

OLAP에 의한 자료의 기억공간을 줄이기 위한 한가지 방법

손 명 숙

선행연구[1, 2]에서는 OLAP체계가 MOLAP체계에 기초한것으로 하여 방대한 기억용량이 요구되며 이러한 다차원자료의 기억용량을 줄이기 위한 차원결합방식을 리용하였다.

그러나 자료요소들을 집계하고 빈값을 가지는 자료들을 삭제하는 방식으로 하지 못하고있다.

논문에서는 기하급수적으로 증대되는 다차원자료의 기억공간을 줄이기 위한 한가지 방법을 제안하였다.

1. 다차원자료의 분석에 대한 정의

정의 모임 $D = \{D_1, D_2, \dots, D_n\}$ 이라고 하고 속성모임 $A = \{A_1, A_2, \dots, A_n\}$ 을 D_1, D_2, \dots, D_n 의 부분속성모임이라고 하자.

T 를 자료요소들의 모임이라고 할 때 $\forall t \in T, t \in D_1 \times D_2 \times \dots \times D_n$ 인 자료모임의 자료요소수를 N 이라고 하면 자료공간 D 는 $D_1 \times D_2 \times \dots \times D_n$ 으로 정의되며 이 자료공간의 크기는 $\|D\| = \|D_1\| \times \|D_2\| \times \dots \times \|D_n\|$ 이다. 이때 자료공간 D 의 밀도는 $r = \frac{N}{\|D\|}$ 이다.

2. 다차원자료의 자료기억공간을 줄이기 위한 무리짓기방법

다차원자료의 기억공간을 줄이기 위하여 분포된 자료값들을 집계하여 집계된 자료값이 령인 불필요한 자료값들을 제거하기 위한 무리짓기를 진행한다. 이 무리짓기는 류사도 척도를 리용하여 류사성행렬을 만들고 매 무리에서의 대표점들을 리용하여 류사한 행들을 이동시켜 집계된 자료요소들을 정렬시킨다. 이렇게 자료요소들을 정렬시킨 다음 높은 밀도에 기초하여 무리들을 결합시킨다.

실례로 어떤 상업부문에서 판매자료에 대한 차원모임을 다음과 같이 구조화할수 있다.

차원모임 $D = \{\text{상품, 판매날자, 매대, 구입대상}\}$

$f(\text{상품}) = \{\text{편, 류, 파, 상품규격}\}$

$f(\text{판매날자}) = \{\text{년, 분기, 월}\}$

$f(\text{매대}) = \{\text{전문화에 따르는 매대그룹, 각 매대명, 매대별판매원수, \dots}\}$

$f(\text{구입대상}) = \{\text{지역별에 따르는 구입대상그룹, 구입대상명, \dots}\}$

이때 속성모임 A 는 매 $f(d_i)$ 들의 합모임으로 된다.

$$A = \bigcup_{\forall d_i \in D} f(d_i)$$

집계된 판매자료는 분석차원에 따라 자료립방체를 구성하며 이에 기초하여 초기자료행렬을 계산한다. 이때 초기자료행렬의 행은 차원모임에서 하나의 차원씩 선택하여 구성하며 열은 하나의 차원에 대한 속성으로 이루어진다. 여기서 행렬의 자료원소는 링아닌것들로써 초기자료를 만들고 집계한다.

무리짓기는 ISODATA법에 의하여 초기무리개수와 초기무리 C_i 와 무리중심값 m_i 를 선택하고 규칙 $\|x - m_i\| = \min\|x - m_j\|$ 에 따라 초기무리들로 분류한다.

류사성행렬은 류사성척도로서 유클리드거리를 리용하여 만든다.

매 무리중심들사이거리 $D_{ij} = \|m_i - m_j\|$ 를 계산하고 일정한 턱값보다 작으면 무리들끼리 결합을 진행한다.

ISODATA법에 의한 무리짓기과정은 그림 1, 2, 3과 같다. 여기서 색칠한 칸은 판매수량값을 기억시킨 부분이다.

11월						
10월						
9월						
8월						
7월						
6월						
5월						
4월						
3월						
2월						
1월						
	어린이 솜옷	어른 남솜옷	어른 녀솜옷	어른 내의류	어른 양복류	어린이 내의류

그림 1. 정돈되지 않은 상태의 자료요소들

11월						
8월						
6월						
5월						
3월						
1월						
9월						
2월						
10월						
7월						
4월						
	어린이 솜옷	어른 남솜옷	어른 녀솜옷	어른 내의류	어른 양복류	어린이 내의류

그림 2. 날자차원에 관하여 집계된 자료요소들

11월						
8월						
6월						
5월						
3월						
1월						
9월						
2월						
10월						
7월						
4월						
	어른 양복류	어린이 내의류	어른 내의류	어른 남솜옷	어린이 솜옷	어른 녀솜옷

그림 3. 정돈된 상태의 집계자료요소들

자료무리짓기는 날자차원(세포축)에 대한 무리분석을 진행하여 류사한 행렬을 이동시킨 다음 날자차원을 제외한 모든 차원의 결합으로 이루어진 가로축에 대하여 류사한 행들을 집계하여 자료요소들을 정렬시킨다. 무리짓기조작은 높은 밀도를 가진 무리로 결합시킨다.

그림 3에서 보는바와 같이 마지막단계는 3개의 영역으로 나타난다.

무리들을 통합시키는 과정은 다음과 같다.

단계 1 차원 $n_i, n_j (i \neq j)$ 에 대하여 속성값들의 집계값이 같은 행들에 대하여 정돈한다.

단계 2 집계된 값들의 모임을 무리 $c_i (i = 1, \dots, k)$ 로 표시한다.

무리사이 행간격에 따라 통합한다.

① $d(c_i, c_j) = \min d(c_r, c_s)$ 인 $c_i, c_j (i < j)$ 를 찾는다.

② If $d(c_i, c_j) \leq M$ then $c_i = c_i \cup c_j$ 라고 하고 c_j 를 제거한다.

이와 같이 두 무리들사이 행간격이 최소로 되는것들끼리 통합한다.

그림 2와 같이 자료요소들을 최소한 집계하여 다차원자료의 기억공간을 줄인다.

맺는 말

론문에서 제안한 방법은 선행방법보다 외부기억공간의 리용률이 높다. 특히 분석차원이 많아지고 그에 따르는 자료요소들이 추가된다면 기억용량이 기하급수적으로 증대되는 것과 관련하여 무리짓기방법을 리용하여 다차원자료의 기억밀도를 높여 기억공간을 줄일 수 있다.

참고 문헌

[1] Robert Wrembel et al.; Data Warehouses and OLAP, IRM Press Inc, 28~32, 2008.

[2] H. M. Liao et al.; Journal of Computer Science and Technology, 23, 6, 905, 2011.

주체107(2018)년 2월 5일 원고접수

A Method for Reducing the Memory Space of Data by OLAP

Son Myong Suk

In this paper, we proposed a method for reducing the memory space of multi-dimension data increasing by geometrical progression using the clustering method.

Key words: OLAP, memory space