



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0000610
(43) 공개일자 2018년01월03일

(51) 국제특허분류(Int. Cl.)

G06F 17/30 (2006.01)

(52) CPC특허분류

G06F 17/30539 (2013.01)

G06F 17/30958 (2013.01)

(21) 출원번호 10-2016-0078849

(22) 출원일자 2016년06월23일

심사청구일자 2016년06월23일

(71) 출원인

세종대학교산학협력단

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

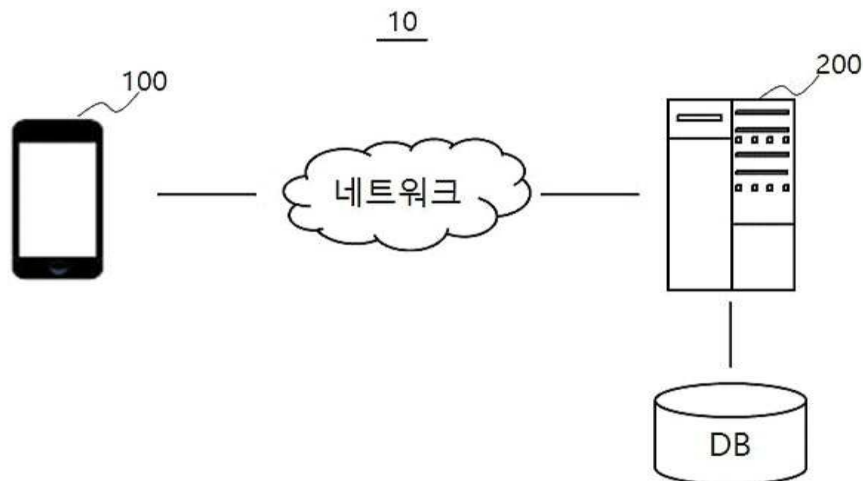
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법 및 이를 수행하는 회귀 그래프 패턴 마이닝 장치

(57) 요약

회귀 그래프 패턴을 마이닝하는 방법은 회귀 그래프 패턴 마이닝 장치에서 수행된다. 상기 방법은 (a) 그래프 데이터베이스 내의 그래프 데이터와 그래프 요소에 대한 중요도 정보를 스캔하여 유효하지 않은 그래프 요소를 제거하는 단계, (b) 상기 그래프 요소의 유효 정보를 기초로 회귀 그래프 패턴 마이닝을 위한 유효 그래프 요소 리스트를 구축하고, 상기 유효 그래프 요소 리스트를 기반으로 회귀 그래프 패턴을 확장하여 가중화 회귀 그래프 후보 패턴을 추출하는 단계 및 (c) 상기 가중화 회귀 그래프 후보 패턴을 기초로 유효한 가중화 회귀 그래프 패턴을 선별하여 가중화 회귀 그래프 패턴 마이닝 결과를 획득하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

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)

정우성

경기도 의정부시 천보로 569번길 18 (동아빌라 30
1호)

이주대

서울특별시 송파구 위례광장로 185 (장지동, 위례
신도시 송파푸르지오 107동 304호)

명세서

청구범위

청구항 1

회귀 그래프 패턴 마이닝 장치에서 수행되는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법에 있어서,

- (a) 그래프 데이터베이스 내의 그래프 데이터와 그래프 요소에 대한 중요도 정보를 스캔하여 유효하지 않은 그래프 요소를 제거하는 단계;
- (b) 상기 그래프 요소의 유효 정보를 기초로 회귀 그래프 패턴 마이닝을 위한 유효 그래프 요소 리스트를 구축하고, 상기 유효 그래프 요소 리스트를 기반으로 회귀 그래프 패턴을 확장하여 가중화 회귀 그래프 후보 패턴을 추출하는 단계; 및
- (c) 상기 가중화 회귀 그래프 후보 패턴을 기초로 유효한 가중화 회귀 그래프 패턴을 선별하여 가중화 회귀 그래프 패턴 마이닝 결과를 획득하는 단계를 포함하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 2

제1항에 있어서, 상기 (a) 단계는

상기 그래프 데이터베이스의 스캔을 통해 상기 그래프 요소에 해당하는 정점과 간선 정보의 지지도 정보를 누적하고, 그래프 트랜잭션 내 중복되는 상기 그래프 요소의 중복 카운트를 제거하는 단계를 포함하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 3

제2항에 있어서, 상기 (a) 단계는

상기 누적된 그래프 요소를 가중치 과추정 제약조건 및 다중 최소 지지도 제약조건을 기반으로 제거하는 단계를 더 포함하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 4

제1항에 있어서, 상기 (b) 단계는

상기 그래프 요소 중 적어도 하나의 유효한 그래프 요소를 선별하여 상기 유효 그래프 요소 리스트를 구축하는 단계를 포함하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 5

제1항에서 있어서, 상기 (c) 단계는

상기 가중화 회귀 그래프 후보 패턴에 대한 최대 가중치 정보를 이용하여 최대 가중화 지지도 값을 계산하고, 상기 가중화 회귀 그래프 후보 패턴의 극최소 지지도 임계값과 비교하는 단계를 포함하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 6

제5항에 있어서, 상기 (c) 단계는

상기 가중화 회귀 그래프 후보 패턴에 대한 실제 가중화 지지도 값을 계산하고, 상기 가중화 회귀 그래프 후보 패턴의 극최소 지지도 임계값과 비교하는 단계를 더 포함하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 7

제5항에 있어서, 상기 최대 가중치 정보는

상기 간선이 가질 수 있는 가중치 중 가장 큰 값에 해당하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 8

제5항에 있어서, 상기 극최소 지지도 임계값은

상기 그래프 요소에 대한 고유의 임계값 중 가장 작은 값에 해당하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 9

제1항에 있어서, 상기 가중화 회귀 그래프 패턴은

상기 가중화 회귀 그래프 후보 패턴의 중요도와 회귀도 특성을 고려한 패턴에 해당하는 것을 특징으로 하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법.

청구항 10

그래프 데이터베이스 내의 그래프 데이터와 그래프 요소에 대한 중요도 정보를 스캔하는 트랜잭션 스캔부;

상기 스캔된 내용으로부터 유효하지 않은 상기 그래프 요소를 제거하고, 상기 그래프 요소의 유효 정보를 기초로 회귀 그래프 패턴 마이닝을 위한 유효 그래프 요소 리스트를 구축하는 트랜잭션 처리부;

상기 유효 그래프 요소 리스트를 기반으로 회귀 그래프 패턴을 확장하여 가중화 회귀 그래프 후보 패턴을 추출하는 후보 패턴 추출부; 및

상기 가중화 회귀 그래프 후보 패턴을 기초로 유효한 가중화 회귀 그래프 패턴을 선별하여 가중화 회귀 그래프 패턴 마이닝 결과를 획득하는 유효 패턴 판단부를 포함하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법을 수행하는 회귀 그래프 패턴 마이닝 장치.

발명의 설명**기술 분야**

[0001]

본 발명은 회귀 그래프 패턴 마이닝 기술에 관한 것으로, 보다 상세하게는 복잡한 형태로 이루어진 그래프의 정점과 간선의 상호 연관관계 및 순서 등의 복잡한 특성 정보를 고려할 수 있는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법 및 이를 수행하는 회귀 그래프 패턴 마이닝 장치에 관한 것이다.

