

전자학습환경에서 무게화련관규칙발굴을 리용한 학습과정분석의 한가지 방법

어남일, 김철혁

선행연구[1]에서는 자료창고(Data Warehouse)를 구축하고 매 항목들의 출현빈도뿐 아니라 그 중요도에 따라 무게를 결정하고 련관규칙을 발굴하는 방법을 제안하였다.

선행연구[2]에서는 자료발굴(Data Mining)의 응용분야와 자료전처리수법, 련관규칙발굴과 무리화방법들을 제안하였다.

론문에서는 전자학습환경에서 진행되는 원격교육에서 자료발굴기술을 리용하여 학생의 학습활동기록자료기지에서 자료창고를 구축하고 련관규칙을 생성하여 학습과정을 분석하는 방법을 제안하였다.

1. 전자학습환경에서 자료창고구성과 업무자료기지구축

전자학습이란 정보통신기술을 기초로 하는 학습에 중점을 둔 교육 및 훈련체계를 말한다.

전자학습(Electronic Learning)환경에서 학생의 학습활동을 분석하기 위하여 학습과정을 일반적으로 서술하면 《누가, 언제, 어디서, 무엇을 얼마만큼 학습하였는가?》라는 질문에 대한 대답으로 된다. 이로부터 사용자, 시간, 장소, 대상, 학습량에 따라 학생의 학습활동을 특징지을수 있다. 이러한 특성값들을 리용하여 학습과정을 분석할수 있다.

학습활동기록자료기지에서 학습활동을 특징짓는 정보들을 차원과 항목들로 표시하고 련관규칙을 발굴하여 학습과정을 분석할수 있다.

련관규칙발굴을 위하여 자료창고를 구축하고 업무자료기지를 구축한다.

n 차원으로 구성된 자료기지를 표현하는 자료립방체를 $C(D_1, D_2, \dots, D_n)$ 으로 표시한다. 여기서 D_i 는 자료립방체를 구성하는 $i(i = \overline{1, n})$ 번째 차원이다.

현실자료모임인 학습활동기록자료기지의 자료들에 대하여 자료립방체를 다음과 같은 차원들로 구성한다.

$$C(D_{\text{date}}, D_{\text{user}}, D_{\text{time}}, D_{\text{place}}, D_{\text{quantity}})$$

여기서 D_{date} 는 학습날자차원, D_{user} 는 사용자차원, D_{time} 은 학습시간차원, D_{place} 는 학습장소차원, D_{quantity} 는 학습량차원이다.

차원표는 차원정보와 그 차원에 속하는 항목정보(I), 사실자료(F)로 구성되었다.

주어진 차원들에 대하여 자료창고구축을 위하여 학습날자에 따르는 사용자차원표와 시간장소차원표를 구성한다.

어떤 날자에 학습을 진행한 사용자들의 수를 m 이라고 할 때 학습을 진행한 사용자 i 의 사용자차원표는 다음과 같다.

$$DT_{\text{user}}^i(I_u, I_d, F_u^i, F_d), i = \overline{1, m}$$

여기서 I_u 는 사용자 i 의 항목정보, I_d 는 학습날자항목정보, F_u^i 는 사용자 i 의 사실자료정보, F_d 는 학습날자정보이다.

학습을 진행한 사용자 i 의 시간장소차원표는 다음과 같다.

$$DT_{\text{timeplace}}^i(I_t, I_p, I_q, F_t, F_p, F_q), i = \overline{1, m}$$

여기서 I_t 는 사용자 i 의 학습시간항목정보, I_p 는 사용자 i 의 학습장소항목정보, I_q 는 사용자 i 의 학습량항목정보, F_t 는 사용자 i 의 학습시간정보, F_p 는 사용자 i 의 학습장소정보, F_q 는 사용자 i 의 학습량정보이다.

구성된 차원표의 사실자료정보들에 대한 전처리조작을 진행한다.

사용자차원표의 전처리는 다음과 같이 진행한다.

단계 1 사용자마다당은 식별자마다당이 중복이 되지 않는것을 고려하여 그대로 리용한다.

$$F_u^i = \{\text{사용자 } i \text{의 식별자}\}$$

단계 2 학습날자에 대하여 휴식일인가, 아닌가를 평가하고 부호화를 진행한다.

$$F_d = \begin{cases} D_0, \text{학습날자} <> \text{휴식일} \\ D_1, \text{학습날자} <> \text{휴식일} \end{cases}$$

시간장소차원표의 전처리는 다음과 같이 진행한다.

