)$$

т.е. эффективность нередукционных методов в этом подклассе выше, чем редукционных.

Доказательство. Вычислительная сложность редукционного метода из класса F^{lp} на каждой из подкубовых структур OLAP-гиперкуба задается формулой

$$f(n_i, a) = n_i(\log_a n_i)^{-1} = \frac{n_i}{\log_a n_i}, \quad i = 1, \dots, k, \quad k \geq 2,$$

где $n = n_1 + \dots + n_k$, и, следовательно, при любом $2 \leq k \leq n$ для каждого i имеем $n_i < n$, откуда

$$\frac{n_1}{\log_a n_1} > \frac{n_1}{\log_a n}, \quad \dots, \quad \frac{n_k}{\log_a n_k} > \frac{n_k}{\log_a n}. \quad (10)$$

Складывая неравенства (10) находим, что

$$\frac{n_1}{\log_a n_1} + \dots + \frac{n_k}{\log_a n_k} > \frac{n_1 + \dots + n_k}{\log_a n} = \frac{n}{\log_a n}. \quad (11)$$

Принимая во внимание (11), получаем

$$f^R(n, a) = \sum_{i=1}^k f(n_i, a) > f(n, a),$$

что и требовалось доказать. Теорема 1 доказана.

Отметим, что приведенное доказательство распространяется на любой редукционный алгоритм из рассматриваемого подкласса, и, таким образом, доказанный в теореме 1 критерий эффективности $f^R > f$, определяет преимущество нередукционных методов анализа OLAP-гиперкубов данных над всеми редукционными методами в подклассе F^{lp} вида (8) методов полиномиально-логарифмической степени сложности.

3. Количественное сравнение эффективности методов в классе полиномиальной степени сложности

Приведенные в предыдущих разделах статьи критерии эффективности выявляют преимущество одних методов над другими (редукционных над нередукционными, и наоборот), но не дают ответа на вопрос: на сколько, или, точнее, во сколько раз одни методы эффективнее других.

В данном разделе статьи будет установлена количественная верхняя оценка уменьшения сложности редукционных методов анализа OLAP-кубов многомерных данных, принадлежащих классу F^p полиномиальной степени сложности, при декомпозиции исходного гиперкуба на два подкуба, по сравнению с нередукционными методами.

Для сравнения эффективностей редукционных и нередукционных методов одного класса вначале определим относительную эффективность методов следующим образом.

Определение 1. Относительной эффективностью редукционного метода $f^R \in F$ из класса сложности F по отношению к нередукционному методу $f \in F$ из того же класса сложности F называется отношение их вычислительных сложностей

$$E_F(f, f^R) = \frac{f}{f^R}. \quad (12)$$

Когда из контекста ясно о каких классах сложности идет речь будем обозначать относительную эффективность просто E .

З а м е ч а н и е 1. Таким образом, критерии эффективности редукционных методов по отношению к нередукционным методам в терминах относительной эффективности будут выглядеть следующим образом: $E > 1$ — редукционный метод эффективнее, $E < 1$ — эффективнее нередукционный метод, $E = 1$ — оба метода одинаково эффективны.

Итак, рассмотрим при $k = 2$ множество подкубовых структур $\mathcal{L}_2 \triangleq \{L_1, L_2\}$ разбиений гиперкуба $K_n H_m$ на два подкуба $K_{n_1} H_m$ и $K_{n_2} H_m$, состоящих соответственно из n_1 и n_2 ($n_1 + n_2 = n$) подкубов исходной решетки L . Скажем, что $\{n_1, n_2\} \in \mathcal{L}_2$, если $n_1 + n_2 = n$, т. е. множество \mathcal{L}_2 состоит из всевозможных таких пар, дающих в сумме n .

Вычислительная сложность методов из класса полиномиальной степени сложности задается следующими формулами:

- для нередукционного метода $f \in F^p$ имеем $f(n, p) = n^p$;
- для редукционного метода $f^R \in F^p$ при декомпозиции на две подкубовые структуры: $f^R(n, p) = f(n_1, p) + f(n_2, p) = n_1^p + n_2^p$, где $\{n_1, n_2\} \in \mathcal{L}_2$ (в этом случае скажем, что редукционные методы определены на множестве \mathcal{L}_2 и введем для них обозначение $f_{\mathcal{L}_2}^R(\{n_1, n_2\})$).

