문항 응답 데이터에서 문항간 연관규칙의 질적 향상을 위한 도구 개발

곽은영† · 김현철††

요 약

본 논문은 연관규칙 마이닝을 이용하여 성취도 평가 결과인 문항 응답 데이터를 대상으로 의미 있는 문항간 관련성을 찾아낼 수 있는 도구를 개발하는데 연구의 목적이 있다. 제안된 도구는 의미 없는 데이터들을 제거하여 보다 더 흥미(interestingness)있는 연관규칙을 생성하도록 하며, 이러한 결과는 교수-학습 방법이나 문제은행의 질을 향상시키는데 필요한 많은 정보를 제공할 수 있을 것이다. 이를 위하여 임의의 문항 응답 실험 데이터 집합을 생성하고 정보이론(Information Theory) 기반의 surprisal 이라는 도구를 개발하여 의미 없는 데이트를 제거한 후, 연관규칙을 추출하였다. 실험 데이터는 특정 문항간 관계가 의도적으로 빈발 생성되도록 만들어지며, 추출된 연관규칙이 그러한 문항간 관계를 적절히 반영하고 있는지의 여부를 평가하고, 원본 데이터와 지지도(support) 기반으로 추출된 연관규칙과 비교함으로써 surprisal 도구의 타당성을 증명하였다.

키워드 : 문항분석, 데이터마이닝, 연관규칙

A Measure for Improvement in Quality of Association Rules in the Item Response Dataset

Eun-Young Kwak† , Hyeoncheol Kim††

ABSTRACT

In this paper, we introduce a new measure called surprisal that estimates the informativeness of transactional instances and attributes in the item response dataset and improve the quality of association rules. In order to this, we set artificial dataset and eliminate noisy and uninformative data using the surprisal first, and then generate association rules between items. And we compare the association rules from the dataset after surprisal-based pruning with support-based pruning and original dataset unpruned. Experimental result that the surprisal-based pruning improves quality of association rules in question item response datasets significantly.

Keywords : Item Analysis, Data Mining, Association Rule

1. 서 론

연관규칙 마이닝은 소비자의 구매패턴을 찾아내는 장바구니 분석에서 자주 사용되어지는 데이터 마이닝 기법 중의 하나이며, 최근에는 다른 여러 분야에서도 활용 가치가 높아지고 있다. 발견된 규

† 정 회 원: 고려대학교 컴퓨터교육과 박사과정(교신저자)
†† 종신회원: 고려대학교 컴퓨터교육과 교수
논문접수: 2006년 10월 23일, 심사완료: 2007년 1월 25일

칙의 유용성은 흥미도(interestingness)로 측정되어지며, 각 분야(domain)의 특성을 잘 반영하는 흥미로운 규칙을 추출하는 것이 중요하다.

Apriori는 트랜잭션(transaction) 데이터 집합에서 빈발항목을 찾아내는데 유용한 알고리즘으로서 발견된 연관규칙의 흥미도 평가는 일반적으로 지지도(support)와 신뢰도(confidence)로 평가되어진다[6]. 흥미도란 연관규칙이 얼마나 의미 있고 유용한지를 나타내는 정도이며, 발견된 연관규칙들이 모두 흥미 있는 것이 아니기 때문에 객관적인 흥미도를 측정하는 것이 중요하다. 그러나 지지도와 신뢰도만으로는 규칙들의 흥미도를 측정하기가 충분치 않기 때문에 흥미도를 측정할 수 있는 여러 가지 도구들이 사용되어진다. 그러나 이러한 도구들은 대개 빈발항목 집합을 분석함으로써 마케팅 분야에서 일반적으로 사용되어지는 흥미도 측정 도구이므로 다른 분야에서는 무의미한 연관규칙을 생성할 수 있는 문제점이 있다.

