DOI: 10.3745/KIPSTD.2009.16D.5.691

시간 속성을 갖는 이벤트의 의미있는 희소 관계에 기반한 연관 규칙 탐사

한 대 영[†] · 김 대 인[††] · 김 재 인[†] · 송 명 진[†] · 황 부 현[†††]

요 약

이벤트는 환자의 증상과 같이 시간 속성을 갖는 하나의 흐름을 의미하며 인터벌 이벤트는 시작과 종료 시점에 대한 시간 간격을 갖는다. 그리고 시간 데이터마이닝에 대한 많은 연구가 있었지만 환자 이력, 구매자 이력, 로그 이력과 같은 인터벌 이벤트에 대한 지식 탐사 방법에 대한 연구는 미흡하다. 이 논문에서는 이벤트들의 인과 관계에 대한 연관 규칙을 탐사하고 이 규칙에 기반하여 결과 이벤트 발생을 예측하는 시간 데이터마이닝 방법을 제안한다. 제안 방법은 이벤트 시간 속성을 사용하여 인터벌 이벤트로 요약하고 이벤트들의 인과 관계를 탐사하여 이벤트 발생을 예측한다. 성능평가를 통하여 제안 방법은 다양한 지지도를 적용하여 발생 빈도에 상관없이 이벤트 발생에 높은 영향을 주는 의미있는 희소 관계를 발견함으로써 기존의 데이터마이닝 기법에 비하여 보다 우수한 정보를 탐사할 수 있다.

키워드 : 시간 속성, 인터벌 이벤트, 인과 관계, 연관 규칙, 의미있는 희소 관계

Finding Association Rules based on the Significant Rare Relation of Events with Time Attribute

DaeYoung Han[†] · Daeln Kim[††] · Jaeln Kim[†] · MyungJin Song[†] · BuHyun Hwang[†††]

ABSTRACT

An event means a flow which has a time attribute such as the a symptom of patients, an interval event has the time period between the start-time-point and the end-time-point. Although there are many studies for temporal data mining, they do not deal with discovering knowledge from interval event such as patient histories and purchase histories. In this paper, we suggest a method of temporal data mining that finds association rules of event causal relationships and predicts an occurrence of effect event based on discovered rules. Our method can predict the occurrence of an event by summarizing an interval event using the time attribute of an event and finding the causal relationship of event. As a result of simulation, this method can discover better knowledge than others by considering a lot of supports of an event and finding the significant rare relation on interval events which means an essential cause of an event, regardless of an occurrence support of an event in comparison with conventional data mining techniques.

Keywords : Temporal Property, Interval Event, Causal Relationship, Association Rule, Significant Rare Relation

1. 서 론

데이터마이닝이란 축적된 데이터에 숨겨진 유용한 정보를 발굴하는 것으로 순차 패턴, 유사 시퀀스, 시간 관계 규칙과 같이 시간에 대한 유용한 정보를 탐사하는 시간 데이터마이닝에 대한 연구가 활발하게 진행되고 있다[1-4]. 시간 데이터마이닝에서 환자 이력, 구매자 이력, 로그 이력에 대한 이벤트는 발생 시점 및 종료 시점과 같은 시간 속성을 갖는 객체의 흐름(flow)으로 이벤트는 같은 이벤트 타입을 갖는 이벤트 시퀀스로 요약된다[5]. 그리고 축적된 이벤트 시퀀스 정보에 대한 시간 데이터마이닝을 통하여 이벤트 발생의 인과 관계에 대한 연관 규칙을 탐사하고 과거의 이벤트를 바탕으로 미래 발생할 이벤트를 예측한다[6]. 그러나 기존의 연관 규칙 탐사 방법은 이벤트 발생에 대한 지지도만을 고려하여 빈발 이벤트(frequent event)에 대한 연관 규칙만을 탐사한다. 그러나 이벤트 중에는 발생 빈도가 낮더라도 특정 이벤트와 높은 확률로 함께 발생하거나 영향을 받는 이벤트가 존재한다[7]. 특히 이러한 이벤트가 응용에 따라 중

※ 본 논문은 2009년도 학술진흥재단 기본연구지원사업(2009-0076136)에 의하여 연구되었음.
† 준 회 원 : 전남대학교 전자컴퓨터공학부 석사과정
†† 정 회 원 : 전남대학교 전자컴퓨터공학부 시간강사
††† 종신회원 : 전남대학교 전자컴퓨터공학부 교수
논문접수 : 2009년 4월 20일
수 정 일 : 1차 2009년 7월 9일, 2차 2009년 7월 29일
심사완료 : 2009년 7월 29일

요한 의미를 갖는 이벤트(significant event)라면 이러한 이벤트를 포함하는 연관 규칙은 다른 규칙보다 더욱 유용한 정보로 활용될 수 있다.

이 논문의 제안 방법은 시간 속성을 갖는 이벤트의 지속성을 고려하여 이벤트 요약 정보인 인터벌 이벤트를 구성한다. 인터벌 이벤트는 시간 간격 정보를 갖는 이벤트 요약 정보로 이벤트 타입과 이벤트 시간 시점 및 종료 시점을 포함한다. 예를 들어 한 고객에게 특정 이벤트 A가 단위시간 1, 2, 12에 발생한 경우 이벤트의 지속성을 고려하지 않는다면 A는 $(A,[1,12])$와 같은 인터벌 이벤트로 요약된다. 그러나 단위시간 단위(granularity)가 월(month)이라면 2월과 12월 사이에는 많은 시간 간격이 존재하며 A는 1월부터 12월까지 지속된 것이 아니라 1월부터 2월, 그리고 12월에 독립적으로 발생된 것으로 간주하는 것이 바람직하다[5]. 또한 제안 방법은 다양한 지지도를 고려하여 인터벌 이벤트에 대한 연관 규칙을 탐사한다. 기존의 연관 규칙 탐사 방법은 빈발 이벤트만을 고려하여 연관 규칙을 탐사하지만 이벤트 발생 빈도가 이벤트의 중요도를 의미하지는 않으므로 이 논문에서는 다양한 지지도를 적용하여 연관 규칙을 탐사한다. 본 논문의 구성은 다음과 같다. 2절에서는 시간 데이터마이닝에 대한 관련 연구를 기술하고, 3절에서는 제안하는 의미있는 희소 관계 탐사 방법에 관련된 정의와 알고리즘을 기술한다. 4절에서는 실험을 통하여 제안 방법의 성능을 분석하고, 끝으로 5절에서는 결론 및 향후 연구 방향을 기술한다.

