

МЕТОДИ ТА АЛГОРИТМИ ПРОЕКТУВАННЯ

УДК 004.89

Є.Н. Федорчук

Національний університет “Львівська політехніка”,
кафедра програмного забезпечення

АЛГОРИТМИ І ТЕХНОЛОГІЇ ПОШУКУ КЛАСТЕРІВ НА ОСНОВІ ОПТИМІЗАЦІЙНИХ КРИТЕРІЇВ ДЛЯ ТОВАРНИХ БАЗ ДАНИХ

© Федорчук Є.Н., 2007

Розглянуто алгоритми пошуку цінових кластерів на основі оптимізаційних критеріїв. Задача пошуку в дискретній області цін зводиться до умовної задачі в неперервній області. Така апроксимація забезпечує швидкий пошук кластерів з контрольованою похибкою. Наведено алгоритм, технологію і результати пошуку кластерів для бази даних Oracle .

Algorithms of searching for price clusters on the base optimisation criterions are considered. Problem of searching in the discrete price area is bring to the conditional problem in the unceasing area. Such aproximation ensures a quick searching for an clusters with the control mistake. Algorithm, technology and results of the searching for an clusters for Oracle database are described.

Вступ. Важливу роль в прикладному аналізі бізнес-процесів відіграють методи інтелектуального аналізу даних (ІАД). Сучасні інформаційні технології та засоби активно використовують методи ІАД, зокрема кластерний аналіз. Такі продукти Microsoft, як Microsoft SQL Server 7.0, Microsoft Dynamic AX.4.0 містять відповідні процедури пошуку кластерів.

Прикладний аналіз товарних операцій в бізнесі передбачає широке коло задач пошуку необхідного асортименту в групах товарів, які належать до задач кластеризації. При кластеризації товарів використовуються досить умовні принципи кластеризації. В умовах широкої і різномірної номенклатури товарів доцільнішим є управління асортиментом на рівні сегментів, ніж управління на рівні номенклатури. Основною проблемою для пошуку кластерів у великих базах товарних даних є формування пошукових критеріїв. Існує низка пошукових задач, які можна звести до дискретних задач оптимізації. Особливістю цих задач є їхня комбінаторна складність, обумовлена великою кількістю різнотипних елементів.

Класи задач пошуку кластерів. Розглянемо задачу кластеризації як задачу пошуку асортименту товарів з врахуванням їх необхідної кількості. Є групи дискретних елементів, які мають одну основну характеристику, наприклад, вартість. Необхідно вибрати таку сукупність кластер – елементів з усіх груп, щоб вона відповідала заданій вартості.

Математична постановка такої задачі як задачі неперервної оптимізації може мати вигляд:
знайти

$$\min \Phi = \left(\sum_{i=1}^n k_i c_i - C_{\text{зад}} \right)^2, \quad (1)$$

де n – кількість груп; k_i – кількість елементів в i -й групі; c_i – вартість i -го елемента; $C_{\text{зад}}$ – задана вартість набору. Обмеження задачі містять:

➤ обмеження на вартість елементів в i -й групі:

$$c_{\min}^i \leq c^i \leq c_{\max}^i; \quad (2)$$

➤ обмеження на кількість елементів в i -й групі:

$$1 \leq k^i \leq k_{\max}^i \quad (3)$$

Після знаходження розв'язку задачі значення c^i уточнюються шляхом пошуку у масиві вартостей найближчих дискретних значень. Значення k^i – кількості елементів у i -й групі, заокруглюються до найближчого цілого.

Якщо в задачі (1) всі значення k_i дорівнюють одиниці, то задача є одновимірною.

Розглянемо приклади одновимірних задач та їх математичну постановку у вигляді (1)–(3).

Задача оптимального наповнення споживчого кошика є типовою економічною задачею. Її вигляд: знайти

$$\min \Phi = \left(\sum_{i=1}^n k_i c_i - C_{\text{зад}} \right)^2, \quad (4)$$

де n – кількість груп; k_i – кількість елементів в i -й продуктивній групі; c_i – вартість i -го продуктового елемента; $C_{\text{зад}}$ – задана вартість споживчого кошика. Кількість елементів, заданих в одиницях ваги, об'єму, дорівнює 1. Обмеження задачі мають вигляд (2). Розв'язання задачі – кластер – дає оптимальний вміст споживчого кошика із врахуванням розкиду цін на основні товари у межах міста чи регіону. Таке моделювання задачі дає змогу прогнозувати купівельну спроможність населення.