특히, 본 연구의 실험 대상인 문항응답 데이터에서는 각 문항의 난이도와 학생의 점수(성취능력)라는 요인들에 의하여 발생할 수 있는 빈발항목 집합으로부터 무의미한 연관규칙이 생성될 수 있다. 즉, 단순히 빈발항목 집합으로부터 연관규칙을 찾아내는 방법만을 사용하는 경우에는 다수의 무의미한 규칙들이 생성 될 수 있는데, 그 이유는 다음과 같다.

- 난이도가 쉬운 문항은 정답률이 높기 때문에 빈발항목집합으로 구성되어지고, 대부분의 연관규칙은 문항의 내용이나 평가하고자하는 의미와는 상관없이 난이도가 쉬운 문항과 발생되게 된다. 따라서 빈발항목 집합으로부터 추출된 연관규칙은 문항들 간의 의미 있는 연관성을 나타내는 것이 아니다.
- 동일 문항에 대하여 점수가 높은 학생과 점수가 낮은 학생이 정답 반응한 의미는 다른데, 문항응답 데이터 집합은 각 학생들의 점수에 대한 이러한 정보를 가지고 있다. 그러나 빈발 항목집합으로 연관규칙을 찾는 경우에는 이러한 특성을 반영하지 못하게 된다.

이에 본 연구에서는 학교에서 실시되는 성취도 평가에 대한 응답 데이터를 대상으로 문항간 연관규칙을 분석함으로써 특정 문항에 대한 정답 반응이 다른 문항의 정답 반응과 연관되어 있는 의미있는 패턴을 찾음으로써 연관규칙의 질을 향상시킬 수 있는 새로운 도구를 개발하였다. 실험에 사용된 문항응답 데이터는 <표 1>과 같다. 각 트랜잭션은 각 문항에 대한 한 학생의 응답 데이터이고, 속성(attribute)은 한 문항에 대한 모든 학생들의 응답 데이터를 나타내며 각 문항에 대하여 정답을 한 경우는 1, 오답을 한 경우는 0으로 표시하였다.

본 연구의 실험 대상인 문항 응답 데이터에서는 각 문항의 난이도와 학생의 점수(성취능력)라는 요인들에 의하여 발생할 수 있는 무의미한 연관규칙을 감소시키는데 초점을 두고 있다. 따라서 본 연구에서는 정보이론(Information Theory)기반의 도구인 *surprisal*을 이용하여 문항응답 데이터 집합 내의 모든 개체와 속성에 대한 정보성(informativeness)을 평가하고 무의미한 데이터들은 제거한 후, 연관규칙을 추출하여 문항간 연관성을 분석하였다. 이러한 결과는 실제 학교 현장에서 특정 영역에 대한 성취도 평가나 교수-학습 방법 연구, 문제은행 구성 등에 도움을 줄 수 있을 것이다.

<표 1> 문항응답 데이터

		Question Item						
		I_1	I_2	...	I_j	...	I_{m-1}	I_m
Student	S_1	1	1	...	0	...	1	1
	S_2	1	0	...	1		1	1
	...							
	S_3	1	0	...	1	...	0	1
	...							
	S_n	1	0	...	1	...	0	1

2. 관련연구

연관규칙의 흥미도 측정에 관한 연구들과 흥미도 측정 도구의 효과에 대한 논의는 계속적으로 이루어지고 있는데, 이 연구들은 주로 연관규칙을 추출하고 난 후에 다양한 도구를 사용하여 흥미도를 측정하거나 무의미한 규칙들을 가지치기(prune)하고 흥미로운 규칙을 생성하기 위한 알고리즘을 고

안하는 것이다[1][2][3][8][9][10][12][13].