2. 관련 연구

시간 속성을 갖는 데이터로부터 지식을 탐사하는 시간 데이터마이닝은 순차 패턴, 유사 시퀀스, 시간 규칙 탐사 기법으로 분류된다[8-12]. [1,2]에서는 순차 패턴 탐사기법으로 에피소드 기법을 제안하였다. 에피소드란 밀접히 관련된 이벤트 시퀀스를 의미하며 에피소드 기법은 최소 빈발 임계값을 만족하는 시간 윈도우의 에피소드에 대한 연관 규칙을 추출한다. 예를 들어 이벤트 시퀀스 (A, B, D, A, C, B, D, A, D)에서 윈도우 사이즈가 3초이고 최소 빈발 임계값이 50%이면 시퀀스 (A, B)는 전체 3개 중 2개의 윈도우에서 발생하므로 빈발한 에피소드로 추출된다. [11, 12]에서는 연관 규칙 탐사, 분류, 특성화와 같은 기존의 데이터마이닝 기법을 확장한 시간 연관 규칙 탐사 기법을 제안하였다. 시간 연관 규칙 탐사는 순환적으로 반복되는 연관 규칙을 발견하기 위한 순환 연관 관계 탐사[11]와 캘린더 형태로 표현된 시간 패턴에 대한 연관 규칙을 발견하는 캘린더 연관 관계 탐사[12]로 나누어진다. [11]에서 제안한 시간 데이터마이닝 방법은 이벤트 발생 시점만을 고려하여 연관 관계를 탐사하며 인터벌 이벤트를 고려하지 않는다. [12]에서는 캘린더 C를 인터벌 집합으로 정의하고 $\{(s_1,e_1),(s_2,e_2),...,(s_l,e_k)\}$형태로 기술하며 인터벌 이벤트를 고려한 연관 규칙을 탐사한다. 그러나 [12]의 방법에서의 인터벌 이벤트는 이벤트의 지속성을 고려하지 않는다.

[5]에서는 이벤트 지속성을 고려하여 이벤트 시퀀스를 구성하는 시간 연관 규칙 탐사 방법을 제안하였다. 제안 방법은 이벤트의 지속성을 고려하여 인터벌 이벤트를 요약함으로써 보다 합리적인(reasonable) 인터벌 요약 정보를 구성하고 여기에 기반하여 연관 규칙을 탐사한다. 그러나 [5]의 방법은 지지도 이상의 발생 빈도를 갖는 빈발 이벤트에 대한 연관 규칙만을 탐사하므로 빈발하지 못한 희소 이벤트(rare event)에 대한 연관 규칙은 탐사할 수 없는 문제가 있다. 그러나 희소 이벤트 중에는 발생 빈도가 낮더라도 다른 희소 이벤트와 높은 확률로 함께 발생하는 이벤트가 존재한다[7]. 그리고 이러한 이벤트가 중요한 의미를 갖고나 다른 이벤트 발생에 영향을 주는 주요한 원인 이벤트라면 이러한 이벤트에 대한 연관 규칙은 다른 규칙보다 보다 유용한 정보로 활용될 수 있다.

이 연구에서는 인터벌 이벤트 인과 관계 규칙을 찾는 시간 데이터마이닝 방법을 제안한다. 제안하는 방법은 이벤트 지속성을 고려하여 이벤트 시퀀스를 요약하며 이벤트 발생 빈도에 대한 지지도를 고려하여 빈발 이벤트에 대한 연관 관계를 탐사한다. 또한 제안 방법은 신뢰도에 기반을 둔 다양한 지지도를 정의하여 적용함으로써 발생 빈도에 상관없이 이벤트 발생에 높은 영향을 주는 의미있는 희소 관계(significant rare relation) 정보를 추출함으로써 다양한 인터벌 이벤트 인과 관계 규칙을 탐사한다.

3. 연관 규칙 탐사

이벤트 발생 빈도는 이벤트의 중요도를 의미하지 않으므로 빈발 이벤트만 고려하여 연관 규칙을 탐사하는 것은 바람직하지 않다[7]. 이 절에서는 이벤트 지속성을 고려한 인터벌 이벤트 구성 및 제안 방법에 대한 기본 개념을 정의하고 빈발하지는 않지만 특정 이벤트와 높은 확률로 함께 발생되는 의미있는 희소 관계 탐사에 대하여 기술한다.

3.1 기본개념

실세계에는 시간 속성을 갖는 이벤트에 대한 다양한 응용이 존재한다. 예를 들어 환자가 주기적으로 진찰을 받는 경우 환자의 진료 카드는 다양한 증상을 포함하며 각각의 증상은 시간 속성(진료 시간) t를 갖는 이벤트 E로 정의되며 순서쌍 (E,t)로 기술한다. 그리고 한 환자에 대한 진찰 기록은 시간 속성을 갖는 트랜잭션으로 간주된다[5]. 같은 환자에 대한 트랜잭션 집합에서 요약되는 인터벌 이벤트는 시작 시점과 종료 시점 정보를 포함하며 정의 3.1과 같다. 그리고 이벤트의 지속성 여부를 판단하기 위한 기준으로 정의 3.2와 같은 이벤트 독립성 임계값을 적용하여 인터벌 이벤트를 요약한다.

[정의 3.1] (인터벌 이벤트) 인터벌 이벤트는 시간 간격을 갖는 이벤트 요약 정보로 이벤트 E에 대하여 $(E,[vs,ve])$로 표현된다. vs와 ve는 E의 시작 시점과 종료 시점으로 E

가 vs와 ve 동안에 발생되었음을 의미하며, E의 시작 시점과 종료 시점은 각각 $E.vs$와 $E.ve$로 기술한다.

[정의 3.2] (이벤트 독립성 임계값) 이벤트 독립성 임계값은 같은 고객의 트랜잭션 집합에 존재하는 이벤트 시퀀스 ES의 지속성을 판단하기 위한 임계값이다. $ES=\{(E_m,T_1)(E_m,T_2)(E_m,T_3)...(E_m,T_n)\}$에서 각 이벤트 타입 E_m의 발생 시간 간격이 이벤트 독립성 임계값 δ보다 작으면($\forall E_m, (E_m.T_{i+1}-E_m.T_i)\leq\delta$, where $1\leq i\leq n-1$) 이벤트 시퀀스는 지속적임을 의미한다.

제안 방법은 이벤트 독립성 임계값 δ을 정의하여 δ보다 큰 인터벌을 갖는 이벤트 시퀀스는 독립적인 이벤트로 간주하여 두 개의 서브 시퀀스로 분리한다. 즉 특정 환자가 같은 증상을 보이더라도 장시간 동안 증상이 없는 시간 간격이 존재하면 증상이 지속되었다기 보다는 증상 종료 후 동일한 증상이 재발한 것으로 간주하는 것이 합리적이다. 지속성을 고려하여 분리된 인터벌 시퀀스는 인터벌 이벤트로 요약되며 인터벌 이벤트 x와 y에는 [5]에서와 같이 Allen의 연산자에 근거한 표 1과 같은 인터벌 관계가 존재한다.

<표 1>의 인터벌 관계는 이벤트 발생에 대한 원인과 결과 분석 정보로 활용 가능하다[5]. 즉 과거에 발생한 이벤트들의 인터벌 관계에 기반을 두고 미래에 발생할 이벤트를 예측할 수 있으므로 인터벌 관계에 대한 연관 규칙 탐사는 매우 중요한다. 그러나 [4,5]와 같은 대부분의 시간 데이터 마이닝 기법들은 빈발 이벤트만을 고려하여 인터벌 관계를 탐사한다. 그러나 어떠한 이벤트는 자주 발생하지는 않지만 특정 이벤트에 많은 영향을 미치는 경우가 있으며 이벤트 발생 정도는 이벤트의 중요성 및 발생 영향력을 의미하지 않는다. 따라서 제안 방법은 다양한 지지도를 사용하여 빈발 이벤트와 빈번하지는 않지만 특정 이벤트와 함께 발생하는 이벤트 쌍에 대한 의미있는 희소 이벤트(significant rare event)에 존재하는 인터벌 관계를 탐사한다. 또한 제안 방법은 이벤트 발생에 대한 영향력을 분석하여 정의된 임계값 이상의 영향력을 미치는 인터벌 관계를 추출한다. 제안 방법에서 연관 규칙 탐사에 적용하는 지지도는 정의 3.3과 같다.