배경 기술

- [0003] 종래의 빈발 패턴 마이닝 기술은 단순한 아이템으로 나열된 데이터로부터 사용자에게 의해 설정된 최소 지지도 임계값 이상으로 빈번하게 발생하는 패턴 정보를 추출한다. 종래의 빈발 패턴 마이닝 기술은 오직 데이터의 빈발한 정도만을 마이닝 요소로 설정하여 이용하며 단순한 아이템의 나열로 표현되는 데이터베이스 처리에 한정된다. 즉, 종래의 빈발 패턴 마이닝 기술은 아이템의 지지도 정보 외에 아이템 고유의 중요도 정보나 패턴의 희귀도 특성 같은 요소들을 고려하지 못하는 문제점을 가지고 있다.
- [0004] 한국등록특허 제10-1594345호는 빈발 패턴 마이닝 방법에 관한 것으로, 주어진 데이터베이스를 스캔하여 글로벌 FP-Tree를 생성하고, 지역 FP-Tree를 재귀적으로 구축하면서 빈발 패턴을 생성하며, 해당 패턴을 top-k 리스트에 삽입함으로써 유효 패턴 정보를 관리하는 기술을 개시한다.
- [0005] 한국등록특허 제10-1432710호는 지지도 연관성을 갖는 빈발 부분 그래프의 마이닝 방법에 관한 것으로, 주어진 그래프 데이터베이스 스캔을 통해 유효한 그래프 요소 정보를 파악하고, 단계별로 그래프 패턴을 확장해나가면서 추출되는 그래프 패턴의 지지도 연관성을 계산하여 일정 수준 이상의 연관성을 갖는 패턴을 선별하고 추출하는 기술을 개시한다.
- [0006] M. Samiullah, C.F. Ahmed, A. Fariha, M.R. Islam, N. Lachiche, "Mining frequent correlated graphs with a new measure", Expert Systems with Applications, 2014는 추출되는 각 그래프 패턴을 구성하는 요소인 정점과 간선의 연관도 특성을 파악하여 주어진 임계값 이상의 연관도를 가지는 패턴을 유효 마이닝 결과로 고려하는 방법에 관한 것으로, 이러한 방법은 그래프 요소들 각각에 대한 고유의 최소 지지도 임계값을 고려하지 못하기 때문에 추출되는 패턴의 희귀도 특성을 고려할 수 없다.

선행기술문헌

특허문헌

- [0008] (특허문헌 0001) 한국등록특허 제10-1594345호 (2016.02.05 등록)
- (특허문헌 0002) 한국등록특허 제10-1432710호 (2014.08.14 등록)

비특허문헌

- [0009] (비특허문헌 0001) M. Samiullah, C.F. Ahmed, A. Fariha, M.R. Islam, N. Lachiche, "Mining frequent correlated graphs with a new measure", Expert Systems with Applications, 2014

발명의 내용

해결하려는 과제

- [0010] 본 발명의 일 실시예는 그래프의 정점과 간선의 상호 연관관계 및 순서의 복잡한 특성 정보를 고려하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 희귀 그래프 패턴을 마이닝하기 위한 방법 및 이를 수행하는 희귀 그래프 패턴 마이닝 장치를 제공하고자 한다.
- [0011] 본 발명의 일 실시예는 그래프 패턴별로 가지는 중요도 특성을 차등 적용하여 유효한 패턴을 선별하는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 희귀 그래프 패턴을 마이닝하기 위한 방법 및 이를 수행하는 희귀 그래프 패턴 마이닝 장치를 제공하고자 한다.

과제의 해결 수단

- [0013] 실시예들 중에서, 희귀 그래프 패턴을 마이닝하는 방법은 희귀 그래프 패턴 마이닝 장치에서 수행된다. 상기 방법은 (a) 그래프 데이터베이스 내의 그래프 데이터와 그래프 요소에 대한 중요도 정보를 스캔하여 유효하지 않

은 그래프 요소를 제거하는 단계, (b) 상기 그래프 요소의 유효 정보를 기초로 회귀 그래프 패턴 마이닝을 위한 유효 그래프 요소 리스트를 구축하고, 상기 유효 그래프 요소 리스트를 기반으로 회귀 그래프 패턴을 확장하여 가중화 회귀 그래프 후보 패턴을 추출하는 단계 및 (c) 상기 가중화 회귀 그래프 후보 패턴을 기초로 유효한 가중화 회귀 그래프 패턴을 선별하여 가중화 회귀 그래프 패턴 마이닝 결과를 획득하는 단계를 포함한다.

[0014] 상기 (a) 단계는 상기 그래프 데이터베이스의 스캔을 통해 상기 그래프 요소에 해당하는 정점과 간선 정보의 지지도 정보를 누적하고, 그래프 트랜잭션 내 중복되는 상기 그래프 요소의 중복 카운트를 제거하는 단계를 포함할 수 있다. 상기 (a) 단계는 상기 누적된 그래프 요소를 가중치 과추정 제약조건 및 다중 최소 지지도 제약조건을 기반으로 제거하는 단계를 더 포함할 수 있다.

[0015] 상기 (b) 단계는 상기 그래프 요소 중 적어도 하나의 유효한 그래프 요소를 선별하여 상기 유효 그래프 요소 리스트를 구축하는 단계를 포함할 수 있다.

[0016] 상기 (c) 단계는 상기 가중화 회귀 그래프 후보 패턴에 대한 최대 가중치 정보를 이용하여 최대 가중화 지지도 값을 계산하고, 상기 가중화 회귀 그래프 후보 패턴의 극최소 지지도 임계값과 비교하는 단계를 포함할 수 있다. 상기 (c) 단계는 상기 가중화 회귀 그래프 후보 패턴에 대한 실제 가중화 지지도 값을 계산하고, 상기 가중화 회귀 그래프 후보 패턴의 극최소 지지도 임계값과 비교하는 단계를 더 포함할 수 있다.

[0017] 상기 최대 가중치 정보는 상기 간선이 가질 수 있는 가중치 중 가장 큰 값에 해당할 수 있다.

[0018] 상기 극최소 지지도 임계값은 상기 그래프 요소에 대한 고유의 임계값 중 가장 작은 값에 해당할 수 있다.

[0019] 상기 가중화 회귀 그래프 패턴은 상기 가중화 회귀 그래프 후보 패턴의 중요도와 회귀도 특성을 고려한 패턴에 해당할 수 있다.

[0020] 실시예들 중에서, 회귀 그래프 패턴 마이닝 장치는 그래프 데이터베이스 내의 그래프 데이터와 그래프 요소에 대한 중요도 정보를 스캔하는 트랜잭션 스캔부, 상기 스캔된 내용으로부터 유효하지 않은 상기 그래프 요소를 제거하고, 상기 그래프 요소의 유효 정보를 기초로 회귀 그래프 패턴 마이닝을 위한 유효 그래프 요소 리스트를 구축하는 트랜잭션 처리부, 상기 유효 그래프 요소 리스트를 기반으로 회귀 그래프 패턴을 확장하여 가중화 회귀 그래프 후보 패턴을 추출하는 후보 패턴 추출부 및 상기 가중화 회귀 그래프 후보 패턴을 기초로 유효한 가중화 회귀 그래프 패턴을 선별하여 가중화 회귀 그래프 패턴 마이닝 결과를 획득하는 유효 패턴 판단부를 포함한다.

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 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법 및 이를 수행하는 회귀 그래프 패턴 마이닝 장치는 그래프의 정점과 간선의 상호 연관관계 및 순서의 복잡한 특성 정보를 고려할 수 있다.

[0024] 본 발명의 일 실시예에 따른 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하기 위한 방법 및 이를 수행하는 회귀 그래프 패턴 마이닝 장치는 그래프 패턴별로 가지는 중요도 특성을 차등 적용하여 유효한 패턴을 선별할 수 있다.