대부분의 연구들은 장바구니 분석과 같은 정량적 항목들간의 이진연관규칙(boolean association rule)인 단일차원연관규칙(single level association rule)을 다루거나 서로 다른 계층의 항목과 속성을 참조하여 연관성을 설명하는 다중차원연관규칙(multi level association rule)을 다루는 것이다[4][11]. 이러한 연구들의 대상인 실험 데이터집합은 각 항목이 가지고 있는 속성간의 연관성이 상위 계층 항목간의 이진 연관성 결과에 영향을 주지 않는다. 그러나 본 연구의 대상인 문항 응답 데이터는 각 문항마다 가지고 있는 난이도와 각 학생의 점수라는 속성에 의하여 문항간 연관성의 의미가 달라지기 때문에 기존의 마이닝 방법과 흥미도 측정 도구만으로는 유용한 연관규칙을 발견 할 수 없게 된다.

3. 흥미도 측정도구

일반적으로 Apriori 알고리즘 같은 연관규칙 마이닝은 생성된 연관규칙의 흥미도를 지지도와 신뢰도로 판단한다. 지지도와 신뢰도는 두 항목이 함께 나타나는 연관성을 빈발 정도로 측정하는 기본적인 도구이며, 상관도(correlation), 확신도(conviction), Gini index, IS, PS 등과 같은 도구들도 사용된다[1][2][7][8].

본 연구에서는 제안된 *surprisal*로 무의미한 데이터를 먼저 제거한 후, 추출된 연관규칙의 흥미도를 비교하기 위하여 지지도, 신뢰도, 상관도, 확신도 도구를 사용하였다. 지지도와 신뢰도는 데이터 마이닝시 기본적으로 사용되는 도구이고, 상관도는 항목간의 연관성을 측정하며 확신도는 인과관계를 측정하기 위하여 사용하였다.

3.1 지지도와 신뢰도

지지도는 주어진 데이터 집합 내에서 함께 동시에 발생하는 항목집합의 빈발정도를 나타내는 도구로서 최소 지지도 임계값 이상의 항목집합을 빈발항목집합으로 간주한다.

$$\text{Support}(A,B) = P(A \wedge B)$$

최소 지지도 임계값으로 빈발항목집합이 생성되고, 각 k개의 빈발항목집합으로부터 $\sum_{i=1}^{k-1} {}^{k}_{i}$ 개의 연관규칙이 생성되어진다.

신뢰도는 특정 연관규칙이 발생될 수 있는 조건부 확률로서 최소 지지도 임계값으로 생성된 연관규칙의 흥미도를 측정한다,

$$\text{Confidence}(A \Rightarrow B) = P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

일반적으로 지지도와 신뢰도 값으로 연관규칙의 흥미도를 평가하는데, 이 때 흥미도란 빈발정도로서 항목간의 의미있는 관계를 나타내는 것은 아니다. 그러므로 추출된 연관규칙의 흥미도를 지지도와 신뢰도 값으로만 평가하게 되면 무의미한 규칙들이 생성될 수 있는데, 이 문제는 상관도과 확신도의 값을 비교함으로써 어느 정도 해결될 수 있다.

3.2 상관도와 확신도

상관도는 빈발항목집합의 항목간 의존도를 나타낸다.

$$\text{Correlation}(A,B) = \frac{P(A \wedge B)}{P(A) \cdot P(B)}$$

Corr(A,B)의 값이 1이면, A와 B는 각각 독립적으로 상관관계가 없으며, 1보다 크면 A로 인하여 B가 발생되는 긍정적 관계라고 볼 수 있다.

확신도는 연관규칙 A⇒B의 성립은 ¬(A∧¬B)과 같다는 관계로부터 측정되어 질 수 있다.

$$\text{Conviction}(A \Rightarrow B) = \frac{P(A) \cdot P(\neg B)}{P(A \wedge \neg B)}$$

conviction 값이 1과 같으면 두 항목간의 관계는 독립적이라 볼 수 있다.

