

연관규칙을 이용한 지능적 질의 처리 시스템 *

이동하, 김성민, 남도원, 이진영

포항공과대학교 전자계산학과 지능정보시스템연구실

Intelligent Query System using Association Rule

Dong-Ha Lee, Sung-Min Kim, Do-Won Nam, Jeon-Young Lee

Dept. of Computer Science and Engineering POSTECH Pohang Kyungbuk 790-784 KOREA

{dongha, smkim, irene, jeon}@white.postech.ac.kr

개 요

연관규칙을 응용하여 데이터베이스 질의를 지능적으로 처리하는 시스템을 소개한다. 데이터베이스의 질의는 사용자의 의도를 반영하여 전체 데이터베이스 중 관심있는 부분을 지정하는 역할을 한다고 볼 수 있다. 질의내의 항목을 대상으로 연관규칙을 응용하는 것이 가능하다면, 사용자의 관심부분에 대한 규칙을 추출할 수 있다. 사용자질의로부터 추출한 연관규칙은 지능적 질의처리에 활용된다. 지능적 질의처리로는 확장질의, 비교질의, 제안질의 등으로, 사용자의 원래 질의에 부가적인 정보를 추가하여 사용자의 의사결정을 돕는다. 또, 사용자의 데이터베이스 이용 패턴은 사용자 그룹으로 일반화될 수 있다.

키워드: 지식탐사와 데이터 마이닝(Knowledge Discovery and Data Mining), 지능 질의 처리(Intelligent Query Processing)

1 서론

대용량의 데이터베이스분야에 기존의 인공지능, 기계학습의 연구를 접목시키는 분야로 지식탐사(knowledge discovery and data mining)가 생겨났다. 지식탐사에는 다양한 지식패턴을 데이터베이스에서 추출하고 응용하는데, 대표적인 것으로는 클러스터링(clustering)정보나, 클래스의 특성추출(class characterization), 연관 규칙(association rule) 추출등을 들 수 있다. 본 연구에서는 연관규칙을 이용하여 지능적 처리를 수행하는 시스템의 구현을 목적으로 한다.

연관규칙(Agrawal *et al.*, 1993), (Savasere *et al.*, 1995)은 그 형

태의 명료함과 잠재적인 응용분야의 다양함으로 인해 최근 많은 연구가 수행되었다.

연관규칙은 처음 소개된 후(Agrawal *et al.*, 1993), 대용량의 데이터베이스에 응용되기 위한 기법을 적용하는 등의 연구방향(Agrawal *et al.*, 1994), (Houtsma, M. and A. Swami, 1995), 성능을 위주로 고려한 해성이나, 통계적인 정보를 이용하는 연구방향(Park *et al.*, 1997), (Agrawal *et al.*, 1996) 그리고 연관규칙의 표현력을 높여보려는 연구(Srikant and Agrawal, 1996)등이 있었다.

최근 들어 연관 규칙에 대한 관심은 더욱 더 높아지고 많은 연구가 되고 있는데, 이런 배경에는 연관규칙이 항목의 집합들로 이루어진 데이터베이스에서 항목들간의 존재 의존

* 이 논문은 POSCO의 기술개발연구비 지원으로 수행하였음(과제번호 1UD9701601)

성을 반영하는 지식이기 때문이다. 즉, A 항목은 B 항목과 잘 나타난다는 정보를 가지고 있는 연관 규칙은 카탈로그 배포, 상품 진열 등의 결정에 중요한 정보를 가지고 있으며, 다양한 시장 데이터베이스(market database)에 적용되어 성공적인 결과를 얻고 있다.

본 연구에서는 연관규칙을 기존의 관계형 데이터베이스를 대상으로 하는 지능적 질의 처리(Gaasterland *et al.* 1992),(Han *et al.* 1996)에 응용하여, 사용자의 의도와 잠재적 정보 요구를 예측하고, 미리 제공해주는 시스템을 설계한다. 지능적 질의처리란 사용자의 질의에 대한 답을 데이터베이스에서만 추출하는 것이 아니고, 기존의 지식을 활용하여, 부가적이거나 참고적인 답을 같이 제출하거나, 질의 자체를 수정하여 효과적인 의사결정을 돕는 질의 처리 방식이다.