Проектування конфігурації комп'ютера. Виберемо головним критерієм конфігурації комп'ютера його вартість як суму вартостей комплектувальних елементів. Дискретну задачу оптимізації такого критерію можна звести до задачі неперервної параметричної мінімізації функції

$$\Phi = \left(\sum_{i=1}^n c_i - C_k \right)^2 \quad (5)$$

де c_i – вартості для n – складових елементів комп'ютера; C_k – задана вартість комп'ютера при прямих обмеженнях на окремі вартості

$$c_{i \min} \leq c_i \leq c_{i \max},$$

де $c_{i \min}$ та $c_{i \max}$ – мінімальні та максимальні значення вартості елемента в його групі.

У розглянутій постановці задачі можна розв'язувати як оптимізаційні задачі неперервної оптимізації ітераційними алгоритмами. Відповідний алгоритм і технологія в середовищі Excel, призначені для пошуку конфігурації комп'ютера, розглянуті в роботі [1]. У цій роботі пропонується пошуковий алгоритм без ітераційної процедури для пошуку кластерів у базах даних.

Аналіз пошуку кластерів на основі розбиття області обмежень. Задачу пошуку кластера в базі даних сформулюємо як задачу пошуку цінового кластера заданої вартості $C_{\text{зад}}$. Кластер утворює сукупність товарів, яка формується вибором окремого товару з кожної групи. Нехай в базі є n груп товарів. У кожній групі є більше одного товару. Для кожної i -ї групи можна задати границі неперервної вартості:

$$C_{i \min}, C_{i \max}; \quad i \in [1, n], \quad (6)$$

де $C_{i \min}$, $C_{i \max}$ – мінімальна та максимальна вартість окремого товару в i -й групі відповідно. Такі границі для всіх груп товарів утворюють неперервну область обмежень задачі пошуку. Сформулюємо лему.

Лема. Для заданої неперервної області обмежень можна обчислити як мінімум один кластер

із заданою апріорі вартістю $C_{\text{зад}} = \sum_{i=1}^n C_i$ на основі відношення

$$p = \left(\sum_{i=1}^n C_{i \max} - \sum_{i=1}^n C_{i \min} \right) / \left(\sum_{i=1}^n C_i - \sum_{i=1}^n C_{i \min} \right) \quad (7)$$

де p – коефіцієнт розбиття області обмежень, $p \geq 1$.

Доведення. Розіб'ємо кожний інтервал вартостей (6) за допомогою відношення:

$$(C_{i \max} - C_{i \min}) = p * (C_i - C_{i \min}).$$

Запишемо послідовно

$$\begin{aligned} (C_{1,\max} - C_{1,\min}) &= p^* (C_1 - C_{1,\min}); \\ (C_{2,\max} - C_{2,\min}) &= p^* (C_2 - C_{2,\min}); \\ (C_{3,\max} - C_{3,\min}) &= p^* (C_3 - C_{3,\min}); \\ &\dots; \\ (C_{n,\max} - C_{n,\min}) &= p^* (C_n - C_{n,\min}). \end{aligned}$$

Підсумувавши послідовно всі вирази, отримаємо

$$\left(\sum_{i=1}^n C_{i,\max} - \sum_{i=1}^n C_{i,\min} \right) = p \left(\sum_{i=1}^n C_i - \sum_{i=1}^n C_{i,\min} \right).$$

Звідси знаходимо

$$p = \left(\sum_{i=1}^n C_{i,\max} - \sum_{i=1}^n C_{i,\min} \right) / \left(\sum_{i=1}^n C_i - \sum_{i=1}^n C_{i,\min} \right) \quad (8)$$

Цей вираз повністю збігається з умовою леми, отже, вона доведена.

Висновки. 1. Знайдений кластер $C_{\text{зад}} = \sum_{i=1}^n C_i$ у неперевній області цін є апроксимацією для дискретної області цін в базі даних. За його допомогою будують пошуковий критерій для обчислення реального кластера $C_{\text{зад}}$ у вихідній товарній базі.