4. 정보이론 기반의 도구 개발

"surprisal" 은 다음과 같이 정의되어진다.

$$u = -\log_2(P)$$

예를 들면, 상자 안에 있는 1,2,3이라 표시된 3개의 공을 꺼내는 경우, 3번 공이 나올 확률 P의 값이 0이란 정보를 알고 있을 때 3번 공이 나온다면 그 사건은 매우 "surprise" 한 것으로서 u는 ∞의 값을 가지게 된다. 반면에 3번 공이 나올 확률 P가 1이라는 정보를 알고 있을 때 3번 공이 나오면 그 사건은 당연한 것이기 때문에 "surprise" 하지 않으며, u 값은 0이 된다[5].

본 실험에서는 <표 1>의 문항 응답 데이터에서 각 문항들을 정답률에 의하여 정렬하고 각각의 학생 S_i에 대하여 각 문항에 정 또는 오로 응답한 것에 대한 "surprisal" 값을 구하였다. 만약, 어떤 학생의 총 정답 문항수가 k개라는 것을 안다면, 그 학생은 총 문항 m개 중 정답률이 높은 문항 순으로 k개를 맞혔다고 추측할 수 있다. 실제로 그 학생이 정답률이 높은 문항 순으로 k개를 맞혔다면 이러한 사실은 전혀 "surprise" 하다고 볼 수 없다. 반면에 정답 문항 k개 중 정답률이 낮은 문항의 개수가 많을수록, 그 사실은 "surprise" 하다고 볼 수 있다. 이러한 사실을 기반으로 "surprisal" 정도를 측정할 수 있다.

학생 S_i가 특정 문항 I_j를 맞힐 확률을 $P(B_{i,j})$라 정의한다. $P(S_i)$는 학생 S_i의 모든 문항에 대한 $P(B_{i,j})$이며, $P(I_j)$는 문항 I_j의 모든 학생에 대한 $P(B_{i,j})$로 정의한다. 이것을 확률적으로 계산한다면 $P(I_j)$는 문항 I_j의 정답률로, $P(S_i)$는 학생 S_i의 점수가 발생할 확률로서 구할 수 있다.

$$P(S_i) = \frac{\sum_1^j P(B_{i,j})}{m}$$

$$P(I_j) = \frac{\sum_1^i P(B_{i,j})}{n}$$

P(M)는 모든 i , j에 대한 $P(B_{i,j})$의 평균으로 정의하며 다음과 같이 구한다.

$$P(M) = \frac{\sum_1^i \sum_1^j P(B_{i,j})}{n \cdot m} = \frac{\sum_1^j P(I_j)}{m} = \frac{\sum_1^i P(S_i)}{n}$$

u 값을 구하기 위하여 $P(B_{i,j})$를 구하는데, 다음과 같이 식을 계산할 수 있다.

$$\frac{P(S_i)}{P(M)} = \frac{\sum_1^j P(B_{i,j})/m}{P(I_j)/m}$$

$$\sum_1^j P(B_{i,j}) = \frac{P(S_i) \cdot \sum_1^j P(I_j)}{P(M)}$$

이 때, i번째 학생에 대하여 $P(B_{i,j}) \propto P(I_j)$ 이므로 다음과 같이 $P(B_{i,j})$를 구할 수 있다.

$$P(B_{i,j}) = \frac{P(S_i)}{P(M)} \cdot P(I_j) \qquad (1)$$

$P(B_{i,j})$는 확률 값으로 1보다 작거나 1 값만을 가질 수 있는데, $\frac{P(S_i)}{P(M)} > 1$ 이면 $P(B_{i,j})$는 1보다 큰 값을 가지게 된다. 이 경우에는 1보다 작은 값을 가지는 다른 문항에 대한 $P(B_{i,j})$에게 $(1-P(B_{i,j}))$ 값만큼 주게 된다. 즉, 한 학생에 대하여 모든 문항에 대한 $P(B_{i,j})$ 값이 구해지고, 1보다 큰 값을 가지는 모든 $P(B_{i,j})$에 대하여 $(P(B_{i,j}) -1)$의 총합이 구해진다. 그 후, 1보다 작은 $P(B_{i,j})$들 중 가장 큰 값을 가지는 값에 우선적으로 $(1-P(B_{i,j}))$만큼 더해져서 1값을 만들게 되는 것이다.