단계 1 장소정보에 대하여 전처리를 진행한다.

$$F_p = \{p | p = \overline{1, 4}\}$$

여기서 $p=1$ 은 지능화말단, $p=2$ 는 사무실, $p=3$ 은 집, $p=4$ 는 전자열람실이다.

단계 2 학습시간정보에 대하여 2h간격으로 나누어 전처리를 진행한다.

$$F_t = \{t | t = \overline{1, 24/2}\}$$

여기서 $t=1$ 은 0-2시, $t=2$ 는 2-4시, ..., $t=12$ 는 22-24시이다.

단계 3 학습량은 단계 2의 주어진 시간구간에 학습을 진행하여 통과된 학습항목개수(강의, 자체평가, 연습 등)로 전처리를 진행한다.

$$F_q = \{q_t\}, t = \overline{1, 24/2}$$

여기서 q_t 는 t 시간구간동안 진행한 학습항목개수이다.

단계 4 시간장소차원표의 시간, 장소, 학습량차원의 사실자료정보들이 시간에 따라서 련관관계를 가지고있으므로 차원표를 시간구간에 따라 구성한다.

$$DT_{\text{timeplace}}^i(I_{\text{tpq}}^1, I_{\text{tpq}}^2, \dots, I_{\text{tpq}}^{12}, F_{\text{tpq}}^1, F_{\text{tpq}}^2, \dots, F_{\text{tpq}}^{12}), i = \overline{1, m}$$

여기서 $I_{\text{tpq}}^j (j = \overline{1, 12})$ 는 j 번째 시간구간에서의 시간장소차원항목정보이고 $F_{\text{tpq}}^j (j = \overline{1, 12})$ 는 j 번째 시간구간에서 시간, 장소, 학습량정보이다. 그리고

$$F_{\text{tpq}}^j = \left(T_{F_t^j} P_{F_p^j} Q_{F_q^j} \right)$$

이다. 여기서 F_t^j 는 j 번째 시간구간의 학습시간정보, F_p^j 는 j 번째 시간구간의 학습장소정보, F_q^j 는 j 번째 시간구간의 학습량정보이다.

전처리조작을 진행하여 부호화된 차원표와 항목정보들을 리용하여 업무자료기지를 구축한다.

구축된 업무자료기지를 다음의 표에 보여주었다.

표. 구축된 업무자료기지

TID	사용자	날자	항목 1	항목 2	항목 3	...	항목 11	항목 12
1	user 1	D_0	$T_1P_3Q_5$	$T_2P_3Q_8$...	$T_{11}P_1Q_2$	
2	user 2	D_0	$T_1P_4Q_7$		$T_3P_4Q_4$...		$T_{12}P_4Q_8$
3	user 5	D_0		$T_2P_3Q_2$	$T_3P_3Q_3$...		$T_{12}P_3Q_2$
4	user 1	D_1	$T_1P_1Q_6$	$T_2P_1Q_7$...	$T_{11}P_3Q_3$	
5	user 3	D_0	$T_1P_2Q_3$		$T_3P_2Q_6$...		$T_{12}P_2Q_1$
...

표에서 $T_1P_3Q_5$ 는 시간 0-2시, 학습장소 집, 학습량 5개를 의미하며 TID는 매 사용자의 학습날자를 의미한다.

2. 무게화련관규칙발굴

업무자료기지의 차원, 항목들에 무게화를 진행하여 다차원무게화자료구조를 실현한다. 차원정보와 항목정보들에 대하여 학습과정분석과 사용자관심에 기초하여 설정순위와 중요성, 빈발정도에 따라 무게를 결정한다.

차원과 항목의 무게값을 업무자료기지의 차원가운데서 차원설정순위와 항목의 계층준위에 따라 중요도를 결정한다.

사용자차원표의 사용자항목의 무게는 사용자에게 의하여 학습과정이 진행되므로 학습한 사용자들의 무게를 다같이 1로 정한다. 즉 $W_{\text{user}}^i = 1, i = \overline{1, m}$ 이다. 여기서 m 은 어떤 학습날자에 학습한 사용자수이다.

사용자차원표의 학습날자항목의 무게는 휴식일에 대하여 1:2의 비율로 할당한다.

$$W_{\text{user}}^i = \begin{cases} 1, F_d = D_0 \\ 0.5, F_d = D_1 \end{cases}$$

여기서 F_d 는 학습날자사실자료정보이다.

시간장소차원표에서는 시간, 장소, 학습량정보에 따라 항목이 결정되므로 매 정보들 사이의 연관관계를 고려하여 무게를 결정한다.

