



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년05월29일
(11) 등록번호 10-1735285
(24) 등록일자 2017년05월08일

(51) 국제특허분류(Int. Cl.)

G06F 17/30 (2006.01)

(52) CPC특허분류

G06F 17/30539 (2013.01)

G06F 17/30227 (2013.01)

(21) 출원번호 10-2016-0066788

(22) 출원일자 2016년05월30일

심사청구일자 2016년05월30일

(56) 선행기술조사문헌

KR1020130060796 A

윤은일 외 6명. 이레이저블 패턴들을 마이닝하기 위한 최신 알고리즘들의 분석 및 성능평가. 2015년도 한국인터넷정보학회 춘계학술발표대회 논문집. 2015년 5월, 제16권1호, page 85-86.

KR1020130088584 A

(73) 특허권자

세종대학교 산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

윤은일

서울특별시 성북구 종암로5길 7 (종암동)

이강인

충청북도 청주시 서원구 모충로12번길 4, 103동 905호 (개신동, 송학아파트)

(뒷면에 계속)

(74) 대리인

정부연

전체 청구항 수 : 총 10 항

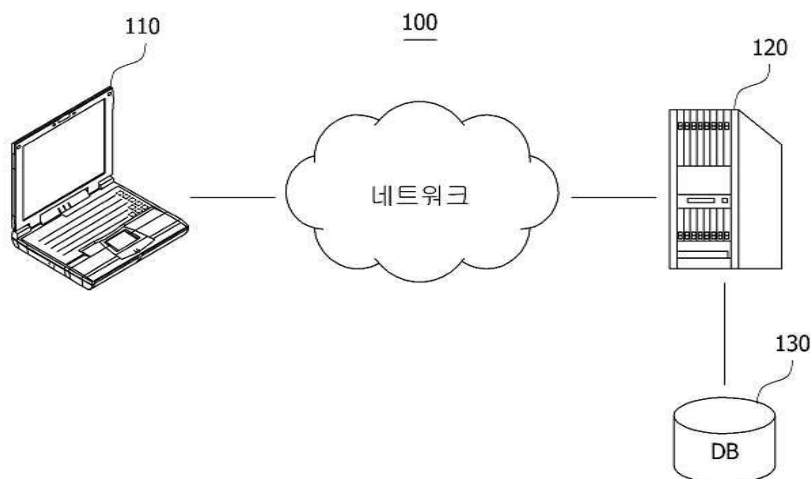
심사관 : 최정권

(54) 발명의 명칭 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법

(57) 요약

가중화 이레이저블 패턴을 마이닝하는 방법은 가중화 이레이저블 패턴 마이닝 장치에서 수행된다. 상기 방법은 (a) 데이터베이스 내의 아이템 가중치 정보를 스캔하여 가중화 이레이저블 후보 아이템을 추출하는 단계, (b) 상기 가중화 이레이저블 후보 아이템의 유효 정보를 기초로 가중화 이레이저블 패턴 마이닝을 위한 가중화 이레이저블 패턴 트리를 도출하는 단계 및 (c) 상기 가중화 이레이저블 패턴 트리를 기반으로 노드 순서 리스트를 구축하고, 재귀적 방식 기반으로 상기 가중화 이레이저블 패턴을 확장하여 가중화 이레이저블 패턴 마이닝 결과를 획득하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06F 17/30327 (2013.01)

G06F 2216/03 (2013.01)

(72) 발명자

양홍모

충청북도 청주시 상당구 수영로 350, 104동 302호
(금천동, 장자마을1단지부영아파트)

김동규

경기도 김포시 전원로 28, 114동 1304호 (장기동,
전원마을월드1단지아파트)

공지예외적용 : 있음

명세서

청구범위

청구항 1

가중화 이레이저블 패턴 마이닝 장치에서 수행되는 가중화 이레이저블 패턴을 마이닝하는 방법에 있어서,

- (a) 데이터베이스 내의 아이템 가중치 정보를 스캔하여 가중화 이레이저블 후보 아이템을 추출하는 단계;
- (b) 상기 가중화 이레이저블 후보 아이템의 유효 정보를 기초로 가중화 이레이저블 패턴 마이닝을 위한 가중화 이레이저블 패턴 트리를 도출하는 단계; 및
- (c) 상기 가중화 이레이저블 패턴 트리를 기반으로 노드 순서 리스트를 구축하고, 재귀적 방식 기반으로 상기 가중화 이레이저블 패턴을 확장하여 가중화 이레이저블 패턴 마이닝 결과를 획득하는 단계를 포함하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 2

제1항에 있어서, 상기 (a) 단계는

상기 가중화 이레이저블 후보 아이템에 대한 지지도 또는 빈도수를 내림차순으로 계산하는 단계를 포함하는 것을 특징으로 하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 3

제2항에 있어서, 상기 (a) 단계는

상기 계산된 정보와 가중치 정보를 기초로 상기 데이터베이스를 구성하는 아이템들에서 가중화 이레이저블 후보 패턴 정보를 선별하는 단계를 더 포함하는 것을 특징으로 하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 4

제1항에 있어서, 상기 (b) 단계는

상기 데이터베이스의 첫 번째 스캔을 통해 정렬된 아이템들에서 유효하지 않은 아이템들을 제거하고, 두 번째 스캔을 통해 가중화 이레이저블 후보 패턴 정보 중 적어도 하나의 유효한 데이터를 선별하여 상기 가중화 이레이저블 패턴 트리를 구축하는 단계를 포함하는 것을 특징으로 하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 5

제1항에 있어서, 상기 (c) 단계는

상기 구축된 가중화 이레이저블 패턴 트리의 각 노드에 대한 전위 순회 및 후위 순회 순서를 계산하는 단계를 포함하는 것을 특징으로 하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 6

제5항에 있어서, 상기 (c) 단계는

상기 구축된 가중화 이레이저블 패턴 트리의 각 노드에 대해 얻어진 전위 순회 및 후위 순회 정보와 이윤 값을 이용하여 상기 노드 순서 리스트를 구축하는 단계를 더 포함하는 것을 특징으로 하는 가중화 이레이저블 패턴을

마이닝하는 방법.

청구항 7

제1항에 있어서, 상기 가중화 이레이저블 패턴은

상기 데이터베이스 내 각각의 아이템에 고유의 가중치 값을 부여한 정보를 기초로 아이템 중요도 순위가 낮은 패턴에 해당하는 것을 특징으로 하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 8

제1항에 있어서, 상기 노드 순서 리스트는

상기 패턴의 확장을 위해 사용하는 해시리스트 형태의 자료구조로, 상기 가중화 이레이저블 패턴 트리의 각 노드의 정보를 저장하는 것을 특징으로 하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 9

제8항에 있어서, 상기 각 노드의 정보는

전위 순회 값, 후위 순회 값 및 이윤 값을 포함하고,

상기 노드 순서 리스트의 인덱스는 상기 전위 순회 값에 해당하며 상기 노드 순서 리스트의 원소는 {상기 전위 순회 값, 상기 후위 순회 값, 상기 이윤 값}을 가지는 것을 특징으로 하는 가중화 이레이저블 패턴을 마이닝하는 방법.

청구항 10

가중화 이레이저블 패턴 마이닝 장치에 있어서,

(a) 데이터베이스 내의 아이템 가중치 정보를 스캔하여 가중화 이레이저블 후보 아이템을 추출하고, (b) 상기 가중화 이레이저블 후보 아이템의 유효 정보를 기초로 가중화 이레이저블 패턴 마이닝을 위한 가중화 이레이저블 패턴 트리를 도출하며, (c) 상기 가중화 이레이저블 패턴 트리를 기반으로 노드 순서 리스트를 구축하고, 재귀적 방식 기반으로 상기 가중화 이레이저블 패턴을 확장하여 가중화 이레이저블 패턴 마이닝 결과를 획득하는 가중화 이레이저블 패턴 마이닝 장치.

발명의 설명

기술 분야

[0001] 본 발명은 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 기술에 관한 것으로, 보다 상세하게는 데이터베이스 내의 각 아이템에 가중치 조건을 부여함으로써 알고리즘의 가용성을 향상시킬 수 있는 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법에 관한 것이다.

배경 기술

[0003] 종래의 빈발 패턴 마이닝 기술은 아이템 또는 다수의 아이템들로 구성되는 패턴의 지지도 정보만을 고려하고, 주어진 최소 지지도 임계 값과 비교하여 단순히 아이템 또는 패턴이 빈발한지 아닌지를 판단한다. 또한, 종래의 이레이저블 패턴 마이닝 방법들은 데이터베이스를 구성하는 다양한 아이템들 고유의 가중치 요소를 마이닝 과정에 적용할 수 없다는 문제점을 가지고 있다. 종래의 빈발 패턴 마이닝 기술은 데이터베이스로부터 유효한 패턴들을 포함하는 트랜잭션을 검출하기 위하여 반복적인 스캔작업을 수행한다. 즉, 종래의 빈발 패턴 마이닝 기술

은 원하는 패턴 결과를 얻기까지 반복적인 스캔 작업을 통해 막대한 연산 오버헤드를 발생시키는 문제점을 가지고 있다.

[0004] 한국등록특허 제10-1443285호는 유용성 높은 패턴의 마이닝 방법에 관한 것으로, 데이터베이스로부터 관독된 트랜잭션의 유용성을 계산하여 트랜잭션 내 아이템을 정렬하고, 아이템 노드의 빈도수와 수량 설정에 따라 헤더 테이블을 갱신하여 전역 유용성 패턴 트리를 구축하고, 헤더 테이블을 트랜잭션 가중화 유용성의 내림차순으로 재정렬하는 기술을 개시한다.

[0005] Z.H. Deng, X.R. Xu, "Fast mining erasable itemsets using NC_sets", Expert Systems with Applications, 2012는 이레이저블 패턴 마이닝의 기본 개념을 적용한 트리 기반의 마이닝 하는 방법에 관한 기술로, 단순히 주어진 데이터베이스 내 각 트랜잭션의 이온 값을 고려하여 패턴의 게인 값을 도출하고, 이를 기초로 주어진 최대 게인 임계 값과 비교하여 각 패턴이 유효한지 아닌지를 판별하기 때문에 유용한 이레이저블 패턴 결과를 선별적으로 추출할 수 없다

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국등록특허 제10-1443285호 (2014.09.22 등록)

비특허문헌

[0009] (비특허문헌 0001) Z.H. Deng, X.R. Xu, "Fast mining erasable itemsets using NC_sets", Expert Systems with Applications, 2012

발명의 내용

해결하려는 과제

[0010] 본 발명의 일 실시예는 이레이저블 패턴 마이닝을 위한 기준 값으로 설정되는 최대 게인 임계 값 이하의 게인 값을 갖는 패턴을 유효한 이레이저블 패턴으로 추출하는 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법을 제공하고자 한다.

[0011] 본 발명의 일 실시예는 가중치 적용에 따른 패턴 손실 없이 실제 유효한 가중화 이레이저블 패턴을 선별적으로 추출하는 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법을 제공하고자 한다.

[0012] 본 발명의 일 실시예는 분할정복 기반의 재귀적 호출 방식을 기초로 패턴을 확장하는 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법을 제공하고자 한다.

과제의 해결 수단

[0014] 실시예들 중에서, 가중화 이레이저블 패턴을 마이닝하는 방법은 가중화 이레이저블 패턴 마이닝 장치에서 수행된다. 상기 방법은 (a) 데이터베이스 내의 아이템 가중치 정보를 스캔하여 가중화 이레이저블 후보 아이템을 추출하는 단계, (b) 상기 가중화 이레이저블 후보 아이템의 유효 정보를 기초로 가중화 이레이저블 패턴 마이닝을 위한 가중화 이레이저블 패턴 트리를 도출하는 단계 및 (c) 상기 가중화 이레이저블 패턴 트리를 기반으로 노드 순서 리스트를 구축하고, 재귀적 방식 기반으로 상기 가중화 이레이저블 패턴을 확장하여 가중화 이레이저블 패턴 마이닝 결과를 획득하는 단계를 포함한다.