[정의 3.3] (지지도) 지지도는 0과 1사이의 실수로 다음과 같다.

〈표 1〉 인터벌 관계

관계	조건	표현
before(x,y)	$x.ve < y.vs$	$x\ \overrightarrow{b}\ y$
equals(x,y)	$(x.vs=y.vs) \cap (x.ve=y.ve)$	$x\ \overrightarrow{e}\ y$
meets(x,y)	$x.ve = y.vs$	$x\ \overrightarrow{m}\ y$
overlaps(x,y)	$(x.vs<y.vs) \cap (x.ve<y.ve)$	$x\ \overrightarrow{o}\ y$
during(x,y)	$(x.vs<y.vs) \cap (x.ve>y.ve)$	$x\ \overrightarrow{d}\ y$

- 1차 지지도 (*1st support*) : 빈발 이벤트 탐사를 위한 최소 지지도로 전체 고객 중 이벤트 A가 발생한 고객 비율을 의미한다.
- 2차 지지도 (*2st support*) : 의미있는 희소 이벤트 탐사를 위한 최소 지지도로 이벤트 A가 발생한 고객 중 이벤트 B가 함께 발생한 고객 비율을 의미한다($|A|\geq|B|$).
- 1차 관계 지지도 (*1st relation support*) : 전체 고객 중 이벤트 A와 B에 대한 인터벌 관계 $IR(A,B)$가 존재하는 고객 비율을 의미한다.
- 2차 관계 지지도 (*2st relation support*) : 이벤트 A와 B가 함께 발생한 고객 중 A와 B에 대한 인터벌 관계 $IR(A,B)$가 존재하는 고객 비율을 의미한다.
- 3차 관계 지지도 (*3st relation support*) : 이벤트 B가 발생한 고객 중 이벤트 A와 B에 대한 인터벌 관계 $IR(A,B)$가 존재하는 고객 비율을 의미한다.

$$\text{1차지지도}=\frac{|A|}{|C_{ust}|},\ \text{2차지지도}=\frac{|(A\cup B)|}{MAX(|A|,|B|)},$$

$$\text{1차관계지지도}=\frac{|IR(A,B)|}{|C_{ust}|},$$

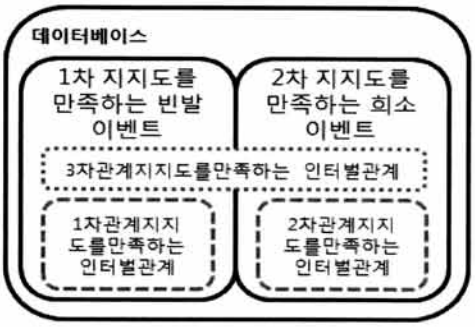
$$\text{2차관계지지도}=\frac{|IR(A,B)|}{|(A\cup B)|},$$

$$\text{3차관계지지도}=\frac{|IR(A,B)|}{|B|}$$

$|C_{ust}|$, $|A|$, $|A\cup B|$, $|MAX(|A|,|B|)|$, $|IR(A,B)|$는 각각 전체 고객 수, 이벤트 A가 발생한 고객 수, 이벤트 A와 B가 동시에 발생한 고객 수, 이벤트 A 또는 B중 더 많은 고객에게서 발생된 이벤트가 발생한 고객 수, 이벤트 A와 B에 대한 인터벌 관계 $IR(A,B)$를 갖는 고객 수를 의미한다.

1차 지지도는 [4,5]와 같은 인터벌 데이터마이닝에서 적용하는 지지도로 전체 고객 중 이벤트 A가 발생한 고객수가 1차 지지도보다 큰 경우 A는 빈발 이벤트로 추출된다. 그러므로 1차 지지도 이상의 확률로 발생된 빈발 이벤트는 전체 고객에게서 자주 발생된 이벤트를 의미한다. 2차 지지도는 자주 발생하지는 않지만 높은 확률로 함께 발생하는 이벤트 쌍을 탐사하기 위한 지지도로 이벤트 쌍 (A,B)가 전체 고객에게서 자주 발생되지는 않지만 이벤트 A가 발생한 고객 중 이벤트 B가 함께 발생한 고객 수가 2차 지지도 이상의 확률로 발생한 경우를 의미하며 (A,B)는 응집력이 높은 이벤트를 의미한다. 3차 지지도는 이벤트 발생에 높은 영향력을 주는 이벤트 쌍 (A,B)을 탐사하기 위한 지지도를 의미한다. 그리고 정의 3의 지지도를 만족하는 이벤트와 인터벌 관계에 대한 포함 관계는 (그림 1)과 같다.

(그림 1)과 같이 전체 고객에게서 빈발하게 발생하여 정의된 1차 지지도 이상의 확률로 발생된 빈발 이벤트와 이들 사이에 존재하는 인터벌 관계 중 1차 관계 지지도 이상의 확률로 발생된 빈발 인터벌 관계로 탐사된다. 또한 전체 고객에게서 자주 발생되지는 않지만 특정 이벤트와 2차 지지도 이상의 확률로 함께 발생된 의미있는 희소 이벤트와 이



(그림 1) 지지도 포함 관계

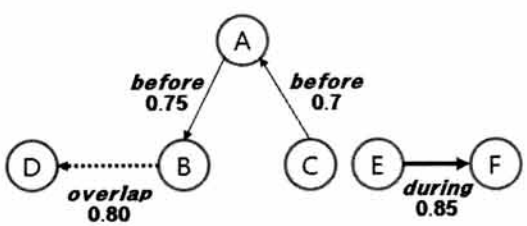
들 사이에 존재하는 관계 중 2차 관계 지지도 이상의 확률로 함께 발생된 의미있는 인터벌 관계로 탐사된다. 그리고 빈발 이벤트가 특정 희소 이벤트 발생에 높은 영향을 미치는 경우 빈발 이벤트와 희소 이벤트 사이에 존재하는 인터벌 관계 중 3차 관계 지지도 이상의 확률로 발생된 이벤트 발생에 영향을 주는 인터벌 관계로 탐사한다. 제안 방법은 이러한 다양한 인터벌 관계를 탐사하며 제안 방법에서만 탐사할 수 있는 의미있는 희소 관계는 정의 3.4와 같다.

[정의 3.4] (의미있는 희소 관계) 의미있는 희소 관계란 이벤트 발생 빈도가 1차 지지도보다 작지만 두 이벤트가 매우 높은 확률로 함께 발생하는 이벤트 쌍 (A,B)에 존재하거나 이벤트 발생에 높은 영향력을 주는 이벤트 쌍 (A,B)에 존재하는 인터벌 관계로 A와 B에 대한 인터벌 관계 $IR(A,B)$이 전체 고객 중 2차 관계 지지도 또는 3차 관계 지지도 이상으로 발생한 관계를 의미한다.

정의 3.3의 지지도를 적용하여 탐사된 인터벌 관계는 정의 3.5와 같은 *IRR-Graph*(Interval Relation Rules-Graph)로 표현되어 이벤트 발생에 대한 원인 및 결과 분석을 위한 연관 규칙 탐사에 활용되며 *IRR-Graph*는 정의 3.5와 같다.

