

인터벌이벤트의 인과관계에 기초한 영향력 예측 기법

송명진, 김대인, 황부현
전남대학교 전자컴퓨터공학부
e-mail : audwls0324@nate.com

A Method for Predicting Effect based on the Causal relations of Interval Events

Myung-Jin Song, Dae-In Kim, Bu-Hyun Hwang
Dept of Electronics and Computer Engineering,
Chonnam National University

요 약

이벤트 사이에 존재하는 연관 정보를 탐사함으로써 발생 가능한 이벤트를 예측할 수 있다. 그러나 기존의 시간 데이터마이닝 기법은 어느 정도 영향을 주고받았는지에 대한 영향력은 예측할 수 없다. 본 논문에서는 인터벌이벤트 사이에 존재하는 연관 정보를 탐사하고 탐사된 규칙에 대한 영향력을 측정할 수 있는 방법을 제안한다. 제안 방법은 이벤트 지속성을 고려하여 인터벌이벤트를 구성하고 빈발 이벤트 사이에 존재하는 연관 정보에 대한 영향력 정도를 측정한다. 그리고 이벤트 발생에 대한 주요한 원인이벤트를 탐사함으로써 이벤트 인과관계에 대한 다양한 정보를 제공할 수 있다.

1. 서론

데이터마이닝은 “데이터에 잠재된 유용한 지식을 추출하는 기술”로 정의된다. 최근에는 시간 간격을 갖는 인터벌데이터에 잠재된 유용한 시간 정보에 대한 연관규칙을 탐사하는 기법에 대한 연구가 진행되고 있다[1]. 인터벌이벤트 사이에 존재하는 연관규칙을 탐사하는 것은 미래에 발생 가능한 이벤트를 미리 예측할 수 있으므로 중요하다 [2]. 본 논문에서는 이벤트 발생에 대한 다수의 원인이벤트가 존재하는 경우, 이벤트 시간 정보에 대한 연관규칙을 탐사하고 원인이벤트의 영향력 정도를 측정하는 기법을 제안한다. 이러한 정보는 이벤트 발생에 대한 주요한 원인 이벤트를 예측함으로써 비상 상황에 대처할 수 있는 중요한 정보로 활용 가능하다.

2. 관련 연구

시간 데이터로부터 의미 있는 지식을 탐사하기 위한 시간 데이터마이닝에 대한 연구가 진행되고 있으며 이러한 기법들은 크게 순차패턴 탐사, 유사 시계열 탐사 그리고 시간 규칙 탐사로 분류된다[3-5].

[1]에서는 이벤트 지속성과 시간 간격을 고려하여 이벤트시퀀스를 구성하고 이에 기초하여 인터벌이벤트로 요약함으로써 시간관계에 대한 연관규칙을 탐사하는 기법을 제안하였다. 제안방법은 이벤트 지속성을 고려함으로써 합리적인 인터벌이벤트를 구성하고, 인터벌이벤트의 발생빈도를 기준하여 빈발 이벤트들 사이에 존재하는 연관규칙을 탐사한다. 그러나 이러한 기법들은 이벤트들 사이에 존

재하는 시간 정보는 탐사할 수 있지만, 이벤트 발생에 대한 인과관계가 존재하는 경우 이벤트들이 어느 정도 영향을 주고받았는지에 대한 영향력은 측정할 수 없다. 이 논문에서는 인터벌이벤트 사이에서 원인이벤트가 결과이벤트에 미친 영향력 정도를 측정하는 기법을 제안한다. 그리고 분석 결과를 통하여 미래 발생 가능한 이벤트 예측 정보로 활용가능하다.