단순한 시장 데이터베이스를 대상으로하는 연관규칙의 기존 알고리즘을 이용하여 지능적 질의처리를 수행하기에는 문제가 있다. 지식 탐사 분야에서 도입된 중요한 지식 패턴 중의 하나인 연관 규칙은 본래, 값의 집합에서 동시에 나타나는 값간의 의존성을 반영하는 패턴이기 때문에, 관계형 데이터베이스에서의 데이터 표현 방식인 튜플 구조나 질의에 대해서는 직접 적용이 안되기 때문이다. 먼저 질의 표현 자체가 지식탐사의 대상으로서는 너무 복잡한 표현이기 때문에, 질의 표현을 변형하여 관계형 데이터베이스의 튜플형태로 변환시킨다. 그러면 관계형 데이터베이스에서 연관규칙을 추출하는 문제만 남는다.

이를 기술하기 위해, 본 연구에서는 관계형 데이터베이스에서의 연관규칙 추출방법, 질의로부터 연관규칙 추출방법 그리고, 이를 이용한 지능적 질의 처리 기법에 대해 단계적으로 설명한다.

2 지식탐사와 연관규칙

2.1 지식탐사

지식 탐사(knowledge discovery)란 대용량의 단편적인 데이터베이스에서 유용한 정보를 추출하고 이용하는 일련의 과정을 일컫는 용어이다. 최근 이 분야는 데이터베이스, 연역 데이터베이스, 인공지능, 기계 학습, 통계학 등 기존의 여러 연구 분야의 기술들을 기반으로 독자적인 연구 분야로 구축되

고 있다.

지식 탐사를 통해 흥미 있는 지식, 데이터의 규칙성, 높은 수준의 정보들이 추출될 수 있고, 다른 각도에서 조사될 수 있다. 이 과정에서 데이터베이스는 지식의 생산과 검증을 위한 정보의 제공자 역할을 한다.

지식 탐사의 응용 분야로는 정보 관리, 질의 처리, 의사결정, 프로세스 제어 등등 다양하며 최근 발전하는 온라인 정보서비스와 웹정보 서비스에도 보다 나은 정보의 제공과 사용자의 행동 양식을 파악하는 등에 지식 탐사 기술은 중요한 역할을 수행할 것으로 예상되고 있다.

데이터 마이닝의 결과로써 나타나는 지식 단편의 종류로는 연관 규칙, 일반화된 데이터베이스 내의 테이블, 분류 또는 클러스터링 규칙 등등이 있다.

$L_1 := \{\text{frequent 1-item sets}\};$

for ($k := 2; L_{k-1} \neq \emptyset ; k++$) **do begin**

$C_k :=$ New candidates of size k generated from $L_{k-1};$

forall transactions t in database **do begin**

$C_t = \text{subset}(C_k, t)$

Increment the count of all candidates in C_k that are contained in $t;$

end

$L_k :=$ All candidates in C_k with minimum support;

end

Answer := $\cup_k L_k ;$

그림 1. 충분 지지 항목 추출

2.2 연관규칙

연관규칙은 $a \Rightarrow b$ 의 형태를 갖는 패턴으로서, a 와 b 는 항목의 집합이다. 이 $a \Rightarrow b$ 형태의 연관규칙이 갖는 의미는 a 항목 집합이 나타날 때는 b 항목 집합도 동반하여 나타나는 경향이 있다는 뜻이다.

연관 규칙의 구체적인 예는 다음과 같다. "빵을 구매하는 사람의 40%는 우유도 구매하며, 전체 트랜잭션의 2%는 빵과 우유를 포함하고 있다". 여기서 40%는 이 규칙의 신뢰도(confidence)라고 불리고, 2%는 이 규칙의 지지도(support)라고 불린다. 연관 규칙 탐사는 주어진 데이터베이스에서 사용자

가 미리 정의한 최소 신뢰도와 최소 지지도를 초과하는 모든 연관 규칙을 찾아내는 것이다. 이러한 연관 규칙을 강한 규칙(strong rule)이라고 한다. 신뢰도는 규칙의 강도를 뜻하고, 지지도와 우세도는 그 규칙이 갖는 통계적 중요성을 반영한다.