[정의 3.5] (*IRR-Graph*) *IRR-Graph*는 인터벌 이벤트와 인터벌 관계를 각각 노드와 에지로 갖는 방향(direct) 그래프로 각각의 에지는 인터벌 관계의 발생 빈도에 대한 지지도를 가중치로 갖는다. *IRR-Graph*에서 실선(solid line) 에지는 빈발 이벤트에 대한 인터벌 관계, 굵은선(bold line) 에지는 의미있는 희소 이벤트 쌍에 대한 인터벌 관계를, 그리고 점선(dotted line) 에지는 특정 이벤트 발생에 높은 영향력을 미치는 이벤트에 대한 인터벌 관계를 의미한다.

(그림 2)에서 실선 에지로 표현된 인터벌 관계 *before*(A,B)와 *before*(C,A)는 각각 빈발 이벤트 C와 A, 그리고 A와 B에 대한 인터벌 관계를 의미한다. 굵은선 에지로 표현된 인터벌 관계 *during*(E,F)는 전체 고객에게서 이벤트 E와 F는 자주 발생되지 않지만 E와 F는 높은 확률로 함께 발생하며 그 중 85%의 확률로 *during*(E,F) 관계가 존재함을 의미한다. 또한 점선 에지로 표현된 인터벌 관계 *overlap*(B,D)는 이벤트 D는 자주 발생되지는 않지만 D가 발생하는 경우 중 80%의 확률로 *overlap*(B,D) 관계에 의하여 발생하므로 B는 D 발생에 높은 영향력을 주는 이벤트임을 알 수 있다. 최종적으로 추출된 인터벌 관계 규칙에 대한 *IRR-Graph*의 화살표 방향을 따라서 $C\overrightarrow{b}A\overrightarrow{b}B(\overrightarrow{o}D)$와 $(E\overrightarrow{d}F)$와 같은 시간 순서에 대한 인터벌 관계 규칙(( )안의 정보는 의미있는 희소 관계 규칙) 패스를 추출하여 이벤트 발생에 대한 다양한 인과 관계를 분석한다.



(그림 2) *IRR-Graph*

3.2 알고리즘

제안 방법은 빈발 이벤트와 빈발 인터벌 관계 탐사, 의미있는 희소 관계 탐사, 그리고 추출된 인터벌 관계에 기반을 둔 이벤트 발생에 대한 인과 관계 분석 단계로 구성된다. 첫번째 단계는 데이터베이스를 고객 아이디로 정렬하고 1차 지지도 이상의 발생 빈도를 갖는 빈발 이벤트와 이들 사이에 존재하는 1차 관계 지지도 이상의 발생 빈도를 갖는 빈발 인터벌 관계를 추출한다. 두 번째 단계는 이벤트 발생 빈도가 1차 지지도를 만족하지 못하지만 2차 지지도 이상의 확률로 함께 발생한 이벤트 쌍을 의미있는 희소 이벤트로 추출한다. 그리고 의미있는 희소 이벤트에 존재하는 인터벌 관계 중 2차 관계 지지도 이상의 발생 빈도를 갖는 의미있는 희소 관계로 추출한다. 또한 정의된 1차 관계 지지도와 2차 관계 지지도를 만족하지 않지만 이벤트 발생에 대한 영향력을 고려하여 3차 관계 지지도 이상의 발생 빈도를 갖는 인터벌 관계를 추출한다. 세번째 단계는 추출된 인터벌 관계에 대한 *IRR-Graph*를 작성하여 이벤트 발생에 대한 인과 관계를 분석한다.

3.2.1 빈발 인터벌 관계 탐사 알고리즘

제안 방법은 빈발 인터벌 관계를 탐사하기 위하여 (그림 3)(a)와 같이 데이터베이스를 고객 아이디와 시간 순서로 정렬하고 고객별로 이벤트 발생 빈도를 계산한다. (그림 3)에서 1차 지지도가 50%라면 전체 고객 5명 중 3명 이상의 고객에서 발생한 이벤트가 (그림 3)(b)와 같이 빈발 이벤트로 추출된다. 그리고 이벤트 독립성 임계값이 3이라면 같은 이벤트 타입에 속하는 이벤트 시퀀스 중 인터벌이 3보다 큰 경우 독립적인 이벤트로 간주하여 같은 이벤트 타입을 갖더라도 이벤트 시퀀스를 나누어 (그림 3)(c)와 같은 인터벌 이벤트를 구성한다. 그리고 1차 관계 지지도가 40%라면 추출되는 빈발 인터벌 관계는 (그림 4)와 같으며, 인터벌 이벤트 구성 과정 및 빈발 인터벌 관계 탐사 과정은 각각 알고리즘 3.1과 알고리즘 3.2와 같다.

ID	time	event
10	1 3 4 5 6 9 10 11 12	A A,B A,F B,F B, A,D D A,C,D C
20	1 3 4 6 7 8 10	A,C A,C D D A B B,C

ID	time	event
30	1 3 4 7 9 10 11	A A,D A,D B,C B,C B C
40	1 2 3 4 5 8	E,G G A A,E,F B,F F
50	1 2 3 4	A,D D,E,G G E

(a) 정렬된 데이터베이스

event	support
A	5
B	4
C	3
D	4
E	2
F	2
G	2

1차 지지도 >= 50%

event	support
A	5
B	4
C	3
D	4

(b) 빈발 이벤트 항목

ID	Interval event
10	(A,[1,4])(A,[9,11])(B,[3,6])(C,[11,12])(D,[9,11])
20	(A,[1,3])(A,[7,7])(B,[8,10])(C,[1,3])(C,[10,10])(D,[4,6])
30	(A,[1,4])(B,[7,10]),(C,[7,11])(D,[3,4])
40	(A,[3,4])(B,[5,5])
50	(A,[1,1])(D[1,3])

(c) 인터벌 이벤트

(그림 3) 인터벌이벤트 구성

Interval relation	support
***before*(A,B)**	3
***before*(A,C)**	3
***before*(A,D)**	2
***before*(D,B)**	2
***before*(D,C)**	2

(그림 4) 빈발 인터벌 관계

[알고리즘 3.1] 인터벌 이벤트 구성 알고리즘

Input : 이벤트 시퀀스 집합 ISS, 이벤트 독립성 임계값 δ
Output : 인터벌 이벤트 집합 IES

1. 이벤트 타입 E_i의 이벤트 시퀀스 집합 $ISS(E_i) = \emptyset$
2. For $ISS(E_i)$의 시퀀스 s에 대하여 // 이벤트 독립성 임계값을 고려한 이벤트 시퀀스 구성
   Step1 : s의 첫 이벤트 e_1을 e_s로 설정
   Step2 : e_s부터 스캔하여 두 이벤트 e_{s+i}와 e_{s+i+1}의 시간정보 t_{s+i}와 t_{s+i+1}의 인터벌이 δ보다 크면 서브 시퀀스 $<e_s,e_{s+1},...,e_{s+i}>$를 독립 시퀀스로 간주하여 $ISS(E_i)$에 추가
   Step3 : $s+i+1$이 E_i의 이벤트 시퀀스 개수보다 작으면 e_s를 e_{s+i+1}로 설정 후 *Step2* 반복
1. 인터벌 이벤트 집합 $IES = \emptyset$
2. For $ISS(E_i)$의 독립 시퀀스 s'에 대하여 // 이벤트 타입 E_i에 대한 인터벌 이벤트 집합 IES 구성
   s'에 포함된 이벤트 시퀀스의 첫 이벤트 e_1과 마지막 이벤트 e_n의 시간 정보를 각각 인터벌 이벤트의 시작 시점 vs와 종료 시점 ve로 설정하여 인터벌 이벤트 $(E_i,[vs,ve])$를 IES에 추가
3. return IES;

제안 방법에서 알고리즘 3.1은 수행하여 인터벌 이벤트의 연속성 여부를 판단하는 과정으로 정의된 이벤트 독립성 임계값 이상의 인터벌을 갖는 이벤트 시퀀스는 같은 이벤트 타입에 속하더라도 독립된 두 개의 인터벌 이벤트로 분리된다.

