УДК 681.3 : 004.82

# МЕТОД БИНАРНОЙ ДЕКОМПОЗИЦИИ ФУНКЦИОНАЛЬНЫХ ПРЕДИКАТОВ

А.А. Иванилов
(Харьковский национальный университет радиоэлектроники)

*Предложен метод бинарной декомпозиции функциональных предикатов на основе средств теории нормализации реляционных отношений, позволяющий обрабатывать формулу или отношение не целиком, а по частям.*

***бинарная декомпозиция, функциональный предикат***

**Введение.** Решение уравнений алгебры предикатов необходимо для моделирования различных функций интеллекта, в особенности для моделирования естественного языка [1, 2]. Декомпозиция уравнений важна как один из этапов процесса решения уравнений алгебры предикатов, поскольку в ряде случаев систему простых уравнений решать значительно легче, чем эквивалентное ей одно сложное уравнение. Декомпозиция уравнений также может служить мощным средством упрощения и сокращения записи уравнений алгебры предикатов [1].

Разработка методов декомпозиции ведется в работах [1, 3], но полностью эта задача ещё не решена. В работах [4, 5] изложена концепция логической сети, как устройства для решения систем уравнений алгебры предикатов с двумя переменными (бинарные уравнения). Для бинарной декомпозиции (что необходимо для построения логических сетей) пока существует лишь стандартный метод [3], который является довольно громоздким. Между тем, многие языковые модели, описанные с помощью алгебры предикатов [2, 6], представлены многоместными функциональными предикатами. Чтобы построить логические сети для этих моделей, необходимо выполнить их бинарную декомпозицию.

С другой стороны, между реляционной алгеброй [8] и кванторной алгеброй [7] существует глубокая аналогия: реляционные отношения и предикаты можно рассматривать практически как одно и то же, реляционная операция естественного соединения в точности соответствует операции конъюнкции предикатов и т. д. Теорию нормализации реляционных отношений можно рассматривать как теорию декомпозиции без потерь информации (т.е. как теорию эквивалентной декомпозиции), основанную на проекции как операторе декомпозиции и на естественном соединении как соответствующем операторе композиции. Отсюда наш интерес к теории нормализации отношений. Поэтому **целью данной работы** является разработка практического метода бинарной декомпозиции функциональных предикатов на основе средств теории нормализации реляционных отношений.

**Декомпозиция.** Рассмотрим уравнение P = 1. Пусть формула P выражается в виде

$$P = P_1 \wedge P_2 \wedge \ldots \wedge P_m, \tag{1}$$

где $P, P_1, P_2, \ldots, P_m$ – формулы алгебры предикатов (в дальнейшем просто предикаты). Очевидно, что исходное уравнение можно заменить эквивалентной ему системой уравнений $P_1 = 1$, $P_2 = 1$, , …, $P_m = 1$ тогда и только тогда, когда выполняется условие (1). Преобразование уравнения P = 1 в эквивалентную систему $P_1 = 1$, $P_2 = 1$, , …, $P_m = 1$ называется *декомпозицией* этого уравнения; обратный процесс – преобразование системы в одно уравнение – называется *композицией* уравнения.

Чтобы выполнить декомпозицию, надо исходный предикат P представить конъюнкцией более простых предикатов $P_1, P_2, \ldots, P_m$ (1); в этом смысле будем говорить о *декомпозиции предиката* P. Поэтому практически результатом декомпозиции можно считать не систему уравнений $P_1 = 1$, $P_2 = 1$, , …, $P_m = 1$, а совокупность предикатов $P_1, P_2, \ldots, P_m$. В данной работе «более простые» уравнения – это уравнения меньшей размерности, т.е. каждый из предикатов $P_1, P_2, \ldots, P_m$ зависит от меньшего числа переменных, чем P.

Декомпозиция вида (1) называется *эквивалентной*. Наряду с эквивалентной декомпозицией на практике часто встречается *неэквивалентная* декомпозиция [1, С. 84], когда в конечной системе уравнений присутствуют вспомогательные переменные [3]. В этом случае исходное уравнение не будет эквивалентно полученной системе уравнений. Исходный предикат связан с конечными предикатами соотношением

$$P = \exists y_1 \exists y_2 \ldots \exists y_k (P_1 \wedge P_2 \wedge \ldots \wedge P_m), \tag{2}$$

где $y_1, y_2, \ldots, y_k$ – вспомогательные переменные.

В связи с концепцией логических сетей возникла задача, когда исходное уравнение требуется представить системой, где размерность каждого уравнения равна двум. Как уравнения с двумя переменными, так и предикаты, зависящие от двух переменных, называются *бинарными*. При этом бинарные предикаты, фигурирующие в уравнениях системы, могут быть связаны с исходным предикатом равенством (1), либо (2). В этом случае будем говорить о *бинарной декомпозиции* уравнения (или предиката).