(1)의 공식으로부터 학생 S_i가 문항 I_j에 정답

또는 오답한 것에 대한 surprisal u 값을 구할 수 있다.

$$u(B_{i,j}) = -\log P(B_{i,j})$$

각 학생 S_i 또는 각 문항 I_j에 대한 surprisal은 특정 i 또는 j에 대한 $u(B_{i,j})$의 평균으로 구할 수 있다.

$$u(S_i) = \frac{\sum_{1}^{j} u(B_{i,j})}{m}$$

$$u(I_j) = \frac{\sum_{1}^{i} u(B_{i,j})}{n}$$

5. 실험결과

제안된 surprisal u의 타당성을 증명하기 위하여 임의의 데이터 집합을 만들어 실험하였다. 본 실험에서는 surprisal u로 각각의 학생(개체:instance), 문항(속성:attribute)에 대한 흥미도 또는 의미성을 측정하여 원본 데이터집합에서 무의미한 규칙을 생성할 수 있는 학생과 문항 데이터들을 먼저 제거한 후, 연관규칙을 추출하였다. 이렇게 추출된 연관규칙들을 원본 데이터 집합과 기준 지지도 값 이상의 항목으로 구성된 데이터집합으로부터 추출된 연관규칙들과 비교하였다.

실험 문항응답 데이터 집합은 1,0000명의 학생이 10개 문항에 대하여 응답한 것으로 난수 발생에 의하여 임의로 생성되도록 하였다. 각 문항의 정답률은 10~90%로 임의로 생성되도록 하였으며, 학생들의 점수 $P(S_i)$는 정규분포를 따르도록 하였다. <표 2>는 원본 데이터 집합의 각 문항에 대한 정답률과 점수 분포이다. 실험에서 특정 두 개 문항의 연관성을 의도적으로 조성하고 surprisal로 무의미한 데이터를 제거하고 난 후, 정제된 데이터 집합에서 연관규칙을 추출하였다. 추출된 연관규칙이 의도한 두 문항의 관계를 반영하는지의 여부를 평가하기 위하여 문항 6번과 2번, 문항 10번과 4번이 각각 90%의 연관성이 발생하도록 실험 데이터 집합을 만들었다. 이것은 문항 6을 맞힌 학생의 90%가 문항 2번도 맞히고, 문항 10번을 맞힌 학생의 90%가 문항 4번도 맞힌다는 것을 의미한다. 따라서 surprisal u의 타당성은 추출된 연관규칙에서 두개의 연관규칙 $I_6 \rightarrow I_2$과 $I_{10} \rightarrow I_4$이 발견되는 것으로서 증명되어질 수 있다.

본 실험은 다음과 같이 진행되었다: (1) 우선, 최소 지지도와 surprisal 값으로 원본 데이터 집합의 데이터를 제거하여 정제한다; (2) 신뢰도, 상관도, 확신도 도구를 사용하여 정제된 데이터 집합으로부터 연관규칙을 추출한다; (3) 정제된 데이터 집합과 원본 데이터 집합으로부터 추출된 연관규칙을 비교한다.

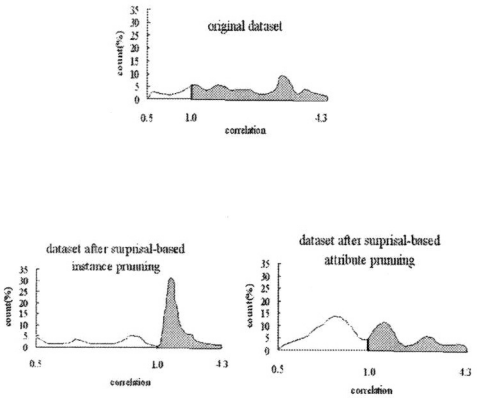
<표 2> 실험 데이터 집합의 문항 정답률과 점수 분포

문항	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
정답률	9018 (90%)	6913 (70%)	7010 (70%)	6004 (60%)	5992 (60%)	4920 (50%)	5064 (50%)	2888 (30%)	1063 (10%)	996 (10%)

