



(19) 대한민국특허청(KR)

(12) 공개특허공보(A)

(11) 공개번호

10-2016-0093237

(43) 공개일자

2016년08월08일

(51) 국제특허분류(Int. Cl.)

G06F 17/30 (2006.01)

(52) CPC특허분류

G06F 17/30539 (2013.01)

G06F 17/30501 (2013.01)

(21) 출원번호

10-2015-0013980

(22) 출원일자

2015년01월29일

심사청구일자

2015년01월29일

(71) 출원인

충북대학교 산학협력단

충청북도 청주시 서원구 충대로 1 (개신동)

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

윤은일

서울특별시 성북구 종암로5길 7 (종암동)

양홍모

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

(뒷면에 계속)

(74) 대리인

정부연

전체 청구항 수 : 총 19 항

(54) 발명의 명칭 데이터 스트림 상에서 가중화 최대 빈발 패턴을 마이닝하기 위한 슬라이딩 윈도우 기반 빈발 패턴 관리 방법

(57) 요약

슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 방법은 (a)트랜잭션이 특정 배치(Batch) 크기만큼 수집되었는지 여부를 스캔하는 단계, (b)상기 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우(Sliding Window)에 삽입하는 단계, (c)상기 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신하는 단계 및 (d)상기 트리 자료 구조에 포함된 아이템 각각에 대한 지지도를 기초로 상기 지지도가 내림차순으로 정렬되도록 상기 트리 자료 구조를 재구축하는 단계를 포함한다.

대표도 - 도1

100



(52) CPC특허분류

Y10S 707/99936 (2013.01)

Y10S 707/99937 (2013.01)

(72) 발명자

이강인

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

김동규

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

류근호

충청북도 청주시 서원구 충대로 1 충북대학교 43동
201호

명세서

청구범위

청구항 1

- (a) 트랜잭션이 특정 배치(Batch) 크기만큼 수집되었는지 여부를 스캔하는 단계;
- (b) 상기 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우(Sliding Window)에 삽입하는 단계;
- (c) 상기 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신하는 단계; 및
- (d) 상기 트리 자료 구조에 포함된 아이템 각각에 대한 지지도를 기초로 상기 지지도가 내림차순으로 정렬되도록 상기 트리 자료 구조를 재구축하는 단계를 포함하는 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 방법.

청구항 2

제1항에 있어서, 상기 (b)단계는

상기 수집된 트랜잭션을 상기 슬라이딩 윈도우에 삽입할 수 있는지 여부를 확인하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 3

제2항에 있어서, 상기 (b)단계는

상기 수집된 트랜잭션을 상기 슬라이딩 윈도우에 삽입할 수 없는 경우에는 상기 슬라이딩 윈도우에 삽입되어 있는 적어도 하나의 배치 중에서 가장 오래된 배치를 삭제하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 4

제1항에 있어서, 상기 (c)단계는

기 구축되어 있는 트리 자료 구조가 있는지 여부를 확인하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 5

제4항에 있어서, 상기 (c) 단계는

상기 기 구축되어 있는 트리 자료 구조가 있는 경우에는 상기 트랜잭션에 포함된 아이템을 상기 트리 자료 구조에 삽입하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 6

제4항에 있어서, 상기 (c)단계는

상기 기 구축되어 있는 트리 자료 구조가 없는 경우에는 상기 트랜잭션을 기초로 새로운 트리 자료 구조를 생성하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 7

제1항에 있어서, 상기 (d)단계는

상기 트리 자료 구조의 단말 노드로부터 루트 노드까지의 경로에 포함되어 있는 아이템이 지지도를 기초로 내림차순으로 정렬되어 있는지 여부를 확인하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 8

제7항에 있어서, 상기 (d)단계는

상기 아이템이 지지도를 기초로 내림차순으로 정렬되어 있지 않은 경우에는 해당 경로를 상기 트리 자료 구조로부터 분리하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 9

제8항에 있어서, 상기 (d)단계는

상기 분리한 경로에 포함되어있는 아이템을 지지도를 기초로 내림차순으로 정렬하고 상기 트리 자료 구조에 삽입하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 10

제9항에 있어서, 상기 (d)단계는

상기 내림차순으로 정렬하는 과정에서 상기 지지도가 0이 된 아이템이 있는 경우에는 해당 아이템에 대한 노드를 삭제하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 11