В соответствии с понятием конечного отношения определенного на доменах (как правило, бесконечных) под *конечным предикатом* будем понимать предикат на произвольной (конечной или бесконечной) области, но соответствующий конечному отношению на этой области. (Заметим, что такая точка зрения на конечный предикат отличается от изложенной в [4], где конечным предикатом называется предикат, заданный на конечной области). Предикат $P(x_1, x_2, \ldots, x_n)$, заданный на области $M_1 \times M_2 \times \ldots \times M_n$, по аналогии с заголовком реляционного отношения будем обозначать $P(x_1 : M_1,\ x_2 : M_2,\ \ldots,\ x_n : M_n)$. В дальнейшем, говоря о некотором предикате, всегда будем иметь в виду конечный предикат на области.

**Кванторное описание реляционных понятий и теорем.** Приведем описание связи между предикатами и реляционными отношениями, некоторыми предикатными и реляционными операциями, а так же сформулируем в терминах кванторной алгебры понятия функциональной, многозначной зависимостей и зависимости соединения, сформулируем основные теоремы теории нормализации. Именно благодаря этой связи можно применить нормализацию для «эквивалентной» декомпозиции предикатов.

Каждому предикату можно сопоставить реляционное отношение по следующему правилу. Пусть задан предикат $P(x_1 : M_1,\ x_2 : M_2,\ \ldots,\ x_n : M_n)$. Тогда *соответствующее* предикату P отношение R будет иметь заголовок $\{x_1 : M_1,\ x_2 : M_2,\ \ldots,\ x_n : M_n\}$ и тело, которое определяется следующим образом: отношение R содержит кортеж $\{x_1 : a_1,\ x_2 : a_2,\ \ldots,\ x_n : a_n\}$ тогда и только тогда, когда СДНФ предиката P содержит конституэнту единицы $x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$, где $a_1 \in M_1,\ a_2 \in M_2,\ \ldots,\ a_n \in M_n$.

Точно так же можно сопоставить каждому отношению свой предикат. Единственная разница между ними – это наличие порядка на множестве переменных у предиката. Но во многих практических задачах, например, при описании моделей естественного языка, порядок переменных в предикате не имеет значения и выбирается произвольно. Поэтому предикат алгебры предикатов и реляционное отношение практически можно считать изоморфными понятиями.

При описании декомпозиции отношений с целью получения нормальных форм используются две операции реляционной алгебры – операция проекции и операция естественного соединения. Первая используется как оператор декомпозиции, вторая – как оператор композиции отношений. Поэтому выразим эти две операции в терминах кванторной алгебры.

Операция естественного соединения отношений соответствует конъюнкции предикатов. Пусть отношениям $R_1$ и $R_2$ соответствуют предикаты $P_1$ и $P_2$. Тогда отношению $R_1$ JOIN $R_2$ соответствует предикат $P_1 \wedge P_2$.

Операция проекции отношения соответствует квантору существования от соответствующего предиката. Пусть отношению R с атрибутами

$\{x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n\}$ соответствует предикат $P(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$. Тогда отношению $R[y_1, y_2, \ldots, y_n]$ соответствует предикат

$$P[y_1, y_2, \ldots, y_n] = \exists x_1 \exists x_2 \ldots \exists x_m P(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n).$$

Обозначение операции проекции во многих случаях короче соответствующего ему кванторного выражения. Поэтому будем использовать эту операцию в кванторной алгебре и говорить, что предикат $P[y_1, y_2, \ldots, y_n]$ является проекцией предиката P по переменным $\{y_1, y_2, \ldots, y_n\}$.

Приведем определение функциональной зависимости (ФЗ). Пусть R – это отношение, а X и Y – произвольные подмножества атрибутов отношения R. Тогда Y *функционально зависимо* от X, что обозначается как $X \to Y$, тогда и только тогда, когда каждое значение множества X отношения R связано в точности с одним значением множества Y отношения R.

ФЗ $X \to Y$ в отношении R будет выполняться тогда и только тогда, когда соответствующий ему предикат P(X:S, Y:T) удовлетворяет *условию однозначности*

$$\forall A \in S\, \forall B_1, B_2 \in T\, (P(A, B_1) \wedge P(A, B_2) \supset D(B_1, B_2)),$$

где D(Y, Z) – предикат равенства элементов Y и Z. В этом случае будем говорить, что в предикате P(X, Y) выполняется ФЗ $X \to Y$.

Приведем определение многозначной зависимости (МЗ). Пусть X, Y и Z являются произвольными подмножествами множества атрибутов отношения R. Тогда X *многозначно зависит* от Z, что обозначается записью $Z \to\!\!> X$, тогда и только тогда, когда множество значений X, соответствующее заданной паре (значение Z, значение Y) отношения R, зависит только от Z, но не зависит от Y.