3. 인터벌 연관규칙 탐사

시간속성을 갖는 이벤트 e 는 현실세계에서 다양한 형태로 존재하며 $e=(E, t)$ 로 기술한다. 여기에서 E 는 이벤트 종류, t 는 이벤트 발생 시점을 의미한다. 그리고 이벤트의 지속 여부를 고려하여 합리적인 인터벌이벤트를 구축하기 위하여 이벤트 독립성임계값을 적용한다. 이벤트시퀀스에 포함된 이벤트의 발생 간격이 정의된 이벤트 독립성임계값 ϵ 보다 크다면 독립적인 관계로 간주하여 같은 이벤트 종류를 갖더라도 두 개의 시퀀스로 분리된다. 그리고 각각의 이벤트시퀀스는 시간간격을 가진 인터벌이벤트 $e'=(E, [vs, ve])$ 로 정의되며 vs 와 ve 는 각각 이벤트 시작 시점과 종료 시점을 의미한다[1].

서로 다른 인터벌이벤트 x 와 y 는 이벤트관계집합 $\Omega = \{(x,y) \mid x,y \in \text{Interval Event}, x \neq y\}$ 에 포함되는 임의의 사건 조합 (x, y) 은 이항 시간 관계 $R(x, y)$ 을 갖는다. 그리고 $R(x, y)$ 은 Allen이 제안한 시간간격 연산자를 참조하여 표1과 같은 관계로 요약가능하다.

〈표1〉 인터벌 관계

관계	조건	표현
$before(x,y)$	$x.ve \leq y.us$	$x \rightarrow_b y$
$during(x,y)$	$(x.us \leq y.us) \wedge (y.ve \leq x.ve)$	$x \rightarrow_d y$
$overlaps(x,y)$	$(x.us \leq y.us) \wedge (x.ve \leq y.ve)$	$x \rightarrow_o y$

표 1의 인터벌 관계는 이벤트 x 와 y 간의 시간관계를 표현하며, 인터벌이벤트 발생에 대한 인과관계 분석을 위한 정보로 사용된다. 인터벌 관계는 주어진 지지도를 만족하는 빈발이벤트들 사이에 존재하는 시간정보로 전체 고객에게 빈발하게 발생한 이벤트들 사이에 존재하는 시간 정보를 탐사한다. 본 논문에서는 신뢰도에 기초하여 인터벌 관계에 대한 영향력을 분석한다. 그러나 인터벌 관계 $X \rightarrow A$ 에서 X 에 대한 A 의 신뢰도만을 고려한다면 A 가 X 에 의하여 얼마나 영향을 받았는지는 예측할 수 없다. 예를 들어 $sup(X)=0.5$, $sup(Y)=0.8$, $sup(X \rightarrow A)=0.5$ 일 때 신뢰도만을 고려하여 이벤트 발생 영향력을 계산하면, 신뢰도 $conf(X \rightarrow A)$ 는 $sup(X \rightarrow A)/sup(X)$ 에 의하여 1로 측정된다 (0.5/0.5). 그리고 $X \rightarrow A$ 의 신뢰도가 1이므로 X 는 A 를 발생시키는 100% 원인이벤트라고 예상할 수 있다. 그러나 이것은 실제로 B 는 0.3 만큼은 A 가 아닌 다른 원인이벤트 Y 에 의하여 발생하였으므로 이벤트들 사이에 존재하는 인터벌관계에 대한 정확한 영향력은 측정할 수 없다. 그러므로 본 논문에서는 인터벌관계에 포함된 두 개의 이벤트 모두에 대한 신뢰도에 기초하여 이벤트 사이의 영향력을 식 1과 같이 측정한다. 이벤트 X 와 이벤트 A 사이에 인터벌관계 $X \rightarrow A$ 가 있을 때 X 가 A 에 미치는 영향력 $Eff(X \rightarrow A)$ 는 다음과 같다.

$$Eff(X \rightarrow A) = \frac{1}{2} \left(\frac{sup(X \rightarrow A)}{sup(X)} + \frac{sup(X \rightarrow A)}{sup(A)} \right) \quad (1)$$

식 1에서 $sup(X)$ 는 이벤트 X 에 대한 지지도를, 기호 \rightarrow 는 이벤트 X 가 이벤트 A 보다 선행하여 발생하였으며 이들 사이에 인터벌관계가 있음을 의미한다. 그리고 $\frac{sup(X \rightarrow A)}{sup(X)}$ 는 이벤트 X 가 이벤트 A 발생에 어느 정도 영향을 주었는가를, $\frac{sup(X \rightarrow A)}{sup(A)}$ 는 이벤트 A 가 이벤트 X 로부터 어느 정도 영향을 받았는지를 의미한다. 제안 방법은 두 개의 지지도를 모두 고려하여 식 1과 같이 인터벌관계에 포함된 이벤트들이 얼마나 영향을 주고받았는지를 측정한다.