점수	0	1	2	3	4	5	6	7	8	9	10
학생수	200	200	200	1,400	1,600	2,300	2,700	900	100	200	200

<그림 1>은 원본 데이터 집합과 surprisal로 정제된 데이터 집합의 각 문항간 상관도 값의 분포로서, surprisal 값으로 정제된 데이터 집합이 원본 데이터 집합보다 1보다 큰 상관도 값을 더 많이 가지고 있는 것을 볼 수 있다. 이것은 surprisal로 정제된 데이터 집합에서 의미 있는 연관성이 더 많이 발생한 것을 의미한다. 학생, 문항 데이터를 정제하기 위하여 각각의 surprisal 최소 임계값을 0.5, 0.3으로 정하였고, 최소 지지도

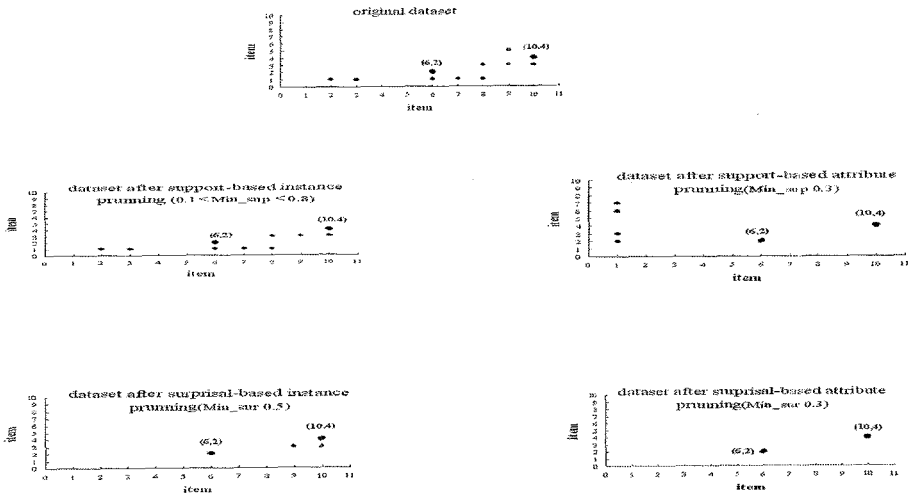
임계값은 각각 0.2와 0.3으로 하였다. Apriori 알고리즘을 이용하여 연관규칙을 생성하고 지지도 신뢰도, 상관도, 확신도 도구로 각 순위를 비교하였다.



<그림 1> 원본 데이터 집합과 surprisal로 정제된 데이터 집합의 문항간 상관도 분포

<표 3>은 원본 데이터 집합으로부터 추출된 연관규칙들을 각 도구로서 비교한 것이다. 실험에서 의도적으로 연관성을 부여한 규칙 (I_6, I_2)과 (I_{10}, I_4)는 지지도와 신뢰도 값으로는 낮은 순위이지만, 상관도와 확신도 값은 상대적으로 높은 순위에 속하고 있다. 그러나 다른 무의미한 규칙들도 생성되어진 것을 볼 수 있다.