Доказано, что для данного отношения R{X, Y, Z} многозначная зависимость $Z \to\!\!> X$ выполняется тогда и только тогда, когда также выполняется многозначная зависимость $Z \to\!\!> Y$. Таким образом, многозначные зависимости всегда образуют связанные пары, что обозначается в виде $Z \to\!\!> X \mid Y$.

МЗ $Z \to\!\!> X \mid Y$ в отношении R{X, Y, Z} будет выполняться тогда и только тогда, когда соответствующий ему предикат P(X, Y, Z) можно выразить в виде

$$P(X, Y, Z) = \bigvee_{C \in U} Z^C P(X, Y, C),$$

где $P(X, Y, C) = A^C(X) \wedge B^C(Y)$ для каждого $C \in N$. В этом случае будем говорить, что в предикате P{X, Y, Z} выполняется МЗ $Z \to\!\!> X \mid Y$.

Приведем определение зависимости соединения (ЗС). Пусть R является отношением, а A, B, …, Z – произвольными подмножествами множества атрибутов отношения R. Отношение R удовлетворяет *зависимости* *соединения* *(A, B, …, Z) тогда и только тогда, когда оно равносильно соединению своих проекций с подмножествами атрибутов A, B, …, Z.

ЗС *(A, B, …, Z) в отношении R будет выполняться тогда и только тогда, когда соответствующий ему предикат P равносилен конъюнкции своих проекций с подмножествами переменных A, B, …, Z.

В этом случае будем говорить, что предикат P удовлетворяет зависимости соединения *(A, B, …, Z).

С учетом описанных понятий сформулируем две основные теоремы, которые являются основой теории нормализации.

*Теорема Хеза*. Если предикат P(X, Y, Z) удовлетворяет функциональной зависимости $X \to Y$, то P равен конъюнкции его проекций по переменным {X, Y} и {X, Z}.

*Теорема Фейгина*. Предикат P(X, Y, Z) равен соединению своих проекций по переменным {X, Y} и {X, Z} тогда и только тогда, когда для предиката P выполняется многозначная зависимость $X \to> Y \mid Z$.

Теорему Фейгина можно сформулировать также следующим образом.

Предикат P(X, Y, Z) удовлетворят зависимости соединения *({X, Y}, {X, Z}) тогда и только тогда, когда он удовлетворяет многозначной зависимости $X \to> Y \mid Z$.

**Метод бинарной декомпозиции функциональных предикатов.** Пусть задан предикат $P(x_1 : S_1, x_2 : S_2, \ldots, x_m : S_m, z_1 : T_1, z_2 : T_2, \ldots, z_n : T_n)$ с ФЗ $\{x_1, x_2, \ldots, x_m\} \to \{z_1, z_2, \ldots, z_n\}$, $m, n \in N$, $N = \{1, 2, \ldots\}$ – натуральный ряд. Необходимо выполнить бинарную декомпозицию предиката $P(x_1, x_2, \ldots, x_m, z_1, z_2, \ldots, z_n)$ за счет использования одной вспомогательной переменной.

Сначала рассмотрим наиболее общий случай: $m \geq 2$, n – любое, и предикат P не содержит никаких других зависимостей (функциональных, многозначных или зависимостей соединения).

Введем обозначения $\overline{x} = (x_1, x_2, \ldots, x_m)$ и $\overline{z} = (z_1, z_2, \ldots, z_n)$.

В соответствии со стандартным методом, описанном в [3], каждому кортежу $(\overline{x}, \overline{z})$ соответствующего предикату P отношения в произвольном порядке взаимно однозначно сопоставляется его имя $y \in N$ (N – некоторое счетное множество, в качестве которого здесь выбран натуральный ряд):

$$\begin{aligned} y^1 &= x_1^{a_{11}} x_2^{a_{12}} \ldots x_m^{a_{1m}} z_1^{c_{11}} z_2^{c_{12}} \ldots z_n^{c_{1n}}; \\ y^2 &= x_1^{a_{21}} x_2^{a_{22}} \ldots x_m^{a_{2m}} z_1^{c_{21}} z_2^{c_{22}} \ldots z_n^{c_{2n}}; \\ &\ldots \\ y^r &= x_1^{a_{r1}} x_2^{a_{r2}} \ldots x_m^{a_{rm}} z_1^{c_{r1}} z_2^{c_{r2}} \ldots z_n^{c_{rn}}, \end{aligned} \qquad (3)$$

где r – число кортежей отношения, соответствующего предикату P.

