



ПОШУК ФУНКЦІОНАЛЬНИХ ЗАЛЕЖНОСТЕЙ ЗАСОБАМИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Таняньський О.С., Яковлева О.В.

Харківський національний університет радіоелектроніки

В реляційних базах даних (БД) даталогічне або логічне проектування призводить до розробки схеми БД, тобто сукупності схем відношень, які адекватно моделюють абстрактні об'єкти предметної області і семантичні зв'язки між цими об'єктами. Основою аналізу коректності схеми є так звані функціональні залежності (ФЗ) між атрибутами БД. Деякі залежності між атрибутами стосунків є небажаними із-за побічних ефектів і аномалій, які вони викликають при модифікації БД. При цьому під процесом модифікації БД розуміється внесення нових даних у БД або видалення деяких даних з БД, а також оновлення значень деяких атрибутів.

Основою проектування реляційних БД є безліч функціональних залежностей, заданих на схемі відношення. Проте при інтеграції або реструктуризації БД, а також при виявленні взаємовідносин між даними необхідно розглядати зворотне завдання, а саме завдання пошуку можливих ФЗ, аналізуючи екземпляр БД.

Для вирішення такого завдання використовуються методи і алгоритми інтелектуальної обробки даних [1]. В даному випадку інтелектуальний аналіз БД представляє процес витягання знання про закономірності поведінки накопичених даних. Такі питання визначають круг завдань, пов'язаних з визначенням наближених ФЗ.

Щоб виявити ФЗ, які задовольняють деяким закономірностям, можна скористатися методом розділення кортежів на групи по відмінності значень кожного атрибуту. Для кожного атрибуту кількість груп відповідає кількості різних ознак (значень). Кожну групу називають класом еквівалентності.

Нехай задано відношення $R(A, B, C, D)$, представлене на рисунку 1.

A	B	C	D
a	b_1	c	d_1
a_1	b_1	c_1	d_1
a_1	b	c_1	d_2
a	b_2	c	d_2

Рис. 1 - Відношення R .

У атрибуті A значення " a_1 " є в кортежах 1 і 4, таким чином, вони формують клас еквівалентності $\{1, 2\}$. Так само значення " a_2 " є присутнім в кортежах 2, 3. Отже, класи еквівалентності по атрибуту A відбудуться з двох класів еквівалентності $\mathcal{R}_A = \{\{1, 2\}, \{2, 3\}\}$. Можна будувати класи еквівалентності для складених атрибутів, наприклад, для B і D класи еквівалентності приймуть вигляд $\mathcal{R}_{B,D} = \{\{1, 2\}, \{3\}, \{4\}\}$.



Після детального аналізу класів еквівалентності можна визначити, принаймні, наближені ФЗ. Класи еквівалентності в $\mathfrak{R}$ визначають класи еквівалентності в $\mathfrak{R}'$, якщо кожен клас еквівалентності в $\mathfrak{R}$ є підмножиною деякого класу еквівалентності в $\mathfrak{R}'$. Тобто ФЗ $X \rightarrow Y$ існує, якщо $\mathfrak{R}$ визначає $\mathfrak{R}'$. У відношенні R (рис. 1) окрім розглянутих класів еквівалентності можна виділити наступні класи:

$$\begin{aligned}\mathfrak{R}_B &= \mathfrak{R}_{B,D} = \{\{1, 2\}, \{3\}, \{4\}\}, \\ \mathfrak{R}_C &= \mathfrak{R}_{A,C} = \{\{1, 4\}, \{2, 3\}\}, \\ \mathfrak{R}_D &= \{\{1, 2\}, \{2, 3\}\}, \\ \mathfrak{R}_{A,B} &= \mathfrak{R}_{A,B,C} = \mathfrak{R}_{B,C} = \mathfrak{R}_{C,D} = \mathfrak{R}_{B,C,D} = \mathfrak{R}_{A,C,D} = \mathfrak{R}_{A,D} = \{\{1\}, \{2\}, \{3\}, \{4\}\}.\end{aligned}$$

Таким чином, безліч нетривіальних функціональних залежностей для відношення R відповідатиме

$$F = \{A \rightarrow C; B \rightarrow D; A, B \rightarrow C; A, B \rightarrow D; A, B, C \rightarrow D; B, C \rightarrow A; B, C \rightarrow D; \\ C, D \rightarrow A; C, D \rightarrow B; B, D, C \rightarrow A; A, C, D \rightarrow B; A, D \rightarrow B; A, D \rightarrow C\}.$$

Очевидно, що функціональні залежності є наближеними і можуть бути змінені залежно від тих даних, що вводяться.

Очевидним способом скорочення існуючого набору функціональних залежностей є виключення з нього тривіальних залежностей. Залежність називається тривіальною, якщо вона не може не виконуватися.

Як мається на увазі в самій їх назві, з практичної точки зору подібні залежності не представляють значного інтересу, на відміну від нетривіальних залежностей, які дійсно є обмеженнями цілісності в повному розумінні цього поняття. Проте з точки зору формальної теорії залежностей необхідно враховувати усі залежності, як тривіальні, так і нетривіальні.

Спираючись на теорію функціональних залежностей, можна припустити, що для детальнішого аналізу отриманих ФЗ необхідно застосувати алгоритми пошуку замикання і базису безлічі F . З іншого боку, виділення з наближених функціональних залежностей реальних є завданням нетривіальним. Існуючі методи спираються на ті, що припускають похибки в отриманих ФЗ. Таким чином, подальші дослідження мають бути спрямовані на поліпшення коефіцієнтів помилок.

1. Базы данных. Интеллектуальная обработка информации [Текст] : учеб. / В.В. Корнеев, А.Ф. Гареев, С.В. Васютин, В.В. Райх. – М. : Издатель Молгачева С.В., Издатель Нолидж, 2001. – 496 с.

2. Базы данных. Модели, разработка, реализация [Текст]: учеб. / Т.С. Карпова – М.: НОУ «Интуит», 2016. – 403 с.

3. Основы использования и проектирования баз данных [Текст]: учеб. / В.М. Илюшечкин – М.: Издательство Юрайт, 2014. – 213 с.