제1항에 있어서, 상기 (d)단계는

상기 재구성된 트리 자료 구조를 기초로 노드 링크를 재연결하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 12

제1항에 있어서,

(e) 상기 트리 자료 구조를 기초로 가중화 최대 빈발 패턴을 추출하는 단계를 더 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 13

제12항에 있어서, 상기 (e)단계는

상기 트리 자료 구조를 기초로 임계값을 만족시키는 잠재적 가중화 지지도를 가지는 아이템으로 구성된 조건적 트리 자료 구조를 재귀적으로 생성하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 14

제13항에 있어서, 상기 (e) 단계는

상기 조건적 트리 자료 구조가 단일 경로로 구성되는 경우에는 상기 단일 경로에 포함되는 아이টে를 기초로 후보 빈발 패턴을 생성하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 15

제14항에 있어서, 상기 (e) 단계는

상기 후보 빈발 패턴의 실제 가중화 지지도가 임계값을 만족하는 경우에는 상기 후보 빈발 패턴을 상기 가중화 최대 빈발 패턴으로 추출하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 16

제1항에 있어서, 상기 트리 자료 구조는

상기 슬라이딩 윈도우에 포함되어 있는 아이টে 각각에 대한 아이টে 이름, 가중치, 지지도 및 노트 링크를 저장 및 관리하는 테이블 및 상기 아이টে이 실제로 저장되는 트리를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 17

제16항에 있어서, 상기 트리는

상기 아이টে 이름 및 지지도를 포함하는 노트 및 상기 아이টে 이름, 지지도 및 상기 아이টে이 포함되어 있던 배치에 대한 부분 지지도 정보를 포함하는 단말 노트를 포함하는 것을 특징으로 하는 최신 빈발 패턴 관리 방법.

청구항 18

트랜잭션이 특정 배치 크기만큼 수집되었는지 여부를 스캔하는 스캔부;

상기 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우에 삽입하는 슬라이딩 윈도우 삽입부;

상기 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신하는 트리 자료 구조 갱신부; 및

상기 트리 자료 구조에 포함된 아이টে 각각에 대한 지지도를 기초로 상기 지지도가 내림차순으로 정렬되도록 상기 트리 자료 구조를 재구축하는 트리 자료 구조 재구축부를 포함하는 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 서버.

청구항 19

트랜잭션이 특정 배치 크기만큼 수집되었는지 여부를 스캔하는 기능;

상기 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우에 삽입하는 기능;

상기 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신하는 기능; 및

상기 트리 자료 구조에 포함된 아이টে 각각에 대한 지지도를 기초로 상기 지지도가 내림차순으로 정렬되도록 상기 트리 자료 구조를 재구축하는 기능을 포함하는 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 방법에 관한 컴퓨터 프로그램을 기록한 기록매체.

발명의 설명

기술 분야

- [0001] 본 발명은 슬라이딩 윈도우 기반 빈발 패턴 관리 기술에 관한 것으로서, 보다 상세하게는, 수집된 트랜잭션을 슬라이딩 윈도우에 삽입하고 트리 자료 구조를 갱신 및 재구축하는 것을 통하여 최신 빈발 패턴만을 관리할 수 있는 슬라이딩 윈도우 기반 빈발 패턴 관리 방법에 관한 것이다.

배경 기술

- [0002] 일반적으로, 데이터 스트림 상의 패턴 마이닝은 지속적으로 축적되는 데이터 스트림 환경에서 필요한 패턴들을 찾아낼 수는 있지만 오래 전에 입력된 데이터와 최근에 입력된 데이터를 동일 선상에서 고려한다. 또한, 데이터 스트림의 특성에 따라 시간이 지날수록 누적된 데이터의 크기가 점차적으로 증가하기 때문에 대규모의 데이터를 모두 마이닝 하는데 막대한 시간과 노력이 들게 되고 신뢰도가 떨어지게 된다.
- [0003] 한국등록특허 제10-0812378호는 지속적으로 발생하는 트랜잭션 데이터 집합인 데이터 스트림 환경에서 빈발항목 집합 탐색을 위한 축약형 전위 트리를 이용한 빈발항목집합 탐색 방법에 관한 것으로서, 데이터 스트림 환경에서 빈발 항목을 탐색하는데 효과적인 데이터 구조를 정의하고 이 데이터 구조를 이용하여 필요한 정보를 탐색한다.
- [0004] 한국등록특허 제10-0913027호는 응용 도메인에서 지속적으로 발생하는 트랜잭션들로 구성되는 비한정적인 데이터 집합을 대상으로 새로운 트랜잭션이 생성됨에 따라 해당 트랜잭션을 한번만 읽어서 데이터 마이닝을 수행하여 해당 트랜잭션을 포함하는 증가된 데이터 집합에 내재된 정보를 분석하는 데이터 마이닝 방법에 관한 것이다.