연관 규칙 탐사의 응용 분야는 개인 구매도 분석, 상품의 교차 매매(cross-marketing), 카탈로그 디자인, 열가 매출품(loss leader)분석, 상품 진열, 구매 성향에 따른 고객 분류 등등 다양하다.

연관규칙추출문제는 주어진 시장 데이터베이스에서 정해진 신뢰도와 지지도를 만족하는 모든 연관규칙을 찾아내는 것이다. 이를 위해서는 다수항목집합을 추출하고, 다수항목집합을 이용해 연관규칙을 조합해내는 두 가지 단계로 알고리즘이 구성되어 있다.

Procedure Discovering Rules

```
forall large itemsets  $l_k, k > 1$  do
  call GenRules( $l_k, l_k$ );
```

Procedure GenRules(l_k, A_m)

```
A = { (m-1)-itemsets that are subset of  $a_m$  };
forall  $a_{m-1}$  in A do begin
  conf = support( $l_k$ ) / support( $a_{m-1}$ );
  if (conf >= min_conf) then begin
    output the rule  $a_{m-1} \Rightarrow (l_k - a_{m-1})$ ;
    if (m-1 > 1) then
      call genrules( $l_k, a_{m-1}$ );
  end
end
```

그림 2. 연관규칙생성

```
{Korea, Foreign}  $\subset$  Any
{Seoul, Not_Seoul}  $\subset$  Korea
{Kyungbuk, Kyungnam, Kangwon,...}  $\subset$  Not_Seoul
{Chunchon, Wonju, ...}  $\subset$  Kangwon
```

그림 3. "출생지" 속성에 주어진 개념 계층 트리

다수항목집합이란 데이터베이스에서 동시에 나타나는 항목들의 집합이다. 그림 1 에서 다수항목집합을 추출하는 기본 알고리즘을 보였다. 다수항목집합을 추출할 때 여러 번의

Procedure MLLA(database DB, concept hierarchies A)

```
 $L_1 :=$  {frequent 1-item sets};
for (  $k := 2; L_{k-1} \neq \emptyset ; k++$ ) begin
   $C_k :=$  New candidates of size  $k$  generated from  $L_{k-1}$ .
  forall transactions  $t$  in database do begin
    Add all ancestors of each item in  $t$  to  $t$ , removing
    any duplicates.
    Increment the count of all candidates in  $C_k$  that
    are contained in  $t$ .
  end
   $L_k :=$  All candidates in  $C_k$  with minimum support.
end
Answer :=  $\cup_k L_k$ ;
```

그림 4. 다단계 충분 지지 항목 추출

데이터베이스 전체를 읽어들이는 과정이 필요하다. 일단 다수항목집합이 추출되면, 다수항목집합으로부터 연관규칙을 추출하는 단계에서는 데이터베이스를 검색하지 않는다. 연관규칙 $a \Rightarrow b$ 가 성립하기위해서 지지도와 신뢰도만 조사하는데 신뢰도는 $\text{support}(ab) / \text{support}(a)$ 로 표현되기 때문이다.

2.3 다단계 연관규칙

항목들간에 계층적 개념이 포함되는 경우가 있다. 이러한 정보를 표현하는 방법중의 하나는 개념 계층 트리라고 불리는 데이터 구조이다.

다단계 연관규칙(Han and Fu, 1995),(Srikant and Agrawal, 1995)이란 항목들에 주어진 개념 계층 트리를 이용하여 여러 단계의 일반화된 연관규칙을 말한다. 그림 3 에 주어진 출생지 정보를 이용한다면, 각 개인의 정보로부터 얻는 연관규칙보다 상위개념의 일반화된 정보를 추출 할 수 있다.

그림 4. 에 기본적인 다단계 연관 규칙 추출을 위한 충분한 지지도를 가진 항목 추출 기법을 보인다. 주어진 개념 계층은 각 속성들에게 주어질 수 있다. 여러 회사의 제품 각각의 연관 관계보다는 모든 회사를 묶은 우유와 빵간의 연관

관계에서 유용한 결과가 나올 가능성이 있다.