Заметим, что вычислительная сложность нередукционного метода фиксирована, в то время как вычислительная сложность редукционного метода зависит от разбиений на подкубовые структуры. Таким образом, наибольшая относительная эффективность редукционного метода по сравнению с нередукционным методом достигается тогда, когда найдено разбиение с минимальной вычислительной сложностью редукционного метода.

Определим верхнюю границу относительной эффективности (уменьшения вычислительной сложности) редукционного метода $f^R \in F^p$ из класса полиномиальной степени сложности, определенного на множестве \mathcal{L}_2 по отношению к нередукционному методу $f \in F^p$ из класса полиномиальной степени сложности как

$$\overline{E}_{F^p}(f, f^R) = \max_{\{n_1, n_2\} \in \mathcal{L}_2} E = \frac{f}{\min_{\{n_1, n_2\} \in \mathcal{L}_2} f_{\mathcal{L}_2}^R(\{n_1, n_2\})}. \quad (13)$$

Далее будем использовать обозначение просто \overline{E} , так как из контекста ясно, о чем идет речь.

Для редукционных методов на множестве \mathcal{L}_2 справедливо следующее утверждение.

Теорема 2. Пусть число n является четным — $n = 2m$. Тогда для любого $m > 1$ верхняя граница относительной эффективности достигается на \mathcal{L}_2 при $n_1 = n_2 = m$, причем

$$\overline{E} = 2^{p-1}, \quad (14)$$

и не зависит от m .

Доказательство. Относительная эффективность для любой пары $\{n_1, n_2\} \in \mathcal{L}_2$ определяется по формуле

$$E = \frac{n^p}{n_1^p + n_2^p}. \quad (15)$$

Пусть для определенности всегда $n_1 = \max\{n_1, n_2\}$. Положим $c = n_1(n_2)^{-1}$. Тогда $c \geq 1$ для любой пары $\{n_1, n_2\} \in \mathcal{L}_2$.

1. Рассмотрим вначале случай, когда степень p равна двум ($p = 2$). Для величины E в силу (15) в данном случае справедливы соотношения:

$$E = \frac{n^2}{n_1^2 + n_2^2} = \frac{(c+1)^2}{c^2 + 1} = \frac{c^2 + 2c + 1}{c^2 + 1} = 1 + \frac{2c}{c^2 + 1}. \quad (16)$$

Поскольку при любом $c \geq 1$ всегда $c^2 + 1 \geq 2c$, учитывая равенства (16), получаем, что

$$E \leq 2, \quad (17)$$

причем равенство $E = 2$ (E принимает максимальное значение) достигается в (17) в случае, когда $c = 1$, т.е. при $n_1 = n_2 = m$. Таким образом, $\overline{E} = 2^1 = 2^{2-1} = 2^{p-1}$, что и требовалось доказать.

2. Рассмотрим теперь случай произвольных натуральных чисел $p \geq 3$. Имеем в силу (15) и выбора $c \geq 1$

$$E^{-1} = \frac{n_1^p + n_2^p}{n^p} = \frac{c^p + 1}{(c+1)^p} = \left(\frac{c}{c+1}\right)^p + \frac{1}{(c+1)^p} = \left(1 - \frac{1}{c+1}\right)^p + \frac{1}{(c+1)^p}. \quad (18)$$

Положим $u = 1/(c+1)$. Учитывая (17), (18) находим, что

$$E^{-1} = (1-u)^p + u^p. \quad (19)$$

Используя (19), найдем первую производную

$$\frac{dE^{-1}}{du} = -p(1-u)^{p-1} + pu^{p-1}. \quad (20)$$

Учитывая (20), находим

$$\frac{dE^{-1}}{du} = 0 \text{ при } u = \frac{1}{2}. \quad (21)$$

Найдя вторую производную

$$\frac{d^2E^{-1}}{du^2} = p(p-1)(1-u)^{p-2} + p(p-1)u^{p-2}$$

видим, что при любом $p \geq 3$ она положительна:

$$\left. \frac{d^2E^{-1}}{du^2} \right|_{u=1/2} = 2p(p-1) \left(\frac{1}{2}\right)^{p-2} > 0,$$