선행기술문헌

특허문헌

- [0005] (특허문헌 0001) 한국등록특허 제10-0812378호
(특허문헌 0002) 한국등록특허 제10-0913027호

발명의 내용

해결하려는 과제

- [0006] 본 발명의 일 실시예는 수집된 트랜잭션을 슬라이딩 윈도우에 삽입하여 최신의 데이터만을 유지할 수 있는 최신 빈발 패턴 관리 방법을 제공하고자 한다.
- [0007] 본 발명의 일 실시예는 생성된 트리 자료 구조를 아이템의 지지도 내림차순으로 정렬되도록 재구축하여 최신의 데이터가 트리 자료 구조를 통하여 효율적으로 저장되고 표현될 수 있는 최신 빈발 패턴 관리 방법을 제공하고자 한다.
- [0008] 본 발명의 일 실시예는 패턴의 최대성을 고려하여 대표 패턴만을 선별적으로 추출할 수 있는 최신 빈발 패턴 관리 방법을 제공하고자 한다.

과제의 해결 수단

- [0009] 실시예들 중에서, 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 방법은 (a)트랜잭션이 특정 배치(Batch) 크기만큼 수집되었는지 여부를 스캔하는 단계, (b)상기 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우(Sliding Window)에 삽입하는 단계, (c)상기 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신하는 단계 및 (d)상기 트리 자료 구조에 포함된 아이тем 각각에 대한 지지도를 기초로 상기 지지도가 내림차순

으로 정렬되도록 상기 트리 자료 구조를 재구축하는 단계를 포함한다.