[알고리즘 3.2] 빈발 인터벌 관계 탐사 알고리즘

Input : 고객 데이터베이스 DB, 1차 지지도 SUP_{1st}, 1차 관계 지지도 SUP_{1stR}
Output : 빈발 인터벌 관계 집합 $IERS_{freq}$

1. 고객 데이터베이스 DB를 고객 아이디와 시간으로 정렬
2. For DB의 고객 C_{ust}에 대하여 이벤트 발생 빈도 기록
3. 빈발 이벤트 집합 $FES = \emptyset$, 희소 이벤트 집합 $RES = \emptyset$
4. For 이벤트 타입 E_i에 대하여
   If ($\frac{|E_i|}{|C_{ust}|} \geq$) then 이벤트 타입 E_i를 FES에 추가
   else 이벤트 타입 E_i를 RES에 추가
5. For 고객 C_{ust}에 대하여
   Step1 : For FES의 이벤트 타입 E_i에 대하여 이벤트 시퀀스 집합 $ISS(C_{ust})=\{ES(C_{ust},E_i)\}$ 구성
   Step2 : 빈발 인터벌 이벤트 집합 IES_f = 인터벌 이벤트 구성 알고리즘 수행 결과(인터벌 이벤트 집합 IES)
   Step3 : For IES_f의 인터벌 이벤트에 대한 인터벌 이벤트 관계 집합 $IERS$ 계산
   Step4 : For $IERS$에 포함된 인터벌 이벤트 관계 발생 빈도 기록
6. 빈발 인터벌 관계 $IERS_{freq} = \emptyset$
7. For $IERS$에 포함된 인터벌 이벤트 관계 IER_i에 대하여 // 빈발 인터벌 관계 집합 추출
   If ($\frac{|IER_i|}{|C_{ust}|} \geq$) then 인터벌 이벤트 관계 IER_i를 $IERS_{freq}$에 추가

제안 방법에서 알고리즘 3.2는 1차 지지도 이상의 발생 횟수를 갖는 이벤트를 빈발 이벤트로 추출하고 빈발 이벤트들 사이에 존재하는 인터벌 관계 중 1차 관계 지지도 이상의 발생 횟수를 갖는 인터벌 관계를 빈발 인터벌 관계로 추출한다.

3.2.2 의미있는 희소 관계 탐사 알고리즘

제안 방법은 [4,5]와 같은 인터벌 데이터마이닝에서 탐사할 수 없는 자주 발생되지는 않지만 높은 확률로 함께 발생하는 희소 이벤트와 이들 사이에 존재하는 인터벌 관계를 각각 2차 지지도와 2차 관계 지지도를 사용하여 탐사한다. (그림 3)에서 빈발 이벤트가 아닌 이벤트 E, F, G는 자주 발생하지 않는 희소 이벤트이지만 발생하는 경우 높은 확률로 함께 발생하는 이벤트 쌍에 대한 비율이 2차 지지도보다 큰 경우 의미있는 희소 이벤트로 추출된다. 그리고 의미있는 희소 이벤트에 대한 인터벌 관계가 2차 관계 지지도보다 큰 경우 (그림 5)와 같은 의미있는 희소 관계로 추출된다.

(그림 5)에서 희소 이벤트 E, F, G의 발생 빈도와 E와 G가 함께 발생한 고객 수 $|E \cup G|$는 2이므로 정의된 2차 지지도와 2차 관계 지지도가 70%라면 그림 5(a)와 같이 이벤트 쌍 (E,G)는 의미있는 희소 이벤트로 추출된다. 그리고 E와 G 사이에 존재하는 인터벌 이벤트는 그림 5(b)와 같으며 2차 관계 지지도 70% 이상이므로(2/2=100%) *during*(E,G)는 의미있는 희소 관계로 추출된다. 제안 방법에서 의미있는 희소 관계 탐사 과정은 알고리즘 3.3과 같다.

event	support
E	2
F	2
G	2
E∪F	1
E∪G	2
F∪G	1

2차 지지도 >= 70%

IE∪GI/Max(IEI, IGI)= 2/Max(2,2) = 100%

(a) 의미있는 희소 이벤트 : (E,G)

ID	Interval event
40	(E,[1,4])(G,[1,2])
50	(E,[2,4])(G,[2,3])

(b) 인터벌 이벤트

2차 관계 지지도 >= 70%

(c) 의미있는 희소 관계 : *during(E,G)*

(그림 5) 희소 관계 탐사

[알고리즘 3.3] 2차 관계 지지도에 기반한 희소 관계 탐사 알고리즘

Input : 희소 이벤트 집합 RES, 2차 지지도 SUP_{2st}, 2차 관계 지지도 SUP_{2stR}

Output : 의미있는 희소 관계 집합 $IERS_{sig}$

1. For 고객 C_{ust}에 대하여
   For 희소 이벤트 집합 RES에 포함된 이벤트 타입 E_i와 $E_j(i \neq j)$로 구성된 이벤트 쌍 (E_i,E_j)의 발생 빈도 기록
2. 의미있는 희소 이벤트 집합 $SRES$ = Ø
3. For $SRES$에 포함된 희소 이벤트 쌍 (E_i,E_j)에 대하여
   If ($\frac{|(E_i \cup E_j)|}{MAX(|E_i|,|E_j|)} \geq$) then 희소 이벤트 쌍 (E_i,E_j)를 $SRES$에 추가
4. For 고객 C_{ust}에 대하여
   Step1 : For $SRES$의 이벤트 타입 E_i에 대하여 이벤트 시퀀스 집합 $ISS(C_{ust})$={ES(C_{ust},E_i)} 구성
   Step2 : 의미있는 희소 인터벌 이벤트 집합 IES_r = 인터벌 이벤트 구성 알고리즘 수행 결과
   Step3 : For IES_r의 인터벌 이벤트에 대하여 인터벌 이벤트 관계 집합 $IERS$ 계산
   Step4 : For $IERS$에 포함된 인터벌 이벤트 관계 발생 빈도 기록
5. 의미있는 희소 관계 집합 $IERS_{sig}$ = Ø
6. For $IERS$에 포함된 인터벌 이벤트 관계 IER_i에 대하여 // 의미있는 희소 관계 집합 추출
   If ($\frac{|IER_i|}{|(E_x \cup E_y)|} \geq$) then 인터벌 이벤트 관계 IER_i을 $IERS_{sig}$에 추가