2. Точність обчислення кластера визначається точністю апроксимації дискретних цін. Її можна регулювати, корегуючи пошуковий критерій.

Алгоритм пошуку кластерів. За результатами аналізу можна побудувати алгоритм пошуку кластера із заданою вартістю $C_{\text{зад}}$. Алгоритм розв'язання розглянутих вищих задач містить дві групи операцій [2]. Перша група виконує:

- аналіз бази даних задач і визначення кількості груп елементів;
- вибір C_{\min} , C_{\max} та $C_{\text{сум,мін}}$ та $C_{\text{сум,мах}}$ у групах;
- визначення необхідної вартості кластера $C_{\text{зад}}$;
- обчислення коефіцієнта розбиття p .

Друга група виконує:

- обчислення вартостей C_i для елементів кластера;
- формування пошукового критерію для дискретних вартостей у кожній товарній групі;
- пошук елементів кластера в базі даних товарів за результатами обчислень.

Технологія та результати пошуку цінових кластерів для бази даних в середовищі Oracle. Розглянутий алгоритм був реалізований в середовищі Oracle для пошуку кластерів в задачах аналізу споживчого кошика та в задачі проектування конфігурації комп'ютера.

Схему пошуку кластерів подано на рисунку. Ця схема дає змогу забезпечити оптимальну швидкодію пошуку за рахунок опрацювання тільки пошукового образу бази – сформованих таблиць. Ці таблиці є вхідними даними для пошуку. Операції розбиття області обмежень та пошуку ґрунтуються на запитах, описаних за допомогою мови SQL-запитів.



Схема алгоритму для пошуку кластерів

У табл. 1 подано приклади вмісту бази даних та результати пошуку двох кластерів для задачі аналізу вмісту продуктового кошика вартістю 90грн. Ціни подано для одиниці товару. Для пошуку п'яти кластерів сумарний час роботи алгоритму становить 0,59 сек для процесора із тактовою частотою 2 ГГц. Складність цієї задачі визначається як 7^6 можливих варіантів кластерів.

Таблиця 1

Результати пошуку кластерів у продуктовому кошику

Приклади груп товарів у базі даних				Приклад вмісту м'ясної групи товарів			
№ з/п	Групи товарів	Ціна, мін, грн.	Ціна, мах, грн.	№ з/п	Група	Ціна	Тип
1	Крупи	1,9	3,3	1	М'ясо	20,1	Курятина
2	М'ясо	20,1	48,4	2	М'ясо	22,25	Шинка
3	Молочні/Інше	1,8	9,25	3	М'ясо	30,7	Яловичина
4	Овочі	1	4,15	4	М'ясо	37,5	Свинина
5	Риба	4,5	15,25	5	М'ясо	40,13	Ковбаса
6	Фрукти	5	15	6	М'ясо	48,4	Ковбаса 'Домашня'
7	Хліб	1,5	2,5				

Приклад знайденого кластера ціною 78 грн.				Приклад знайденого кластера ціною 64.4 грн.			
№ з/п	Групи товарів	Назви продуктів	Ціна, грн.	№ з/п	Групи товарів	Назви продуктів	Ціна, грн.
1	Крупи	Макарони	3	1	Крупи	Борошно	2.8
2	М'ясо	Ковбаса	40.13	2	М'ясо	Яловичина	40.13
3	Молочні/Інше	Цукор	4.4	3	Молочні/Інше	Масло	3.6
4	Овочі	Баклажани	3.5	4	Овочі	Кабачки	3
5	Риба	Щупак	13.15	5	Риба	Сайра	12.3
6	Фрукти	Персики	12	6	Фрукти	Мандарини	10
7	Хліб	Рум'яний	2.3	7	Хліб	3 висівками	2

У табл. 2 подано приклад вмісту кластера – конфігурації комп'ютера. Для бази даних, яка містила інформацію про 2000 тис. записів – елементів конфігурації комп'ютера, задано вартість конфігурації 3000 грн. Результати пошуку 2-х кластерів : час пошуку – 0,45 с, знайдена вартість – 3252 грн. Похибка становить 252 грн. і є на рівні 10 %.

