

처리고속화를 위한 자료조직방법

배원철, 한용환

실천에서는 계층모형의 나무구조를 이루면서 계층적으로 련관이 이루어지는 자료들이 많이 제기되고있다.

실례로 교무행정관리업무에서 교수자들의 과정안집행상태, 학부강좌구분, 타학부들과의 강의구분자료와 교수업무, 교육강령집행에 필요한 교수집행정형산출에서 리용되는 개인별, 강좌별, 과목별구분자료 등을 들수 있다.

우리는 관계형자료기지설계의 일반적인 방법[1-4]에서와 같이 여러개의 관계표에 나누어 관리하여야 할 계층자료들을 선행한 방법과는 달리 1개의 관계표안에 구축하고 기록들사이 련관을 설정하는 방법으로 표현하고 관리하려고 한다.

이를 위하여 하나의 관계표안에 구축하려고 하는 계층자료들에 대하여 그것들사이 련관관계를 다음과 같이 설정한다.

이제 관계표 R_1 이 다음과 같은 n 개 기록들의 자료모임으로 이루어졌다고 하자.

$$R_1 = \{r_1, r_2, r_3, \dots, r_n\}$$

이때 계층모형의 정의로부터 알수 있는것처럼 계층모형을 이루는 자료들사이에는 《선조-후손》의 계층련관관계가 존재하며 그러한 계층자료들은 몇번째 계층의 자료인가를 나타내는 계층준위속성을 가지게 된다.

따라서 관계표의 기록순서는 의미가 없지만 계층자료들의 모임에서는 요소들에 다음과 같은 순서관계를 설정하게 된다.

$r_k, r_j \in R_1$ 인 $\forall r_k, r_j$ 에 대하여 $k < j$ 이면 r_k 의 계층준위 $> r_j$ 의 계층준위

한편 r_k 와 r_j 가 《선조-후손》의 계층련관관계를 가질 때 즉 r_j 가 r_k 의 부모 또는 선조마디로 될 때

$$r_k \mapsto r_j$$

로 표시하기로 한다. 이때 r_k 가 r_j 로부터 직접 계승된 후손마디가 아니라 p 번 계승되어 생겨난 후손마디로 될 때 r_k 는 r_j 로부터 생겨난 $p-1$ 개의 또 다른 선조마디를 가지게 된다. 즉

$$r_k \mapsto r_k^1, r_k^1 \mapsto r_k^2, \dots, r_k^{p-1} \mapsto r_j$$

여기서 r_k^i 는 r_k 의 i 번째 선조마디이다.

이러한 r_k 와 r_j 사이의 계층련관관계를

$$r_k \xrightarrow{P} r_j$$

로 표시하면 p 는 두 자료사이의 계층련관속성으로 된다. 즉 $p=1$ 이면 r_k 는 r_j 의 부모마

디로 되며 $p > 1$ 이면 p 번째 선조마디로 된다.

이러한 계층자료들에 대하여 임의의 두 자료의 쌍들로 다음의 모임을 구성할수 있다.

$$p_1 \times R_1 = \{(r_1, r_2), (r_1, r_3), \dots, (r_1, r_n), \dots, (r_{n-1}, r_n)\}$$

한편 관계표 R_1 의 식별자마당을 ID 라고 할 때 n 개 기록의 식별자마당값들의 모임을

$$ID = \{ID_1, ID_2, \dots, ID_n\}$$

으로 표시할수 있다.

정의 1 $F: R_1 \times R_1 \rightarrow \{0, ID\}$ 를 계층련관함수라고 한다.

이때 모임 $R_1 \times R_1$ 의 요소 (r_k, r_j) 에 대하여

$$\begin{aligned} r_k &\stackrel{P}{\mapsto} r_j \text{이면} & F(r_k, r_j) &= ID_j (k \geq j) \\ r_k &\mapsto r_j \text{이면} & F(r_k, r_j) &= 0 (k \geq j) \\ k &< j \text{ 이면} & F(r_k, r_j) &= 0 \end{aligned}$$

으로 되며 이 함수의 값을 보고 입력파라미터로 설정된 두 자료사이에 계층련관관계가 존재하는가 하지 않는가를 알수 있다.

정의 2 계층련관함수의 값을 요소로 하는 다음의 행렬

$$C = \{C_{kj} \mid C_{kj} = F(r_k, r_j)\}$$

를 계층련관행렬이라고 한다.

계층모형을 이루는 자료들사이 련관관계를 반영하는 계층련관행렬은 그 정의로부터 다음과 같은 성질을 만족시킨다.

$$\textcircled{1} C_{k1} = ID_1$$

계층모형은 하나의 나무구조를 이루는것으로 하여 모든 자료는 뿌리마디와 계층련관 관계를 가지게 된다. 즉

$$r_k \in R_1 \text{인 } \forall r_k (k = \overline{2, n}) \text{에 대하여}$$

$$\begin{aligned} &P \\ r_k &\mapsto r_1 \end{aligned}$$

의 계층련관관계가 성립한다. 여기서 뿌리마디는 임의의 계층자료에 대하여 제일 웃준위 선조마디로 되므로 이때 p 값은 k 번째 자료가 가지는 계층련관속성들중에서 최대로 되며 곧 그 자료의 계층준위속성으로 된다. 따라서 행렬에서 1렬의 값은 뿌리마디기록의 식별자 값 ID_1 로 된다.