제안 방법에서 알고리즘 3.3은 이벤트 발생 횟수가 1차 지지도를 만족하지 않지만 특정 이벤트가 함께 발생한 확률이 2차 지지도 이상인 이벤트를 의미있는 희소 이벤트로 정의하고 이들 사이에 존재하는 인터벌 관계 중 2차 관계 지지도 이상의 발생 횟수를 갖는 인터벌 관계를 의미있는 희소 관계로 추출한다. 또한 제안 방법은 1차 관계 지지도와 2차 관계 지지도를 만족하지 못하지만 특정 이벤트 발생에 영향력이 큰 인터벌 관계를 의미있는 희소 관계로 탐사한다. (그림 3)에서 이벤트 F를 포함하는 인터벌 관계는 1차 관계 지지도와 2차 관계 지지도를 모두 만족하지 못한다. 그러나 F는 두 명의 고객(아이디 10과 40)에게서만 발생되었고 두 명의 고객 모두 이벤트 A와 F에 대한 인터벌 관계 *meet(A,F)*가 존재한다. 따라서 제안 방법은 (그림 6)과 같이 이벤트 발생에 많은 영향을 주는 이벤트에 존재하는 인터벌 관계를 이벤트 발생 인과 관계 분석을 위한 정보로 추출한다. 3차 관계 지지도를 사용한 의미있는 희소 관계 탐사 과정은 알고리즘 3.4와 같다.

ID	Interval event
10	(A,[1,4])(B,[3,6])(F,[4,5])
40	(A,[3,4])(B,[5,5])(E,[1,4])(F,[4,8])(G,[1,2])

(a) 인터벌 이벤트

ID	Interval relation
10	*meet(A,F),during(B,F)*
40	*meet(A,F)*

(b) 인터벌 관계

Interval relation	support
meet(A,F)	2

(c) 의미있는 희소 관계 : *meet(A,F)*

(그림 6) 영향력 관계 탐사

제안 방법에서 알고리즘 3.4는 빈발 이벤트와 의미있는 희소 이벤트들 사이에 존재하는 인터벌 관계 중 3차 관계 지지도 이상의 발생 횟수를 갖는 인터벌 관계를 의미있는 희소 관계로 추출한다.

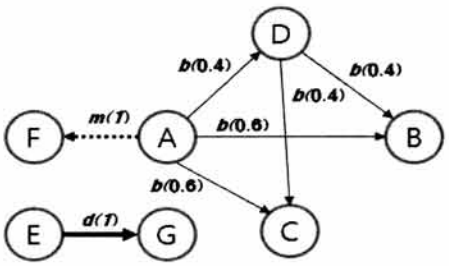
[알고리즘 3.4] 3차 관계 지지도에 기반한 희소 관계 탐사 알고리즘

Input : 빈발 이벤트에 대한 인터벌 이벤트 집합 IES_f, 의미있는 희소 이벤트에 대한 인터벌 이벤트 집합 IES_r,
의미있는 희소 관계 집합 $IERS_{sig}$, 3차 관계 지지도 SUP_{3stR}

Output : 의미있는 희소 관계 집합 $IERS_{sig}$

1. For 고객 C_{ust}에 대하여
   Step1 : 빈발 이벤트에 대한 인터벌 이벤트 집합 IES_f의 인터벌 이벤트 E_f와 희소 이벤트에 대한 인터벌 이벤트 집합 IES_r의 인터벌 이벤트 E_r사이에 존재하는 인터벌 관계 $(E_f \overrightarrow{R} E_r)$에 대하여 인터벌 이벤트 관계 집합 $IERS$ 계산
   Step2 : For $IERS$에 포함된 인터벌 이벤트 관계의 발생 빈도 기록
2. For 인터벌 이벤트 관계 집합 $IERS$에 포함된 인터벌 이벤트 관계 IER_i에 대하여
   If ($\frac{|IER_i|}{|E_y|} \geq$) then 인터벌 이벤트 관계 IER_i을 $IERS_{sig}$에 추가

// 3차 관계 지지도 이상의 발생 빈도를 갖는 의미있는 희소 관계 집합 추출



(그림 7) *IRR-Graph*

3.2.3 이벤트 발생 예측

제안 방법은 탐사된 인터벌 관계에 기준을 둔 연관 규칙을 사용하여 이벤트 발생에 대한 인과 관계를 분석하기 위하여 *IRR-Graph*를 구성한다. 제안 방법을 통하여 추출된 (그림 3)의 이벤트에 존재하는 인터벌 관계에 대한 *IRR-Graph*는 (그림 7)과 같다.

(그림 7)에서 실선 에지로 표현된 인터벌 관계에 의하여 이벤트 A 발생 후 최소 40% 이상의 확률로 이벤트 B, C, D중의 하나가 발생함을 알 수 있다. 그리고 이벤트 D 발생 후 이벤트 B 또는 C가 발생되므로 $A\overrightarrow{b}D\overrightarrow{b}(B \text{or } C)$와 같은 인터벌 관계 규칙 패스를 얻을 수 있다. 또한 그림 7에서 굵은선 에지로 표현된 인터벌 관계 *during*(E,G)는 통하여 이벤트 E와 G는 전체 고객에게서 자주 발생되지는 않았지만 굵은선 에지의 가중치가 1이므로 이벤트 E가 발생하는 도중에 반드시 G가 발생함을 알 수 있다. 또한 실선 에지로 표현된 인터벌 관계 *meet*(A,F)에 의하여 이벤트 A가 발생하는 경우 이벤트 F가 발생하는 것은 아니지만 F는 높은 확률로 이벤트 A의 종료 시점에만 발생하므로 F는 A에 영향을 받는 이벤트임을 알 수 있다. 이 논문에서 제안한 인터벌 관계 탐사 방법은 다양한 지지도를 정의하여 적용함으로써 발생 빈도가 높은 이벤트에 대한 인터벌 관계뿐만 아니라 특정 이벤트 사이에 높은 확률로 발생하거나 발생 빈도에 상관없이 이벤트 발생에 많은 영향력을 미치는 인터벌 관계를 탐사한다. 그리고 이러한 인터벌 관계는 이벤트 발생에 높은 영향을 주는 인과 관계 정보 등에 활용가능하며 보다 중요한 의미를 갖는다.

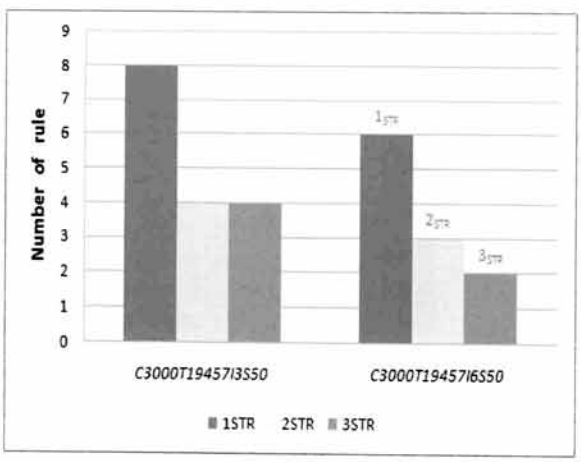
4. 성능 평가

이 절에서는 제안 방법의 성능 분석을 실시하며 모든 실험은 MS XP OS에서 512MB RAM, Pentium 4(3GHz)시스템 환경에서 Visual C로 프로그램을 작성하여 수행한다. 실험 데이터는 8개의 이벤트를 포함하는 고객 트랜잭션을 사용하며 C, T, I, S의 네 가지 기호로 표현된다(C는 전체 고객 수, T는 전체 트랜잭션 수, I는 이벤트 독립성 임계값, S는 1차 지지도를 의미한다). *C3000T19457I3S40*은 3,000명의 고객에 대한 19,457개의 트랜잭션에 이벤트 독립성 임계값과 1차 지지도를 각각 3과 40%로 적용한 실험 데이터를 의미한다. 그리고 다(C는 전체발생 빈도는 작지만 중요도가 높은 의미있는 희소 관계를 대한한다). C3정의 3.3에서 정의한 다양한 지지도를 [7]에서와 같이 ($SUP_{2st} \geqq SUP_{1st}$), ($SUP_{3stR} \geqq SUP_{2stR} \geqq SUP_{1stR}$) 비율로 적용한다. 그리고 실험을 통하여 제안 방법은 기존의 방법에 비하여 다양한 연관 규칙을 탐사할 수 있음을 실험으로 보인다.