[0015] 상기 (a) 단계는 상기 가중화 이레이저블 후보 아이템에 대한 지지도 또는 빈도수를 내림차순으로 계산하는 단계를 포함할 수 있다. 상기 (a) 단계는 상기 계산된 정보와 가중치 정보를 기초로 상기 데이터베이스를 구성하는 아이템들에서 가중화 이레이저블 후보 패턴 정보를 선별하는 단계를 더 포함할 수 있다.

[0016] 상기 (b) 단계는 상기 데이터베이스의 첫 번째 스캔을 통해 정렬된 아이템들에서 유효하지 않은 아이템들을 제거하고, 두 번째 스캔을 통해 가중화 이레이저블 후보 패턴 정보 중 적어도 하나의 유효한 데이터를 선별하여

상기 가중화 이레이저블 패턴 트리를 구축하는 단계를 포함할 수 있다.

- [0017] 상기 (c) 단계는 상기 구축된 가중화 이레이저블 패턴 트리의 각 노드에 대한 전위 순회 및 후위 순회 순서를 계산하는 단계를 포함할 수 있다. 상기 (c) 단계는 상기 구축된 가중화 이레이저블 패턴 트리의 각 노드에 대해 얻어진 전위 순회 및 후위 순회 정보와 이윤 값을 이용하여 상기 노드 순서 리스트를 구축하는 단계를 더 포함할 수 있다.
- [0018] 상기 가중화 이레이저블 패턴은 상기 데이터베이스 내 각각의 아이템에 고유의 가중치 값을 부여한 정보를 기초로 아이템 중요도 순위가 낮은 패턴에 해당할 수 있다.
- [0019] 상기 노드 순서 리스트는 상기 패턴의 확장을 위해 사용하는 해시리스트 형태의 자료구조로, 상기 가중화 이레이저블 패턴 트리의 각 노드의 정보를 저장할 수 있다. 상기 각 노드의 정보는 전위 순회 값, 후위 순회 값 및 이윤 값을 포함할 수 있고, 상기 노드 순서 리스트의 인덱스는 상기 전위 순위 값에 해당하며 상기 노드 순서 리스트의 원소는 {상기 전위 순회 값, 상기 후위 순회 값, 상기 이윤 값}을 가질 수 있다.
- [0020] 실시예들 중에서, 가중화 이레이저블 패턴 마이닝 장치는 (a) 데이터베이스 내의 아이템 가중치 정보를 스캔하여 가중화 이레이저블 후보 아이템을 추출하고, (b) 상기 가중화 이레이저블 후보 아이템의 유효 정보를 기초로 가중화 이레이저블 패턴 마이닝을 위한 가중화 이레이저블 패턴 트리를 도출하며, (c) 상기 가중화 이레이저블 패턴 트리를 기반으로 노드 순서 리스트를 구축하고, 재귀적 방식 기반으로 상기 가중화 이레이저블 패턴을 확장하여 가중화 이레이저블 패턴 마이닝 결과를 획득한다.

발명의 효과

- [0022] 개시된 기술은 다음의 효과를 가질 수 있다. 다만, 특정 실시예가 다음의 효과를 전부 포함하여야 한다거나 다음의 효과만을 포함하여야 한다는 의미는 아니므로, 개시된 기술의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.
- [0023] 본 발명의 일 실시예에 따른 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법은 이레이저블 패턴 마이닝을 위한 기준 값으로 설정되는 최대 개인 임계 값 이하의 개인 값을 갖는 패턴을 유효한 이레이저블 패턴으로 추출할 수 있다.
- [0024] 본 발명의 일 실시예에 따른 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법은 가중치 적용에 따른 패턴 손실 없이 실제 유효한 가중화 이레이저블 패턴을 선별적으로 추출할 수 있다.
- [0025] 본 발명의 일 실시예에 따른 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝하는 방법은 분할정복 기반의 재귀적 호출 방식을 기초로 패턴을 확장할 수 있다.

도면의 간단한 설명

- [0027] 도 1은 본 발명의 일 실시예에 따른 가중화 이레이저블 패턴 마이닝 시스템을 설명하는 블록도이다.
- 도 2는 도 1에 있는 가중화 이레이저블 패턴 마이닝 장치에서 수행되는 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝 하는 과정을 설명하는 순서도이다.
- 도 3은 본 발명의 일 실시예에 따른 가중화 이레이저블 패턴 트리와 노드 순서 리스트를 구축하는 과정을 설명하는 순서도이다.
- 도 4는 가중화 이레이저블 패턴 마이닝에 이용되는 데이터베이스와 가중치 정보를 설명하는 도면이다.
- 도 5는 최소 지지도 임계 값이 30%일 때 구축되는 가중화 이레이저블 패턴 트리와 그에 해당하는 노드 순서 리스트 정보를 설명하는 도면이다.
- 도 6은 분할정복 방식에 기반한 재귀적 방법을 통한 가중화 이레이저블 패턴 마이닝 과정을 설명하는 순서도이다.
- 도 7은 최소 지지도 임계 값이 30%일 때 마이닝 되는 가중화 이레이저블 패턴의 패턴 조합 과정을 설명하는 도면이다.
- 도 8 내지 도 13은 본 발명의 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마

이닝하는 과정을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0028] 본 발명에 관한 설명은 구조적 내지 기능적 설명을 위한 실시예에 불과하므로, 본 발명의 권리범위는 본문에 설명된 실시예에 의하여 제한되는 것으로 해석되어서는 아니 된다. 즉, 실시예에는 다양한 변경이 가능하고 여러 가지 형태를 가질 수 있으므로 본 발명의 권리범위는 기술적 사상을 실현할 수 있는 균등물들을 포함하는 것으로 이해되어야 한다. 또한, 본 발명에서 제시된 목적 또는 효과는 특정 실시예가 이를 전부 포함하여야 한다거나 그러한 효과만을 포함하여야 한다는 의미는 아니므로, 본 발명의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.
- [0029] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0030] "제1", "제2" 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위한 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.
- [0031] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결될 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다고 언급된 때에는 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 한편, 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0032] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함하다" 또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0033] 각 단계들에 있어 식별부호(예를 들어, a, b, c 등)는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 단계들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.
- [0034] 본 발명은 컴퓨터가 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현될 수 있고, 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있으며, 또한, 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.
- [0035] 여기서 사용되는 모든 용어들은 다르게 정의되지 않는 한, 본 발명이 속하는 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한 이상적이거나 과도하게 형식적인 의미를 지니는 것으로 해석될 수 없다.
- [0037] 도 1은 본 발명의 일 실시예에 따른 가중화 이레이저블 패턴 마이닝 시스템을 설명하는 블록도이다.
- [0038] 도 1을 참조하면, 가중화 이레이저블 패턴 마이닝 시스템(100)은 사용자 단말(110), 가중화 이레이저블 패턴 마이닝 장치(120) 및 데이터베이스(130)를 포함한다.
- [0039] 사용자 단말(110)은 사용자의 제어에 의해 가중화 이레이저블 패턴 마이닝 장치(120)에 접속하여 가중화 이레이저블 패턴의 마이닝을 요청할 수 있다. 여기에서, 가중화 이레이저블 패턴은 데이터베이스(130) 내 각각의 아이템에 고유의 가중치 값을 부여한 정보를 기초로 아이템 중요도 순위가 낮은 패턴에 해당할 수 있다. 사용자 단말(110)은 가중화 이레이저블 패턴 마이닝 장치(120)에서 제공된 마이닝 결과 정보 또는 데이터 그룹의 특성 정보를 화면에 디스플레이 할 수 있다. 사용자 단말(110)은 모바일 단말기, 태블릿 PC, 랩탑 PC 또는 데스크탑 PC에 해당할 수 있다.
- [0040] 가중화 이레이저블 패턴 마이닝 장치(120)는 사용자 단말(110)의 요청에 따라 데이터베이스(130)에 저장된 데이

터에 대한 중요 패턴 마이닝을 수행할 수 있다. 보다 구체적으로, 가중화 이레이저블 패턴 마이닝 장치(120)는 사용자 데이터의 생성 시각을 기초로 특정 시간 범위 내의 사용자 데이터를 추출하고, 추출된 사용자 데이터를 아이템 헤더 테이블 및 가중화 이레이저블 패턴 트리를 생성 및 갱신할 수 있다. 가중화 이레이저블 패턴 마이닝 장치(120)는 가중화 이레이저블 패턴 트리 경로의 아이템 링크를 기초로 해당 경로에 대한 특성 값을 추출하고 가중화 이레이저블 패턴 트리 경로의 대표 특성 값을 도출하여, 패턴 마이닝의 결과를 분석할 수 있다. 가중화 이레이저블 패턴 마이닝 장치(120)는 마이닝 결과 정보 또는 데이터 그룹의 특성 정보를 사용자 단말(110)에 제공할 수 있다.

- [0042] 도 2는 도 1에 있는 가중화 이레이저블 패턴 마이닝 장치에서 수행되는 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마이닝 하는 과정을 설명하는 순서도이다.
- [0043] 도 2를 참조하면, 가중화 이레이저블 패턴 마이닝 장치(120)는 데이터베이스(130)와 데이터베이스 내의 아이템 가중치 정보를 스캔하고(단계 S210), 데이터베이스 내 포함되는 데이터의 상세 정보를 파악하여 불필요한 정보를 제거하며, 마이닝을 위한 필수적인 데이터를 선별하여 가중화 이레이저블 후보 아이템을 추출할 수 있다(단계 S220).
- [0044] 가중화 이레이저블 패턴 마이닝 장치(120)는 추출 과정에서 추출된 가중화 이레이저블 후보 아이템의 유효 정보를 기초로 가중화 이레이저블 패턴 마이닝을 위한 트리 구조를 구축할 수 있고(단계 S230), 이를 통해 가중화 이레이저블 패턴 트리를 도출할 수 있다(단계 S235).
- [0045] 가중화 이레이저블 패턴 마이닝 장치(120)는 도출된 가중화 이레이저블 패턴 트리를 기반으로 노드 순서 리스트를 구축할 수 있고(단계 S240), 이를 통해 노드 순서 리스트를 도출할 수 있다(단계 S245).
- [0046] 가중화 이레이저블 패턴 마이닝 장치(120)는 가중화 이레이저블 패턴 트리 구조와 노드 순서 리스트에 대한 구축이 완료되면, 재귀적 방식 기반으로 가중화 이레이저블 패턴을 확장할 수 있고(단계 S250), 그 결과로 가중화 이레이저블 패턴 마이닝 결과를 획득할 수 있다(단계 S260).
- [0048] 도 3은 본 발명의 일 실시예에 따른 가중화 이레이저블 패턴 트리와 노드 순서 리스트를 구축하는 과정을 설명하는 순서도이다.
- [0049] 도 3을 참조하면, 가중화 이레이저블 패턴 마이닝 장치(120)는 먼저 데이터베이스 내의 데이터(즉, 가중화 이레이저블 후보 아이템)를 한번 스캔하여 데이터에 대한 지지도 또는 데이터 빈도수를 내림차순으로 계산하고(단계 S310), 계산된 정보와 가중치 정보를 기초로 데이터베이스를 구성하는 아이템들 중 가중화 이레이저블 후보 패턴 정보를 선별할 수 있다. 이때, 가중화 이레이저블 패턴 마이닝 장치(120)는 가중치 요소의 적용에 따른 패턴 손실을 막기 위해 과소추정 제약조건을 이용한 근사치 패턴 프루닝 기법을 활용할 수 있다.
- [0050] 가중화 이레이저블 패턴 마이닝 장치(120)는 선별된 가중화 이레이저블 후보 패턴 정보 중 적어도 하나의 유효한 데이터를 선별하여 저장하기 위한 빈 가중화 이레이저블 패턴 트리를 생성하고, 두 번째 데이터베이스 스캔을 통해 가중화 이레이저블 패턴 트리를 구축할 수 있다(단계 S320).
- [0051] 가중화 이레이저블 패턴 마이닝 장치(120)는 가중화 이레이저블 패턴 트리 구축 과정을 데이터베이스 내 각 트랜잭션을 전처리하고 삽입하는 순서로 이루어지고, 더 이상 처리할 트랜잭션이 없을 때까지 해당 작업을 반복해서 수행할 수 있다(단계 S330). 가중화 이레이저블 패턴 마이닝 장치(120)는 만약 처리할 트랜잭션이 있는 경우, 현재 트랜잭션에 대해 첫 번째 스캔과정을 통해 얻어진 지지도(또는 빈도수) 내림차순에 따라 트랜잭션의 아이템들을 정렬하고, 이 과정에서 유효하지 않은 아이템들을 제거할 수 있다.
- [0052] 또한, 가중화 이레이저블 패턴 마이닝 장치(120)는 상기 작업 과정을 통해 전처리 된 트랜잭션을 빈 가중화 이레이저블 패턴 트리에 삽입할 수 있고(단계 S340), 삽입 후에는 추가된 데이터에 대한 헤더 테이블 정보를 갱신할 수 있다(단계 S350).
- [0053] 가중화 이레이저블 패턴 마이닝 장치(120)는 상기 작업이 더 이상 처리할 트랜잭션이 없을 때까지 반복하고 모든 트랜잭션에 대한 작업이 마무리되면 현재 데이터베이스에 대한 완전한 가중화 이레이저블 패턴 트리를 획득할 수 있다(단계 S360).
- [0054] 가중화 이레이저블 패턴 마이닝 장치(120)는 가중화 이레이저블 패턴 트리 구축 작업이 마무리 되면, 구축된 가중화 이레이저블 패턴 트리의 각 노드에 대한 전위 순회 및 후위 순회 순서를 계산할 수 있고(단계 S370), 각 노드에 대해 얻어진 전위 순회 및 후위 순회 정보와 이윤 값을 이용하여 노드 순서 리스트를 구축할 수 있다(단

계 S380). 구축된 가중화 이레이저블 패턴 트리에 해당하는 노드 순서 리스트를 획득할 수 있다(단계 S390).