следовательно, в точке $u = 1/2$ функция E^{-1} как функция от u достигает своего минимума, равного

$$E^{-1} \Big|_{u=1/2} = \left(\frac{1}{2}\right)^p + \left(\frac{1}{2}\right)^p = 2 \cdot 2^{-p} = 2^{-(p-1)}, \quad (22)$$

соответственно в этой же точке значение E максимально, откуда получаем, что

$$\bar{E} = 2^{p-1}.$$

Теперь, функция E^{-1} как функция от c достигает своего минимума (и, соответственно, значение E максимально), в силу выбора u , в точке $1/(c+1) = 1/2$, т.е. при $c = 1$.

Из (22) имеем

$$E^{-1}(1) = 2^{-(p-1)},$$

следовательно,

$$E(1) = 2^{(p-1)}, \quad (23)$$

и, так как в силу выбора, $c = 1$ при $n_1 = n_2$, и $n_1 + n_2 = n = 2m$, видим, что относительная эффективность E принимает значение (23) при $n_1 = m$, $n_2 = m$, т.е. для $\{m, m\} \in \mathcal{L}_2$, и это значение равно \bar{E} .

Таким образом установлено, что верхняя граница относительной эффективности $\bar{E} = 2^{(p-1)}$, и она достигается при $n_1 = n_2 = m$.

Теорема 2 полностью доказана.

Теорема 2 устанавливает точную верхнюю границу относительной эффективности (уменьшения вычислительной сложности) редукционных методов анализа OLAP-гиперкуба данных по отношению к нередукционным методам в классе полиномиальной степени сложности при декомпозиции исходного гиперкуба на подкубовую структуру из двух кубов. Теорема 2 утверждает, что в условиях теоремы редукционные методы эффективнее нередукционных методов не более, чем в 2^{p-1} раз, и их максимальная эффективность достигается

при декомпозиции гиперкуба на два подкуба одинаковой размерности, равной размерности гиперкуба, и у которых размерности структур данных совпадают с их геометрической размерностью. При этом максимально достижимая эффективность редуцированных методов по отношению к нередуцированными не зависит от размерности кубов, а зависит от степени полиномиальной сложности метода, и чем сложнее методы, тем большую эффективность дает применение редуцированных методов при такой декомпозиции данных.

Заключение

В настоящей работе исследованы математические методы анализа процессов редукиции больших OLAP-гиперкубов многомерных данных на подкубовые компоненты. Установлены критерии уменьшения (увеличения) вычислительной производительности решения задач редукиции гиперкубов данных по сравнению с традиционными нередукиционными методами анализа информации в аналитических OLAP-системах. Для специального подкласса методов редукиции полиномиальной степени сложности получены количественные верхние границы уменьшения их вычислительной сложности по сравнению с недекомпозиционными методами анализа OLAP-гиперкубов.

Представленные в данной работе результаты по анализу процессов редукиции гиперкубов многомерных данных используются в междисциплинарном проекте РФФИ, выполняемом совместно медиками и специалистами в области информационных технологий, использующих принципы и методы виртуального моделирования искусственных биологических органов на основе OLAP-моделей, где OLAP применяется при описании больших иерархий из множеств продуктовых наборов, используемых пациентами при приеме пищи. Кроме того, это важно при on-line мониторинге огромного количества пациентов (миллионы) при анализе региональных медицинских данных или данных по странам и континентам.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, грант 19-07-00686а.