[실험 1] 8개의 이벤트를 포함하는 *C3000T19457I3S50*과 *C3000T19457I6S50* 집합을 적용하여 추출되는 인터벌 관계 수를 비교하며 1차 관계 지지도, 2차 지지도, 2차 관계 지지도, 3차 관계 지지도는 각각 20%, 60%, 25%, 35%로 적용한다. 그리고 실험 1의 두 개의 데이터 집합은 이벤트 독립성 임계값이 각각 3과 6이므로 같은 트랜잭션에서을 사용하더라도 이벤트 시퀀스의 연속성 여부에 따라 서로 다른 인터벌 이벤트가 추출된다. 실험 결과 실험 1의 이벤트 집합에서 1차 지지도 50%를 만족하는 빈발 이벤트로 1,500명 이상의 고객에게서 발생한 이벤트 A, B, E, F가 추출되며 600번 이상 발생한 인터벌 관계는 정의된 1차 관계 지지도를 만족하므로 빈발 인터벌 관계로 추출된다.

제안 방법은 의미있는 희소 관계를 탐사하기 위하여 2차 지지도 이상의 확률로 함께 발생하는 이벤트 쌍을 의미있는 희소 이벤트로 탐사한다. 실험 결과 두 개의 이벤트 쌍 (C, D)와 (G, H)가 2차 지지도를 만족하는 의미있는 희소 이벤트로 추출된다. 또한 제안 방법은 3차 관계 지지도를 적용하여 이벤트 발생에 미치는 영향력이 높은 이벤트에 대한 인터벌 관계를 탐사하며, 실험 1에서 추출되는 최종적인 인터벌 연관 규칙은 (그림 8)과 같다.

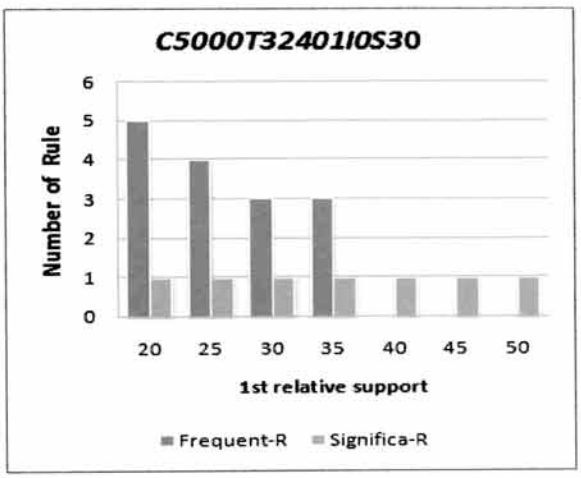
실험 결과 [4,5]와 같은 이벤트 발생 빈도만을 고려하는 방법은 두 개의 실험 데이터 집합에서 각각 8개와 6개 (1STR)의 연관 규칙을 탐사한다. 그러나 제안 방법은 의미있는 희소 관계 규칙을 탐사할 수 있으므로 빈발 이벤트에 대한 연관 규칙뿐만 아니라 두 개의 실험 데이터 집합에서 각각 8개와 5개의 의미있는 희소 관계(2STR+3STR)에 대한 연관 규칙을 추가로 탐사한다.



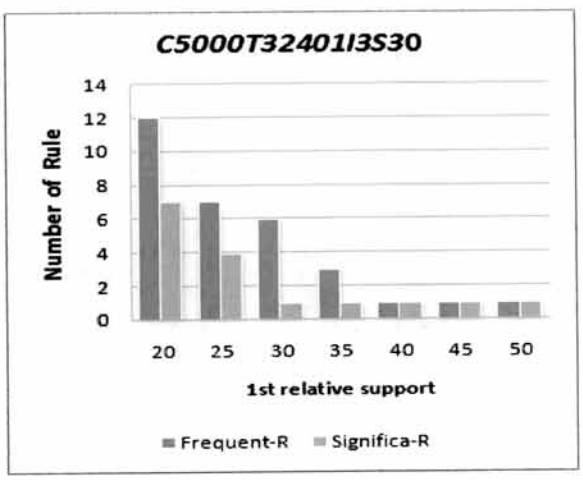
(그림 8) 실험 1의 인터벌 연관 규칙

[실험 2] 8개의 이벤트를 포함하는 *C5000T32401I0S30*, *C5000T32401I3S30*, 그리고 *C5000T32401I6S30* 집합을 적용하여 추출되는 인터벌 연관 규칙을 비교한다. *C5000T32401I0S30* 집합은 이벤트 독립성 임계값 *I*가 0이므로 인터벌 이벤트의 연속성을 고려하지 않는 경우를 의미하며, 실험 2는 실험 1보다 1차 지지도를 낮게 적용하여 추출되는 빈발 이벤트 수를 증가시켰다. 그리고 실험 2는 1차 관계 지지도를 20%에서 50%까지 증가함에(scale-up) 따라 추출되는 인터벌 연관 규칙을 분석한다. 또한 2차 지지도는 60%, 그리고 2차 관계 지지도와 3차 관계 지지도는 1차 관계 지지도 증가에 따라 각각 5%씩 추가로 증가하여 실험에 적용한다. (그림 9), (그림 10), 그리고 (그림 11)은 각각의 데이터 집합에서 추출되는 인터벌 연관 규칙 수를 보여준다.