3 관계형 데이터베이스에서의 연관규칙

여러 테이블(또는 객체형 데이터베이스의 경우 클래스)들에 걸친 연관 규칙을 추출하는 알고리즘으로 ICAR(Inter-Class Association Rules)알고리즘이 제안되었다(Lee *et al.* 1997). 그러나 ICAR의 객체 추상화 방법은 bottom-up 방식이고, 객체 추상화의 각 단계에서 속성위주일반화기법에 의존하므로 결과로 얻는 객체 계층 구조의 일관성이 부족하다. 본 논문에서는 ID3(Quinlan, 1986)와 유사한, 엔트로피이론에 기반한 객체 계층 구조를 추출한다. 튜플 추상화 기법은 각 테이블 내의 튜플을 객체로 추상화하는 단계, 추상화된 객체를 합병해서 대상 트랜잭션을 생성하는 단계, 그리고 합병된 트랜잭션에 대해 다단계 연관 규칙 추출 알고리즘을 적용하는 단계로 구성되어 있다.

sid	Name	Age	Birth_place
9325d07	Dong-Ha Lee	30	Seoul
9425d05	Dong-Yal Seo	29	ChunChon
9625m20	Do-Won Nam	25	TaeGu
...

그림 5. student 테이블

3.1 객체 추상화

이 단계의 목적은 기본 테이블(base table: 실제 데이터베이스에 저장되어 있는 관계형 테이블)의 각각의 튜플을 객체로 추상화하고 이들 객체를 계층 구조에 포함시키는 단계이다. 예를 들어 다음과 같은 어느 학과의 학생에 대한 테이블이 있다고 가정한다. 이 테이블을 top-down 기법을 이용하여 그림 6 와 같은 student 객체의 계층구조를 얻고자 하는 것이 객체 추상화 단계의 목표이다.

이러한 객체 계층구조를 추출하기 위해, '바람직한' 객체 계층구조로서는 두가지 판단지표가 있을 수 있다. 첫번째는 계층구조내의 분기수(branch number)이다. 분기수는 결과로 얻는 연관규칙의 개수와 관련이 있다. 가지수가 너무 많은 경우 결과로 얻은 객체 계층구조의 너비가 너무 넓어지고, 최종적으로 얻는 연관규칙이 너무 구체적인 사례가 많아서, support 를 대부분 만족시키지 못하는 경우가 생긴다.

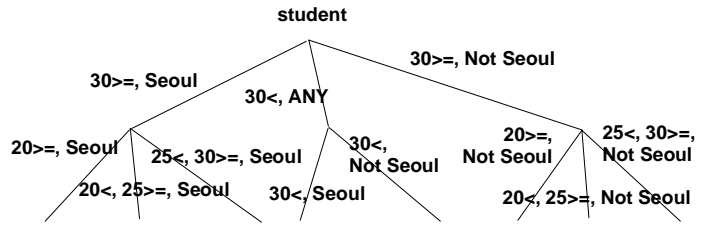


그림 6. student 테이블의 객체 계층 구조

두번째는 객체 계층 구조내에 student 테이블의 튜플을 배열하여 보았을 때의 가급적 균일하게 분포되어야 한다는 것이다. 이러한 균형(balance)문제는 결과로 얻는 객체 계층 구조의 일관성에도 관련이 있다.

엔트로피(entropy)는 다음과 같이 정의된다

$$\text{Entropy}(S) = -P(\text{pos}) \log P(\text{pos}) - P(\text{neg}) \log P(\text{neg})$$

N개의 클래스로 구분되는 경우는 다음과 같이 정의된다.

$$\text{Entropy}(S) = -\sum P(C_i) \log P(C_i)$$

Student 테이블의 객체 계층 구조에서, 정보이득(information gain)이 가장 높은 쪽으로 분기를 하면 된다.

일반적인 정보이론에서 주어진 속성 A와 집합 S에서 정보이득 G(S,A)은 다음과 같이 정의된다.

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \text{Entropy}(S_v) |S_v|/|S|$$

이러한 정의를 이용하여 튜플을 일반화하는 BTA 알고리즘을 그림 7에 보였다. student 테이블 내의 튜플을 표현하는 객체는 그림 6의 생성된 객체 계층 구조를 따라 일반화될 수 있다. 즉 sid 를 9325d07로 갖는 학생은 CC(25<, 30>=, Seoul)로 이름지어지는 상위 개념 클래스(concept class)를 가지고, 또 이 개념 클래스는 CC(30>=, Seoul)로 이름지어지는 상위 개념 클래스로 일반화된다.

