

## 무계불은 빈발항목에 의한 학생성적예측의 한가지 방법

김 순 실

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《교육방법을 혁신하는데 맞게 학교들에서 시험방법, 학생들의 실력평가방법을 개선 하여야 합니다.》(《김정일선집》 증보판 제23권 399페이지)

교육부문에서는 교원의 교수활동과 학생들의 학습활동에 관한 수많은 자료들이 축적되고있다. 그러나 자료처리에서는 아직까지 초보적인 자료구분과 간단한 통계단계에 머무르고있으며 교육자료에 대한 구체적인 연구[1-3]가 광범히 진행되지 못하고있다.

론문에서는 자료발굴기술을 리용하여 학생정보로부터 최종성적을 예측하기 위한 한가지 방법을 제안하였다.

### 1. 자료립방체모형에 기초한 성적자료부호화

교육정보화체계의 성적자료로부터 자료발굴기술을 리용하여 앞으로의 성적을 예측하고 부족점을 제때에 찾아 극복대책을 세워나가는것은 교원의 교수활동에서 매우 중요한 의의를 가진다.

교육자료인 학생성적의 특징은 다음과 같다.

첫째로, 성적자료들은 다차원자료모형에 기초한다. 차원으로 《등급》, 《회수》, 《항목》들을 설정한다.(그림 1)

둘째로, 매 차원들은 일련의 속성과 속성값을 가진다.

셋째로, 자료가 다차원이고 매 차원은 계층구조를 가지므로 이것은 다차원다층자료(자료립방체)모형에 기초한다.

넷째로, 차원정보와 사실정보를 리용하여 주제실현을 진행한다. 성적자료창고를 구축하기 위하여 교육정보화체계의 평상시성적상세항목에서 해당한 과목을 선택하고 개별적인 학생에 대하여 선택검열, 과제수행, 문제풀이, 실험실습, 검열시험, 학과경연에 따르는 성적자료들을 수집한다.

수집한 성적자료는 다차원다층의 계층구조를 가지므로 다음의 급수층들로 구성한다.(그림 2)

그림 2에서 보는것처럼 윗층에서 아래층으로 내려가면서 보다 세분화되고 매 항목이 더 특징적인 의미를 담는다.

따라서 교육자료발굴에 리용할수 있는 수집된 자료가운데서 성적에 영향을 미치는

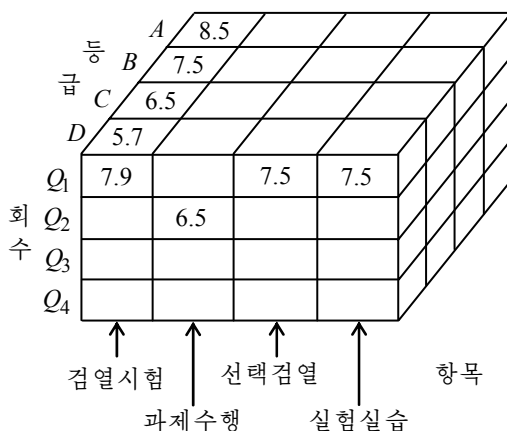


그림 1. 성적자료모형

항목(선택검열(B), 과제수행(C), 문제풀이(D), 실험실습(E), 검열시험(F), 학과경연(G), 종합 성적(H))들과 등급(최우등(A), 우등(B), 보통(C), 락제(D)), 급수(1, 2, 3)를 표 1과 같은 수자로 부호화한다.