- [0010] 일 실시예에서, 상기 (b)단계는 상기 수집된 트랜잭션을 상기 슬라이딩 윈도우에 삽입할 수 있는지 여부를 확인하는 단계를 포함할 수 있다.
- [0011] 일 실시예에서, 상기 (b)단계는 상기 수집된 트랜잭션을 상기 슬라이딩 윈도우에 삽입할 수 없는 경우에는 상기 슬라이딩 윈도우에 삽입되어 있는 적어도 하나의 배치 중에서 가장 오래된 배치를 삭제하는 단계를 포함할 수 있다.
- [0012] 일 실시예에서, 상기 (c)단계는 기 구축되어 있는 트리 자료 구조가 있는지 여부를 확인하는 단계를 포함할 수 있다.
- [0013] 일 실시예에서, 상기 (c) 단계는 상기 기 구축되어 있는 트리 자료 구조가 있는 경우에는 상기 트랜잭션에 포함된 아이템을 상기 트리 자료 구조에 삽입하는 단계를 포함할 수 있다.
- [0014] 일 실시예에서, 상기 (c)단계는 상기 기 구축되어 있는 트리 자료 구조가 없는 경우에는 상기 트랜잭션을 기초로 새로운 트리 자료 구조를 생성하는 단계를 포함할 수 있다.
- [0015] 일 실시예에서, 상기 (d)단계는 상기 트리 자료 구조의 단말 노드로부터 루트 노트까지의 경로에 포함되어 있는 아이템이 지지도를 기초로 내림차순으로 정렬되어 있는지 여부를 확인하는 단계를 포함할 수 있다.
- [0016] 일 실시예에서, 상기 (d)단계는 상기 아이템이 지지도를 기초로 내림차순으로 정렬되어 있지 않은 경우에는 해당 경로를 상기 트리 자료 구조로부터 분리하는 단계를 포함할 수 다.
- [0017] 일 실시예에서, 상기 (d)단계는 상기 분리한 경로에 포함되어있는 아이템을 지지도를 기초로 내림차순으로 정렬하고 상기 트리 자료 구조에 삽입하는 단계를 포함할 수 있다.
- [0018] 일 실시예에서, 상기 (d)단계는 상기 내림차순으로 정렬하는 과정에서 상기 지지도가 0이 된 아이템이 있는 경우에는 해당 아이템에 대한 노드를 삭제하는 단계를 포함할 수 있다.
- [0019] 일 실시예에서, 상기 (d)단계는 상기 재구축된 트리 자료 구조를 기초로 노드 링크를 재연결하는 단계를 포함할 수 있다.
- [0020] 일 실시예에서, (e) 상기 트리 자료 구조를 기초로 가중화 최대 빈발 패턴을 추출하는 단계를 더 포함할 수 있다.
- [0021] 일 실시예에서, 상기 (e)단계는 상기 트리 자료 구조를 기초로 임계값을 만족시키는 잠재적 가중화 지지도를 가지는 아이템으로 구성된 조건적 트리 자료 구조를 재귀적으로 생성하는 단계를 포함할 수 있다.
- [0022] 일 실시예에서, 상기 (e) 단계는 상기 조건적 트리 자료 구조가 단일 경로로 구성되는 경우에는 상기 단일 경로에 포함되는 아이템을 기초로 후보 빈발 패턴을 생성하는 단계를 포함할 수 있다.
- [0023] 일 실시예에서, 상기 (e) 단계는 상기 후보 빈발 패턴의 실제 가중화 지지도가 임계값을 만족하는 경우에는 상기 후보 빈발 패턴을 상기 가중화 최대 빈발 패턴으로 추출하는 단계를 포함할 수 있다.
- [0024] 일 실시예에서, 상기 트리 자료 구조는 상기 슬라이딩 윈도우에 포함되어 있는 아이템 각각에 대한 아이템 이름, 가중치, 지지도 및 노드 링크를 저장 및 관리하는 테이블 및 상기 아이템이 실제로 저장되는 트리를 포함할 수 있다.
- [0025] 일 실시예에서, 상기 트리는 상기 아이템 이름 및 지지도를 포함하는 노드 및 상기 아이템 이름, 지지도 및 상기 아이템이 포함되어 있던 배치에 대한 부분 지지도 정보를 포함하는 단말 노드를 포함할 수 있다.
- [0026] 실시예들 중에서, 최신 빈발 패턴 관리 서버는 트랜잭션이 특정 배치 크기만큼 수집되었는지 여부를 스캔하는 스캔부, 상기 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우에 삽입하는 슬라이딩 윈도우 삽입부, 상기 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신하는 트리 자료 구조 갱신부 및 상기 트리 자료 구조에 포함된 아이템 각각에 대한 지지도를 기초로 상기 지지도가 내림차순으로 정렬되도록 상기 트리 자료 구조를 재구축하는 트리 자료 구조 재구축부를 포함한다.
- [0027] 실시예들 중에서, 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 방법에 관한 컴퓨터 프로그램을 기록한 기록매체는 트랜잭션이 특정 배치 크기만큼 수집되었는지 여부를 스캔하는 기능, 상기 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우에 삽입하는 기능, 상기 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조

를 갱신하는 기능 및 상기 트리 자료 구조에 포함된 아이템 각각에 대한 지지도를 기초로 상기 지지도가 내림차순으로 정렬되도록 상기 트리 자료 구조를 재구축하는 기능을 포함한다.

발명의 효과

- [0028] 본 발명의 일 실시예에 따른 최신 빈발 패턴 관리 방법은 수집된 트랜잭션을 슬라이딩 윈도우에 삽입하여 최신의 데이터만을 유지할 수 있다.
- [0029] 본 발명의 일 실시예에 따른 최신 빈발 패턴 관리 방법은 생성된 트리 자료 구조를 아이템의 지지도 내림차순으로 정렬되도록 재구축하여 최신의 데이터가 트리 자료 구조를 통하여 효율적으로 저장되고 표현될 수 있다.
- [0030] 본 발명의 일 실시예에 따른 최신 빈발 패턴 관리 방법은 패턴의 최대성을 고려하여 대표 패턴만을 선별적으로 추출할 수 있다.

도면의 간단한 설명

- [0031] 도 1은 본 발명의 일 실시예에 따른 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 시스템에 관한 블록도이다.
- 도 2는 도 1의 최신 빈발 패턴 관리 서버에 관한 블록도이다.
- 도 3은 도 1의 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 시스템에서 수행되는 최신 빈발 패턴 관리 방법에 관한 흐름도이다.
- 도 4는 도 3의 최신 빈발 패턴 관리 방법에 관한 보다 구체적인 흐름도이다.
- 도 5는 최대 빈발 패턴 추출 방법에 관한 보다 구체적인 흐름도이다.
- 도 6은 슬라이딩 윈도우의 흐름에 관한 예시이다.
- 도 7은 트리 자료 구조 및 트리 자료 구조 재구축에 관한 예시이다.