[0055] 여기에서, 이러한 관계 특성을 이용하는 이유는 다음과 같다. 구축된 가중화 이레이저블 패턴 트리의 각 경로는 하나의 전처리 된 트랜잭션을 의미한다. 따라서, 만약 어떠한 추가적인 고려사항 없이 구축된 트리에 대해 마이닝 작업을 수행한다면, 패턴의 게인 값을 구하는 과정에서 같은 트랜잭션에 속하는 두 개 이상의 아이템들의 이윤 값들이 중복해서 합산될 수 있다. 반면에 패턴의 게인 값을 구하는 과정에서 노드 간의 조상-후손 관계 특성을 고려하여 어떤 두 개의 노드가 서로 조상-후손의 관계를 가진다면, 해당 노드들 각각의 이윤 값을 중복해 고려하지 않음으로써 잘못된 게인 값 계산을 막고 정확한 가중화 이레이저블 패턴 마이닝 작업을 수행할 수 있다.

[0056]

도 4는 가중화 이레이저블 패턴 마이닝에 이용되는 데이터베이스와 가중치 정보를 설명하는 도면이다.

[0058]

도 4를 참조하면, 도 4(a)는 데이터베이스를 설명하는 도면이고, 도 4(b)는 데이터베이스 내의 모든 아이템들의 가중치 정보를 설명하는 도면이다.

[0059]

도 4(a)의 데이터베이스(130)는 트랜잭션을 구별하는 구분자 PID(410), 트랜잭션을 구성하는 아이템들(420), 해당 트랜잭션의 이윤 값(430)을 포함하고, 이윤 값(430)은 각종 화폐 단위나 기타 이윤을 나타낼 수 있는 기호로 표현될 수 있다.

[0060]

도 4(a)의 데이터베이스(130)가 복수의 트랜잭션들로 구성되어 있을 때, 데이터베이스 D는 $D = \{P_1, P_2, P_3, \dots, P_n\}$ 으로 정의된다. 각각의 트랜잭션 P_k (k 는 1에서 n 사이의 값)는 고유의 키 값, 이윤 값 및 아이템들의 집합($I = \{i_1, i_2, i_3, \dots, i_m\}$)으로 구성된다. 예를 들어, 도 4(a)의 데이터베이스에서 아이템 B, D를 통해 트랜잭션 P_8 이 구성되며, 트랜잭션 P_8 을 통해 얻은 이윤 값은 50이다.

[0061]

패턴 X의 게인 값은 패턴 X를 한 아이템 집합이라고 할 때, 다음과 같이 정의된다.

수학식 1

$$Gain(X) = \sum_{\{P_k | X \cap P_k \text{ Items} \neq \emptyset\}} P_k \cdot profit$$

[0062]

예를 들어, 도 4(a)의 데이터베이스에서 $Gain(A) = 50 + 250 = 300$ 이고, $Gain(J) = 50 + 250 + 100 = 400$ 이며, $Gain(AJ) = 50 + 250 + 100 = 400$ 이다.

[0064]

최대 게인 임계 값(Maximum gain threshold(δ))은 데이터베이스의 전체 트랜잭션들의 이윤 값 총 합에 퍼센테이지로 주어진다. 그러면 최대 게인 임계 값(Maximum gain threshold)은 다음과 같이 정의 된다.

수학식 2

$$maximum\ gain\ threshold(\delta) = \delta \times \sum_{P_k \in D} P_k \cdot profit$$

[0065]

예를 들어, 도 4(a)의 데이터베이스에서 최대 게인 임계 값(δ)이 30%일 때, $maximum\ gain\ threshold(0.3) = 0.3 \times 5250 = 1575$ 이다.

[0067]

이레이저블 패턴(Erasable pattern)은 주어진 최대 게인 임계 값(δ)이 있을 때, 어떤 패턴 X가 이레이저블하다는 것은 패턴 X의 게인 값이 최대 게인 임계 값보다 작거나 같다는 뜻이다. 즉, 패턴 X가 이레이저블 패턴이라면 패턴 X는 다음 수식을 만족한다.

수학식 3

$$\text{maximum gain threshold}(\delta) \geq \text{Gain}(X)$$

[0068]

[0069] 예를 들어, 도 4(a)의 데이터베이스에서 최대 게인 임계 값이 0.3일 때, Gain(A)는 300 (<1575)이므로 패턴 A는 이레이저블 패턴이고, 패턴 AJ 또한 이레이저블 하다. 반대로, Gain(E)와 Gain(EH)의 값은 각각 1750, 2550으로 최대 게인 임계 값(1575)보다 큰 값을 가지므로 언이레이저블(unerasable) 패턴이다.

[0070] 안티-모노톤 법칙(Anti-monotone property)은 만약 패턴 Y가 패턴 X의 상위 패턴일 때, 패턴 X가 언이레이저블 하다면 패턴 Y 또한 언이레이저블 하다. 예를 들어, 도 4(a)의 데이터베이스에서 최대 게인 임계 값이 0.3일 때, 패턴 EH는 패턴 E의 상위 패턴이고, 패턴 E는 Gain(E) = 1750 (>1575)하기 때문에 언이레이저블 하다. 그러므로 패턴 EH 또한 언이레이저블 하다.

[0071] 또한, 안티-모노톤 법칙은 전통적인 이레이저블 패턴 마이닝의 핵심적인 규칙으로서, 패턴을 조합하거나 확장할 때 효율적으로 처리할 수 있다.

[0073] 도 4(b)의 아이템(420)의 가중치 요소는 각각의 아이템에 대한 고유의 가중치가 할당될 수 있다(440).

[0074] 도 4(b)의 아이템의 가중치(Weight of an item)는 상품의 가격과 같은 사용자에게 의해 특정된 값이다. 데이터베이스에 포함되는 모든 아이템들은 자신의 고유한 가중치를 할당 받을 수 있다. 특정 아이템 i_p 의 가중치인 $W(i_p)$ 은 양의 실수 값으로 표현된다. 예를 들어, 아이템 A의 가중치는 $W(A) = 0.8$ 과 같은 값을 취할 수 있고, 아이템 B와 G의 가중치는 $W(B) = 0.4$, $W(G) = 1.0$ 으로 표현할 수 있다.

[0075] 상품 가중치 $PW(P_q)$ 는 다음과 같이 정의 된다.

수학식 4

$$PW(P_q) = \sum_{i_p \in P_q} W(i_p)$$

[0076]

[0077] 예를 들어, P_4 의 상품 가중치 $PW(P_4) = W(C) + W(D) + W(H) = 0.4 + 0.4 + 0.4 = 1.2$ 이고, P_6 의 상품 가중치는 $PW(P_6) = W(A) + W(E) + W(F) + W(H) + W(I) + W(J) = 0.8 + 0.4 + 1.0 + 0.4 + 0.2 + 0.6 = 3.4$ 이다.

[0078] 본 발명의 일 실시예는 데이터베이스 내의 각 아이템들의 중요도를 판단하기 위해 아이템들의 가중치 합을 이용함으로써 각각의 트랜잭션에서 아이템들이 가지는 생산기여도를 계산한다. 생산기여도는 다음과 같이 정의된다.

[0079] 예를 들어, $P_q = \{i_1, i_2, \dots, i_n\}$ 가 아이템들로 구성된 한 트랜잭션이라고 하자. 아이템들의 고유 가중치를 이용함으로써 P_q 에 포함된 각 아이템들이 트랜잭션 P_q 를 구성하는데 얼마만큼의 기여도가 있는지 계산할 수 있다. 즉, 아이템의 기여도는 그 아이템이 특정 트랜잭션에서 차지하는 중요도의 비율을 의미한다. 트랜잭션 P_q 안의 아이템 i_p (P_q 의 원소)에 대한 생산기여도 $MCR(i_p, P_q)$ 는 다음의 수식을 통해 계산할 수 있다.

수학식 5

$$MCR(i_p, P_q) = \frac{W(i_p)}{PW(P_q)}$$

[0080]

[0081] 수학식 5를 이용하여 데이터베이스 내의 트랜잭션에 포함된 아이템의 생산기여도를 파악한다. 생산기여도는 아이템의 가중치가 존재한다면 절대 0의 값을 가질 수 없다. 수학식 5를 통해 계산된 결과는 0보다 크고 1보다 작거나 같은($0 < \text{생산기여도}(i_p, P) \leq 1$) 양의 실수 형태를 가진다.

[0082] 예를 들어, 도 4(a)의 데이터베이스 안의 트랜잭션들 내의 각 아이템 생산기여도는 도 4(b)의 각 아이템별 가중치와 아이템 생산기여도를 이용하여 다음과 같이 구할 수 있다. 트랜잭션 P_1 의 가중치 $PW(P_1)$ 는 1.8이다. 트

랜잭션 P₁에서 MCR(A, P₁)은 0.8 / 1.8 = 0.444이고, MCR(B, P₁)는 0.4 / 1.8 = 0.222이며, MCR(J, P₁)는 0.333이다. 트랜잭션 P₂의 가중치 PW(P₂)는 0.8이다. 트랜잭션 P₂의 아이템 E의 생산기여도 MCR(E, P₂)는 0.4 / 0.8 = 0.5이고, MCR(H, P₂)는 0.4 / 0.8 = 0.5이다. 이를 통해 도 4(a)의 데이터베이스 내의 모든 각각의 아이템들에 대해서 계산할 수 있다.

[0083] 또한, 아이템의 가중치를 통해서 얻어진 생산기여도들은 데이터베이스가 스캔 되는 동안 계산되게 된다. 각 아이템은 해당 아이템이 포함된 트랜잭션들의 수만큼 생산기여도 값을 가지게 되며, 이 값들을 아이템별로 그룹화하여 평균을 계산할 수 있고, 계산결과를 다음과 같이 정의한다.

[0084] 아이템의 평균 생산기여도는 아이템 아이템별 가중치와 각 트랜잭션의 정보를 통해 구한 기여도의 평균 값이다. 평균 생산기여도는 데이터베이스 내의 각각의 아이템에 대해 구하며, 각 아이템의 생산기여도와 다음의 수식을 이용함으로써 해당 아이템에 대해서 공통적으로 적용될 수 있는 값이 계산된다.

수학식 6

$$MCR_{mean}(i_p) = \frac{\sum_{i_p \in P_q \in D} \text{생산기여도}(i_p, P_q)}{\# \text{ of } \forall P_r \ni i_p}$$

[0085]

[0086] 예를 들어, MCR_{mean}(A)는 (MCR(A, P₁) + MCR(A, P₆)) / 2 = (0.444 + 0.235) / 2 = 0.340이고, MCR_{mean}(B)는 (MCR(B, P₁) + MCR(B, P₅) + MCR(B, P₈) + MCR(B, P₉)) / 4 = (0.222 + 0.286 + 0.500 + 0.286) / 4 = 0.323이다.