Выражение (3) представляет предикат $Py(y,\bar{x},\bar{z})$, который, очевидно, содержит пару ФЗ $y \leftrightarrow \{x_1,x_2,\dots,x_m,z_1,z_2,\dots,z_n\}$. В соответствующем этому предикату отношении $Py\{y,x_1,x_2,\dots,x_m,z_1,z_2,\dots,z_n\}$ атрибут y является потенциальным ключом, что означает, что каждый атрибут этого отношения функционально зависит от y. Предикат $Py(y,\bar{x},\bar{z})$, согласно теореме Хеза, удовлетворяет зависимости соединения $*(\{y,x_1\}$, $\{y,x_2\}$, …, $\{y,x_m\}$, $\{y,z_1\}$, $\{y,z_2\}$, …, $\{y,z_n\})$. В соответствии с данной ЗС выполняется декомпозиция предиката Py на бинарные проекции. При этом предикат Py связан с исходным P соотношением

$$P(\bar{x},\bar{z}) = \exists y Py(y,\bar{x},\bar{z}).$$

Проблема этой декомпозиции в том, что при достаточно большом по числу кортежей отношении $P\{x_1,x_2,\dots,x_m,z_1,z_2,\dots,z_n\}$ переменная y принимает такое же число значений в рамках предиката Py и его проекций. Это создает несколько проблем. Во-первых, формулы полученных в результате декомпозиции бинарных предикатов получаются большими, т.к. в каждом из них присутствует переменная y. Во-вторых, большое число значений переменной y усложняет схемотехническую реализацию логической сети, построенной на базе такой бинарной модели. Кроме того, при декомпозиции предиката $P(\bar{x},\bar{z})$ никак не используется ФЗ $\{x_1,x_2,\dots,x_m\} \to \{z_1,z_2,\dots,z_n\}$.

Для бинарной декомпозиции $P(\bar{x},\bar{z})$ можно ввести вспомогательную переменную более рационально, т.е. так, чтобы при минимальном числе значений переменной y в предикате $Py(y,\bar{x},\bar{z})$ выполнялась МЗ $y \to x_1 | x_2 | \dots | x_m | z_1 | z_2 | \dots | z_n$. Тогда по теореме Фейгина будет выполняться зависимость соединения $*(\{y,x_1\}$, $\{y,x_2\}$, …, $\{y,x_m\}$, $\{y,z_1\}$, $\{y,z_2\}$, …, $\{y,z_n\})$. На основе этой ЗС выполняется бинарная декомпозиция. В рамках полученных формул переменная y будет пробегать меньше значений, чем при стандартном методе. Однако во многих случаях, когда описываемое предикатом отношение велико, этот метод может оказаться трудноосуществимым, поскольку придется обрабатывать одновременно всю формулу исходного предиката или всю таблицу соответствующего ему отношения.

Предлагаемый ниже метод основан на простом факте: область значений (понимаемая в этой работе как множество всех образов функции) конечной дискретной функции всегда не больше по количеству элементов, чем область её определения. (А такие функции можно представлять реляционными отношениями и описывать конечными предикатами). Рассмотрим $P(\bar{x},\bar{z})$ с ФЗ $\{x_1,x_2,\dots,x_m\} \to \{z_1,z_2,\dots,z_n\}$ как функцию $p: A \to C$, где

$$A = \{\bar{x} \mid P[x_1, x_2, \ldots, x_m] = 1\};$$
$$C = \{\bar{z} \mid P[z_1, z_2, \ldots, z_n] = 1\} \qquad (4)$$

(эти обозначения не раз понадобятся в дальнейшем). $A \subseteq S_1 \times S_2 \times \ldots \times S_m$, $C \subseteq T_1 \times T_2 \times \ldots \times T_n$, $|A| = r$, $|C| = l$. Так как $\forall \bar{c} \in C \;\; \exists \bar{a} \in A$ такой, что $\bar{c} = p(\bar{a})$, то $p: A \to C$ – сюръекция. Для любого предиката $P(\bar{x}, \bar{z})$ с ФЗ $\{x_1, x_2, \ldots, x_m\} \to \{z_1, z_2, \ldots, z_n\}$ справедливо неравенство $r \geq l$, поскольку в разных кортежах $(\bar{x}, \bar{z})$ значения $\bar{z}$ могут повторяться, а значения $\bar{x}$ – нет.

В силу ФЗ $\{x_1, x_2, \ldots, x_m\} \to \{z_1, z_2, \ldots, z_n\}$ число кортежей в соответствующем предикату $P(\bar{x}, \bar{z})$ отношении будет равно $|A|$, т.е. r. Поэтому при стандартном методе бинарной декомпозиции предиката $P(\bar{x}, \bar{z})$ число значений вспомогательной переменной y в рамках полученных формул (обозначим это число k) будет равно количеству всех элементов из области определения функции $p: A \to C$, т.е. k = r.