도면의 간단한 설명

[0026] 도 1은 본 발명의 일 실시예에 따른 회귀 그래프 패턴 마이닝 시스템을 설명하는 도면이다.

도 2는 도 1에 있는 회귀 그래프 패턴 마이닝 장치의 구성을 나타내는 블록도이다.

도 3은 도 1에 있는 회귀 그래프 패턴 마이닝 장치에서 수행되는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하는 과정을 설명하는 순서도이다.

도 4는 다중 최소 지지도 기반의 가중화 회귀 그래프 패턴 마이닝을 위한 그래프 데이터베이스의 예를 설명하는 도면이다.

도 5는 그래프 데이터베이스 스캔을 통해 다중 최소 지지도 제약조건 및 중요도 값을 고려하여 유효 그래프 요소 데이터를 선별하는 과정을 설명하는 순서도이다.

도 6은 다중 최소 지지도 및 가중치 과추정 제약조건을 만족하는 유효한 정점 및 간선 정보를 기반으로 그래프 패턴 확장을 수행하는 과정을 설명하는 순서도이다.

도 7은 가중화 회귀 그래프 패턴 마이닝에서 그래프 대칭성을 활용한 경로 확장 과정을 설명하는 도면이다.

도 8은 다중 최소 지지도 및 가중치를 적용한 그래프 패턴의 지지도 계산 및 단계별 유효 패턴 판별 과정을 설명하는 순서도이다.

도 9는 가중화 회귀 그래프 패턴 마이닝 알고리즘을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0027] 본 발명에 관한 설명은 구조적 내지 기능적 설명을 위한 실시예에 불과하므로, 본 발명의 권리범위는 본문에 설명된 실시예에 의하여 제한되는 것으로 해석되어서는 아니 된다. 즉, 실시예는 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.
- [0028] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0029] "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.
- [0030] 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.
- [0031] 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.
- [0032] 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.
- [0033] 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.
- [0034] 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.

- [0036] 도 1은 본 발명의 일 실시예에 따른 회귀 그래프 패턴 마이닝 시스템을 설명하는 도면이다.
- [0037] 도 1을 참조하면, 회귀 그래프 패턴 마이닝 시스템(10)은 사용자 단말(100), 회귀 그래프 패턴 마이닝 장치(200) 및 그래프 데이터베이스(300)를 포함하고, 사용자 단말(100)과 회귀 그래프 패턴 마이닝 장치(200)는 네트워크를 통해 연결될 수 있다.
- [0038] 사용자 단말(100)은 회귀 그래프 패턴 마이닝 장치(200)에 접속하여 회귀 그래프 패턴의 마이닝을 요청할 수 있다. 사용자 단말(100)은 회귀 그래프 패턴 마이닝 장치(200)로부터 가중화 회귀 그래프 패턴 결과를 수신하여 디스플레이 할 수 있다. 사용자 단말(100)은 회귀 그래프 패턴 마이닝 장치(200)와 연결된 컴퓨팅 장치에 해당하고, 예를 들어, 데스크톱, 노트북, 태블릿 PC 또는 스마트폰 등으로 구현될 수 있다.
- [0039] 회귀 그래프 패턴 마이닝 장치(200)는 그래프 트랜잭션에 대한 그래프 데이터 및 중요도 정보를 저장해둔 그래프 데이터베이스(300)와 연동될 수 있고, 적어도 하나의 사용자 단말(100)과 연결되어 가중화 회귀 그래프 패턴 결과를 사용자 단말(100)에 제공할 수 있다.
- [0041] 도 2는 도 1에 있는 회귀 그래프 패턴 마이닝 장치의 구성을 나타내는 블록도이다.
- [0042] 도 2를 참조하면, 회귀 그래프 패턴 마이닝 장치(200)는 트랜잭션 스캔부(210), 트랜잭션 처리부(220), 후보 패턴 추출부(230), 유효 패턴 판단부(240) 및 제어부(250)를 포함한다.
- [0043] 트랜잭션 스캔부(210)는 그래프 데이터베이스(300)와 연동하여 해당 정보들을 스캔할 수 있다. 보다 구체적으로, 트랜잭션 스캔부(210)는 그래프 데이터베이스(300) 내의 각 그래프 트랜잭션에 대한 그래프 데이터와 중요도 정보를 스캔할 수 있다.
- [0044] 트랜잭션 처리부(220)는 스캔한 정보들을 기초로 다중 최소 지지도 제약조건을 고려하여 그래프 요소를 제거할 수 있다. 보다 구체적으로, 트랜잭션 처리부(220)는 다중 최소 지지도 제약조건을 만족하지 않으면 그래프 요소를 제거할 수 있다.
- [0045] 후보 패턴 추출부(230)는 그래프 패턴을 확장하여 가중화 회귀 그래프 후보 패턴을 추출할 수 있다. 유효 패턴 판단부(240)는 그래프 패턴의 중요도와 회귀도 특성을 고려하여 가중화 회귀 그래프 패턴의 집합을 마이닝의 결과로 도출할 수 있다.
- [0046] 제어부(250)는 회귀 그래프 패턴 마이닝 장치(200)의 전체적인 동작을 제어하고, 트랜잭션 스캔부(210), 트랜잭션 처리부(220), 후보 패턴 추출부(230) 및 유효 패턴 판단부(240) 간의 제어 흐름 또는 데이터 흐름을 제어할 수 있다.
- [0048] 도 3은 도 1에 있는 회귀 그래프 패턴 마이닝 장치에서 수행되는 그래프 요소의 다양한 중요도를 고려하여 다중 최소 지지도 기반의 회귀 그래프 패턴을 마이닝하는 과정을 설명하는 순서도이다.
- [0049] 도 3을 참조하면, 트랜잭션 스캔부(210)는 그래프 데이터베이스(300)를 스캔할 수 있다. 보다 구체적으로, 트랜잭션 스캔부(210)는 스캔 과정을 통해 정점, 간선 및 정점과 간선들 간의 연관관계 정보로 구성된 그래프 데이터 및 각 그래프 요소에 대한 중요도 정보를 스캔할 수 있다(단계 S310).
- [0050] 트랜잭션 처리부(220)는 각 그래프 트랜잭션에 대한 스캔 작업을 수행함으로써 그래프 데이터베이스(300)를 구성하는 각각의 정점 및 간선에 대한 지지도 정보를 파악할 수 있다. 트랜잭션 처리부(220)는 가중치 과추정 제약조건을 기초로 유효하지 않은 그래프 요소를 제거할 수 있다(단계 S320). 일 실시예에서, 트랜잭션 처리부(220)는 가중치 과추정 제약조건을 적용함으로써 가중치 적용에 따른 안티-모노톤 속성(Anti-monotone property) 위배에 의한 패턴 손실을 막을 수 있다. 여기에서, 안티-모노톤 속성(Anti-monotone property)은 전통적인 데이터 마이닝의 핵심적인 규칙으로서, 패턴을 조합하거나 확장할 때 효율적으로 처리할 수 있다.
- [0051] 트랜잭션 처리부(220)는 다중 최소 지지도 제약조건을 기초로 그래프 요소를 제거할 수 있다. 보다 구체적으로, 트랜잭션 처리부(220)는 사용자에게 의해 설정된 다중 최소 지지도 제약조건 정보로부터 가장 작은 값을 갖는 임계값을 극최소 지지도 임계값으로 설정한 후에 해당 임계값 조건을 만족하지 않는 그래프 요소를 제거할 수 있다(단계 S330).