[0087] 표 1은 아이템들의 생산기여도들을 이용하여 평균 생산기여도들을 측정된 결과이다.

[0088] [표 1. 아이템별 평균 생산기여도]

Item	A	B	C	D	E	F	G	H	I	J
평균 생산기여도	0.340	0.323	0.326	0.380	0.297	0.294	0.833	0.309	0.142	0.313

[0089]

[0090] 패턴의 평균 생산기여도는 한 패턴 $X = \{i_1, i_2, \dots, i_r\}$ 의 평균 생산기여도 AMCR(X)는 패턴 X에 포함된 아이템들의 생산기여도를 이용함으로써 다음과 같이 계산된다.

수학식 7

$$MCR_{mean}(X) = \frac{\sum_{p=1}^r MCR_{mean}(i_p)}{r}$$

[0091]

[0092] 예를 들어, AMCR(AB)는 (AMCR(A) + AMCR(B)) / 2 = (0.340 + 0.323) / 2 = 0.332이고, AMCR(GI)는 (AMCR(G) + AMCR(I)) / 2 = (0.833 + 0.142) / 2 = 0.488 이다. 또한, 길이가 2 이상인 한 패턴의 평균 생산기여도는 그 패턴의 모든 서브 패턴들의 평균 생산기여도의 평균과 같다.

[0093] 가중화 게인 값(Weighted gain)은 한 패턴 X의 가중화 게인 값 WG(X)은 그 패턴의 게인 값에서 패턴의 평균 생산기여도 값을 나눠서 계산한다. WG(X)는 다음의 수식을 이용해 계산된다.

수학식 8

$$WG(X) = \frac{Gain(X)}{MCR_{mean}(X)}$$

[0094]

[0095] 예를 들어, WG(A)는 300 / 0.340 = 882.4이고, WG(B)는 1850 / 0.323 = 5727.6이며, WG(AB)는 2100 / 0.332 =

6325.3이다.

[0096] 본 발명은 단편적인 접근방식인 게인이 아니라 게인과 함께 아이템의 가중치를 모두 고려하는 가중화 게인 값을 이용하여 패턴을 마이닝할 수 있다.

[0097] 가중화 이레이저블 패턴(Weighted erasable pattern)은 가중치가 적용된 환경에서 주어진 최대 게인 임계 값(δ)이 있을 때, 어떤 패턴 X가 이레이저블하다는 것은 다음과 같이 패턴 X의 가중화 게인 값이 최대 게인 임계 값보다 작거나 같다는 뜻이다. 즉, 수학적 9를 만족하는 패턴 X를 가중화 이레이저블 패턴이라 한다.

수학적 9

$$\text{Maximum gain threshold}(\delta) \geq WG(X)$$

[0098]

[0099] 예를 들어, 최대 게인 임계 값(δ)이 0.3일 때, 도 4(a)의 데이터베이스 안에서 발견 가능한 가중화 이레이저블 패턴들을 마이닝 해보자. 먼저, 길이가 1인 패턴들(또는 아이템들)의 가중화 게인 값을 측정해보자. WG(A)는 882.4, WG(B)는 5727.6, WG(C)는 6595.1, WG(D)는 1447.4, WG(E)는 5892.3, WG(F)는 850.3, WG(G)는 3001.2, WG(H)는 3883.5, WG(I)는 32394.4 및 WG(J)는 1278.0이다. 이 아이템들 중에서 가중화 이레이저블 패턴들 (≤ 1575)은 총 4개 A, D, F, J이다. 수학적 8을 이용해 길이가 2 이상인 패턴들도 추출이 가능하다. 기준을 A로 갖는 패턴들을 계산해보자. WG(AB)=6325.3, WG(AC)=7357.4, WG(AD)=1611.1, WG(AE)=5642.6, WG(AF)=946.4, WG(AG)=4770.0, WG(AH)=3846.2, WG(AI)=19294.6 및 WG(AJ)=1223.2이다. 이 패턴들 중에서 가중화 이레이저블 패턴은 AF와 AJ이다. 이와 같은 방식으로 다른 패턴들도 반복적으로 구해나갈 수 있다.

[0100] 최대 평균 생산기여도 MaxMCR는 데이터베이스 내 각각의 아이템이 가지는 평균 생산기여도 값들 중 가장 큰 값이고, 패턴의 확장을 수행할 때 최대 평균 생산기여도 값을 고려함으로써 패턴이 확장될 때 발생 가능한 패턴손실을 막을 수 있다. 보다 구체적으로, 한 패턴 X의 최대 평균 생산기여도 값 MaxMCR은 평균 생산기여도 값 MCR_mean(X)보다 항상 크거나 같다.

[0101] 최대 가중화 게인 값은 한 패턴 X의 최대 가중화 게인 MWG()은 최대 평균 생산기여도의 MaxMCR와 각 패턴의 게인 값을 이용하여 다음의 수식과 같이 계산한다.

수학적 10

$$MWG(X) = \frac{Gain(X)}{MaxMCR}$$

[0102]

[0103] 계산된 MWG(X) 값이 최대 게인 임계 값 이하의 값을 가지는 경우 가중화 이레이저블 패턴의 후보가 된다. 즉, 한 패턴 X의 최대 가중화 게인 값은 가중화 게인 값보다 항상 작거나 같다.

[0104] 예를 들어, 최대 게인 임계 값(δ)이 0.3일 때, 도 4(a)의 데이터베이스 안에서 발견 가능한 가중화 이레이저블 패턴들의 후보들을 마이닝 할 수 있다. 데이터베이스를 한번 스캔하면서 데이터베이스에 포함되는 모든 아이템들의 정보를 각각 저장한다.

[0105] [표 2. 도 4(a)의 데이터베이스로부터 도출된 아이템들의 정보]

Item	A	B	C	D	E	F	G	H	I	J
Frequency	2	4	4	4	4	1	1	4	5	3
Profit	300	1850	2150	550	1750	250	2500	1200	4600	400
MWG(Item)	360.1	2220.9	2581.0	660.3	2100.8	300.1	3001.2	1440.6	5522.2	480.2

[0106]

[0107] 표 2는 모든 아이들에 대해 저장되는 정보들은 다음과 같다. 아이템 정보는 아이템 이름, 아이템의 지지도(또는 빈도수), 아이템이 등장하는 트랜잭션들의 이윤 값의 총합이다. 첫 번째 데이터베이스 스캔이 끝나면 저장된 정보와 도 4(b)의 아이템별 가중치를 이용하여 각 아이템의 생산기여도 및 평균 생산기여도를 구한다(표 1). 평균 생산기여도 값들 중에서 가장 큰 값인 아이템 G의 0.833을 MaxMCR 값으로 선택한다. 최대 가중화 게인 값의 수

학식 10을 이용함으로써 표 2와 같이 각 아이템의 최대 평균 생산기여도 값 MWG(Item)을 구한다. 임계 값(1575)보다 높은 MWG값을 가지는 아이템들 B, C, E, G, I는 프루닝되고, 가중화 이레이저블 후보 패턴들은 총 5개의 패턴들 A, D, F, H, J 만 남는다.

- [0109] 도 5는 최소 지지도 임계 값이 30%일 때 구축되는 가중화 이레이저블 패턴 트리와 그에 해당하는 노드 순서 리스트 정보를 설명하는 도면이다.
- [0110] 도 5를 참조하면, 가중화 이레이저블 패턴 트리는 데이터베이스의 대표 특징을 저장하는 헤더 테이블(510)과 실제 데이터를 저장하는 프리픽스 트리(520)를 포함한다.
- [0111] 프리픽스 트리(520)는 하나의 루트 노드와 다수의 일반 노드들로 구성된다. 여기에서, 하나의 일반 노드는 노드 이름, 해당 노드의 이윤 값, 노드의 전위 순회 값을 속성으로 포함할 수 있다.
- [0112] 노드 순서 리스트(540)는 프리픽스 트리(520)를 구성하는 각 노드에 대한 전위 순회 및 후위 순회 순서 정보와 각 노드에 해당하는 이윤 정보를 저장할 수 있다.
- [0113] 예를 들어, 도 4(a)의 데이터베이스에 대한 가중화 이레이저블 패턴 트리는 프리픽스 트리(520)로 표현될 수 있다.
- [0114] 도 5(a)는 도 4(a)의 데이터베이스 임계 값이 0.3일 때, 길이가 1인 가중화 이레이저블 후보 패턴들을 마이닝한 후, 후보 패턴들을 이용하여 헤더 테이블(510)과 프리픽스 트리(520)를 구축한 결과이다. 헤더 테이블(510)은 후보 패턴들의 정보를 저장하는 테이블로, 정보는 아이템 이름, 지지도(또는 빈도수), 아이템의 이윤 값 총합을 포함한다. 각 아이템의 이윤 값은 프리픽스 트리(520)에 존재하는 같은 아이템 이름을 가지는 모든 노드들의 값의 총합과 같다. 프리픽스 트리(520)는 아이템들 테이블의 정보를 이용함으로써 다시 한번 데이터베이스를 스캔하는 동안 구축된 프리픽스 트리(520)이다.
- [0115] 헤더 테이블(510)은 길이가 1인 가중화 이레이저블 후보 패턴들을 마이닝한 결과를 지지도(또는 빈도수) 내림차순으로 정렬한 결과이다.
- [0116] 노드 순서 리스트는 효율적인 패턴 확장을 위해 사용하는 해시리스트 형태의 자료구조로, 가중화 이레이저블 패턴 트리의 각 노드의 정보를 저장한다. 여기에서, 각 노드의 정보는 전위 순회 값, 후위 순회 값 및 이윤 값을 포함할 수 있다. 노드 순서 리스트의 인덱스는 전위 순회 값이며, 노드 순서 리스트의 한 원소는 한 노드의 {전위 순회 값, 후위 순회 값, 이윤 값}를 가진다. 즉, 어떤 한 가중화 이레이저블 패턴 트리에서 발견 가능한 노드 순서 리스트의 형식은 다음과 같이 표현된다. NC-list = {<pre-order1, pre-order1, profit1>, <pre-order2, pre-order2, profit2>, ..., <pre-orderk, pre-orderk, profitk>}.
- [0117] 도 5(b)는 가중화 이레이저블 패턴 트리의 각 노드의 정보에 대한 노드 순서 리스트를 설명한다. 예를 들어, 도 5(a)의 가중화 이레이저블 패턴 트리의 각 노드의 정보는 도 5(b)와 같이 노드 순서 리스트로 맵핑된다. 전위 순회 값이 4인 경우, 후위 순회 값은 7이고, 이윤 값은 750이다.
- [0118] 노드 순서 리스트를 이용함으로써 패턴 확장 과정을 수행할 때, 각 노드의 조상-후손 관계를 파악할 수 있다. 가중화 이레이저블 패턴 트리에서 조상-후손 관계는 다음과 같고, 만약 어떤 한 노드 N1가 다른 한 노드 N2의 후손이라면 다음을 만족한다. {한 노드 N1의 전위 순회 값 < N2의 전위 순회 값 및 N1의 후위 순회 값 > N2의 후위 순회 값}
- [0119] 본 발명의 일 실시예는 dMERIT+과 같이 조상-후손 관계를 기반으로 각 가중화 후보 패턴들에 대해 패턴 확장을 수행한다. dMERIT+은 실제 패턴을 확장시킬 때 두 아이템이 서로 같은 트랜잭션에 들어있는지 판단하여 이윤 값이 중복으로 더해지는 것을 방지함으로써 패턴 확장 프로세스를 진행한다.
- [0120] 또한, 기존의 패턴 확장 방법에 노드 순서 리스트를 활용함으로써 dMERIT+에서 사용하던 방법보다 공간적인 효율성을 가진다. 패턴 확장은 다음과 같은 순서로 수행된다. 먼저, 지지도(또는 빈도수) 내림차순으로 정렬된 아이템들 중에서 첫 번째 아이템을 선택한다. 선택된 아이템과 나머지 다른 아이템과 조합하면서 패턴을 생성한다. 길이가 N+1인 패턴을 조합하기 위해 가중화 이레이저블 패턴 알고리즘은 길이가 N인 패턴들 중에 N-1개의 같은 기준 아이템들을 갖는 두 아이템을 이용한다. 조합되는 패턴들의 아이템들 순서는 지지도(또는 빈도수) 내림차순 순서를 기준으로 정렬한다. 또한, 가중화 이레이저블 패턴은 패턴의 게인 값을 효율적으로 계산하기 위해 조합되는 패턴들의 전위 순회 값들의 비교를 통해 아이템들 사이의 조상-후손 관계를 파악한다. 만약 가중화 이레이저블 패턴 트리에서 두 아이템이 포함된 노드들 중에 조상-후손 관계를 가지는 노드들이 있다면, 조합결과 패턴은 조상-후손의 관계를 가지는 전위 순회 값 중에서 패턴에서 마지막에 위치하는 아이템의 전위

순회 값에 대한 정보만을 포함한다. 첫 번째로 선택한 아이템에 대한 패턴 확장이 모두 끝나면, 두 번째 아이템에 대한 패턴 확장을 수행한다. 단, 패턴 확장을 할 때는 기존에 진행한 아이템은 고려하지 않는다.