Задача состоит в том, чтобы для бинарной декомпозиции функционального предиката $P(\bar{x}, \bar{z})$ ввести вспомогательную переменную наименьшей ценой, т.е. так, чтобы она принимала в рамках полученных в результате декомпозиции предикатов как можно меньше значений. Поэтому постараемся соотнести вспомогательную переменную не с областью определения сюръекции $p: A \to C$ (что предлагается сделать стандартным методом), а скорее с областью ее значений. Попытаемся заменить исходный предикат $P(\bar{x}, \bar{z})$ системой бинарных предикатов с переменными $\{y, x_1\}$, $\{y, x_2\}$, …, $\{y, x_m\}$, $\{y, z_1\}$, $\{y, z_2\}$, …, $\{y, z_n\}$ так, чтобы k – число значений переменной y в этих предикатах – было минимальным. Если при стандартном методе бинарной декомпозиции мы всегда имеем k = r (количество элементов области определения A), то сейчас наша задача состоит в том, чтобы оно лежало в пределах $l \leq k \leq r$, причем как можно ближе к l (количество элементов области значений C). Возможна ситуация, когда k = l. Однако для этого, как будет показано дальше, требуется выполнение МЗ $\{z_1, z_2, \ldots, z_n\} \to\!\!> \{x_1, x_2, \ldots, x_m\}$. Поэтому данную ситуацию рассмотрим отдельно.

Всякая функция устанавливает взаимно однозначное соответствие между каждым элементом её области значений и его полным прообразом относительно данной функции (каноническое отображение второго рода по Мальцеву [9]). Множество полных прообразов для всех элементов области значений функции образует разбиение её области определения. Элементы разбиения называют слоями. Для функции $p: A \to C$, описываемой предикатом $P(\bar{x}, \bar{z})$, каноническое отображение второго рода выражается в виде равенств

$$
\begin{aligned}
z_1^{c_{11}} z_2^{c_{12}} \ldots z_n^{c_{1n}} &= P(x_1, x_2, \ldots, x_m, c_{11}, c_{12}, \ldots, c_{1n}); \\
z_1^{c_{21}} z_2^{c_{22}} \ldots z_n^{c_{2n}} &= P(x_1, x_2, \ldots, x_m, c_{21}, c_{22}, \ldots, c_{2n}); \\
&\ldots \\
z_1^{c_{l1}} z_2^{c_{l2}} \ldots z_n^{c_{ln}} &= P(x_1, x_2, \ldots, x_m, c_{l1}, c_{l2}, \ldots, c_{ln}),
\end{aligned}
\qquad (5)
$$

где $(c_{i1}, c_{i2}, \ldots, c_{in}) \in C$, $i = \{1, 2, \ldots, l\}$.

Выражение (5) является явным видом отображения $p : A \to C$. Введем обозначения: $\bar{c}_i = (c_{i1}, c_{i2}, \ldots, c_{in})$, $\bar{z}^{\bar{c}_i} = z_1^{c_{i1}} z_2^{c_{i2}} \ldots z_n^{c_{in}}$, $i = \{1, 2, \ldots, l\}$. Предикат $P(\bar{x}, \bar{c}_i)$ описывает полный прообраз элемента $\bar{c}_i \in C$ при отображении $p : A \to C$. Выражение (5) позволяет рассматривать вектор $\bar{c}_i \in C$ как идентификатор или имя соответствующего ему слоя $\{\bar{x} \mid P(\bar{x}, \bar{c}_i) = 1\}$, $i = \{1, 2, \ldots, l\}$.

Это открывает интересную точку зрения на функциональный предикат $P(\bar{x}, \bar{z})$: он выражает взаимно однозначное соответствие между слоями $\{\bar{x} \mid P(\bar{x}, \bar{c}_i) = 1\}$ и их именами $\bar{c}_i$. Выражение (5) представляет собой явный вид этого соответствия. Такое соотношение в предикате $P(\bar{x}, \bar{z})$ будет выполняться тогда и только тогда, когда выполняется ФЗ $\{x_1, x_2, \ldots, x_m\} \to \{z_1, z_2, \ldots, z_n\}$. Поэтому, сохранив при декомпозиции данное соответствие, мы сохраним выражаемую предикатом $P(\bar{x}, \bar{z})$ функцию $p : A \to C$.

Итак, чтобы выполнить декомпозицию предиката $P(\bar{x}, \bar{z})$ наименьшей ценой предлагается следующий способ действий с учетом сказанного. Во-первых, каждый слой $\{\bar{x} \mid P(\bar{x}, \bar{c}_i) = 1\}$ разбиения представим в виде объединения минимального числа подмножеств $D_{i1}, D_{i2}, \ldots, D_{ik_i}$ из A, т.е.

$$
P(\bar{x}, \bar{c}_i) = D_{i1}(\bar{x}) \vee D_{i2}(\bar{x}) \vee \ldots \vee D_{ik_i}(\bar{x}), \qquad (6)
$$

причем таких, что

$$
\begin{aligned}
D_{i1}(\bar{x}) &= A_1^{i1}(x_1) \wedge A_2^{i1}(x_2) \wedge \ldots \wedge A_m^{i1}(x_m), \\
D_{i2}(\bar{x}) &= A_1^{i2}(x_1) \wedge A_2^{i2}(x_2) \wedge \ldots \wedge A_m^{i2}(x_m), \\
&\ldots \\
D_{ik_i}(\bar{x}) &= A_1^{ik_i}(x_1) \wedge A_2^{ik_i}(x_2) \wedge \ldots \wedge A_m^{ik_i}(x_m),
\end{aligned}
$$

$i = \{1, 2, \ldots, l\}$, $k_1 + k_2 + \ldots + k_l = k$.