본 실험에서는 surprisal 값을 이용하여 각각의 학생 개체에 대한 surprisal 값을 계산하고 0.5보다 작은 값을 가지는 2,155개의 개체는 제거하였다. 그 결과 7,845개의 개체가 정제되었다. 또한 각 문항 속성에 대한 surprisal 값을 구하고, 0.3보다 작은 값을 가지는 문항 1 번과 문항 3번 속성을 제거하였다. <표 4>는 surprisal로 정제한 데이터 집합으로부터 추출된 연관규칙들을 각 도구로 비교한 것으로, 무의미한 연관규칙은 생성되지 않고 임의로 의도한 규칙인 (I_6, I_2)과 (I_{10}, I_4)만이 추출된 것을 볼 수 있다. 또한 지지도를 이용하여 데이터를 정제한 후 연관규칙을 추출하여 비교하였다. <그림 2>는 원본 데이터 집합, 기준 지지도 값으로 정제한 집합, surprisal 값으로 정제한 집합으로부터 추출된 각 연관규칙을 비교한 것이다. surprisal로 정제한 집합으로부터 추출된 연관규칙이 실험데이터에서 의도했던 문항간 연관성을 가장 잘 반영하고 있고, 원본데이터나 지지도 값으로 정제한 집합에서는 무의미한 연관규칙들이 상대적으로 많이 발생한 것을 볼 수 있다.



<그림 2> 원본데이터집합과 support, surprisal로 정제된 집합에서 추출된 연관규칙

<표 3> 원본 데이터 집합으로부터 추출한 연관규칙

순위	support		confidence(≥80%)		correlation(≥1)		conviction(≥1)	
	문항	값	문항	값	문항	값	문항	값
1	2,1	67	7,1	94	**10,4**	1.43	9,3	3.40
2	3,1	60	8,1	93	9,3	1.37	**10,4**	2.80
3	7,1	47	6,1	91	10,3	1.30	10,3	2.38
4	6,1	40	3,1	91	9,5	1.30	**6,2**	2,20
5	**6,2**	39	9,3	90	8,3	1.25	8,3	1.97
6	8,2	27	2,1	89	**6,2**	1.18	9,5	1.95
7	8,3	24	**6,2**	89	7,1	1.06	7,1	1.83
8	10,3	12	10,3	86	8,1	1.05	8,1	1.60
9	**10,4**	12	**10,4**	86	3,1	1.02	6,1	1.21
10	9,5	9	8,3	83	6,1	1.02	3,1	1.21
11	9,3	9	9,5	80	2,1	1.02	2,1	1.03

<표 4> surprisal로 정제된 데이터 집합으로부터 추출한 연관규칙(상: instance 제거, 하: attribute 제거)

순위	support		confidence(≥80%)		correlation(≥1)		conviction(≥1)	
	문항	값	문항	값	문항	값	문항	값
1	**6,2**	30	**6,2**	86	**10,4**	1.45	9,3	2.53
2	10,3	9	9,3	86	9,3	1.34	10,4	2.39
3	**10,4**	9	10,3	82	10,3	1.28	10,3	1.99
4	9,3	6	**10,4**	82	**6,2**	1.16	**6,2**	1.86

순위	support		confidence(≥80%)		correlation(≥1)		conviction(≥1)	
	문항	값	문항	값	문항	값	문항	값
1	**6,2**	39	**6,2**	89	**10,4**	1.43	**10,4**	2.80
2	10,4	12	10,4	86	**6,2**	1.19	**6,2**	1.20

6. 결론 및 제언

본 논문에서는 정보이론 기반의 도구인 surprisal을 제안하였다. surprisal은 데이터 집합 내의 개체(instance)와 속성(attribute)들이 가지고 있는 정보성을 측정하여 정보성이 낮고 무의미한 데이터는 제거함으로써, 생성되어지는 연관규칙의 흥미도를 향상시키게 된다. 특히, 본 실험의 대상 데이터인 문항 응답 데이터와 같이 각 학생 개체의 점수와 문항 속성의 정답률에 대한 정보를 알고 있는 경우, surprisal은 추출되는 연관규칙의 흥미도를 향상시키는 데 효과가 있다.