- [0122] 도 6은 분할정복 방식에 기반한 재귀적 방법을 통한 가중화 이레이저블 패턴 마이닝 과정을 설명하는 순서도이다.
- [0123] 도 5에서 보이는 것과 같이, 데이터와 임계 값에 대해 가중화 이레이저블 패턴 트리와 노드 순서 리스트가 구축되면, 구축된 가중화 이레이저블 패턴 트리 내 헤더테이블(510)의 첫 번째 아이템을 기준으로 설정하고(단계 S610), 설정한 아이템과 함께 조합될 수 있는 아이템이 존재하는지 파악한다(단계 S620). 이때, 아이템이 조합되는 순서는 기존의 가중화 이레이저블 패턴 트리를 구축하는 과정에서 계산했던 지지도(또는 빈도수) 내림차순 정보를 따른다. 이 순서에 따라 아이템 조합과 패턴 확장 작업을 수행하면 전체적으로 가중화 이레이저블 패턴을 추출하기 위해 요구되는 패턴 조합 연산량을 줄일 수 있다.
- [0124] 조합할 아이템이 존재하는 경우에는 기 구축한 노드 순서 리스트를 기반으로 노드 간의 조상-후손 관계를 파악하여 불필요한 이윤 값의 합산을 방지하고, 효율적으로 아이템 조합 작업을 진행하여 가중화 이레이저블 후보 아이템들(또는 길이가 1인 가중화 이레이저블 후보 패턴)로부터 길이가 2인 가중화 이레이저블 후보 패턴을 도출할 수 있다(단계 S630).
- [0125] 여기에서, 본 발명은 아이템 가중치 적용에 따른 패턴 손실을 막기 위해 과소추정 제약조건에 기반한 패턴 프루닝 기법을 적용하고 있기 때문에, 상기 상술한 바와 같이 가중화 이레이저블 패턴에 대한 후보 패턴을 먼저 마이닝하게 되고, 마이닝 된 각 후보 패턴에 대한 실제 가중화 계인 값을 계산하여 주어진 최대 계인 임계 값과 비교함으로써 실제 유효한 가중화 이레이저블 패턴만을 선별하는 작업을 추가적으로 수행한다.
- [0126] 현재 선택된 기준 아이템에 대해 길이가 2인 모든 가능한 가중화 이레이저블 후보 패턴을 추출하게 되면, 추출한 패턴 정보를 기초로 길이가 3인 패턴들을 재귀적인 방식을 통해 다시 마이닝 할 수 있다(단계 S650). 일 실시예에서, 추출한 패턴 정보를 기초로 길이가 4인 패턴들을 재귀적인 방식을 통해 다시 마이닝 할 수 있다.
- [0127] 상기 작업을 더 이상 추출할 수 있는 K길이의 가중화 이레이저블 패턴이 존재하지 않을 때까지 반복하여 모든 작업이 마무리되면, 현재 선택된 기준 아이템의 다음 아이템을 새로운 기준 아이템으로 설정하고, 상기 재귀적 마이닝 작업을 반복해서 수행할 수 있다(단계 S660).
- [0128] 모든 아이템에 대해 더 이상 조합할 아이템이 없을 때까지 작업을 반복하면, 완전한 가중화 이레이저블 패턴 결과를 획득할 수 있다(단계 S640).
- [0130] 도 7은 최소 지지도 임계 값이 30%일 때 마이닝 되는 가중화 이레이저블 패턴의 패턴 조합 과정을 설명하는 도면이다.
- [0131] 도 7을 참조하면, 아이템의 조합 순서는 D, H, J, A, F 순으로 동작한다. 먼저 아이템 D를 기준 아이템으로 설정한 후에 해당하는 조합 작업을 재귀적으로 진행하고(710), 그 후에 아이템 H에 대한 작업(720), 아이템 J에 대한 작업(730), 아이템 A에 대한 작업(740), 아이템 F에 대한 작업(750) 순으로 패턴 조합 작업을 수행한다. 조합 작업이 모두 완료되면 현재 데이터베이스에 대해 임계 값을 만족하는 모든 가능한 가중화 이레이저블 패턴들을 마이닝할 수 있고, 사용자에게 해당 마이닝 결과를 제공할 수 있으며, 사용자는 제공된 마이닝 결과 정보 또는 데이터 그룹의 특성 정보를 기초로 다양한 데이터 분석 및 의사결정 관련 업무를 수행할 수 있다.
- [0132] 예를 들어, 도 5(a)의 가중화 이레이저블 패턴 트리에서 각 아이템별로 노드의 전위 순회 값 정보들로 그룹핑한다. 각 아이템의 정보들을 그룹핑한 결과는 다음과 같다. 아이템 D의 전위 순회 값 집합은 {1}, H의 전위 순회 값 집합은 {2, 4}, J의 전위 순회 값 집합은 {3, 5, 8}, A의 전위 순회 값 집합은 {6, 9}, F의 전위 순회 값 집합은 {7}이다. 길이가 1인 후보 패턴들의 전위 순회 값 집합의 마이닝이 끝난 후, 지지도(또는 빈도수) 내림차순으로 가중화 이레이저블 후보 아이템들을 정렬한다. 정렬된 아이템(D-H-J-A-F)들 중에서, 첫 후보 아이템인 D를 선택하고 이 패턴과 다른 패턴들을 각각 조합한다. 패턴 DH는 아이템들 D와 H의 전위 순회 값 집합들의 비교를 통해 다음과 같이 계산된다. 아이템 D의 전위 순회 값 집합과 아이템 H의 전위 순회 값 2는 서로 조상-후손의 관계이다 (전위 순회 값: $1 > 2$, 후위 순회 값: $3 < 1$). 그러므로, DH의 전위 순회 값 집합은 아이템 H의 전위 순회 값 집합 중에서 값 2가 제외된 {4}이고, 아이템 DH의 계인 값은 아이템 D의 계인 값(550)과 전위 순회 값 4가 가지는 계인 값(750)의 합인 1300 (= 550 + 750)이 된다. DH의 최대 가중화 계인 값은 1560.6이고, 계산된 값은 최대 가중화 임계 값인 1575의 값보다 크기 때문에 가중화 이레이저블 후보 패턴으로 포함되지 않는다.
- [0133] 모든 가중화 이레이저블 후보 아이템들에 대한 패턴 확장 과정이 완료되면, 각각의 패턴들에 대해 실제 이윤

값, 가중화 계인 값을 구한다. 그 이유는 현재 구해진 각 패턴의 MWG값들은 가중화 안티 모노톤 법칙을 위배하지 않기 위해 계산된 방법으로 패턴들의 실제 이윤 값보다 과소 평가된 값들일 수 있기 때문이다. 결과적으로, 조합된 결과들보다 이하의 개수를 가지는 패턴들만이 가중화 이레이저블 패턴의 결과로 반환된다.

- [0135] 도 8 내지 도 13은 본 발명의 과소추정 제약조건 기반의 프루닝 기법을 사용하여 가중화 이레이저블 패턴을 마인닝하는 과정을 나타낸 도면이다.
- [0136] 도 8을 참조하면, 각 아이템의 가중치를 고려한 가중화 이레이저블 패턴(WEP) 마이닝의 과정을 설명한다. 이레이저블 패턴 마이닝 알고리즘들은 데이터베이스로부터 사용자가 처리하는 것을 원하는 패턴들을 더 빠르고 정확하게 검색할 수 있도록 발전해온 방법들이다. 이레이저블 패턴 마이닝을 수행하는 데이터베이스들은 각각의 트랜잭션이 서로 다른 이윤 값을 가질 수 있다.
- [0137] 본 발명의 일 실시예는 트랜잭션들의 이윤 값들뿐만 아니라 트랜잭션을 구성하는 각 아이템들의 가중치까지 고려하여 마이닝을 수행할 수 있다. 즉, 아이템의 가중치는 높는데 반해 트랜잭션의 이윤 값이 낮은 아이템 집합들을 가중화 이레이저블 패턴들로 추출한다. 결과적으로, 본 발명은 아이템 고유의 가중치를 고려하여 패턴들을 선별적으로 마이닝할 수 있다.
- [0138] 데이터베이스는 가중치 정보를 반영하기 위해 각 아이템들에 대한 가중치를 입력으로 받는다. 입력된 정보들을 기초로 가중화 이레이저블 후보 아이템들을 추출한다. 이때, 기존의 이레이저블 패턴 마이닝에서는 고려하지 않던 각 아이템의 가중치 조건을 고려함으로써 실제 제약사항들을 효과적으로 처리할 수 있다. 이후 데이터베이스를 다시 스캔하며 가중화 이레이저블 패턴 트리(WEP-tree)를 구축하고, 트리의 내용을 기초로 노드 순서 리스트(또는 NC-list)를 추출한다. 노드 순서 리스트와 그 후보 아이템들을 이용하여 패턴을 확장하면서 가중화 이레이저블 패턴 집합을 마이닝한다.
- [0139] 본 발명의 일 실시예는 가중화 이레이저블 패턴들을 마이닝하기 위해 다음의 방법들을 순차적으로 진행한다. 첫 번째 단계는 데이터베이스를 스캔하며 길이가 1인 가중화 이레이저블 패턴들과 아이템들을 마이닝한다. 두 번째 단계는 첫 번째 단계에서 구한 패턴들과 데이터베이스를 이용하여 가중화 이레이저블 패턴 트리를 지지도(또는 빈도수) 내림차순으로 구축한다. 세 번째 단계는 구축된 가중화 이레이저블 패턴 트리와 노드 순서 리스트를 이용하여 첫 번째 단계에서 구한 길이가 1인 패턴을 확장시킨다. 여기에서, 패턴의 확장 프로세스에서 가중치 조건에 의한 프루닝 전략들을 통해 효과적으로 가중화 이레이저블 패턴들을 추출할 수 있다.
- [0141] 도 9를 참조하면, 도 4(a)의 트랜잭션 1-6들이 삽입된 가중화 이레이저블 패턴 트리를 설명한다.
- [0142] 도 5(a)의 헤더 테이블을 참고하면, 길이가 1인 가중화 이레이저블 후보 패턴에 해당하는 아이템은 D, H, J, A 및 F이다.
- [0143] 도 9(a)는 도 4(a) 데이터베이스의 트랜잭션 1에서 후보 아이템에 속하는 아이템들은 A와 J이고, 트랜잭션 1의 이윤 값은 50이다. 가중화 이레이저블 패턴 트리의 구축이 완료되지 않았기 때문에 전위 순회 값은 NULL로 설정한다.
- [0144] 도 9(b)는 도 4(a) 데이터베이스의 트랜잭션 2에서 후보 아이템은 H 하나이고, 이윤 값은 150이다. 도 9(c)는 도 4(a) 데이터베이스의 트랜잭션 4의 내용을 삽입한 결과이다. 이때, 트랜잭션 3과 트랜잭션 5의 경우에는 후보 아이템들을 포함하고 있지 않기 때문에 아이템들을 삽입하지 않는다.
- [0145] 도 9(d)는 도 4(a) 데이터베이스의 트랜잭션 6에서 후보 아이템 A, F, J, H를 가지고 있고, 이윤 값은 250을 가진다. 이후의 상품들은 위와 같은 방법으로 진행하고, 그 결과로 만들어지는 가중화 이레이저블 패턴 트리가 도 5의 프리픽스 트리(520)와 같은 노드의 구성을 가진다.
- [0146] 트리의 생성이 끝나면, 트리를 전위 순회 순서로 탐색하며 각 노드에 방문순서를 라벨링한다. 전위 순회 순서로 노드들이 탐색되는 동안 효율적인 패턴확장을 수행하기 위해 전위 순회 값 집합과 노드 순서 리스트가 생성된다. 여기에서, 각 아이템의 전위 순회 값 집합이란 가중화 이레이저블 패턴 트리에서 같은 아이템 이름을 가지는 노드들의 전위 순회 순서 정보를 그룹화 한 것이다. 예를 들어, 도 5의 가중화 이레이저블 패턴 트리에서 아이템 A의 전위 순회 값 집합은 {6, 9}이고 아이템 H의 전위 순회 값 집합은 {2, 4}이다. 각각의 패턴들은 하나의 전위 순회 값 집합을 가질 수 있다. 패턴 확장이 수행되기 전에 각각의 가중화 이레이저블 후보 아이템에 대해 전위 순회 값 집합들을 계산한다
- [0148] 도 10을 참조하면, 가중화 이레이저블 패턴(WEP) 알고리즘의 전체적인 진행 과정을 설명한다. 가중화 이레이저블 패턴 알고리즘은 데이터베이스와 데이터베이스 내의 아이템들 각각의 가중치와 가중화 이레이저블 패턴들을