Список литературы

1. Павловский Ю.Н., Смирнова Т.Г. Проблема декомпозиции в математическом моделировании. М: ФАЗИС, 1998. 272 с.
2. Ёлкин В.И. Редукиция нелинейных управляемых систем: декомпозиция и инвариантность по возмущениям. М: ФАЗИС, 2003. 208 с.
3. Цурков В.Н. Декомпозиция в задачах большой размерности. М.: Наука, 1981. 352 с.
4. Петровский А.Б., Лобанов В.Н. Многокритериальный выбор в пространстве признаков большой размерности: мультимедийная технология ПАКС-М // Искусственный интеллект и принятие решений. 2014. № 3. С. 92–104.

5. Agarwal S., Agrawal R., Deshpande P.M., Gupta A., Naughton J., Ramakrishnan R., Sarawagi S. On the computation of multidimensional aggregates // *Materialized views: techniques, implementations and applications* / Ed. by A. Gupta, I.S. Mumick. Camb.: MIT Press, 1999. Ch. 24. Pp. 506–521. DOI: [10.7551/mitpress/4472.003.0030](https://doi.org/10.7551/mitpress/4472.003.0030)
6. Чубукова И.А. *Data mining: учеб. пособие. 2-е изд.* М.: Бином. Лаборатория знаний, 2008. 382 с.
7. Макаров И.М., Рахманкулов В.З., Ахрем А.А., Ровкин И.О. Исследование свойств гиперкубовых структур в OLAP-системах // *Информационные технологии и вычислительные системы*. 2005. № 2. С. 4–9.
8. Ахрем А.А., Рахманкулов В.З., Южанин К.В. О сложности редукции моделей многомерных данных // *Искусственный интеллект и принятие решений*. 2016. № 4. С. 79–85.
9. Ахрем А.А., Рахманкулов В.З., Южанин К.В. Декомпозиционные методы анализа многомерных данных // *Системные исследования. Методологические проблемы: ежегодник 2015–2018*. Вып. 38. М., 2018. С. 88–97.
10. Ахрем А.А., Носов А.П., Рахманкулов В.З., Южанин К.В. Вычислительная производительность методов редукции гиперкубов многомерных данных аналитических OLAP-систем // *Искусственный интеллект и принятие решений*. 2019. № 4. С. 23–28.
11. Гуров С.И. *Булевы алгебры, упорядоченные множества, решётки. 2-е изд.* М.: URSS, 2013. 352 с.
12. Богомоллов А.М., Салий В.Н. *Алгебраические основы теории дискретных систем*. М.: Наука. ФИЗМАТЛИТ. 1997. 368 с.



Computational Complexity Analysis of Decomposition Methods of OLAP Hyper-cubes of Multidimensional Data

Akhrem A. A.¹, Nosov A. P.^{1,*}, Rakhmankulov V. Z.¹, Yuzhanin K. V.¹

Federal Research Center “Informatics and Control” of RAS, Moscow Russian Federation

*nosov@isa.ru

Keywords: OLAP systems, hypercube of multidimensional data, reduction methods, computational complexity, computational performance

Received: 17.06.2020, Revised: 15.08.2020.

The paper investigates the problems of reduction (decomposition) of multidimensional data models in the form of hypercube OLAP structures. OLAP data processing does not allow changes in the dimension of space. With the increase in data volumes, the productivity of computing cubic structures decreases. Methods for reducing large data cubes to sub-cubes with smaller volumes can solve the problem of reducing computing performance.

The reduction problems are considered for cases when the cube lattice has already determined criteria aggregation, and the cube decomposition into smaller cubes is needed to reduce the computation time of the full lattice when dynamically changing data in the cube.

The objective of the paper is to find conditions for reducing the computational complexity of solving data analysis problems by reduction methods, to obtain exact quantitative boundaries for reducing the complexity of decomposition methods from the class of polynomial degrees of complexity, to establish the nature of the dependence of computational performance on the structural properties of a hypercube, and to determine the quantitative boundaries of computational performance for solving decomposition problems of data aggregation .

The study of the computational complexity of decomposition methods for the analysis of multidimensional hyper-cubes of polynomial-logarithmic and polynomial degrees of complexity is carried out. An exact upper limit is found for reducing the complexity of decomposition methods for analyzing the initial OLAP data hypercube with respect to non-decomposition ones and based on them criteria are proved for the effective application of reduction methods for analyzing hypercube structures in comparison with traditional non-reduction methods.

