

ЗАСТОСУВАННЯ АВТОКОРЕЛЯЦІЙНОЇ ФУНКЦІЇ ДЛЯ КЕРУВАННЯ МАТЕРІАЛІЗОВАНИМИ ПРЕДСТАВЛЕННЯМИ БД

Розгон О.С., Возовиков Ю.М.

Науковий керівник – проф. каф. «Системне програмне забезпечення»,

канд. техн. наук. Кунгурцев О.Б.

Матеріалізовані представлення (МП) дозволяють значно підвищити продуктивність інформаційної системи (ІС) [1]. Але їх використання вимагає аналізу запитів до БД протягом досить тривалого часу. Підвищення ефективності застосування МП можливо, якщо визначені періоди їх найбільш інтенсивного використання (періодичність появи запитів до БД певного типу), що дає змогу вмикати чи вимикати МП. В даний час при використанні МП періодичність запитів не аналізується, тому дослідження у цьому напрямку слід вважати актуальними.

Будемо вважати, що проводиться спостереження за ІС протягом часу T . В

результаті чого отримана множина запитів MZ . З множини MZ виділяється множина запитів типу $Z_i(t)$, для яких можливе створення МП.

Для визначення періодичності появи цих запитів розраховуємо функцію щільності розподілу запитів Y для множини запитів Z_i , як кількість запитів, що спостерігається за одиницю часу δ_t . Для визначення значення слід керуватися наступними міркуваннями. У всякому разі тривалість δ_t не може бути меншою за тривалість запиту Z_i . У той же час, вона не може бути більшою чим одна година, тому що година може бути одиницею виміру роботи працівника, що може привести до появи нових задач (нових запитів). Враховуючи сказане, можна рекомендувати вибір значення δ_t у діапазоні 10 – 30 хвилин.

Для визначення періодичності розподілу запитів пропонується використовувати автокореляційну функцію.

Існують дослідження в області автокореляції, які дозволять вирахувати періодичність появи даних [2].

Алгоритм визначення основного періоду передбачає наступну послідовність дій:

1. Визначення середньоарифметичного значення \bar{Y} елементів функції щільності запиту:

$$\bar{Y} = \sum_{i=1}^n Y_i * \frac{1}{n},$$

де Y_i - щільність розподілу запиту на елементарному інтервалі спостереження

n – кількість елементарних інтервалів в періоді спостереження

2. Обчислення значення функції варіації c_0 :

$$c_0 = \frac{1}{N} \sum_{t=1}^N (Y_t - \bar{Y})^2$$

3. Обчислення значення функції автоковаріації c_h для часового кроку h , де $h = 1, \dots, 4$ елементарні одиниці (елементарна одиниця - 20хв.). Вибір такого діапазону h ґрунтується на специфіці використання МП в прикладній області і виходить з тих припущень, що великий часовий інтервал, наприклад, більше 2 годин буде поглинати час простою БД (обідня перерва, яка є в будь-якій організації), що неприпустимо, оскільки це спотворює майбутню статистику. Тому верхньою межею можна вважати одну годину. Якщо значення елементарного кроку буде менше за 20 хв., це дасть занадто докладну інформацію, в якій немає необхідності. У будь-якому разі треба підтримувати відношення

$$c_h = \frac{1}{N} \sum_{t=1}^{N-h} (Y_t - \bar{Y})(Y_{t+h} - \bar{Y})$$

4. Обчислення коефіцієнта автокореляції r_h на кроці h :

$$r_h = c_h / c_0$$

5. За рекомендаціями [3] високий рівень автокореляції досягається при значенні 0.75 коефіцієнта автокореляції, тому, щоб визначити основний період проводиться порівняння отриманого коефіцієнта з цим значенням. Якщо отримане значення більше, то ми зберігаємо це значення, а часовий крок на якому це значення було отримано оголошуємо основним періодом.

6. Повторюємо кроки 3-5 для кожного h .

7. Повторюємо кроки 1-6 для кожного запиту

Запропонований алгоритм реалізовано у вигляді відповідної програми, що приймає на вході множину моментів з'явлення запитів певного типу, а на виході – період коливання щільності цих запитів. Це дає змогу проектувати механізм вмикання/вимикання для запитів і, таким чином, економити час на обробку запитів, а також пам'ять на зберігання МП.

У теперішній час проводяться дослідження для визначення декількох частот коливання щільності розподілу запитів, що може мати місце у реальних системах (діяльність що активізується по понеділках та наприкінці кожного місяця).

СПИСОК ЛІТЕРАТУРИ:

1. Кунгурцев А.Б., Куок Винь Нгуен Чан. Метод анализа информационной системы для применения материализованных представлений Холодильна техніка і технологія. Одеса, 2005. – 2(94). – с. 102-105.
2. Box, G. E. P., and Jenkins, G. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.
3. Бродский Б. Е. Структурные сдвиги и единичные корни: в нестационарных моделях временных рядов [Электронный ресурс] – 20 с.

Режим доступа:

<http://data.cemi.rssi.ru/GRAF/center/methodology/econometrics/download/tsakh2008.pdf>