마이닝하기 위한 임계 값을 입력 받는다. 특히, 가중치 조건의 처리를 위해 아이тем들의 가중치들을 입력 받는 점이 기존의 알고리즘들과 다른 부분이다. 가중화 이레이저블 패턴 알고리즘은 가중치 조건들을 고려하며 데이터베이스를 마이닝하고, 임계 값 이하의 가중화 이레이저블 패턴들의 집합을 결과로 반환한다. 이 과정을 수행하기 위해 도 10에서는 다양한 변수들이 이용되는 것을 알 수 있으며, 알고리즘의 내용은 다음과 같다. 가중화 이레이저블 패턴은 데이터베이스를 스캔하며 길이가 1인 가중화 이레이저블 후보 아이тем들을 추출해 낸다(line 3). 그 후보 아이тем들 중 실제 가중화 게인 값이 임계 값 이하인 아이тем들은 결과 패턴으로 포함한다(lines 4-8). 가중화 이레이저블 패턴은 추출된 아이тем 정보를 기초로 데이터베이스를 다시 스캔하며 패턴 확장을 효율적으로 진행하기 위해 가중화 이레이저블 패턴 트리를 구축한다(line 9). 해당 가중화 이레이저블 패턴 트리 안의 각 노드가 가지는 순서와 이윤 값(profit) 정보는 노드 순서 리스트와 가중화 이레이저블 후보 아이тем들의 전위 순회 순서 정보로 매핑된다(lines 10-11). 구축된 노드 순서 리스트와 그 후보들의 전위 순회 순서 정보를 이용하여 가중치 조건들을 고려하면서 패턴 확장 과정이 수행된다(line 12).

[0150] 도 11을 참조하면, 길이가 1인 가중화 이레이저블 후보 패턴(또는 후보 아이тем)들을 추출하는 과정을 순차적으로 설명한다. 가중화 이레이저블 후보 패턴을 추출하기 위한 모든 과정은 첫 번째 데이터베이스를 스캔 하는 동안 수행된다(lines 3-24). 데이터베이스의 각 트랜잭션들이 가지는 이윤 값의 합을 구하고(line 4), 그 총합은 사용자에게 입력 받은 임계 값 비율을 기반으로 최대 게인 임계 값을 구할 때 사용된다(line 17). 이전의 알고리즘들과는 다르게, 가중화 이레이저블 패턴은 데이터베이스에서 발견되는 아이тем들의 지지도(또는 빈도수)뿐만 아니라 각 아이тем의 트랜잭션에 대한 생산기여도를 계산하고 이 값을 저장한다. 트랜잭션의 생산정보를 읽는 동안 트랜잭션에 포함된 각 아이тем들의 정보들 아이тем 이름, 지지도(또는 빈도수), 이윤 값 등을 임시로 저장한다(lines 5-15). 아이тем들의 생산기여도들은 추가적인 테이블을 이용하여 각 아이тем을 처리할 때마다 평균 생산기여도로서 계산되고 갱신된다(lines 9 and 13). 만약, 데이터베이스 스캐닝 과정에서 아이тем이 처음 등장했다면 해당 아이тем에 대한 정보를 저장한다(lines 11-14). 데이터베이스 내 모든 트랜잭션들에 대한 정보 업데이트 과정이 끝난 후, 가중화 이레이저블 패턴은 계산된 아이тем들의 평균 생산기여도들 중에서 최대 값을 찾아 변수에 저장한다(line 17). 또한, 가중화 이레이저블 패턴의 마이닝을 위해 최대 게인 임계 값이 계산된다(line 18). 계산된 임계 값을 기반으로 데이터베이스 내 각 저장된 아이тем들 중에서 최대 평균 생산기여도가 임계 값 이하인 아이тем을 추출하고(lines 19-23), 그 아이тем들을 지지도(또는 빈도수) 내림차순으로 정렬한다(line 24). 여기서, 최대 평균 생산기여도가 임계 값 이하인 아이тем은 가중화 이레이저블 후보 아이тем(또는 후보 패턴)으로 선출할 수 있다.

[0152] 도 12를 참조하면, 패턴 추출의 효율성을 극대화할 수 있는 가중화 이레이저블 패턴 트리(WEP-tree)를 생성하는 알고리즘을 설명한다. 가중화 이레이저블 패턴(WEP)은 가중화 이레이저블 후보 아이тем들을 마이닝 한 뒤, 추출된 정보를 기초로 데이터베이스를 한번 더 스캔 하는 동안 가중화 이레이저블 패턴 트리(WEP-tree)를 구축한다(lines 1-13). 이 때, 데이터베이스 내의 각 트랜잭션들 중에 가중화 이레이저블 후보 아이тем들로 추출된 아이тем들만이 트리의 노드로 매핑된다. 각 트랜잭션의 아이тем들은 완전한 트리 구축을 위해 지지도(또는 빈도수) 내림차순으로 정렬되어 삽입된다(lines 3-9). 여기에서, 내림차순은 가중화 이레이저블 후보 아이тем들의 순서일 수도 있다. 데이터베이스 스캐닝 작업이 끝난 후, 트리의 각 노드를 전위 순회 및 후위 순회 순서 방식으로 탐색하며, 각 노드에 전위 순회 순서 정보만 저장한다(line 11). 노드들의 후위 순회 순서 정보는 그 노드 순서 리스트를 생성할 때 사용된다.

[0153] 구축된 가중화 이레이저블 패턴 트리는 패턴 확장을 위해 각 아이тем들의 구성 정보를 저장하게 된다. 가중화 이레이저블 후보 아이тем들의 마이닝과 가중화 이레이저블 패턴 트리 및 노드 순서 리스트의 구축이 끝난 뒤, 구축된 가중화 이레이저블 패턴 트리 및 노드 순서 리스트 자료구조들에 포함된 정보를 이용하여 도 13의 패턴 확장 과정을 수행한다. 패턴 확장 과정에 대한 상세한 설명은 도 13에서 설명한다.

[0155] 도 13을 참조하면, 가중화 이레이저블 패턴 확장 과정을 설명하는 도면이다. 패턴 확장은 이미 추출된 가중화 이레이저블 후보 패턴들의 탐색을 통해 진행된다(lines 2-19). 지지도(또는 빈도수) 내림차순으로 정렬된 그 패턴들은 순서에 따라 기준 패턴을 선택하고, 선택된 패턴의 그 다음의 패턴들과의 조합이 수행된다. 패턴 확장은 선택된 두 N-1 후보 패턴의 아이тем들과 중복되는 아이тем을 제외하고, 길이 N을 가지는 패턴으로 새로운 후보 패턴을 생성하는 것으로 시작된다(line 4). 이후, 두 N-1 후보 패턴의 전위 순회 순서 정보의 값들을 서로 비교하며 새로운 후보 패턴의 게인 값을 계산한다(lines 5-13).

[0156] 보다 구체적으로, 각 전위 순회 순서 정보에 해당하는 노드의 조상-후손 관계를 이용하여 중복되는 트랜잭션의 이윤 값을 제외하는 방법을 사용한다(lines 7-10). 이 계산 과정의 결과로 얻어진 새로운 후보 패턴은 다음 패턴 확장 과정에서 사용하기 위해 리스트에 저장하게 된다(line 14). 또한, 본 발명의 일 실시예는 가중화 이레이

이저블 패턴들의 마이닝을 위해 새로운 후보 패턴과 최대 게인 임계 값 사이의 비교를 진행한다(lines 15-17).

- [0157] 만약, 그 패턴의 가중화 게인 값이 임계 값 이하의 값을 가진다면 결과 패턴에 포함시키고, 그렇지 않다면 단순히 후보 패턴으로만 사용하게 된다. 본 발명은 패턴들의 가중화 게인 값을 계산함으로써, 최대 생산기여도로 인해 발생하는 중복된 이윤 값의 합산 오류를 방지하고 정확한 결과 패턴들을 얻을 수 있다. 모든 후보 패턴들에 대한 조합이 완료되면, 새롭게 획득한 후보 패턴들의 수에 따라 재귀적으로 패턴 확장을 수행한다(lines 20-22). 이 패턴 확장 과정은 모든 기준 아이템들에 대해 처리될 때까지 계속 수행될 수 있다.
- [0159] 상기에서는 본 출원의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 통상의 기술자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 출원을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.
- [0161] [참조문헌]
- [0162] 1. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487-499, Sep. 1994.
- [0163] 2. C.F. Ahmed, S.K. Tanbeer, and B. Jeong, "Efficient Mining of Weighted Frequent Patterns over Data Streams", Proceedings of the 11th IEEE International Conference on High Performance Computing and Communications, pp. 400-406, Jun. 2009.
- [0164] 3. C.F. Ahmed, S.K. Tanbeer, B. Jeong, and Y. Lee, "Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases", IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 12, pp. 1708-1721, Dec. 2009.
- [0165] 4. C.F. Ahmed, S.K. Tanbeer, B. Jeong, Y. Lee, and H. Choi, "Single-pass incremental and interactive mining for weighted frequent patterns", Expert Systems with Applications, vol. 39, no. 9, pp. 7976-7994, Jan. 2012.
- [0166] 5. J. Cai, X. Zhao, and Y. Xun, "Association rule mining method based on weighted frequent pattern tree in mobile computing environment", International Journal of Wireless and Mobile Computing, vol. 6, no. 2, pp. 193-199, 2013.
- [0167] 6. Z. Deng, "Mining Top-Rank-k Erasable Itemsets by PID_lists", International Journal of Intelligent Systems, vol. 28, no. 4, pp. 366-379, Apr. 2013.
- [0168] 7. Z. Deng, G. Fang, and Z. Wang, "MINING ERASABLE ITEMSETS", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, vol. 1, pp. 67-73, Jul. 2009.
- [0169] 8. Z. Deng and X. Xu, "An Efficient Algorithm for Mining Erasable Itemsets", Advanced Data Mining and Applications, pp. 214-225, Nov. 2010.
- [0170] 9. Z. Deng and X. Xu, "Fast mining erasable itemsets using NC_sets", Expert Systems with Applications, vol. 39, no. 4, pp. 4453-4463, Mar. 2012.
- [0171] 10. M. Gan and H. Dai, "Detecting and monitoring abrupt emergences and submergences of episodes over data streams", Information Systems, vol. 39, pp. 277-289, Jan. 2014.
- [0172] 11. G. Grahne and J. Zhu, "Fast Algorithms for Frequent Itemset Mining Using FP-Trees", IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 10, pp. 1347-1362, Oct. 2005.
- [0173] 12. J. Han, J. Pei, Y. Yin, and R. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", Data Mining and Knowledge Discovery, vol. 8, no. 1, pp. 53-87, Jan. 2004.
- [0174] 13. T. Hashem, C.F. Ahmed, M. Samiullah, S. Akther, B. Jeong, and S. Jeon, "An efficient approach for mining cross-level closed itemsets and minimal association rules using closed itemset lattices", Expert Systems with Applications, vol. 41, no. 6, pp. 2914-2938, Jan. 2014.
- [0175] 14. Y. Kim, W. Kim, and U. Kim, "Mining Frequent Itemsets with Normalized Weight in Continuous Data

Streams” , The Journal of Information Processing Systems, vol. 6, no. 1, pp. 79-90, Mar. 2010.