3.2 대상 트랜잭션 데이터베이스 구성

대상 트랜잭션(target transactions)은 실제 마이닝 알고리즘이 수행되는 대상이다. 따라서 ICAR알고리즘의 대상이 되는 테이블들을 조인했을 때 조인이 가능한 튜플의 조합을 결정하고 조인에 참여하는 튜플을 캡슐화한 객체들을 한 테이블에서 하나씩 추출하여 대상 트랜잭션을 생성한다.

4 지능적 질의 처리

Procedure BTA(table T, attributes with concept hierarchies A)

```
Create a node R for the tree;
If the number of tuples in T is less than threshold min_t, or A is
empty Return;
Else begin
  SA ← the attributes from A that classifies T with highest
  information gain and less than threshold max_s;
  Let Ti be the subset of T that are split by SA;
  For each possible table Ti do begin
    Add a new tree branch below R with label of
    corresponding concept of SA;
    NA ← A - attributes from SA whose corresponding
    concept nodes are leaves .
    BTA(Ti, NA);
  End
End
```

그림 7. BTA 알고리즘

예를 들어, 학생 테이블과 과목 테이블과 수강 테이블이 정의되어 있는 관계형 스키마를 그림 8 와 같이 가정한 경우, 대상 트랜잭션은 모두 세 개의 항목만으로 이루어진 (sid, cid, rid)의 형태이다. 대상 트랜잭션 데이터베이스는 이들 세 개의 항목을 가상적으로 결과로 구성된다. 즉 대상 트랜잭션의 수는 기존의 알고리즘과 같이 세 테이블의 조인한 결과 테이블의 튜플 수와 같다.

```
student(sid, name, status, major, gpa,
        birth_date, birth_place, address)
course(cid, name, dept)
register(rid, sno, cno, grade)
```

그림 8. 예제 스키마

이렇게 얻어진 대상 테이블에 이미 알려진 다단계 연관 규칙 탐사 알고리즘을 적용함으로써, 여러 테이블의 항목에서 추출된 연관 규칙을 얻을 수 있다. 이러한 전체적인 알고리즘이 그림 9 에 제시되었다.

4.1 지능적 질의 처리

지능적 질의 처리란 사용자의 질의에 대해 직접적인 응답만을 사용자에게 제공하는 것이 아니라, 시스템이 보유하고 있는 정보를 활용해서 질의의 응답과 함께 보다 많은 정보를 제공을 해 주는 것이다. 관련된 연구로는 협조적 질의 응답(cooperative query answering) (Cuppens and Demolombe, 1988) 과 내재적 응답(Shum and Muntz, 1988)이 있다.

Procedure ICAR2(Tables T, Interesting attributes IA)

```
filter out non interesting attributes from T if it is neither primary
or foreign key ;
For each table t in T do
  extract the generalization path using BTA;
  build target transactions schema collecting the representative
  attributes ;
  build target transactions database by join the encapsulated
  objects
  multi-level association rule mining;
  convert the result to labeled rule;
```

그림 9. ICAR2 알고리즘

최근 지식탐사기법을 지능질의 처리에 응용하려는 연구가 있었다(Han *et al.* 1996). Han의 연구는 자신들이 제안한 지식-부유 데이터 모델 (knowledge-rich data model)이 중심이 되어 일반화기법을 주로 이용하여 지식탐사를 수행한다.

우리는 데이터베이스를 접근하는 사용자의 행동양식 (access pattern)를 분석함으로써 보다 협조적인(cooperative) 형태를 보이는 시스템을 구축하려고 한다.

튜플 형태의 데이터에서 연관규칙을 추출할 수 있으므로, 질의에서 유용한 정보를 추출하여 튜플 형태로 표현할 수 있다면, 질의에서 연관 규칙을 추출할 수 있다. 데이터베이스시스템과 사용자 사이의 질의와 응답을 추상화하여 이를 통해 연관규칙을 추출하는 것이 본 시스템의 목표이다. 이를 위해서 질의패턴과 응답패턴을 정의한다. 사용자가 데이터베이스로부터 원하는 정보를