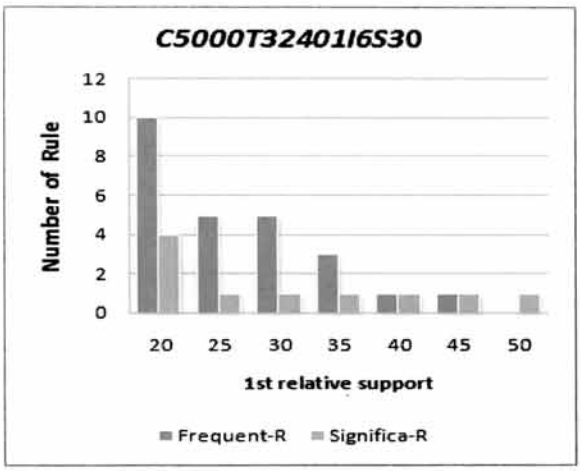
(그림 9)는 인터벌 이벤트 연속성을 고려하지 않는 경우로 같은 이벤트 타입을 갖는 모든 이벤트 시퀀스가 하나의 인터벌 이벤트로 요약되므로 전체 고객에게서 추출된 인터벌 관계와 인터벌 연관 규칙 수는 연속성을 고려하는 경우



(그림 9) 인터벌 연관 규칙 (*I=0*)



(그림 10) 인터벌 연관 규칙 (*I=3*)



(그림 11) 인터벌 연관 규칙 (*I=6*)

에 비하여 감소한다. 그러나 이벤트들 사이에 많은 시간 간격이 존재하는 경우에는 각 이벤트를 독립적인 이벤트로 간주하여 인터벌 이벤트를 구성하는 것이 합리적이다[5]. 그리고 (그림 10)과 (그림 11)은 이벤트 연속성을 고려한 실험 결과를 보여준다. 실험 결과 인터벌 이벤트들 사이에는 빈발하게 발생하는 빈발 인터벌 관계(Frequent-R)에 대한 연관 규칙뿐만 아니라 빈발하지 않더라도 이벤트 발생에 많은 영향력을 미치거나 특정 이벤트와 함께 발생하는 의미있는 희소 관계(Significa-R)에 대한 다양한 연관 규칙이 존재함을 알 수 있다. 그리고 실험 결과 제안 방법은 다양한 지지도를 적용함으로써 이벤트 발생 예측을 위한 다양한 연관 규칙을 탐사할 수 있음을 알 수 있다.

5. 결론 및 향후 연구

이 논문에서는 시간 속성을 갖는 이벤트의 연속성을 고려한 인터벌 이벤트를 구성하고 요약된 정보로부터 미래 발생 예측을 위한 유용한 정보인 인터벌 연관 규칙을 탐사할 수 있는 시간 데이터마이닝 방법을 제안하였다. 제안 방법은 이벤트 독립성 임계값을 적용하여 이벤트 시간 속성에 기반한 인터벌 연속성을 고려함으로써 합리적인 이벤트 시퀀스 요약이 가능하다.

또한 제안 방법은 기존의 시간 데이터마이닝 기법에서 적용하는 이벤트 발생 빈도에 대한 지지도뿐만 아니라 다양한 지지도를 적용함으로써 발생 빈도에 상관없이 특정 이벤트와 높은 확률로 함께 발생하거나 특정 이벤트 발생에 많은 영향력을 미치는 이벤트들 사이에 존재하는 다양한 연관 규칙을 탐사할 수 있다. 특히 제안 방법에서 추출하는 의미있는 희소 관계 규칙은 이벤트 발생에 대한 인과 관계 정보로 미래 발생할 가능성이 높은 이벤트 예측 및 영향력 판단에 적용가능하다.

향후 연구 방향으로 시간 속성을 갖는 이벤트 발생에 대한 다양한 영향력 측정 방법을 제안하고 이에 기반을 두어

다양한 이벤트 발생 영향력 분석 방법에 대하여 연구하고자 한다.

참 고 문 헌

[1] S. Laxman, P. S. Sastry, and K. Unnikrishnan, "Discovering Frequent Episodes and Learning Hidden Markov Models: a Formal Connection," IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.11, pp.1505-1517, Jan., 2006.

[2] S. Laxman, P. S. Sastry, and K. Unnikrishnan, "Discovering Frequent Generalized Episodes when Events Persist for Different Durations," IEEE Transactions on Knowledge and Data Engineering, Vol.19, No.9, pp.1188-1201, Sep., 2007.

[3] J. F. Roddick, K. Hornsby, and M. Spiliopoulou, "Temporal, Spatial and Spatio-Temporal Data Mining Knowledge Discovery Research Bibliography," http://www.cs.flinders.edu.au, Sep., 2001.

[4] E. Winarko, and J. F. Roddick, "ARMADA - An Algorithm for Discovering Richer Relative Temporal Association Rules from Interval-based Data," Journal of Data and Knowledge Engineering Vol.63, pp.76-90, Oct., 2007.

[5] Y. Lee, J. Lee, D. Chai, B. Hwang, and K. Ryu, "Mining Temporal Interval Relation Rules from Temporal Data," Journal of Systems and Software, Vol.82, pp.155-167, 2009.

[6] H. Zhu, and Z. Xu, "An Effective Algorithm for Mining Positive and Negative Association Rules," Proc. of International Conference on Computer Science and Software Engineering, pp.455-458, Dec. 2008.

[7] H. Yun, D. Ha, B. Hwang, and K. Ryu, "Mining Association Rules on Significant Rare Data Using Relative Support," Journal of Systems and Software, Vol.67, pp.181-191, 2003.

[8] R. J. Swargam, M. J. Palakal, "The Role of Least Frequent Item Sets in Association Discovery," Proc. of ICDIM '07. 2nd International Conference, Vol.1, pp.217-223, Oct., 2007.

[9] J. Chung, O. Paek, J. Lee, and K. Ryu, "Temporal Pattern Mining of Moving Objects for Location-Based Service," Proc. of International Conference on Database and Expert Systems Applications Vol.2453, pp.331-340, 2002.

[10] E. Keogh, J. Lin, S. Lee, and H. V. Herle, "Finding the Most Unusual Time Series Subsequences: Algorithms and Applications," Journal of Knowledge and Information Systems Vol.11, No.1, pp.1-27, Dec., 2006.

[11] B. Ozden, S. Ramaswamy, and A. Silberschatz, "Cycle Association Rules," Proc. of International Conference on Data Engineering Orlando, USA, 1998.

[12] S. Ramaswamy, S. Mahajan, and A. Silberschatz, "On the Discovery of Interesting Patterns in Association Rules," Proc. of 24th VLDB Conference New York, USA, pp.368-379, 1998.

한 대 영

e-mail : nara9yo@gmail.com
2008년 전남대학교 전자컴퓨터공학부 (학사)
2008년~현 재 전남대학교 전자컴퓨터공학부 석사과정
관심분야 : 데이터 마이닝, 멀티미디어, 임베디드

김 대 인

e-mail : dikim@chonnam.ac.kr
1998년 전남대학교 전산통계학과 (이학석사)
2006년 전남대학교 전산통계학과 (이학박사)
2004년~현 재 전남대학교 전자컴퓨터공학부 시간강사
관심분야 : 스트림 데이터, 데이터 마이닝, 디지털 콘텐츠

김 재 인

e-mail : sereno3@naver.com
2008년 전남대학교 전자컴퓨터공학부 (학사)
2008년~현 재 전남대학교 전자컴퓨터공학부 석사과정
관심분야 : 스트림 데이터 마이닝, 연속질의, USN응용

송 명 진

e-mail : audwls0324@nate.com
2009년 전남대학교 전자컴퓨터공학부 (학사)
2009년~현 재 전남대학교 전자컴퓨터 공학부 석사과정
관심분야 : 데이터 마이닝, 디지털 콘텐츠

황 부 현

e-mail : bhhwang@chonnam.ac.kr
1978년 숭실대학교 전산학과(학사)
1980년 한국과학기술원 전산학과 (공학석사)
1994년 한국과학기술원 전산학과 (공학박사)
1980년~현 재 전남대학교 전자컴퓨터공학부 교수
관심분야 : 스트림 데이터 마이닝, 분산 시스템, 분산 데이터베이스