Examples of decomposition methods of cube structures are presented, both reducing and increasing computational complexity in comparison with calculations using the full model.

The results obtained can be used in processing and analysis of information arrays of hypercube structures of analytical OLAP systems belonging to the BigData class, or ultra-large computer multidimensional data systems.

References

1. Pavlovskij Yu.N., Smirnova T.G. *Problema dekompozitsii v matematicheskom modelirovanii* [The problem of decomposition in mathematical modeling]. Moscow: FAZIS Publ., 1998. 272 p. (in Russian).
2. Elkin V.I. *Reduktsiia nelinejnykh upravliaemykh system: dekompozitsiia i invariantnost' po vozmushcheniiam* [Reduction of nonlinear controlled systems: decomposition and perturbation invariance]. Moscow: FAZIS Publ., 2003. 208 p. (in Russian).
3. Tsurkov V.N. *Dekompozitsiia v zadachakh bol'shoj razmernosti* [Decomposition in high-dimensional problems]. Moscow: Nauka Publ., 1981. 352 p. (in Russian).
4. Petrovsky A.B., Lobanov V.N. Multiple criteria choice in the attribute space of large dimension: multi-method technology PAKS-M. *Iskusstvennyj intellekt i priniatie reshenij* [Artificial Intelligence and Decision Making], 2014, no. 3, pp. 92–104 (in Russian).
5. Agarwal S., Agrawal R., Deshpande P.M., Gupta A., Naughton J., Ramakrishnan R., Sarawagi S. On the computation of multidimensional aggregates. *Materialized views: techniques, implementations and applications* / Ed. by A. Gupta, I.S. Mumick. Camb.: MIT Press, 1999. Ch. 24. Pp. 506–521. DOI: [10.7551/mitpress/4472.003.0030](https://doi.org/10.7551/mitpress/4472.003.0030)
6. Chubukova I.A. *Data mining* [Data mining]: a textbook. 2nd ed. Moscow: BINOM. Laboratoriia znanij Publ., 2008. 382 p. (in Russian).
7. Makarov I.M., Rakhmankulov V.Z., Akhrem A.A., Rovkin I.O. Investigation of properties of hypercube structures in OLAP systems. *Informatsionnye tekhnologii i vychislitelnye sistemy* [Information Technology and Computing Systems], 2005, no. 2, pp. 4–9 (in Russian).
8. Akhrem A.A., Rakhmankulov V.Z., Yuzhanin K.V. On complexity of the reduction of multidimensional data models. *Scientific and Technical Information Processing*, 2017, vol. 44, no. 6, pp. 406–411. DOI: [10.3103/S0147688217060028](https://doi.org/10.3103/S0147688217060028)
9. Akhrem A.A., Rakhmankulov V.Z., Yuzhanin K.V. Dekompozitsionnye metody analiza mnogomernykh dannykh [Decomposition methods for analyzing multidimensional data]. *Sistemnye issledovaniia. Metodologicheskie problemy: ezhegodnik 2015–2018. Vypusk 38* [System research. Methodological problems: Yearbook 2015–2018. No. 38]. Moscow, 2018. Pp. 88–97 (in Russian).

10. Akhrem A.A., Nosov A.P., Rakhmankulov V.Z., Yuzhanin K.V. Computational performance of hypercube reduction methods for multidimensional data of analytical OLAP system. *Iskusstvennyj intellekt i priniatie reshenij* [Artificial Intelligence and Decision Making], 2019, no. 4, pp. 23–28 (in Russian).
11. Gurov S.I. *Bulevy algebry, uporiadochennye mnozhestva, reshetki* [Boolean algebras, ordered sets, lattices]. 2nd ed. Moscow: URSS Publ., 2013. 352 p. (in Russian).
12. Bogomolov A.M., Salij V.N. *Algebraicheskie osnovy teorii diskretnykh sistem* [Algebraic foundations of the theory of discrete systems]. Moscow: Nauka. Fizmatlit Publ., 1997. 368 p. (in Russian).