Если множество равно объединению некоторых других множеств, то говорят, что они его *покрывают*, а совокупность этих множеств называется его *покрытием*. Подмножества $D_{i1}, D_{i2}, \ldots, D_{ik_i}$ назовем *декартовыми классами*, а их совокупность – *декартовым покрытием* слоя $\{\bar{x} \mid P(\bar{x}, \bar{c}_i) = 1\}$.

Таким образом, предлагается для каждого слоя $\{\bar{x} \mid P(\bar{x}, \bar{c}_i) = 1\}$ разбиения множества A, $i = \{1, 2, \ldots, l\}$, построить минимальное декартово покрытие.

Во-вторых, каждому декартову классу $D_j$ произвольным образом взаимно однозначно сопоставим уникальное имя $j \in B$, B – конечное подмножество некоторого счетного множества; здесь выбран натуральный ряд N. В-третьих, с помощью одного предиката выразим связь между декартовыми классами и их именами, а с помощью второго – связь каждого слоя с теми декартовыми классами, которые его покрывают. Причем соотношение между классами и слоями выражается не непосредственно, а через их уникальные имена: соответственно j и $\bar{c}_i$. Имена декартовых слоев будем выражать с помощью вспомогательной переменной $y \in N$. Тогда первый предикат будет иметь заголовок $P_1(\bar{x}, y)$, а второй – $P_2(y, \bar{z})$. В этих двух предикатах будет содержаться информация о взаимно однозначном соответствии между слоями $\{\bar{x} \mid P(\bar{x}, \bar{c}_i) = 1\}$ разбиения множества A и их именами $\bar{c}_i \in C$, $i = \{1, 2, \ldots, l\}$.

Необходимость в декартовом покрытии для каждого слоя обусловлена стремлением выстроить многозначную зависимость $y \to> x_1 \mid x_2 \mid \ldots \mid x_m$ в предикате $P_1(\bar{x}, y)$. На основе данной МЗ по теореме Фейгина можно будет выполнить бинарную декомпозицию $*(\{y, x_1\}, \{y, x_2\}, \ldots, \{y, x_m\})$. В предикате $P_2(y, \bar{z})$ будет выполняться ФЗ $y \to \{z_1, z_2, \ldots, z_n\}$, поскольку каждому декартову классу будет соответствовать только один слой разбиения. Данная ФЗ влечет зависимости $y \to z_1$, $y \to z_2$, …, $y \to z_n$, что позволяет провести бинарную декомпозицию на основе ЗС $*(\{y, z_1\}, \{y, z_2\}, \ldots, \{y, z_n\})$, которая будет выполняться в силу теоремы Хеза. Предикаты $P_1$ и $P_2$ построены так, что в пределах их формул переменная y принимает минимальное число значений k (т.к. в каждом слое образовано минимально возможное число декартовых классов $k_i$). Вместе $P_1$ и $P_2$ выражают функцию $p: A \to C$.

Формально предикат $P_1(\bar{x}, y)$ выражается в виде

$$\begin{array}{l} y^{j_{i1}} = D_{i1}(\bar{x}); \\ y^{j_{i2}} = D_{i2}(\bar{x}); \\ \quad\ldots \\ y^{j_{ik_i}} = D_{ik_i}(\bar{x}), \end{array} \quad i = \{1, 2, \ldots, l\}.$$

Предикат $P_2(y, \bar{z})$ выражается в виде

$$y^{j_{i1}} \vee y^{j_{i2}} \vee \ldots \vee y^{j_{ik_i}} = \bar{z}^{\bar{c}_i}, \quad i = \{1,2,\ldots,l\}.$$

Нетрудно показать, что предикат P выражается через $P_1$ и $P_2$ в виде

$$P(\bar{x},\bar{z}) = \exists y(P_1(\bar{x},y) \wedge P_2(y,\bar{z})).$$

Действительно,

$$P_1(\bar{x},y) \wedge P_2(y,\bar{z}) = \left( \bigvee_{i=1}^{l} (D_{i1}(\bar{x})y^{j_{i1}} \vee D_{i2}(\bar{x})y^{j_{i2}} \vee \ldots \vee D_{ik_i}(\bar{x})y^{j_{ik_i}}) \right) \wedge$$

$$\wedge \left( \bigvee_{i=1}^{l} (y^{j_{i1}} \vee y^{j_{i2}} \vee \ldots \vee y^{j_{ik_i}})\bar{z}^{\bar{c}_i} \right).$$