발명을 실시하기 위한 구체적인 내용

- [0032] 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.
- [0033] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0034] "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.
- [0035] 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.
- [0036] 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.
- [0037] 각 단계들에 있어 식별부호(예를 들어, a, b, c 등)는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단

계들의 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.

[0038] 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.

[0039] 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.

[0040] 도 1은 본 발명의 일 실시예에 따른 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 시스템에 관한 블록도이다.

[0041] 도 1을 참조하면, 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 시스템(100)은 사용자 단말(110) 및 최신 빈발 패턴 관리 서버(120)를 포함하고, 여기에서, 사용자 단말(110) 및 최신 빈발 패턴 관리 서버(120)는 네트워크(130)를 통하여 연결된다.

[0042] 사용자 단말(110)은 최신 빈발 패턴 관리 서버(120)와 연결되어 최신 빈발 패턴 관리 방법을 수행하는데 대한 데이터를 송수신할 수 있는 컴퓨팅 장치에 해당할 수 있다. 예를 들어, 사용자 단말(110)은 데스크톱(Desktop), 노트북(Laptop), 스마트 폰(Smart phone) 또는 태블릿 PC(Tablet Personal Computer)에 해당할 수 있다. 사용자 단말(110)은 최신 빈발 패턴 관리 서버(120)에 배치(batch)의 크기, 슬라이딩 윈도우(Sliding Window)의 크기 및 트리 자료 구조를 생성, 갱신 및 재구축하기 위하여 요구되는 특정 기준 값을 설정하는데 필요한 특정 정보를 전송할 수 있고 최신 빈발 패턴의 추출 및 출력을 요청할 수 있으며 최신 빈발 패턴 관리 서버(120)로부터 최신 빈발 패턴에 대한 정보를 수신할 수 있다.

[0043] 최신 빈발 패턴 관리 서버(120)는 사용자 단말(110)과 연결되어 특정 작업을 실행할 수 있는 컴퓨팅 장치에 해당할 수 있다. 일 실시예에서, 최신 빈발 패턴 관리 서버(120)는 사용자 단말(110)로부터 수신한 특정 정보를 기초로 최신 빈발 패턴을 트리 자료 구조로 유지 및 관리할 수 있고 트리 자료 구조에서 빈발 패턴을 추출하여 사용자 단말(110)에 제공할 수 있다.

[0044] 도 2는 도 1의 최신 빈발 패턴 관리 서버에 관한 블록도이다.

[0045] 도 2를 참조하면, 최신 빈발 패턴 관리 서버(120)는 스캔부(210), 슬라이딩 윈도우 삽입부(220), 트리 자료 구조 갱신부(230), 트리 자료 구조 재구축부(240) 및 제어부(250)를 포함한다.

[0046] 스캔부(210)는 트랜잭션이 특정 배치(Batch) 크기만큼 수집되었는지 여부를 스캔한다. 일 실시예에서, 사용자는 특정 배치의 크기를 미리 설정할 수 있고, 예를 들어, 트랜잭션 2개를 하나의 배치 크기로 설정할 수 있다. 일 실시예에서, 스캔부(210)는 특정 배치 크기만큼 트랜잭션이 수집되지 않은 경우에는 수집될 때까지 대기하고 주기적으로 스캔하여 트랜잭션이 수집되었는지 여부를 확인할 수 있다.

[0047] 슬라이딩 윈도우 삽입부(220)는 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우(Sliding Window)에 삽입한다. 여기에서, 슬라이딩 윈도우는 적어도 하나의 배치로 구성될 수 있으며 시간이 지남에 따라 데이터를 점진적으로 축적한다. 일 실시예에서, 슬라이딩 윈도우에 포함되는 배치의 크기는 사용자에게 의하여 설정될 수 있다.

[0048] 일 실시예에서, 슬라이딩 윈도우는 적어도 하나의 배치를 포함할 수 있고 적어도 하나의 배치는 적어도 하나의 트랜잭션을 포함할 수 있으며 적어도 하나의 트랜잭션은 적어도 하나의 아이템을 포함할 수 있다.