- [0052] 트랜잭션 처리부(220)는 극최소 지지도 임계값을 설정한 후에 해당 값을 기초로 패턴 손실이 발생하지 않는 범위 내에서 프루닝 작업을 진행할 수 있다. 트랜잭션 처리부(220)는 유효 그래프 요소들을 체계적으로 참조하기 위한 리스트화 작업을 진행할 수 있다. 보다 구체적으로, 트랜잭션 처리부(220)는 프루닝 작업을 진행한 후에 패턴 마이닝 과정에서 체계적인 정보를 참조하기 위해 유효 그래프 요소 리스트를 구축할 수 있다(단계 S340).
- [0053] 후보 패턴 추출부(230)는 단계별 그래프 패턴 확장 과정을 통해 가중화 회귀 그래프에 대한 후보 패턴을 추출할 수 있다(단계 S350). 후보 패턴 추출부(230)는 확장되는 패턴 각각에 대해 단계별 재귀적 수행 방식을 적용한다.
- [0054] 유효 패턴 판단부(240)는 각 단계에 대해 생성되는 가중화 회귀 그래프 후보 패턴에 대해 각각의 실제 가중화 지지도를 계산하고 각 패턴 고유의 임계값과 비교함으로써 유효 패턴 여부를 판단할 수 있다(단계 S360). 유효 패턴 판단부(240)는 유효 패턴 판단 과정을 통해 가중화 회귀 그래프 패턴 결과를 획득할 수 있다(단계 S370).
- [0056] 도 4는 다중 최소 지지도 기반의 가중화 회귀 그래프 패턴 마이닝을 위한 그래프 데이터베이스의 예를 설명하는 도면이다.
- [0057] 도 4를 참조하면, 회귀 그래프 패턴 마이닝 장치(200) 내 구축된 그래프 데이터베이스(300)는 다수의 그래프 트랜잭션들로 구성되고, 적어도 하나의 그래프 트랜잭션은 하나의 그래프 형태를 이루며 다수의 정점들과 간선들로 구성될 수 있다. 여기에서, 각각의 정점과 간선은 고유의 이름을 가지고 있고, 고유의 이름을 통해 구별될 수 있다. 예를 들어, 고유의 이름은 레이블 값에 해당할 수 있다.
- [0058] 일 실시예에서, 회귀 그래프 패턴 마이닝 장치(200) 내 구축된 그래프 데이터베이스(300)의 적어도 하나의 그래프 트랜잭션은 중복되는 다수의 정점들과 간선들이 포함될 수 있고, 중복되는 정점 또는 간선 요소에 대한 지지도가 카운트될 때는 적어도 하나의 그래프 트랜잭션에 대해서 한번만 카운트될 수 있다. 여기에서, 모든 그래프 형태는 경로, 자유 트리, 순환 그래프의 3가지 상태로 표현될 수 있다(410).
- [0059] 그래프 데이터베이스(300)를 구성하고 있는 정점들의 정보는 각각의 정점의 이름, 지지도, 다중 최소 지지도 제약조건을 위한 고유 임계값으로 표현될 수 있고(420), 간선들의 정보는 각각의 간선의 이름, 지지도, 다중 최소 지지도 제약조건을 위한 고유 임계값 및 그래프 요소의 중요도를 표현하기 위한 간선의 가중치 정보로 표현될 수 있다(430).
- [0060] 회귀 그래프 패턴 마이닝 장치(200) 내 그래프 데이터베이스(300)는 그래프를 구성하는 각 그래프 요소마다 다른 중요도를 고려할 수 있다. 회귀 그래프 패턴 마이닝 장치(200)는 그래프의 간선들 각각의 다른 가중치가 부여된 그래프 데이터베이스(300)를 가지고 패턴을 마이닝하는 것을 고려할 수 있고, 이를 통해 노드들 사이를 연결하는 간선들의 지지도와 중요도를 동시에 고려한 가중화 회귀 그래프 패턴 결과를 획득할 수 있다.
- [0061] 그래프 G에 대한 정점 집합이 $V_G = \{v_1, v_2, \dots, v_n\}$ 이고, 간선 집합이 $E_G = \{e_1, e_2, \dots, e_m\}$ 이며, 간선의 가중치 집합이 $W_G = \{w_1, w_2, \dots, w_n\}$ 일 때, 그래프 패턴의 가중치 $W_{sup}(G)$ 는 [수학식 1]을 통해 계산된다. 여기에서, $Sup(G)$ 은 그래프 G의 지지도이고, $Avg(W)$ 은 간선의 가중치 평균 값이다. 즉, 그래프 패턴의 가중치는 아래의 [수학식 1]과 같이 정의된다.
- [0062] [수학식 1]
- [0063]
$$W_{sup}(G) = Sup(G) * Avg(W)$$
- [0065] 그래프 데이터베이스(300)를 구성하는 각각의 요소, 즉, 정점(V)과 간선(E)에 대한 극최소 지지도 임계값들이 필요하다. 그래프 데이터베이스(300) Gtrs가 $GDB = \{Gtr1, Gtr2, \dots, Gtrk\}$ 이고, 별도의 정점 집합 VGDB가 $VGDB = \{v1, v2, \dots, vx\}$ 이며, 별도의 간선 집합 EGDB가 $EGDB = \{e1, e2, \dots, ey\}$ 일 때, 정점(V)과 간선(E)에 대한 극최소 지지도 임계값들은 아래의 [수학식 2]을 통해 사용자가 지정한 값으로 설정된다.

[0066] [수학식 2]

$$\delta_i (1 \leq i \leq x + y)$$

[0067]

[0069] 그래프에 대한 극최소 지지도 임계값은 다음과 같이 정의된다.

[0070] [수학식 3]

$$MGST(G)$$

[0071]

[0072] 여기에서, [수학식 3]은 그래프(G)의 모든 δ 값 중 최소값으로 설정되고, 해당하는 MGST 조건들을 만족하는 가중화 회귀 그래프 패턴들만이 유효한 결과로 도출될 수 있다.

[0073] 보다 구체적으로, 위의 [수학식 1]은 그래프 패턴의 실제 가중화 지지도 대신 그래프 데이터베이스(300) 내에서 가장 큰 간선의 최대 가중화 지지도 값 MaxW를 사용하여 안티-모노톤 속성(Anti-monotone property)을 유지할 수 있고 패턴 손실을 예방할 수 있다. 또한, [수학식 2]는 안티-모노톤 속성(Anti-monotone property)을 위배하지 않는 극최소 지지도 임계값을 설정하여 패턴 마이닝 과정에 적용할 수 있다. 위의 [수학식 1] 및 [수학식 2]을 통해 획득한 그래프 패턴들은 가중화 회귀 그래프 후보 패턴에 해당하고, [수학식 3]의 MGST 조건을 만족하는 그래프 패턴들만이 유효한 결과로 도출될 수 있다.

[0075] 도 5는 그래프 데이터베이스 스캔을 통해 다중 최소 지지도 제약조건 및 중요도 값을 고려하여 유효 그래프 요소 데이터를 선별하는 과정을 설명하는 순서도이다.

[0076] 도 5를 참조하면, 그래프 데이터베이스(300) 내에 스캔할 그래프 트랜잭션이 존재하는지 파악할 수 있고(단계 S510), 처리할 그래프 트랜잭션이 존재한다면 먼저 각 그래프 요소인 정점과 간선에 대해 지지도 정보를 각각 해당하는 요소에 따라 구분하여 누적할 수 있다(단계 S520). 하나의 그래프 트랜잭션 내에서 중복으로 나타나는 정점 또는 간선에 대한 누적 카운트를 제거함으로써 누적 카운트를 방지할 수 있다(단계 S530). 처리된 그래프 트랜잭션에 대한 그래프 구성 정보를 요소 별로 리스트화 하여 저장할 수 있다(단계 S540).