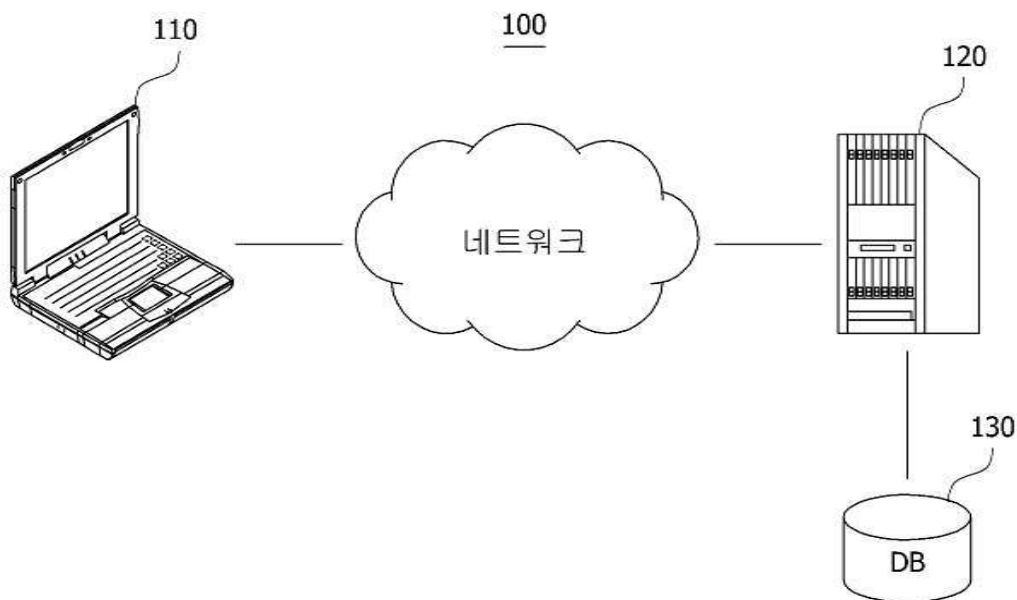
- [0176] 15. T. Le, B. Vo, and F. Coenen, “An Efficient Algorithm for Mining Erasable Itemsets Using the Difference of NC-Sets” , IEEE International Conference on Systems, Man, and Cybernetics, pp. 2270-2274, Oct. 2013.
- [0177] 16. G. Lee, U. Yun, and K. Ryu, “Sliding Window based Weighted Maximal Frequent Pattern Mining over Data Streams” , Expert Systems with Applications, vol. 41, no. 2, pp. 694-708, Feb. 2014.
- [0178] 17. K.W. Lin, M. Hsieh, and V.S. Tseng, “A novel prediction-based strategy for object tracking in sensor networks by mining seamless temporal movement patterns” , Expert Systems with Applications, vol. 37, no. 4, pp. 2799-2807, Apr. 2010.
- [0179] 18. L. Liu, J. Tang, Y. Cheng, A. Agrawal, W. Liao, and A.N. Choudhary, “Mining diabetes complication and treatment patterns for clinical decision support” , 22nd ACM International Conference on Information and Knowledge Management, pp. 279-288, Nov. 2013.
- [0180] 19. G. Nguyen, T. Le, B. Vo, and B. Le, “A New Approach for Mining Top-Rank-k Erasable Itemsets” , Intelligent Information and Database Systems, vol. 8397, pp. 73-82, Apr. 2014.
- [0181] 20. R. Patel, “A Parallel and Distributed Method to mine Erasable Itemsets from High utility patterns” , International Journal of Advanced Research in Computer Science and Electronics Engineering, vol. 1, no. 8, pp. 99-104, Oct. 2012.
- [0182] 21. G. Pyun and U. Yun, “Mining top-k frequent patterns with combination reducing techniques” , Applied Intelligence, DOI. 10.1007/s10489-013-0506-9, Jan. 2014.
- [0183] 22. G. Pyun, U. Yun, and K. Ryu, “Efficient frequent pattern mining based on Linear Prefix Tree” , Knowledge-Based Systems, vol. 55, pp. 125-139, Jan. 2014.
- [0184] 23. F. Rasheed and R. Alhajj, “Periodic pattern analysis of non-uniformly sampled stock market data” , Intelligent Data Analysis, vol. 16, no. 6, pp. 993-1011, Nov. 2012.
- [0185] 24. H. Ryang, U. Yun, and K. Ryu, “Discovering High Utility Itemsets with Multiple Minimum Supports” , Intelligent Data Analysis, In Press, 2014.
- [0186] 25. M. Samiullah, C.F. Ahmed, A. Fariha, M.R. Islam, and N. Lachiche, “Mining frequent correlated graphs with a new measure” , Expert Systems with Applications, vol. 41, no. 4, pp. 1847-1863, Mar. 2014.
- [0187] 26. P. Sarkar and R. Patel, “Taking out High Utility Erasable Itemsets” , International Journal of Computer Application, vol. 3, no. 2, pp. 150-156, Jun. 2012.
- [0188] 27. N. Tatti, “Discovering episodes with compact minimal windows” , Data Mining and Knowledge Discovery, vol. 28, no. 4, pp. 1046-1077, Jul. 2014.
- [0189] 28. S.G. Totad, R.B. Geeta, and P.P. Reddy, “Batch incremental processing for FP-tree construction using FP-Growth algorithm” , Knowledge and Information Systems, vol. 33, no. 2, pp. 475-490, Nov. 2012.
- [0190] 29. V.S. Tseng, B. Shie, C. Wu, and Philip S. Yu, “Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases” , IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 8, pp. 1772-1786, Aug. 2013.
- [0191] 30. K. Yang, T. Hong, G. Lan, and Y. Chen, “A two-phase approach for mining weighted partial periodic patterns” , Engineering Applications of Artificial Intelligence, vol. 30, pp. 225-234, Apr. 2014.
- [0192] 31. U. Yun, “On pushing weight constraints deeply into frequent itemset mining” , Intelligent Data Analysis, vol. 13, no. 2, pp. 359-383, 2009.
- [0193] 32. U. Yun, G. Lee, and K. Ryu, “Mining Maximal Frequent Patterns by Considering Weight Conditions

over Data Streams” , Knowledge-Based Systems, vol. 55, pp. 49-65, Jan. 2014.

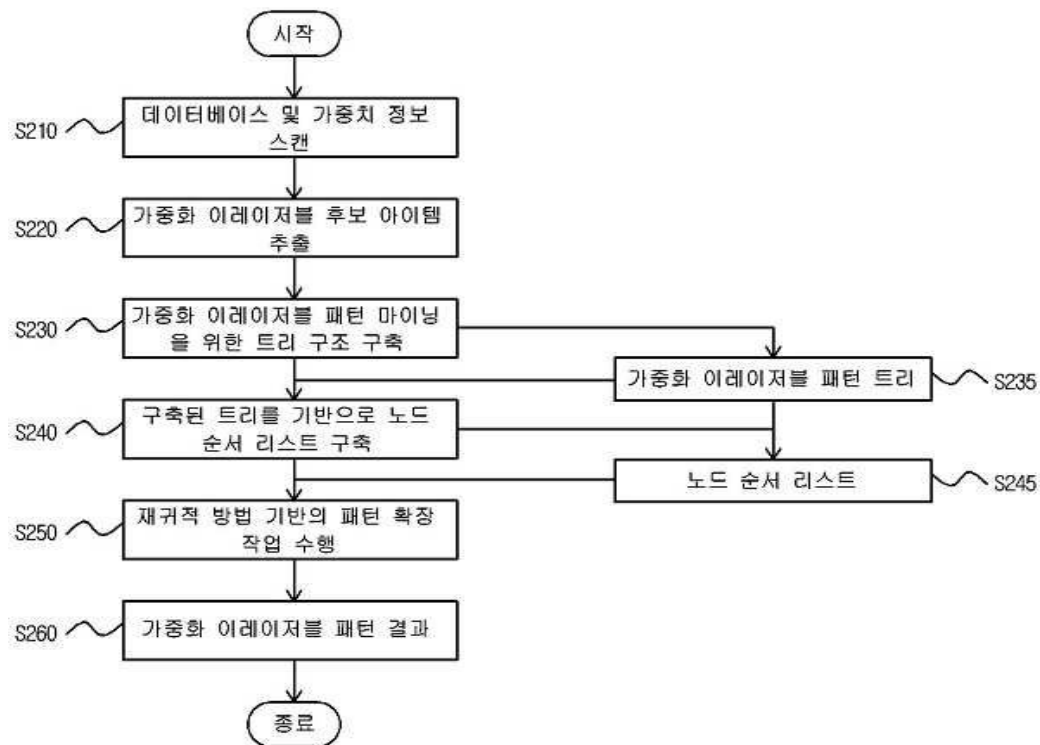
- [0194] 33. U. Yun, H. Ryang, and K. Ryu, “High Utility Itemset Mining with Techniques for Reducing Overestimated Utilities and Pruning Candidates” , Expert Systems with Applications, vol. 41, no. 8, pp. 3861-3878, Jun. 2014.
- [0195] 34. U. Yun and K. Ryu, “Efficient Mining of Maximal Correlated Weight Frequent Patterns” , Intelligent Data Analysis, vol. 17, no. 5, pp. 917-939, Sep. 2013.
- [0196] 35. L. Wang, L. Feng, and M. Wu, “UDS-FIM: An Efficient Algorithm of Frequent Itemsets Mining over Uncertain Transaction Data Streams” , Journal of Software, vol. 9, no. 1, pp. 44-56, Jan. 2014.
- [0197] 36. J. Wang and Y. Zeng, “DSWFP: Efficient Mining of Weighted Frequent Pattern over Data Streams” , Fuzzy Systems and Knowledge Discovery, vol. 2, pp. 942-946, Jul. 2011.

도면

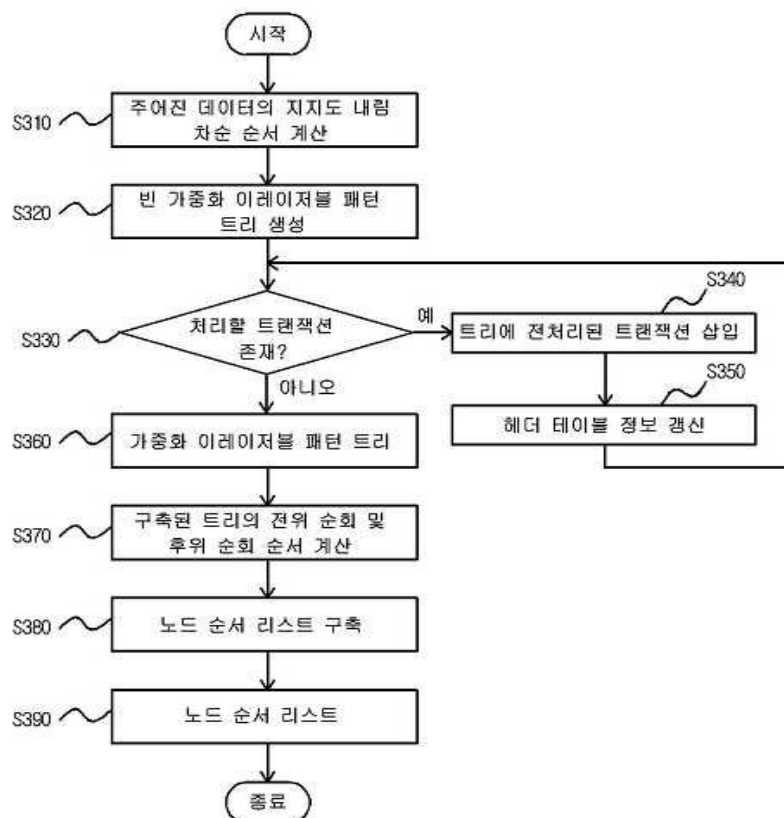
도면1



도면2



도면3



도면4

(a)