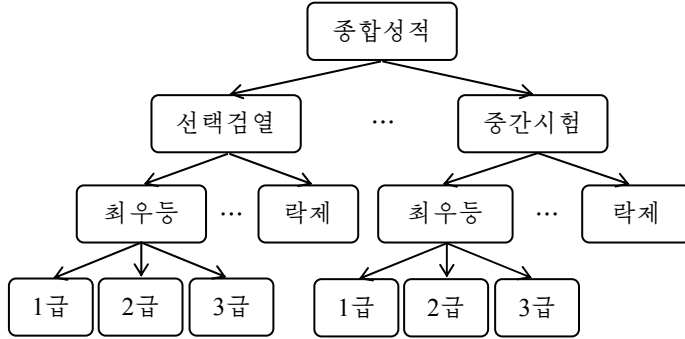


그림 2. 항목에 따르는 등급별 급수충구조

표 1. 항목과 등급, 급수에 따르는 성적자료의 부호화

| 등급         | 급수           | 항목   |      |      |      |      |      |      |
|------------|--------------|------|------|------|------|------|------|------|
|            |              | 선택검열 | 과제수행 | 문제풀이 | 실험실습 | 검열시험 | 학과경연 | 종합성적 |
|            |              | B    | C    | D    | E    | F    | G    | H    |
| A<br>(최우등) | 1(9.50~10.0) | BA1  | CA1  | DA1  | EA1  | FA1  | GA1  | HA1  |
|            | 2(9.0~9.49)  | BA2  | CA2  | DA2  | EA2  | FA2  | GA2  | HA2  |
|            | 3(8.50~8.99) | BA3  | CA3  | DA3  | EA3  | FA3  | GA3  | HA3  |
| B<br>(우등)  | 1(8.0~8.49)  | BB1  | CB1  | DB1  | EB1  | FB1  | GB1  | HB1  |
|            | 2(7.50~7.99) | BB2  | CB2  | DB2  | EB2  | FB2  | GB2  | HB2  |
|            | 3(6.50~7.49) | BB3  | CB3  | DB3  | EB3  | FB3  | GB3  | HB3  |
| C<br>(보통)  | 1(6.00~6.49) | BC1  | CC1  | DC1  | EC1  | FC1  | GC1  | HC1  |
|            | 2(5.50~5.99) | BC2  | CC2  | DC2  | EC2  | FC2  | GC2  | HC2  |
|            | 3(4.50~5.49) | BC3  | CC3  | DC3  | EC3  | FC3  | GC3  | HC3  |
| D(락제)      | 4.49이하       | BD   | CD   | DD   | ED   | FD   | GD   | HD   |

성적자료의 부호화는 결과 행번호를 붙여 항목(B~H)+등급(A~D)+급수(1~3)로 형식화한다. BC1은 선택검열(B)이 보통(C)으로서 1급(6.00~6.49)이라는것을 의미한다.

## 2. 무게붙은 빈발항목을 리용한 학생성적예측방법

일반적으로 학생의 종합성적은 여러가지 세부항목들의 결합으로 평가되는데 개별적 항목의 중요도는 서로 다르므로 보다 정확한 결과를 얻기 위하여서는 매 항목들에 무게 결수를 정확히 설정하여야 한다.

학생종합성적평가를 위한  $n$ 개의 항목이 있다고 하자.

이 경우 항목들의 중요성정도를 서로 2개씩 비교하면 다음의 행렬을 얻는다.

$$A = \begin{pmatrix} w_1/w_1 & \cdots & w_1/w_n \\ \vdots & \ddots & \vdots \\ w_n/w_1 & \cdots & w_n/w_n \end{pmatrix}$$

이제  $i$  째 항목에 대한  $j$  째 항목의 상대적중요성의 비를  $a_{ij}(i, j = \overline{1, n})$  라고 하면

$$a_{ij} \approx w_i / w_j(i, j = \overline{1, n})$$

가 성립한다. 이때  $a_{ij} = 1/a_{ji}$ ,  $a_{ii} = 1(i = \overline{1, n})$  로 된다.

만일  $a_{ij}$  에 대한 추정에서 차이가 있다면  $a_{ij} \neq w_i / w_j$  이다. 이 경우  $a_{ij}w_j - w_i$  의 값은 0으로 되지 않으므로  $a_{ij}(i, j = \overline{1, n})$  들을 서로 다른  $i, j$  들에 대해서도 정확히 추정하자면

$$z = \sum_{i=1}^n \sum_{j=1}^n (a_{ij}w_j - w_i)^2$$

의 값이 최소로 되는 무게  $\{w_1, w_2, \dots, w_n\}$  을 구하여야 한다.

한편  $w_i(i = \overline{1, n})$  들은 무게결수이므로

$$\sum_{i=1}^n w_i = 1, w_i > 0, i = 1, \dots, n$$

과 조건  $\{S = \{w_1, w_2, \dots, w_n\} | \sum_{i=1}^n w_i = 1, w_i > 0\}$  을 만족시키는 최소화문제가 나선다.

$$\min: \sum_{i=1}^n \sum_{j=1}^n (a_{ij}w_j - w_i)^2 = F(w_1, w_2, \dots, w_n)$$

이러한 문제는 등식제한을 가진 최량화문제로 보고 라그랑주함수법을 리용하여 풀수 있다.

라그랑주함수

$$L = \sum_{i=1}^n \sum_{j=1}^n (a_{ij}w_j - w_i)^2 + 2\lambda \left( \sum_{i=1}^n w_i - 1 \right)$$

로부터

$$\frac{\partial L}{\partial w_i} = \sum_{i=1}^n (a_{il}w_l - w_i)a_{il} - \sum_{j=1}^n (a_{ij}w_j - w_i) + \lambda = 0$$

을 얻는다.

보조적으로

$$\sum_{i=1}^n w_i = 1$$

을 고려하면 다음과 같은 비동차선형연립방정식

$$\begin{pmatrix} & & -1 \\ & C & \vdots \\ & & -1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

을 얻는다. 여기서  $C$ 는  $A$ 로부터 주어진다.

이 방정식으로부터 무계결수  $w_i$ 를 구하고 무계붙은 항목들에 의하여 학생성적을 예측하도록 한다.

성적자료기지의 성적표에서 같은 항목들에 대하여서는 산수평균을 하고 종합성적은 무계붙은 항목들의 총합으로 결정한 다음 성적자료의 부호화표(표 1)를 리용하여 표 2와 같은 평가표를 만든다.