[0077] 그래프 데이터베이스(300) 내에 스캔할 그래프 트랜잭션이 존재하지 않는다면 누적한 그래프 요소 정보를 가지고 임계값 조건을 기준으로 불필요한 정점 정보를 삭제할 수 있다(단계 S550). 정점 정보 삭제 과정에서는 임계값으로 극최소 지지도 임계값이 사용되고, 해당 극최소 지지도 임계값은 사용자로부터 주어진 다중 최소 지지도 제약조건을 구성하는 각 요소의 임계값 중 가장 작은 값으로 설정된다. 이때 극최소 지지도 임계값이 아닌 각 요소 고유의 임계값을 사용하여 프루닝 작업을 수행한다면 패턴 확정을 통한 마이닝할 때 안티-모노톤 속성(Anti-monotone property)이 깨지기 때문에 심각한 패턴 손실이 발생할 수 있다. 즉, 극최소 지지도 임계값을 적용함으로써 근사치 기준을 우선적으로 적용하고 각 패턴에 대한 실제 임계값을 적용하여 유효 패턴을 선별하는 과정을 거칠 수 있다.

[0078] 불필요한 정점 정보 삭제가 마무리 되면, 가중치 과추정 제약조건을 기초로 불필요한 간선 정보를 삭제할 수 있다(단계 S560). 간선 정보 삭제 과정 역시 간선 고유의 가중치 정보를 적용한 가중화 지지도 값에 대해 해당 간선 고유의 임계값을 적용하면 안티-모노톤 속성(Anti-monotone property) 위배에 따른 패턴 손실이 초래될 수 있으므로, 간선이 가질 수 있는 최대 가중치 값을 이용한 최대 가중화 지지도 값을 먼저 계산하고, 이를 극최소 지지도 임계값과 비교함으로써 프루닝 작업을 먼저 수행한다. 이때, 조건을 만족하지 않는 간선은 영구적으로 불필요한 데이터가 되므로 프루닝 작업을 진행할 수 있다. 모든 정점 및 간선에 대해 단계별 작업을 적용하면 유효한 그래프 요소 정보를 파악할 수 있다.

[0080] 도 6은 다중 최소 지지도 및 가중치 과추정 제약조건을 만족하는 유효한 정점 및 간선 정보를 기반으로 그래프 패턴 확장을 수행하는 과정을 설명하는 순서도이다.

[0081] 도 6을 참조하면, 회귀 그래프 패턴 마이닝 장치(200)는 그래프 데이터 스캔 작업을 통해 획득한 유효한 정점

정보 집합에 대해 한 번에 하나의 정점씩 각각의 정점을 순차적으로 선택하여 해당 정점을 기준점으로 선택할 수 있고(단계 S610), 해당 기준점으로부터 생성될 수 있는 모든 가능한 확장 그래프 패턴들을 순차적으로 확인할 수 있다(단계 S620).

[0082] 회귀 그래프 패턴 마이닝 장치(200)는 구축한 그래프 요소 리스트를 참조하여 그래프 패턴 확장이 가능한 그래프 형태가 존재하면, 현재 그래프 패턴 상태에서 확장 가능한 그래프 패턴의 수와 종류를 파악할 수 있고(단계 S630). 각각에 대해 다중 최소 지지도 제약조건과 가중치 과추정 제약조건을 적용하여 다시 확장 가능한 가중화 회귀 그래프 패턴의 후보들을 추출할 수 있다(단계 S640). 여기에서, 추출된 가중화 회귀 그래프 후보 패턴들은 다음 단계의 재귀적 확장 작업을 위해 사용될 수 있다.

[0083] 회귀 그래프 패턴 마이닝 장치(200)는 추출된 각 가중화 회귀 그래프 후보 패턴들에 대해 실제 가중화 지지도를 계산할 수 있고, 각 가중화 회귀 그래프 패턴들에 적합하도록 설정된 극최소 지지도 임계값을 비교하여 실제로 유효한 가중화 회귀 그래프 패턴을 선별할 수 있다(단계 S650). 여기에서, 각각의 패턴의 회귀도 특성을 고려한 극최소 지지도 임계값의 계산은 해당 패턴을 구성하는 모든 정점과 간선들에 대한 고유의 임계값 정보를 파악하고, 해당 값들 중 가장 작은 값을 해당 패턴의 극최소 지지도 임계값으로 설정하여 패턴의 회귀도 특성을 고려할 수 있다. 회귀 그래프 패턴 마이닝 장치(200)는 각 가중화 회귀 그래프 후보 패턴에 대한 재귀적 확장 작업을 수행할 수 있다(단계 S660).

[0084] 회귀 그래프 패턴 마이닝 장치(200)는 그래프 패턴 확장이 가능한 그래프 형태가 존재하지 않는다면 처리되지 않은 정점이 존재하는지 파악할 수 있고(단계 S670), 처리되지 않은 정점이 존재하면 해당 정점을 다시 최초의 기준점으로 설정하고 위의 절차들을 반복 수행할 수 있다. 일 실시예에서, 회귀 그래프 패턴 마이닝 장치(200)는 모든 정점에 대해 처리되지 않은 정점이 존재하지 않는다면 모든 패턴 확장 작업이 완료된 상태이므로 현재까지 추출한 다중 최소 지지도 제약조건 및 중요도 제약조건을 만족하는 가중화 회귀 그래프 패턴 정보를 반환할 수 있고, 반환된 가중화 회귀 그래프 패턴 정보를 사용자에게 제공할 수 있다(단계 S680).

[0086] 도 7은 가중화 회귀 그래프 패턴 마이닝에서 그래프 대칭성을 활용한 경로 확장 과정을 설명하는 도면이다.

[0087] 도 7에서, 그래프 패턴의 형태 중 하나인 경로는 특성상 패턴 확장 작업이 양쪽 끝 단에서 수행하게 되고 패턴 확장에 대한 특정 기준 없이 패턴 확장 작업을 수행하면 중복된 그래프 패턴이 생성되는 문제가 발생할 수 있기 때문에, 그래프 경로에서 경로로 패턴이 확장되는 경우에는 한쪽 방향으로만 패턴 확장 작업이 수행되도록 제한할 수 있다(710). 일 실시예에서, 패턴 확장 과정에서 하나의 경로가 더 긴 길이의 경로로 확장되는 것은 두 가지의 경우의 수가 있다(720 및 730). 도 7에서 보이는 것과 같이, 확장될 경로의 대칭성 정보는 확장되는 경로 패턴의 대칭성 부분 중 새로 추가된 정점과 간선 집합을 제외한 나머지 부분에 대해 영향을 끼치므로 확장될 경로의 대칭성 정보를 활용하여 경로의 대칭성을 계산하는 부분에 대한 오버헤드를 줄일 수 있다.

[0089] 도 8은 다중 최소 지지도 및 가중치를 적용한 그래프 패턴의 지지도 계산 및 단계별 유효 패턴 판별 과정을 설명하는 순서도이다.

[0090] 도 8을 참조하면, 회귀 그래프 패턴 마이닝 장치(200)는 최대 가중치 정보를 이용하여 현재 그래프 패턴의 최대 가중화 지지도 값을 계산할 수 있고(단계 S810), 최대 가중치 정보는 그래프 데이터베이스(300) 내 간선들이 가질 수 있는 가중치 값들 중 가장 큰 값으로 설정될 수 있다. 회귀 그래프 패턴 마이닝 장치(200)는 최대 가중화 지지도 값에 대한 계산이 완료되면 극최소 지지도 임계값과 최대 가중화 지지도 값을 비교하는 작업을 수행할 수 있고(단계 S820), 여기에서, 극최소 지지도 임계값은 그래프 데이터베이스(300)를 구성하는 모든 정점과 간선들에 대해 할당된 고유의 임계값 정보들 중 가장 작은 값으로 설정된 값을 의미한다.