이러한 결과는 학교 현장에서 성취도 평가 분석 시, 기존의 평가 도구에서는 발견하지 못했던 문제점들을 해결하는데 도움이 될 것이다. 예를들면, $I_1 \rightarrow I_2$이 발생했을 때, 학생들은 1번 문항의 내용을 유추하여 2번 문항을 정답으로 할 수 있다는 경우를 생각해 볼 수 있다. 표면적으로 1번 문항과 2번 문항의 평가 내용에 대한 학생들의 이해도가 높다고 생각할 수 있지만, 실제로는 그렇지 않다는 것을 알게 된다면, 평가자는 학생들의 학습 내용에 대한 이해도와 학습 방법에 대한 문제점을 파악하여 긍정적으로 해결할 수 있을 것이다. 또한 두 문항이 동시에 출제되지 않도록 문제은행 시스템을 구성할 수도 있을 것이다.

이에, 향후에는 학교 현장에서 실시된 실제 평가 응답 데이터를 대상으로 실험하고, 현장에서의 타당성과 활용성을 입증함으로써 교수-학습 방법, 성취도 평가, 문항평가 등의 교육 분야에 기여할 수 있는 방법을 연구하는 것이 필요하다.

참 고 문 헌

[1] A.A. Freitas(1999). On rule interestingness measures. Knowledge-Based Systems.

[2] A. Silberchatz and A. Tuzhilin(1995). On subjective measures of interestingness in knowledge discovery. Poceeings of the 1st Int. Conf. on Knowledge Discovery and Data Mining.

[3] C. Silverstein, S. Brin, and B. Motwani (1 997). Beyond market baskets-generalizing associationules to dependence rules. Data Mining and Knowledge Discovery.

[4] H. Lu, L. Feng and J. Han(2000), Beyond Intratransation Association Analysis, *ACM Transactions on Information System*, 4:423-454.

[5] M. Tribus(1961). Thermostatics and thermodynamics. D. van Nostrand Company, Inc.

[6] F. Hussain, H. Liu, E. Suzuki, and H. Lu(2000). Exception rule mining with a relative interestingness measure. Proceedings of the 4th Pacic-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications.

[7] J. Han and M. Kamber(2000). Data Mining Concepts and Techniques. Morgan Kaufmann.

[8] P. Tan and V. Krumar(2000). Interestingness measures for association patern. Technical Report TR00-036, Department of Computer Science, University of Minnestota.

[9] P. Tan, V. Krumar, and J. Srivastava (2002). Selecting the right interestingness measure for association patterns.

[10] S. Brin, R. Motwani, and Silverstein. C(1998). Beyond market baskets: Generalizing asso-ciation rule to correlation. Data Mining and Knowledge Discovery.

[11] S. Fortin, L. Liu(1996), An Object-Oriented Approach to Multi-Level Association Rule Ming, *Proceedings of the fifth international conference on Information and knowledge management*, 65~72

[12] T. Brijs, K. Vanhoof, and G(2000). Wets. Dening interestingness for association rules.Intelligent Data Analysis, 229~240.

[13] T. Imielinski, L. Khachyan, A. Abdulghani, and Cubergrades(2000). Generalizing associa-tion rules. Technical report, Dept. Computer Science, Rutgers Univ.

곽 은 영

1992 고려대학교 사범대학 가정교육과(교육학학사)
2001 고려대학교 교육대학원 컴퓨터교육전공(교육학석사)
1993~서울 백암고등학교 교사
2002~현재 고려대학교 대학원 컴퓨터교육학과 박사과정
관심분야: 컴퓨터교육, 데이터마이닝, 평가시스템
E-Mail: key@comedu.korea.ac.kr

김 현 철

1988 고려대학교 전산과학과 학사
1990 미조리 주립대학 (Rolla) (전산학석사)
1998 플로리다 대학 (전산학박사)

1998~1999 삼성 SDS 책임컨설턴트
1999~현재 고려대학교 컴퓨터교육과 교수
관심분야: 컴퓨터교육, 데이터마이닝, 기계학습알고리즘, 바이오인포매틱스
E-Mail: hkim@comedu.korea.ac.kr