표 2. 평가표

| 번호  | 이름  | 등수 | B   | C   | D   | E   | F   | H   |
|-----|-----|----|-----|-----|-----|-----|-----|-----|
| 1   | 강일철 | 6  | BC2 | CA2 | DA2 | EA3 | FA1 | HA3 |
| 2   | 김장훈 | 17 | BC1 | CA2 | DB3 | EA1 | FB2 | HB2 |
| 3   | 김주성 | 16 | BC1 | CA2 | DB1 | EA3 | FB2 | HB2 |
| 4   | 로예령 | 12 | BA2 | CA2 | DA2 | EB2 | FB2 | HB1 |
| 5   | 리경성 | 14 | BA1 | CA2 | DD  | EB2 | FB2 | HB1 |
| 6   | 리광우 | 3  | BA2 | CA2 | DB2 | EB2 | FA2 | HA3 |
| ⋮   | ⋮   | ⋮  | ⋮   | ⋮   | ⋮   | ⋮   | ⋮   | ⋮   |
| 99  | 차광성 | 1  | BA1 | CA2 | DB1 | EB1 | FA2 | HA2 |
| 100 | 함춘국 | 11 | BA2 | CA2 | DB2 | EB2 | FA3 | HA3 |

$S$ 를 교육정보화체계에서 제시된 100개의 실례자료라고 하고  $M$ 을 평가표의 1개 항목 마당이라고 할 때  $S$ 와 관련된  $M$ 의 정보값을 계산하기 위하여 먼저  $S$ 의 종합성적  $H$ 에 기초한 엔트로피를 계산한다.

실험자료에 의하면 100개의 성적모임에서 종합성적을 등급(급수)별로 보면 최우등생이 31(2, 11, 18)명, 우등생이 61(36, 18, 7)명, 보통생이 8(4, 3, 1)명이므로

$$entropy(S) = -\sum_{i=1}^4 p_i \log_2 p_i = 1.260 \ 610 \ 7$$

이다.

자료모임  $S$ 에서 마당  $M$ 과 관련되는 정보값  $Gain(S, M)$ 을 다음과 같이 계산한다.

$$Gain(S, M) = entropy(S) - \sum_{j_M} \frac{|S_j|}{|S|} entropy(S_j)$$

여기서  $S_j$ 는 마당  $M$ 에서 급수가  $j$ 인 성적자료들의 모임이다.

표 2의 매 마당정보값은 표 3과 같다.

표 3. 마당정보값

| 마당정보         | 의 미  | 값         |
|--------------|------|-----------|
| $Gain(S, B)$ | 선택검열 | 0.413 472 |
| $Gain(S, C)$ | 과제수행 | 0.459 853 |
| $Gain(S, D)$ | 문제풀이 | 0.198 937 |
| $Gain(S, E)$ | 실험실습 | 0.211 326 |
| $Gain(S, F)$ | 검열시험 | 0.656 947 |
| $Gain(S, G)$ | 학과경연 | 0.623 582 |

표 3에서와 같이 검열시험  $F$ 가 가장 높은 정보값을 가지므로 검열시험  $F$ 를 뿌리마

디로 할 때 평가경험자료를 4개(최우등, 우등, 보통, 낙제)의 무리로 분류할수 있다.

종합성적 H의 마당값이 하나씩 대응되어있는 무리에 대해서는 분류를 더이상 진행하지 않으며 여러개 대응되어있는 무리에 대해서는 우와 같은 방법으로 분류를 계속한다.

그러면 매 잎마디에 H의 마당값이 하나씩 존재하는 무리들에 대응되는 결정나무가 구성되게 된다.

이 결정나무를 리용하여 IF-THEN규칙형태로 학생평가규칙을 생성한다.

IF F=A1 && C=A2 && B=C2 && D=A2 && E=A3 THEN H=A3

IF F=B2 && C=A2 && B=C1 && D=B3 || D=B1 && E=A1 || E=A3 THEN H=B2

IF F=B2 && C=A2 && B=A1 || B=A2 && D=A2 || D=D && E=E2 THEN H=B1

⋮

## 맺 는 말

무게불은 빈발항목에 의한 자료발굴기술을 리용하여 교육정보화체계로부터 학생성적을 평가하기 위한 규칙을 생성하는 한가지 방법을 제안함으로써 학생들에 대한 최종성적 결과를 예측할수 있게 하였다.

## 참 고 문 헌

- [1] G. S. Abu-Oda, A. M. El-Halees; International Journal of Data Mining & Knowledge Management Process, 5, 1, 15, 2015.
- [2] A. Dutt et al.; International Journal of Information and Electronics Engineering, 5, 2, 105, 2015.
- [3] Amjad Abu Saa; International Journal of Advanced Computer Science and Applications, 7, 5, 212, 2016.

주체109(2020)년 11월 5일 원고접수

## A Method of Student's Mark Estimation by Frequent Item Fixed Weight

*Kim Sun Sil*

In this paper, we suggested a method that created rule for estimating student's mark from education information system using data mining techniques by frequent item fixed weight.

Keywords: data mining, create rule, mark estimation