시간차원의 무게는 다음과 같다.

$$W_{\text{time}} = \begin{cases} 1, \text{분석자요구가 제일 큰 시간} \\ 0.6, \text{분석자요구가 중간인 시간} \\ 0.3, \text{분석자요구가 제일 작은 시간} \end{cases}$$

장소차원의 무게는 다음과 같다.

$$W_{\text{place}} = \{F_p / 4\}$$

여기서 F_p 는 학습장소사실자료정보이다.

학습량차원의 무게는 다음과 같다.

$$W_{\text{quantity}} = \begin{cases} 1, F_q > Q_{\gamma} \\ 0.5, F_q \leq Q_{\gamma} \end{cases}$$

여기서 F_q 는 학습량사실자료정보, Q_7 는 기준학습항목개수이다.

시간장소차원의 무게는 다음과 같다.

$$W_{tpq} = \alpha \times W_{time} + \beta \times W_{place} + \gamma \times W_{quantity}$$

여기서 $\alpha + \beta + \gamma = 1$, $\alpha, \beta, \gamma > 0$ 으로서 α, β, γ 의 비율은 학습과정분석에서 시간, 장소, 학습량이 차지하는 중요도의 비율로 결정한다.

무게화련관규칙발굴과정은 다음과 같다.

단계 1 업무자료기지의 매 차원의 항목들에 대하여 빈발정도와 사용자관심에 따르는 선택순위, 항목의 계층구조에 따라 무게값들을 할당한다.

단계 2 무게화된 항목들로부터 무게화련관규칙의 무게화최소지지도(w_min_sup)를 만족시키는 무게화최대빈발항목모임들을 구한다.

단계 3 무게화최대빈발항목모임의 비지 않는 모든 부분항목들에 대하여 무게화부분항목모임들을 얻는다.

단계 4 무게화부분항목모임으로부터 무게화최소확신도(w_min_conf)를 만족시키는 무게화련관규칙들을 발굴한다.

생성한 련관규칙들로부터 학습과정에 대한 정보들을 분석하고 판단을 진행한다.

실례자료를 놓고 학습과정을 분석한 결과는 다음과 같다.

한달동안 매일 평균 400명의 학생들이 공부한다고 하면 위의 표와 같은 형식으로 업무자료기지를 구성한다.

시간, 장소, 학습량의 비율을 $\alpha : \beta : \gamma = 3 : 2 : 5$, 기준학습항목개수를 $Q_7 = 5$, 무게화최소지지도를 $w_min_sup = 0.5$, 무게화최소확신도를 $w_min_sup = 60\%$ 로 정하고 무게화련관규칙발굴을 진행하면 련관규칙들이 발생한다.

... 《user 1, $D_0 \Rightarrow T_2 P_3 Q_8$ 》, 《user 2, $D_0 \Rightarrow T_1 P_4 Q_7, T_{12} P_4 Q_8$ 》, 《user 3, $D_0 \Rightarrow T_3 P_2 Q_6$ 》, ...

발생되는 규칙들가운데서 유용하게 쓰일수 있는 규칙들을 분석한다.

실례로 규칙 《user 1, $D_0 \Rightarrow T_2 P_3 Q_8$ 》부터는 user 1이 학습날자 D_0 에 학습시간이 2-4시사이, 학습장소는 집에서 학습을 진행하여 통과된 항목개수가 8이라는것을 알수 있다. 이로부터 학습활동기록자료기지에서 분석자요구에 따라 학습과정에 대한 분석결과를 도출할수 있다.

맺 는 말

전자학습환경에서 학생의 학습활동기록자료기지에서 자료발굴기술을 리용하여 자료창고를 구축하고 련관규칙을 생성하여 학습과정을 분석하는 방법을 제안하고 목적하는 지표에 따라 분석자료범위를 최대 10%까지 줄이였다.

참 고 문 헌

- [1] 홍철; 자료발굴, 과학기술출판사, 13~37, 주체102(2013).
- [2] 공혜옥, 김철호; 자료발굴모형, 김일성종합대학출판사, 17~67, 주체107(2018).

주체109(2020)년 5월 5일 원고접수

A Method for Analyzing Study Course Using Mining of Weighted Association Rule in E-Learning Environment

O Nam Il, Kim Chol Hyok

In this paper, we have proposed an analyzing method for the study course by establishing a data warehouse from the database of students' study activity history and creating the association rules using the technique of data mining in the e-learning environment.

Keywords: e-learning, data warehouse, association rule