[0049] 예를 들어, 도 6을 참조하면, 데이터의 스트림 플로우(Stream-flow)(610)에서 트랜잭션(620)은 한번에 발생하는 데이터의 단위로서 적어도 하나의 아이템을 포함할 수 있고 트랜잭션(620) 각각은 트랜잭션 고유의 식별자에 해

당하는 TID를 가진다. 또한, 트랜잭션(620)이 축적됨에 따라 아이템에 대한 가중치 정보(630)가 갱신될 수 있다. 배치(650)는 적어도 하나의 트랜잭션(620)으로 구성될 수 있고 하나의 배치(650)를 구성하는 트랜잭션(620)의 수는 사용자에게 의하여 설정될 수 있다. 슬라이딩 윈도우(640)는 적어도 하나의 배치(650)로 구성될 수 있고, 하나의 슬라이딩 윈도우(640)를 구성하는 배치(650)의 수는 사용자에게 의하여 설정될 수 있다.

[0050] 일 실시예에서, 슬라이딩 윈도우(640)는 배치(650) 단위로 이동할 수 있다. 예를 들어, 도 6을 참조하면, 하나의 슬라이딩 윈도우(640)는 두개의 배치(650)로 구성되고 두개의 배치(650) 각각은 두개의 트랜잭션(620)으로 구성되어 있으므로 하나의 슬라이딩 윈도우(640)는 4개의 트랜잭션(TID: 010, 020, 030, 040)을 포함할 수 있고, 슬라이딩 윈도우(640)에서 오래된 배치가 삭제되는 경우에는 2개의 트랜잭션(TID: 010, 020)이 삭제되고 새로운 2개의 트랜잭션(TID: 050, 060)이 슬라이딩 윈도우(640)에 삽입될 수 있다.

[0051] 일 실시예에서, 슬라이딩 윈도우 삽입부(220)는 수집된 트랜잭션을 슬라이딩 윈도우에 삽입할 수 있는지 여부를 확인할 수 있다. 일 실시예에서, 수집된 트랜잭션을 슬라이딩 윈도우에 삽입할 수 없는 경우에는 슬라이딩 윈도우 삽입부(220)는 슬라이딩 윈도우에 삽입되어 있는 적어도 하나의 배치 중에서 가장 오래된 배치를 삭제하고 수집된 트랜잭션을 삽입할 수 있다. 보다 구체적으로, 슬라이딩 윈도우 삽입부(220)는 수집된 트랜잭션을 슬라이딩 윈도우에 삽입할 수 없는 경우에는 트리 자료 구조에 포함되어 있는 트리의 단말 노드의 정보를 탐색하여 가장 오래된 배치를 확인하고 가장 오래된 배치에 포함된 트랜잭션을 삭제한 후 새로운 트랜잭션을 삽입할 수 있고, 새로운 트랜잭션의 삽입 후에는 단말 노드의 정보를 갱신할 수 있다.

[0052] 트리 자료 구조 갱신부(230)는 슬라이딩 윈도우에 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신한다. 일 실시예에서, 트리 자료 구조는 슬라이딩 윈도우에 포함되어 있는 아이템 각각에 대한 아이템 이름, 가중치, 지지도 및 노드 링크를 저장 및 관리하는 테이블 및 아이템이 실제로 저장되는 트리를 포함할 수 있다. 여기에서, 트리는 아이템 이름 및 지지도를 포함하는 노드 및 아이템 이름, 지지도 및 아이템이 포함되어 있던 배치에 대한 부분 지지도 정보를 포함하는 단말 노드를 포함할 수 있다.

[0053] 일 실시예에서, 트리 자료 구조 갱신부(230)는 기 구축되어 있는 트리 자료 구조가 있는지 여부를 확인할 수 있다. 일 실시예에서, 트리 자료 구조 갱신부(230)는 기 구축되어 있는 트리 자료 구조가 있는 경우에는 트랜잭션에 포함된 아이템을 트리 자료 구조에 삽입할 수 있고, 기 구축되어 있는 트리 자료 구조가 없는 경우에는 트랜잭션을 기초로 새로운 트리 자료 구조를 생성할 수 있다. 예를 들어, 갱신되거나 생성된 트리 자료 구조는 도 7a와 같이 구성될 수 있다. 도 7a를 참조하면, 아이템들 각각에 대한 가중치, 지지도 및 노드 링크가 테이블(710)로서 저장 및 관리되고 아이템들 각각은 트리(710)에 노드(730) 및 단말 노드(740)로 저장될 수 있다.