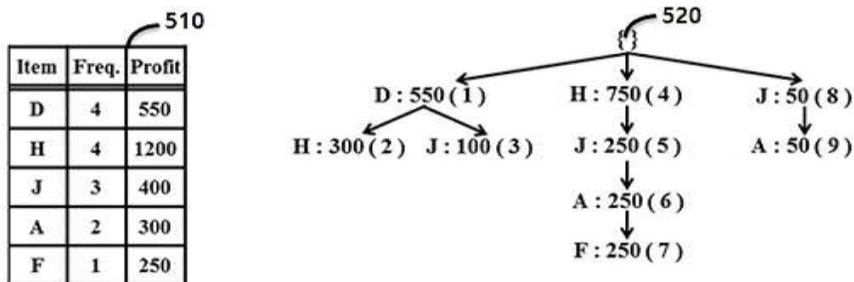
PID	Items	Profit (Million \$)
P_1	A, B, J	50
P_2	E, H	150
P_3	G, I	2500
P_4	C, D, H	300
P_5	B, C, E, I	1250
P_6	A, E, F, H, I, J	250
P_7	C, D, I	100
P_8	B, D	50
P_9	B, C, H, I	500
P_{10}	D, E, J	100

(b)

Item	A	B	C	D	E	F	G	H	I	J
Weight	0.8	0.4	0.4	0.4	0.4	1.0	1.0	0.4	0.2	0.6

도면5

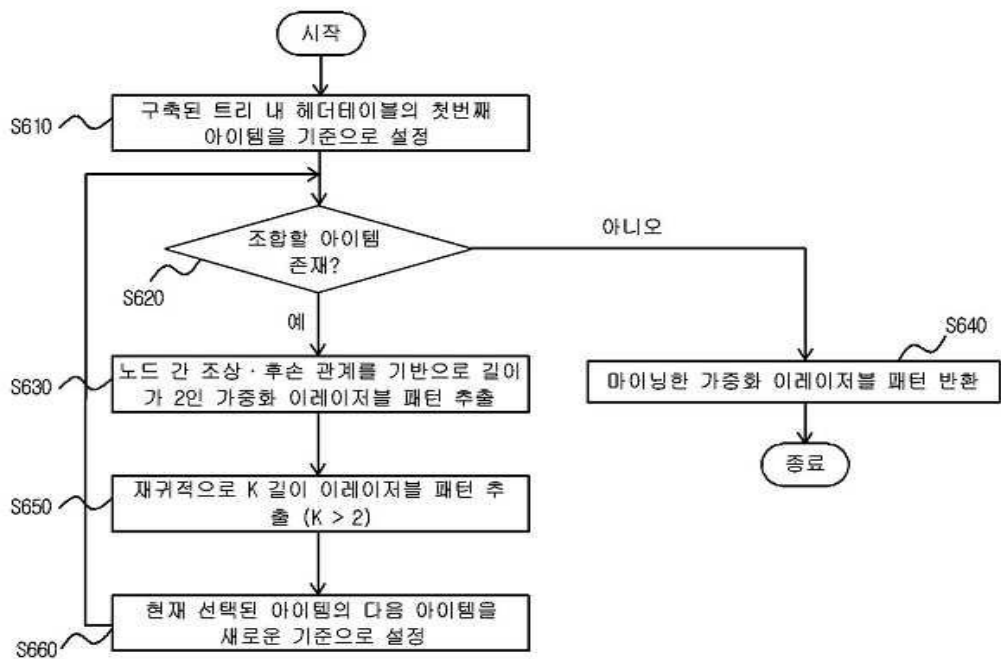
(a)



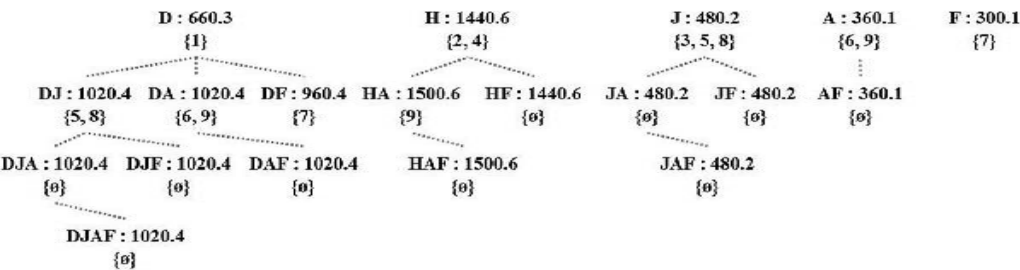
(b)

Pre-order	1	2	3	4	5	6	7	8	9
Post-order	3	1	2	7	6	5	4	9	8
Profit	550	300	100	750	250	250	250	50	50

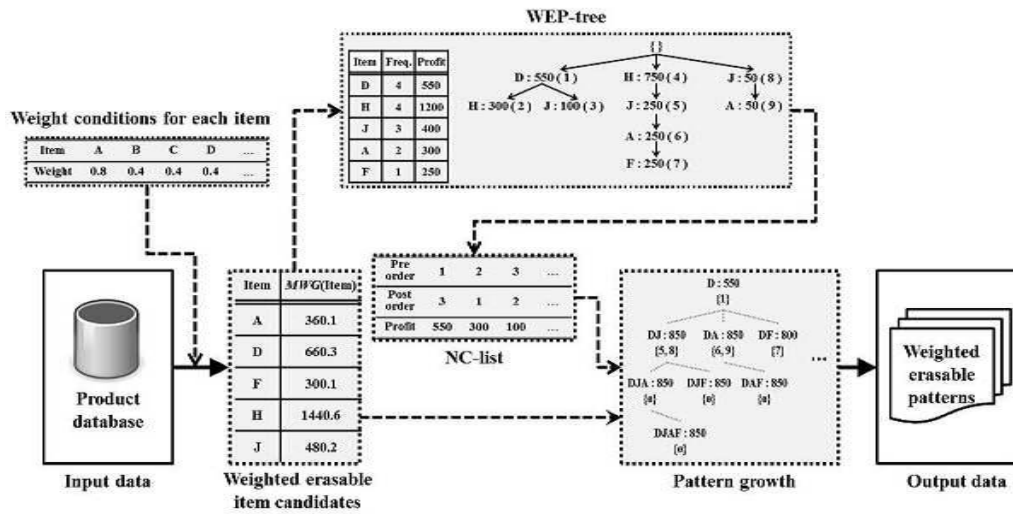
도면6



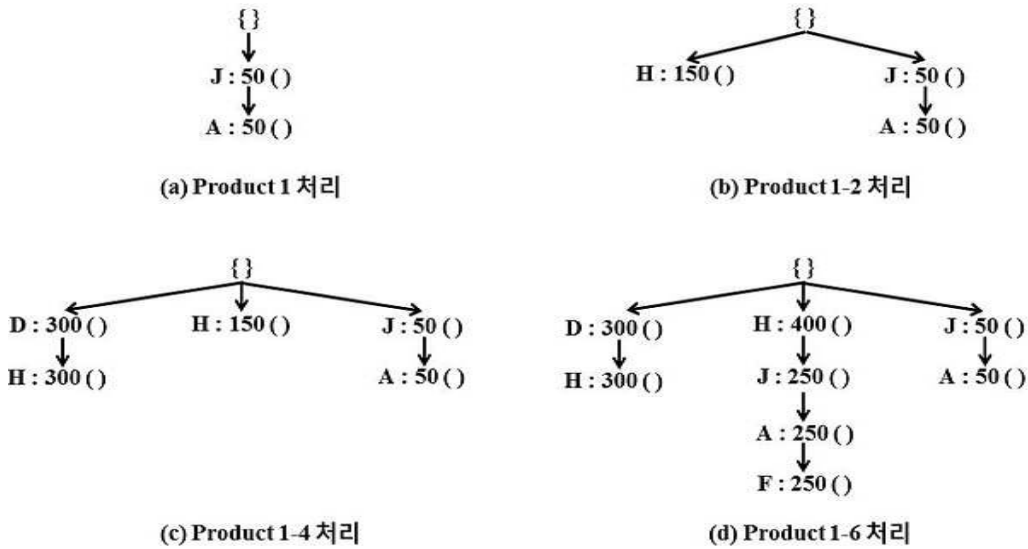
도면7



도면8



도면9



도면10

Input: A product database, D , A weights of items WI , A maximum erasable threshold, δ
Output: A set of weighted erasable patterns, EP
Variables: An WEP-tree, T , A set of 1-pattern candidates, EC , A manufacturing contribution ratio table, $MCRT$, A maximum gain threshold, MGT

Procedure WEP(D, WI, δ)

```

1  begin
2  |  $T \leftarrow \emptyset, EC \leftarrow \emptyset, EP \leftarrow \emptyset, MCRT \leftarrow \emptyset$ 
3  | Call Generate 1-Candidates( $D, WI, EC, MCRT, MGT$ )
4  | foreach  $i_n$  in  $Eone$  do
5  | | if  $WG(i_n) \leq \delta$  then
6  | | |  $EP \leftarrow EP \cup i_n$ 
7  | | end
8  | end
9  | Call Create_WEP-tree( $T, WI, EC, MCRT, MGT$ )
10 | Generate an NC-list,  $NL$ , using information of  $T$ 
11 | Generate a pre order set of each item of  $Eone$ 
12 | Call WEP_Growth( $Eone, EP$ )
13 end
    
```

도면11

```

Gnerate_1-Candidates(D, WI, EC, MCRT)
1  begin
2    A list of items, IL ← ∅
3    foreach a product Pk in D do
4      SUM = SUM + Pk.profit
5      foreach an item in in Pk do
6        if in ∈ IL then
7          Update the frequency of IL[in]
8          IL[in].gain = IL[in].gain + Pk.profit
9          
$$MCRT[i_n].ratio = \frac{(MCRT[i_n].profit * ([i_n].freq - 1)) * \frac{W(i_n)}{PW(P_k)}}{[i_n].freq}$$

10         end
11       else
12         Insert name, profit, frequency(=1) of in into IL
13         MCRT[in] = W(in) / PW(Pk)
14       end
15     end
16   end
17   MCRT.MAXW = the maximum value of manufacturing contribution ratio among values of MCRT
18   Calculate a maximum gain threshold, MGT, using SUM and δ
19   foreach an item IL[ip] in IL do
20     if MWG(IL[ip]) ≤ MGT then
21       EC = EC ∪ IL[ip]
22     end
23   end
24   Sort EC in frequency descending order
25 end

```

도면12

```

Create_WEP-tree(T, WI, EC, MCRT, MGT)
1  begin
2    foreach a product Pk in D do
3      Sort items of Pk in item order of EC
4      foreach an item ip in Pk do
5        if ip ∈ EC then
6          Pk' = Pk' ∪ ip
7        end
8      end
9      Insert Pk' into T
10   end
11   Store pre and post order information into each node of T
12 end

```

도면13

```

WEP_Growth(EC, EP)
1  begin
2    foreach a candidate of the weighted erasable pattern, Ep, in EC do
3      foreach a next candidate of Ep, Er, in EC do
4        Generate a candidate pattern, Ecp, using items of both Ep and Er patterns without the item duplication
5        foreach a pre order, pi, in pre order set of Ep do
6          foreach a pre order, pj, in pre order set of Er do
7            if (pi ≥ pj) || (NL[pi].post ≤ NL[pj].post) then
8              Ecp.gain = Ecp.gain + Er.gain
9              Insert pi into the pre order set of Ecp
10             end
11           end
12         end
13         Ecp.gain = Ecp.gain + Ep.gain
14         ECgrowth = ECgrowth ∪ Ecp
15         if WG(Ecp) ≤ MGT then
16           EP ← EP ∪ Ecp
17         end
18       end
19     end
20     if ECgrowth includes two or more candidate patterns then
21       Call WEP_Growth(ECgrowth, EP)
22     end
23 end

```

【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 9

【변경전】

상기 전위 순위 값에

【변경후】

상기 전위 순회 값에