그리고 인터벌 관계 $X \rightarrow A$ 가 있을 때, 이벤트 X 를 제외하고 이벤트 A 에 영향을 미치는 이벤트 중 최대로 이벤트 A 에 영향력을 미치는 영향력을 제외 최대영향력이라고 정의한다. 그리고 제외 최대영향력 $Excp-Eff(X \rightarrow A)$ 는 식 2와 같다.

$$Excp-Eff(X \rightarrow A) = \max \left\{ \frac{1}{2} \left(\frac{sup(Y \rightarrow A)}{sup(Y)} + \frac{sup(Y \rightarrow A)}{sup(A)} \right) \right\} \quad (2)$$

when $Y \rightarrow A \in FR, X \notin Y$

식 2에서 이벤트 Y 는 이벤트 A 의 발생에 대한 원인 이벤트 X 를 제외한 나머지 원인이벤트를 의미하며, 인터벌 관계 $X \rightarrow A$ 에 대하여 X 가 A 에 미치는 순수영향력 $Pure-Eff(X \rightarrow A)$ 는 식 3과 같다.

$$Pure-Eff(X \rightarrow A) = \frac{1}{2} \left(\frac{sup(X \rightarrow A)}{sup(X)} + \frac{sup(X \rightarrow A)}{sup(A)} \right) - \max \left\{ \frac{1}{2} \left(\frac{sup(Y \rightarrow A)}{sup(Y)} + \frac{sup(Y \rightarrow A)}{sup(A)} \right) \right\} \quad (3)$$

$Pure-Eff(X \rightarrow A)$ 는 이벤트 X 가 이벤트 A 에 미치는 영향력을 계산하고, 이벤트 X 를 제외한 이벤트 A 에 영향을 미치는 다수의 이벤트 중 최대 영향력을 갖는 이벤트를 찾아 그 영향력을 빼줌으로서 순수 영향력을 구할 수 있다. 그리고 이벤트 A 발생에 대한 모든 원인이벤트에 대하여 식 3의 과정을 수행함으로써 A 발생에 영향력을 주는 주요한 원인이벤트를 예측할 수 있다.

4. 결론 및 향후 연구

이 논문에서는 인터벌이벤트 사이에 존재하는 연관 정보를 탐사하고 원인이벤트가 결과이벤트에 미친 영향력 정도를 측정하였다. 제안한 방법은 영향력 정도 측정을 통하여 이벤트 발생에 다수의 원인이벤트가 존재하는 경우, 주요한 원인이벤트를 탐사함으로써 이벤트 발생에 대한 주요 원인이벤트 정보를 제공했다. 향후 연구로 시간속성을 갖는 이벤트 발생 정도에 대한 알고리즘을 개발하고 알고리즘 성능을 평가하는 연구를 진행하고자 한다.

참고문헌

- [1] 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.
- [2] H. Zhu, and Z. Xu, "An Effective Algorithm for Mining Positive and Negative Association Rules", *International Conference on Computer Science and Software Engineering*, pp. 455-458, 2008.
- [3] 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, Issue 3, pp.181-191, Sep. 2003.
- [4] B. Ozden, S. Ramaswamy, and A. Silberschatz, "Cyclic Association Rules", *International Conference on Data Engineering*, Orlando, pp. 412-421, USA, 1998.
- [5] X. Chen, I. Petrounias, H. Heathfield, "Discovering Temporal Association Rules in Temporal Databases", *International Workshop on Applications of Database Technology*, pp. 312-319, 1998.