Воспользуемся законами ложности и идемпотентности, которые в данном случае выразим в виде

$$y^i y^j = 0, \quad i \neq j;$$
$$y^i y^j = y^i, \quad i = j.$$

В результате имеем

$$P_1(\bar{x},y) \wedge P_2(y,\bar{z}) = \bigvee_{i=1}^{l} (D_{i1}(\bar{x})y^{j_{i1}} \vee D_{i2}(\bar{x})y^{j_{i2}} \vee \ldots \vee D_{ik_i}(\bar{x})y^{j_{ik_i}})\bar{z}^{\bar{c}_i}.$$

$B = \{1,2,\ldots,l\} = \{j_{11}, j_{12},\ldots, j_{1k_1}, j_{21}, j_{22},\ldots, j_{2k_2},\ldots, j_{l1}, j_{l2},\ldots, j_{lk_l}\}$. Тогда

$$\exists y(P_1(\bar{x},y) \wedge P_2(y,\bar{z})) = \exists y \in B(P_1(\bar{x},y) \wedge P_2(y,\bar{z})) =$$
$$= D_{11}(\bar{x})\bar{z}^{\bar{c}_1} \vee D_{12}(\bar{x})\bar{z}^{\bar{c}_1} \vee \ldots \vee D_{1k_1}(\bar{x})\bar{z}^{\bar{c}_1} \vee$$
$$\vee D_{21}(\bar{x})\bar{z}^{\bar{c}_2} \vee D_{22}(\bar{x})\bar{z}^{\bar{c}_2} \vee \ldots \vee D_{2k_2}(\bar{x})\bar{z}^{\bar{c}_2} \vee \ldots \vee$$
$$\vee D_{l1}(\bar{x})\bar{z}^{\bar{c}_l} \vee D_{l2}(\bar{x})\bar{z}^{\bar{c}_l} \vee \ldots \vee D_{lk_l}(\bar{x})\bar{z}^{\bar{c}_l} =$$
$$= \bigvee_{i=1}^{l} (D_{i1}(\bar{x}) \vee D_{i2}(\bar{x}) \vee \ldots \vee D_{ik_i}(\bar{x}))\bar{z}^{\bar{c}_i}.$$

Откуда, согласно (6), имеем

$$\exists y(P_1(\bar{x},y) \wedge P_2(y,\bar{z})) = \bigvee_{i=1}^{l} P(\bar{x},\bar{c}_i)\bar{z}^{\bar{c}_i} = P(\bar{x},\bar{z}).$$

Таким образом, исходный предикат $P(\bar{x},\bar{z})$ с ФЗ $\{x_1, x_2,\ldots, x_m\} \to \{z_1, z_2,\ldots, z_n\}$ представлен системой бинарных предикатов с переменными $\{y, x_1\}$, $\{y, x_2\}$, …, $\{y, x_m\}$, $\{y, z_1\}$, $\{y, z_2\}$, …, $\{y, z_n\}$. Причем k – число значений переменной y в этих предикатах – лежит в пределах $|C| \le k \le |A|$, где $p: A \to C$ – функция, описываемая предикатом P.

Рассмотрим случай, когда в предикате $P(\bar{x},\bar{z})$ наряду с ФЗ $\{x_1, x_2,\ldots, x_m\} \to \{z_1, z_2,\ldots, z_n\}$ имеется МЗ $\{z_1, z_2,\ldots, z_n\} \to\!\!\to x_1 | x_2 |\ldots| x_m$, $m, n \ge 2$. В этом случае для бинарной декомпозиции предиката наличие данной функциональной зависимости не требуется, поэтому в описании декомпозиции она никак не учитывается, а бинарная декомпозиция выполняется на основании имеющейся МЗ гораздо проще. Каждому набору $(c_{i1}, c_{i2}, \ldots, c_{in}) \in C$ (4) произвольным образом ставим во взаимно однозначное соответствие свое уникальное имя i:

$$\begin{gathered} y^1 = z_1^{c_{11}} z_2^{c_{12}} \ldots z_n^{c_{1n}}; \\ y^2 = z_1^{c_{21}} z_2^{c_{22}} \ldots z_n^{c_{2n}}; \\ \ldots \\ y^l = z_1^{c_{l1}} z_2^{c_{l2}} \ldots z_n^{c_{ln}}. \end{gathered} \tag{7}$$

Данное выражение представляет собой предикат $P_1(y, \bar{z})$, который содержит ФЗ $y \to \{z_1, z_2, \ldots, z_n\}$.