기술하는 SQL문으로부터 질의패턴을 추출한다. 질의패턴에는 질의가 요청된 시간, 질의를 요청하는 사용자, 질의의 응답으로 선택될 속성명들의 집합, 속성들에 제한된 조건들이 포함되어 있다. 응답패턴은 데이터베이스가

```

discover    association rules
from        query q, user u
where       u.birth_place ="Seoul" and u.u_id =
              q.u_id and ts in [98/01/01,
              98/03/01]
with        support=0.01, confidence=0.4
    
```

그림 10. A pattern discovery query from query pattern

```

ts: time stamp
u_id: user id
at_id: attribute id
cond: constant condition
val: retrieved values
    
```

그림 11. 질의 패턴 튜플과 응답 패턴 튜플

반환하는 질의 응답으로부터 추출되며, 응답이 제출되는 시간, 질의를 요청한 사용자, 질의의 응답으로 선택되는 속성명, 속성에 반환되는 값으로 구성된다. 하나의 질의응답에 대해 여러 개의 응답패턴이 생성된다.

이러한 패턴들은 확장 관계형 데이터모델(*extended relational data model*)에 속한다. 즉, 튜플 추상화기법으로 연관규칙을 추출할 수 있다. 이때 필요한 배경지식으로는 사용자의 계층정보, 속성-테이블정보, 속성들간의 계층정보, 값들간의 계층정보등이다.

4.2 연관규칙을 이용한 지능적 질의 처리

지능적 질의 처리를 위해 사용자 정보 테이블이 존재해야한다. 사용자 정보 테이블에는 나이, 소속, 관심분야로 나누어진 그룹 정보가 있다. 사용자 정보 테이블은 이들 속성들에 정의된 계층 계층 트리를 이용한 미리 튜플 추상화 기법으로 처리된다. 즉 한 사용자는

일반화 트리로 추상화된 객체로 간주되며, 30대 교직원 - 30대 전반 공대교수 - 34세 전산과교수 등과 같은 추상화 경로를 가지게 된다.

또, 속성들간의 계층 구조로는 학사관련 - 학생 테이블 - 학생의 나이 와 같은 추상화 경로를 가지며, 이 계층은 시스템에서 제공하는 계층 계층 트리의 일종이다.

이러한 배경 지식과 질의 패턴, 응답 패턴에 대해 다시 튜플 추상화 기법을 적용하면, 유용한 연관 규칙을 산출할 수 있다. 또 이 때 필요한 패턴 탐사 질의를 SQL 과 유사한 형태로 정의 했다. 그림 10 에 주어진 패턴 탐사 질의는 주변정보를 제공하는 지능질의 기법에 적용될 수 있다.

4.3 지능질의의 확장

보다 지능적인 질의 처리가 가능하다. 질의 패턴 튜플에서 추출한 연관 규칙은 그룹화된 사용자들의 공통적인 경향을 반영할 수 있다. 사용자 그룹의 질의 패턴의 경향이 추출된다면, 이를 이용해서 부가적인 정보를 제공하는 주변 결과 반환(*neighborhood answering*)처리가 효율적이 된다. 또 다른 경우로는, 상대적인 사용자 그룹의 질의 패턴의 경향정보는 비교 질의(*comparison query*)에 이용된다. 예를 들면, 대학원생 그룹은 테이블 A와 B 를 이용한 질의를 자주 이용하고, 대학생 그룹은 테이블 A 와 테이블 C를 조인(*join*)해서 질의를 하는 경향이 있는 경우, 대학원생이 테이블 A와 B를 이용한 질의를 던질 때 대학생이 자주 이용하는 테이블 C에 대한 정보를 부가해서 제공하는 것이다.

질의 패턴 튜플과 응답 패턴 튜플도 확장될 수 있다. 가능한 것 중의 하나는 정-반 응답(*positive and negative answer*) (Yoon and Kerschberg, 1993)를 고려하는 것이다. 질의의 결과는 테이블이고 정(*positive*)결과 이다. 반(*negative*)결과는 질의에 참여하는 모든 테이블을 조인하고, 이를 질의 결과에 투영(*project*)한 뒤, 정결과를 제외한 것이다. 반결과에서 추출되는 패턴들은 질의에서 제외되는, 자주 이용되지 않는 데이터에 대한 패턴이다.