② $C_{kk} = ID_k$ 즉 주대각선에는 해당 순서마디의 식별자값 ID_k 가 놓이게 된다.

③ 계층련관행렬은 아래삼각형행렬이다.

행렬의 웃삼각형을 이루게 되는 요소 C_{kj} 들은 $k < j$ 이므로 계층련관함수의 정의로부터 그 값이 0으로 된다.

그것은 r_k 보다 아래계층준위에 놓이게 되는 r_j 는 r_k 의 부모나 선조마디로 될수 없기 때문이다. 즉

$$k < j \Rightarrow r_k \text{의 계층준위의} > r_j \text{의 계층준위} \Rightarrow r_k \mapsto r_j.$$

따라서 행렬의 웃삼각형부분원소들은 모두 0으로 된다.

④ 임의의 k 번째 행벡토르의 령아닌 요소의 개수는 k 번째 자료의 계층준위속성을 표현한다. 즉 k 번째 행벡토르에 령아닌 요소의 수가 4개 있다면 k 번째 자료는 계층모형의 4번째 계층의 자료라는것을 알수 있다.

⑤ 임의의 k 번째 행벡토르의 령아닌 요소는 k 번째 자료와 계층련관관계를 가지는 자료의 식별자를 표현한다. 즉 k 번째 행벡토르의 령아닌 임의의 요소가 ID_j 라면 ID_j 를 식별자값으로 하는 기록은 k 번째 자료의 임의의 선조마디라는것을 알수 있다.

또한 ID_j 가 령아닌 i 번째 요소라면 ID_j 를 식별자값으로 하는 기록은 k 번째 자료의 i 번째 선조마디라는것도 알수 있다.

⑥ 계층련관관계를 가지는 자료들은 그에 해당하는 행벡토르들이 부분적으로 일치하게 된다. 즉 $r_k \mapsto r_j$ 로 되는 임의의 두 자료 r_k, r_j 에 대하여

$$C_{ki} = C_{ji} (i = \overline{1, j})$$

로 된다.

계층련관행렬의 우와 같은 성질로부터 다음과 같은것을 알수 있다.

첫째로, 계층자료들이 구성하는 계층모형의 최대계층수를 알수 있다.

사실 임의의 k 번째 행벡토르의 령아닌 요소수는 k 번째 자료의 계층준위속성을 나타낸다.(성질 ④)

한편 계층련관행렬은 계층모형의 웃계층으로부터 아래계층으로 내려가는 순서로 자료들의 계층련관정보를 반영한다. 따라서 행렬의 마지막 n 번째 행의 령아닌 요소수는 계층모형의 최대계층수로 된다.

둘째로, 계층련관행렬의 임의의 행(뿌리마디에 해당하는 첫번째 행 제외)에는 주대각선 요소외에 적어도 1개의 령아닌 요소가 존재하게 된다.

계층자료들이 구성하는 계층모형은 그 정의로부터 뿌리마디를 제외한 임의의 자료는 자식 및 후손마디는 가질수도 있고 가지지 않을수도 있지만 하나의 부모마디는 반드시 가지게 된다. 따라서 임의의 계층자료는 적어도 하나의 자료와는 계층련관관계를 가지게 되며 뿌리마디는 적어도 하나이상의 자식 및 후손마디를 가지게 된다. 즉 $\forall r_k (k = \overline{2, n})$ 에 대하여

$$\exists r_j \mid r_k \xrightarrow{P} r_j$$

이며 계층련관행렬인 경우 $i = \overline{2, n}$ 인 $\forall i$ 에 대하여

$$\exists C_{ij} \mid C_{ij} \neq 0 (j < i)$$

이다. 한편 뿌리마디자료 r_1 에 대하여

$$\exists r_j \mid r_k \xrightarrow{P} r_1$$

로 된다.

이로부터 계층자료관리의 조작대상으로 되는 계층자료들의 모임을 다음과 같이 표시할수 있다.

$$R = \{r_i \mid i = \overline{1, n}, \exists r_j, r_i \xrightarrow{P} r_j \vee r_j \xrightarrow{P} r_i\}$$

컴퓨터에서 처리되는 자료구조는 여러가지 형식으로 할수 있으나 우와 같은 6가지 성질을 만족시키도록 계층형식으로 조직하면 그 처리속도를 한층 높일수 있다는것을 증명할수 있다.

참 고 문 헌

- [1] П. В. Горшков; Открытые системы, 4, 21, 2002.
- [2] А. А. Летичевский; Кибернетика, 4, 41, 2008.
- [3] В. Н. Дрибас; Проблемы кибернетики, 28, 269, 2010.
- [4] W. W. Armstrong; Communs. ACM, 9, 1, 13, 2005.

주체104(2015)년 8월 5일 원고접수

A Method of Data Organization for Speeding-up Processing

Pae Won Chol, Han Yong Hwan

We proposed a new method of organizing data that is able to speed-up the processing large scale of data. We introduced the concept of hierarchical correlation function (HCF) and hierarchical correlation matrix (HCM) consisted of HCF and described the properties of HCM and proposed a data structure that satisfies the properties.

Key words: data organization, hierarchical correlation function