Таблиця 2

Результати пошуку кластера для конфігурації комп'ютера

№ з/п	Група	Ціна, грн.	Типи елементів
1	2	3	4
1	Cooler	13	Fan MAX 80x80x25mm BB 4pin
2	Columns	18,72	Колонки 3Nod A-105, 2.0, plastic, RMS: 2x2Вт, 200-20000Гц, black/red, 60x35x160мм
3	Mouse	20,8	Mouse A-4 Tech OP-620 PS/2 оптика
4	Keyboard	23,4	Клавіатура 1807 PS/2 Slim (White)
5	Ethernet card	26	Lan Canyon SN9130 10/100 Realtek Fast Ethernet Adapters with Web-Based Management (10/100M, 100Mbps, Fast Ethernet, PCI)
6	Floppy 3.5	37,44	FDD 1.44 ALPS black
7	Zip-ROM	52	USB Flash drive 512Mb Canyon USB2.0 Aluminium
8	Modem	55,12	Modem int. Micronet Shuttle 3000 (56kbps FaxModem, V.92, Conexant Chipset, Int, Soft) SP3018S V6

1	2	3	4
9	Sound card	57,2	SC C-Media 8738 6каналі PCI
10	CD-ROM	72,8	CD-ROM 52x LG GCR-8523BB
11	Box	88,4	Rise XTC-B2 SS+17SE білий, 4*5.25", 7*3.5", USB2.0+Audio, Structure dim 510*190*450(mm), без БЖ
12	Memory	106,6	SDRAM 128 PC-133 Samsung
13	TV-tuner	182	TN/FM K-World VS-LTV7131RF Philips Chip (MPEG)+FM+w/Nicam & Stereo (9Bit)
14	Scanner	213,2	Сканер ScanExpress 1248UB
15	Printer	218,4	Lexmark Z735 (A4,4800*1200dpi – Black,4800*1200dpi – Color ,15/15ppm USB)
16	Video memory	223,6	VC ASUS ATI Radeon A9250/TD 128 MB 64bit Box with TV-out DVI-I AGP8x
17	Mother board	247	MB FOXCONN 865G7MF-SH-BOX (i865G, s775, FSB800MHz, 2DualDDR, VC, AGP8x, 3PCI, 2xATA, 2xSATA, SB6Ch, 4xUSB2.0, LAN) mATX
18	Hard disk drive HDD	254,8	HITACHI 80Gb 2M cache 7200 ATA-6 (UATA133)
19	Processor	327,6	AMD Sempron CN 3200+ (FSB1600, 256kb, Manila, sAM2) box
20	Monitor	1014	17" MIRAI DML-517N LCD 8ms 300cd/m2 ,400:1 ,silver-black.

Висновки. Запропонований алгоритм є корисним для електронного бізнесу, для моделювання бізнес-процесів, пов'язаних з ціновими кластерами, для прогнозування витрат або прибутку. Економічний ефект полягає у автоматизації рутинних робіт з перебору великої кількості варіантів задач, швидкості пошуку кластерів.

1. Федорчук Є. *Проектування оптимальної за вартістю конфігурації комп'ютера на основі системи Excel // Вісн. Держ. ун-ту "Львівська політехніка". – 2000. – № 392. – С. 3–6.* 2. Федорчук Є.Н., Зарубій Н. *Побудова цінових кластерів на основі оптимізаційних критеріїв для товарних баз даних // Матеріали 2-ї Міжнар. наук.-практ.ї конф. "Науковий потенціал світу – 2005". – 19–30 вересня 2005. – Т. 10. – Дніпропетровськ: Наука і освіта, 2005. – С. 50–51.*

УДК 004.852, 004.942

П.О. Кравець

Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

ІГРОВА ЗАДАЧА ПОБУДОВИ СТОХАСТИЧНО-ОРТОНОРМОВАНИХ СИСТЕМ

© Кравець П.О., 2007

Сформульовано задачу побудови стохастично-ортонормованих систем. Запропоновано ігровий метод та розроблено алгоритм розв'язування задачі. Досліджено вплив параметрів задачі на збіжність ігрового методу.

The task of stochastic orthonormal system construction is formulated. The game method and algorithm of a task solution are developed. The influence of a task parameters on a game method convergence is investigated.

Вступ. У системах автоматизованого проектування, опрацювання і передачі сигналів, стиснення даних, оптимального керування існує необхідність побудови числових ортогональних або ортонормованих систем [1].