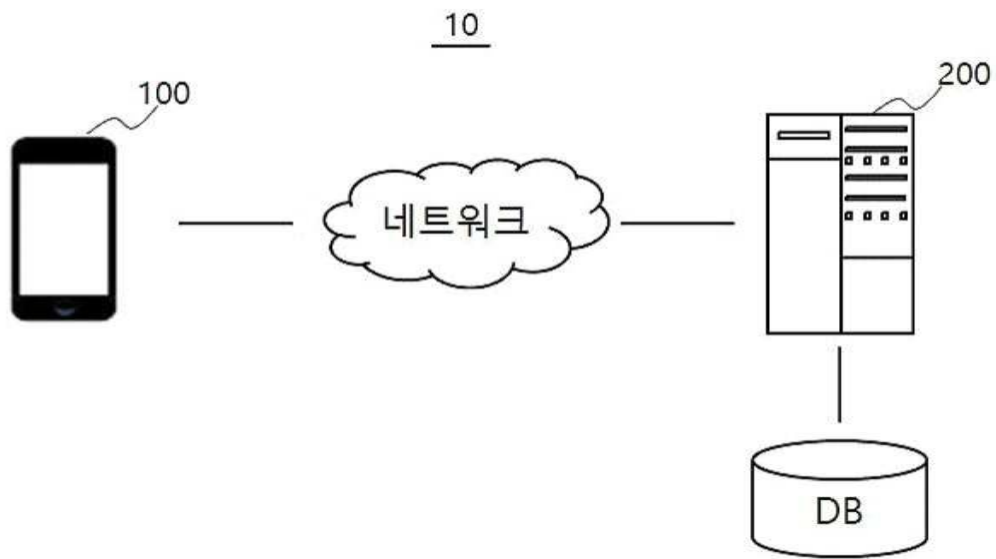
[0091] 회귀 그래프 패턴 마이닝 장치(200)는 최대 가중화 지지도 값이 극최소 지지도 임계값보다 크다면(단계 S830) 현재 그래프 패턴에 대한 실제 가중화 지지도를 계산하고, 계산된 실제 가중화 지지도 값을 해당 패턴의 극최소 지지도 임계값과 비교할 수 있다(단계 S840). 일 실시예에서, 실제 가중화 지지도 값이 극최소 지지도 임계값보다 크다면(단계 S850) 현재 그래프 패턴을 유효한 가중화 회귀 그래프 패턴으로 고려하여 패턴 마이닝 결과값으로 추출할 수 있고(단계 S860), 실제 가중화 지지도 값이 극최소 지지도 임계값보다 크지 않다면(단계 S850) 유효한 패턴 결과가 아니므로 유효 패턴 판별하는 과정을 종료할 수 있다. 여기에서, 그래프 패턴은 가중화 회귀 그래프 후보 패턴의 조건을 만족하는 패턴이 되고, 해당 그래프 패턴은 재귀적 패턴 확장 과정을 위한 요소로

참여할 수 있다.

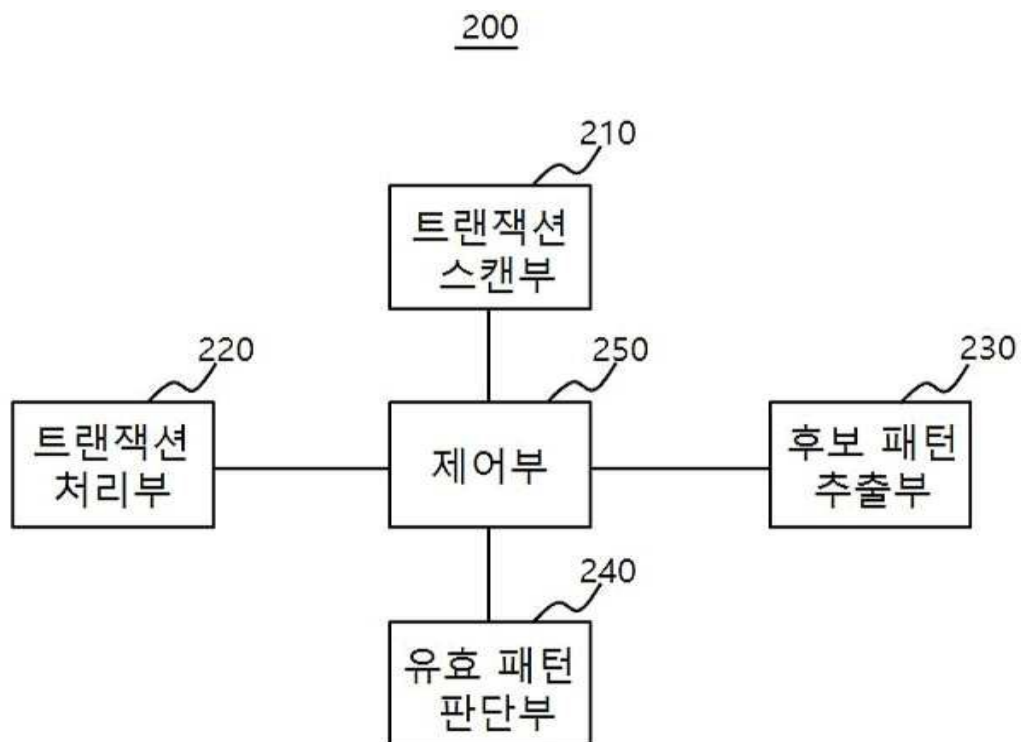
- [0092] 최대 가중화 지지도 값이 극최소 지지도 임계값보다 크지 않다면(단계 S830) 현재 그래프 패턴은 희귀 그래프 패턴 마이닝 과정에서 영구적으로 의미 없는 패턴에 해당하므로 안티-모노톤 속성(Anti-monotone property)에 의해 해당 패턴 및 모든 관련 작업들을 수행하지 않는다(단계 S870).
- [0094] 도 9는 가중화 희귀 그래프 패턴 마이닝 알고리즘을 나타내는 도면이다.
- [0095] 도 9를 참조하면, 가중화 희귀 그래프 알고리즘은 유효한 정점과 간선들을 그래프 데이터베이스(300)로부터 찾은 후에(lines 1-2) 각각의 정점에 대해 가중화 희귀 그래프의 집합 WRGs을 마이닝하기 위한 확장 과정을 수행한다(lines 3-5). 확장 과정에서 가중화 희귀 그래프 알고리즘은 각 간선에 대해 그래프들의 상태를 고려하여 확장시킨다(lines 6-9). 각각의 확장된 그래프에 대해 최대 가중화 지지도가 LMS 보다 작다면 영구적으로 프루닝된다(lines 10). 만약 최대 가중화 지지도에 해당하는 MGST 보다 작지 않다면 유효한 그래프 패턴으로 도출될 수 있다(line 11). 그래프 패턴이 LMS보다 작지 않다면 그래프 패턴에 대한 확장 과정이 재귀적으로 수행된다(lines 12-13).
- [0097] 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.
- [0099] [참조문헌]
- [0100] 1. Hu, Y.H. and Chen, Y.L.: Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. Decision Support Systems. vol. 42, no. 1, pp.1--24 (2006)
- [0101] 2. Kiran, R.U. and Reddy, P.K.: Novel techniques to reduce search space in multiple minimum supports-based frequent pattern mining algorithms. 14th International Conference on Extending Database Technology. pp. 11--20 (2011)
- [0102] 3. Lee, G. and Yun, U.: Mining weighted frequent sub-graphs with weight and support affinities. Proceedings of the 6th Multi-Disciplinary International Workshop on Artificial Intelligence. pp. 227--238 (2012).
- [0103] 4. Liu, B., Hsu, W., and Ma, Y.: Mining association rules with multiple minimum supports. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 337--341 (1999)
- [0104] 5. Nijssen, S. and Kok, J.N.: The Gaston Tool for Frequent Subgraph Mining. Electronic Notes in Theoretical Computer Science. vol. 127, no. 1, pp. 77--87 (2005).
- [0105] 6. Nijssen, S. and Kok, J.N.: A quickstart in frequent structure mining can make a difference, Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 647--652 (2004)