5 결론

연관규칙을 응용하여 데이터베이스 질의를 지능적으로

처리하는 시스템을 소개했다. 데이터베이스의 질의는 사용자의 의도를 반영하여 전체 데이터베이스 중 관심있는 부분을 지정하는 역할을 하므로, 사용자가 관심부분에 대한 규칙을 추출할 수 있다. 사용자질의의 연관규칙은 지능적 질의처리에 활용된다. 지능적 질의처리로는 확장질의, 비교질의, 제안질의 등으로, 사용자의 원래 질의에 부가적인 정보를 추가하여 사용자의 의사결정을 돕는다.

사용자의 데이터베이스 이용 패턴은 사용자 그룹으로 일반화되고, 사용자의 관심부분의 시계열적 추이도 지식탐사의 대상이 될 수 있다.

참조논문

- Agrawal, R., S. Ghosh, T. Imielinski, B. Iyer and A. Swami, "An Interval Classifier for Database Mining Applications," Proc. 18th Int. Conf. VLDB 560-573, 1992.
- Agrawal, R., T. Imielinski and A. Swami, "Mining association rules between sets of items in large database," Proc. 1993 Int. Conf. ACM SIGMOD, 207-216, May 1993.
- Agrawal, R. and R. Srikant, Fast Algorithms for Mining Association Rules, Proc. VLDB Conf. 1994.
- Agrawal, R. and R. Srikant, "Mining Sequential Patterns," Proc. 11th Int. Conf. Data Engineering, 3-14, 1995.
- Agrawal, R., H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo, "Fast discovery of association rules," in U. M. Fayyad et. al. eds, Advances in Knowledge Discovery and Data Mining, Menlo Park, CA, AAAI/MIT Press, 307-328, 1996.
- Cai, Y., N. Cercone, and J. Han, "Attribute-Oriented Induction in Relational Databases," in G. Piatetsky-Shapiro and W. Frawley, Editors, Knowledge Discovery in Databases, Menlo Park, CA, AAAI/MIT Press, 213-228, 1991.
- Cuppens, F. and R. Demolombe, "Cooperative Answering: A Methodology to Provide Intelligent Access to Databases," Proc. 2nd Int. Conf. Expert Database Systems, 621-643, Apr. 1988.
- Han, J. and Y. Fu, "Discovery of Multiple-Level Association Rules from Large Databases," Proc. 21th Int. Conf. VLDB, 420-431, 1995.
- Han, J., Y. Huang, N. Cercone, and Y. Fu, "Intelligent Query Answering by Knowledge Discovery Techniques," IEEE Trans. Knowledge and Data Engineering, Vol. 8. No. 3, 373-390, Jun. 1996.
- Houtsma, M. and A. Swami, Set-Oriented Mining for Association Rules in Relational Database, ICDE 95.
- Gaasterland, T., P. Godfrey and J. Minker, "An Overview of Cooperative Answering," J. Intelligent Information Systems, Kluwer Academic Pub. Vol. 1, No. 2, 123-157, 1992.
- Lee, D.-H., D.-Y. Seo, K.-S. Moon, J.-S. Chang, D.-W. Nam, J.-Y. Lee, "Discovery and Application of Inter-Classes Patterns in Database," Proceedings of the Eighth International Workshop on Database and Expert Systems Applications, Toulouse, France, IEEE Computer Society, pp. 326-331, September 1-2, 1997.
- Park, J.S., M.-S. Chen, and P.S. Yu, Using a Hash-Based Method with

- Transaction Trimming for Mining Association Rules, TKDE 9(5).
- Quinlan, R., "Induction of decision trees," Machine Learning, 1(1), pp. 81-106, 1986.
- Savasere, A., E. Omiecinski and S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases," Proc. 21th Int. Conf. VLDB, 432-444, 1995.
- Shum, C.-D. and R. Muntz, "Implicit Representation for Extensional Answers," Pro. 2nd Int. Conf. Expert Database Systems, 497-522, Apr. 1988.
- Srikant, R. and R. Agrawal, "Mining Generalized Association Rules," Proc. 21th Int. Conf. VLDB, 407-419, 1995.
- Yoon, J. P. and L. Kerschberg, "A Framework for Knowledge Discovery and Evolution in Databases," IEEE Trans. Knowledge and Data Engineering, Vol. 5. No. 6, 973-979, Dec. 1993.