Если в (5) заменить $z_1^{c_{i1}} z_2^{c_{i2}} \ldots z_n^{c_{in}}$ на $y^i$ в соответствии с (7), то будет образован новый предикат $P_2(\bar{x}, y)$, который можно выразить в виде

$$P_2(\bar{x}, y) = \bigvee_{i=1}^{l} y^i P(x_1, x_2, \ldots, x_m, c_{i1}, c_{i2}, \ldots, c_{in}).$$

Так как $P_2(\bar{x}, y)$ был образован из $P(\bar{x}, \bar{z})$ с многозначной зависимостью $\{z_1, z_2, \ldots, z_n\} \to> x_1 \mid x_2 \mid \ldots \mid x_m$ путем взаимно однозначной замены вектора $\bar{z}$ на скаляр y, то предикат $P_2(\bar{x}, y)$ будет содержать МЗ $y \to> x_1 \mid x_2 \mid \ldots \mid x_m$.

Легко доказывается, что предикат P выражается через $P_1$ и $P_2$ в виде

$$P(\bar{x}, \bar{z}) = \exists y (P_1(y, \bar{z}) \wedge P_2(\bar{x}, y)).$$

По теореме Хеза выполняем бинарную декомпозицию $*(\{y, z_1\}, \{y, z_2\}, \ldots, \{y, z_n\})$ предиката $P_1(y, \bar{z})$. По теореме Фейгина выполняем декомпозицию $*(\{y, x_1\}, \{y, x_2\}, \ldots, \{y, x_m\})$ предиката $P_2(\bar{x}, y)$.

Таким образом, только в этом случае число k принимает минимально возможное значение $k = |C|$, где C определяется в соответствии с (4).

Прочие случаи исходных данных для декомпозиции предиката $P(\bar{x}, \bar{z})$ являются тривиальными. Бинарная декомпозиция выполняется либо непосредственно по теореме Хеза, либо непосредственно по теореме Фейгина. Введения вспомогательной переменной не требуется.

Достоинством данного метода является возможность обрабатывать формулу или отношение не целиком, а по частям: для каждого полного прообраза минимальное декартово покрытие подбирается отдельно.

**Заключение.** Все имеющиеся методы неэквивалентной декомпозиции, т.е. декомпозиции с привлечением вспомогательных переменных, ориентированы на статические предикаты или отношения. Описанный метод – не исключение. Между тем отношения в базах данных – это динамические объекты. Поэтому методы неэквивалентной декомпозиции не пригодны для применения в базах данных. Однако они хороши при описании грамматики естественного языка, поскольку грамматику языка по сравнению с промышленными или коммерческими базами данных можно считать статическим объектом.

Дальнейшая разработка методов декомпозиции предикатов на основе теории нормализации имеет хорошие перспективы и принесет не мало плодов в совершенствовании алгебро-логического аппарата, а значит, и для повышения качества моделей интеллектуальных функций. Связь между теорией реляционных баз данных и алгебро-логическим аппаратом следует изучать в дальнейшем, поскольку эти две области могут существенно обогатить друг друга.

## ЛИТЕРАТУРА


1. *Шабанов-Кушнаренко Ю.П. Теория интеллекта. Математические средства. – Х.: Вища шк., 1984. – 144 с.*
2. *Дударь З.В. Математические модели флективной обработки словоформ и их использование в системах автоматической обработки текстов русского языка. Дис ... канд. техн. наук: 05.13.01. – Х., 1984. – 212 с.*
3. *Пронюк А.В. Метод многослойной декомпозиции предикатов и его применение в системах искусственного интеллекта. Дис ... канд. техн. наук: 05.13.23. – Х., 2004. – 191 с.*
4. *О мозгоподобных ЭВМ / М.Ф. Бондаренко, З.В. Дударь, И.А. Ефимова, В.А. Лещинский, С.Ю. Шабанов-Кушнаренко // Радиоэлектроника и информатика. – 2004. – № 4. – С. 83-99.*
5. *Бондаренко М.Ф., Чикина В.А., Шабанов-Кушнаренко Ю.П. Модели языка // Бионика интеллекта. – 2004. – № 1. – С. 27-37.*
6. *Шабанов-Кушнаренко Ю.П. Теория интеллекта. Проблемы и перспективы. – Х.: Вища шк. 1987. – 160 с.*
7. *Алгебра предикатов и предикатных операций / М.Ф. Бондаренко, З.В. Дударь, Н.Т. Процай, В.В. Черкашин, В.А. Чикина, Ю.П. Шабанов-Кушнаренко // Радиоэлектроника и информатика. – 2005. – № 1. – С. 80-87.*
8. *Дейт, К. Введение в системы баз данных, 6-е издание. – К.; М.; СПб.: Издательский дом «Вильямс», 2000. – 848 с.*
9. *Мальцев А.И. Алгебраические системы. – М.: Наука, 1970. – 392 с.*



*Поступила 9.06.2006*

**Рецензент:** доктор технических наук, профессор С.Ю. Шабанов-Кушнаренко, Харьковский национальный университет радиоэлектроники.