도면

도면1



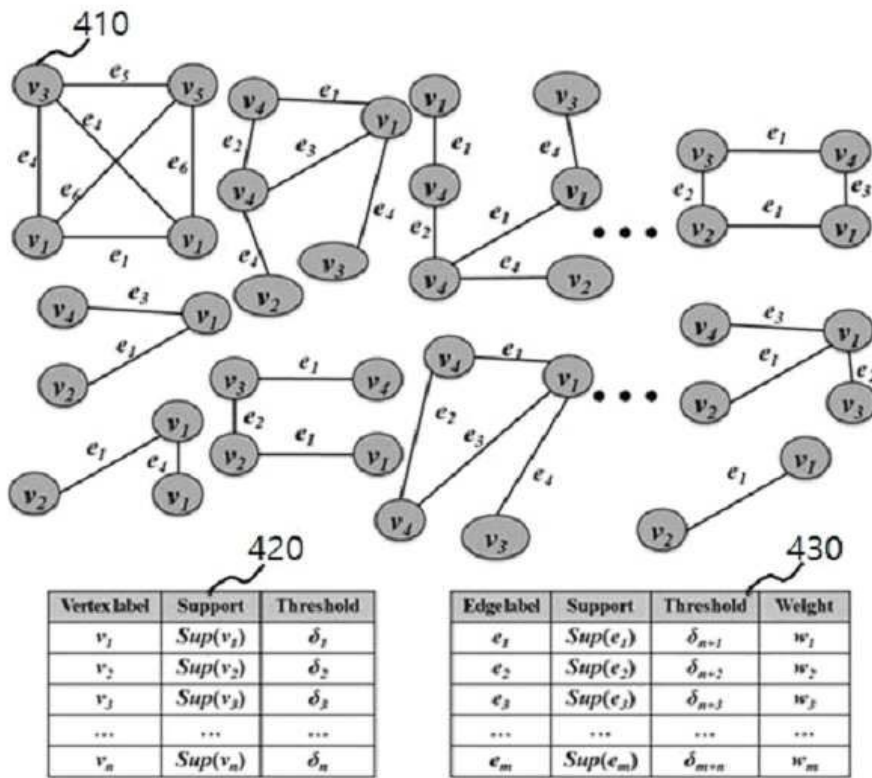
도면2



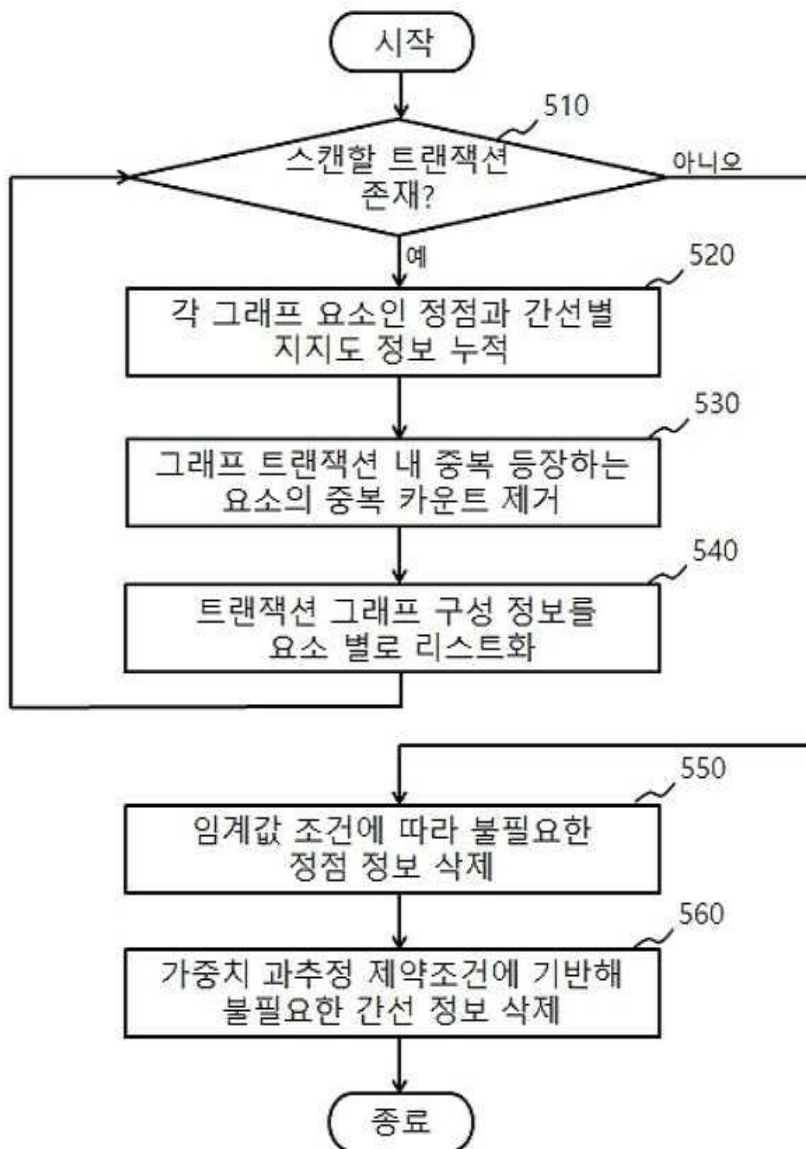
도면3



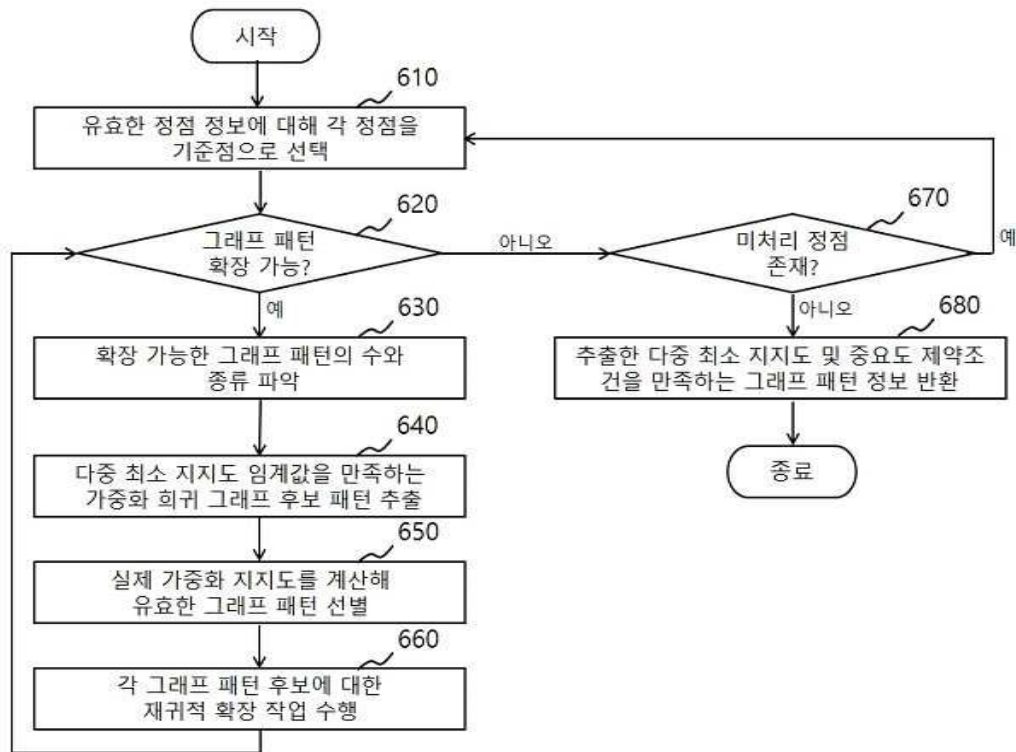
도면4



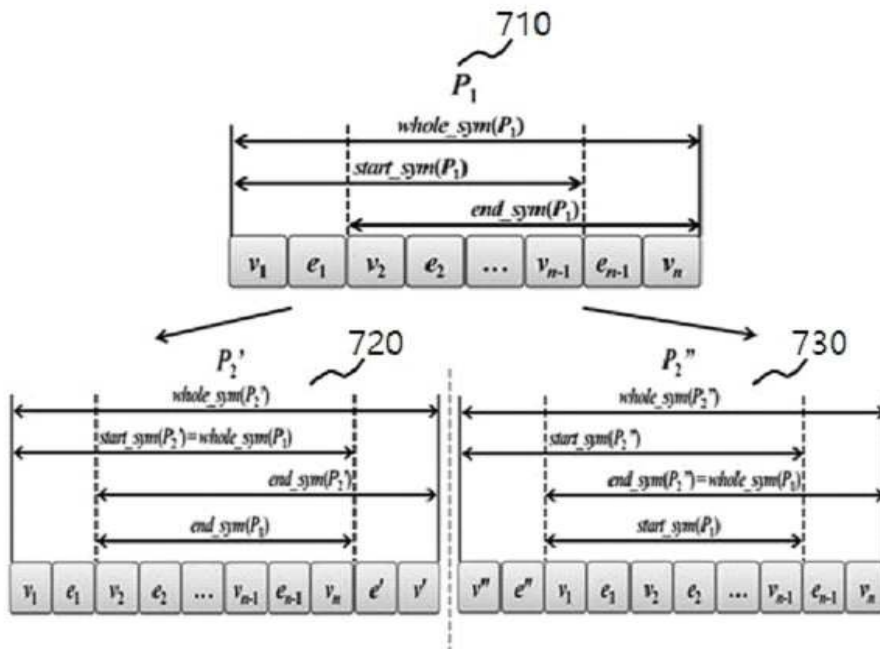
도면5



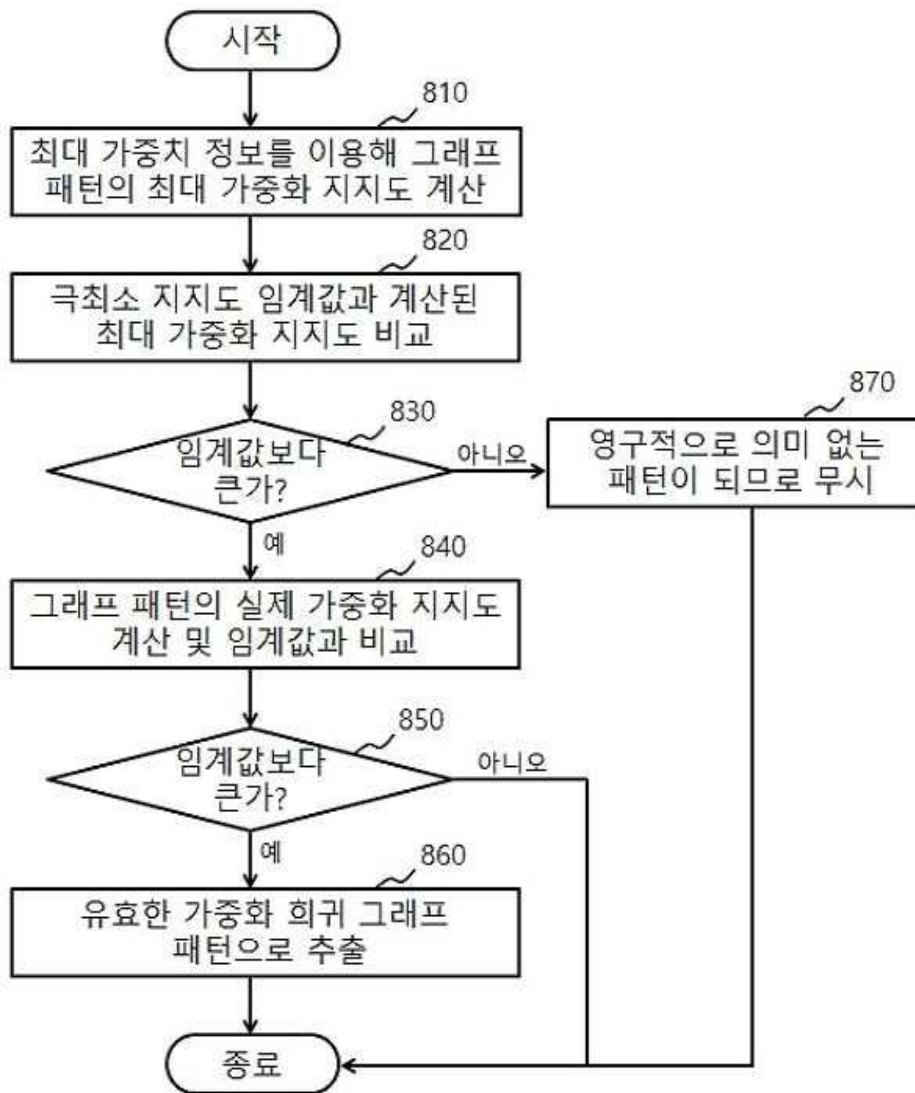
도면6



도면7



도면8



도면9

Input : a graph database, GDB , a list of minimum element support thresholds, L

Output : a set of WRGs, S

Procedure: WRG-Miner

01. a set of vertices, $V \leftarrow$ vertices such that each of their supports $\geq LMS$;

02. a set of edges, $E \leftarrow$ edges such that each of their supports $\ast MaxW \geq LMS$;

03. for each vertex, v_i in V

04. a graph pattern, $G \leftarrow v_i$; a set of edges, $E' \leftarrow$ edges that can be attached to v_i in E ;

05. $S \leftarrow S \cup$ call $WRG-Growth(G, E')$;

Function: WRG-Growth(G, E)

06. for each edge, e_i in E

08. if G is a path or free tree, an expanded graph of G , $G' \leftarrow G \cup e_i$ and the vertex in e_i ;

09. else if G is a cyclic graph, $G' \leftarrow G \cup e_i$; //only a cyclic edge

10. if $Wsup(G') \geq LMS$

11. if $Wsup(G') \geq MGST(G')$, $S \leftarrow S \cup G'$;

12. a set of edges, $E' \leftarrow$ edges that can be attached to G' ;

13. $S \leftarrow S \cup$ call $WRG-Growth(G', E')$;