[0054] 트리 자료 구조 재구축부(240)는 트리 자료 구조에 포함된 아이템 각각에 대한 지지도를 기초로 지지도가 내림차순으로 정렬되도록 트리 자료 구조를 재구축한다. 일 실시예에서, 트리 자료 구조 재구축부(240)는 트리 자료 구조의 단말 노드로부터 루트 노드까지의 경로에 포함되어 있는 아이템이 지지도를 기초로 내림차순으로 정렬되어 있는지 여부를 확인할 수 있다.

[0055] 일 실시예에서, 트리 자료 구조 재구축부(240)는 아이템이 지지도를 기초로 내림차순으로 정렬되어 있지 않은 경우에는 해당 경로를 트리 자료 구조로부터 분리하고 분리한 경로에 포함되어 있는 아이템을 지지도를 기초로 내림차순으로 정렬한 다음 트리 자료 구조에 삽입할 수 있다.

[0056] 일 실시예에서, 트리 자료 구조 재구축부(240)는 트리에 포함되어 있는 모든 단말 노드로부터 루트 노드까지의 경로에 대하여 아이템이 지지도를 기초로 내림차순으로 정렬되어 있는지 여부를 확인할 수 있고 지지도를 기초로 내림차순으로 정렬되어 있지 않은 경로에 대하여는 재구축을 수행하고 노드 링크를 재연결할 수 있다.

[0057] 일 실시예에서, 트리 자료 구조 재구축부(240)는 트리의 재구축 과정에서 지지도가 0이 된 아이템이 있는 경우에는 해당 아이템에 대한 노드를 삭제할 수 있다. 트리 자료 구조 재구축부(240)를 통하여 재구축이 완료된 트리는 모든 경로 각각에 대한 아이템이 지지도를 기초로 내림차순으로 정렬되어 보다 효율적으로 저장하고 표현할 수 있다.

[0058] 예를 들어, 도 7a 및 도 7b를 참조하면, 도 7a는 트리 자료 구조 갱신부(230)를 통하여 갱신되거나 생성된 트리 자료 구조에 해당하고 도 7b는 도 7a의 트리 자료 구조를 기초로 트리 자료 구조 재구축부(240)를 통하여 재구축된 트리 자료 구조에 해당한다. 도 7b를 참조하면, 재구축이 완료된 트리 자료 구조는 아이템의 지지도에 해당하는 Count를 기초로 테이블(750) 및 트리(760)가 정렬될 수 있다.

[0059] 일 실시예에서, 최신 빈발 패턴 관리 서버(120)는 빈발 패턴 추출부를 더 포함할 수 있고, 여기에서, 빈발 패턴 추출부는 사용자로부터 요청이 있는 경우에 트리 자료 구조를 기초로 가중화 최대 빈발 패턴을 추출할 수 있다.

일 실시예에서, 빈발 패턴 추출부는 트리 자료 구조를 기초로 임계값을 만족시키는 잠재적 가중화 지지도를 가지는 아이টে으로 구성된 조건적 트리 자료 구조를 재귀적으로 생성할 수 있다. 여기에서, 잠재적 가중화 지지도는 각 아이테의 지지도와 현재 슬라이딩 윈도우를 구성하고 있는 모든 아이테의 가중치 값 중 가장 큰 값을 곱한 값에 해당하고, 잠재적 가중화 지지도를 이용하는 이유는 각 아이테의 지지도와 해당 아이테의 실제 가중치를 곱한 가중화 지지도는 안티모노톤 속성(Anti-monotone property)을 만족시키지 못하기 때문이다.

[0060] 일 실시예에서, 빈발 패턴 추출부는 잠재적 가중화 지지도를 기초로 특정 임계값 이상의 값을 가지는 아이테에 대하여 지지도를 기초로 내림차순으로 정렬하고 트리를 재탐색하면서 현재의 트리 자료 구조 중 일부 아이테를 포함하는 조건적 트리 자료구조를 생성할 수 있다. 여기에서, 조건적 트리 자료 구조는 사용자에게 의하여 기 설정된 기준에 따라 생성될 수 있고, 분할 정복(Divide and Conquer) 방식을 이용하여 재귀적으로 생성될 수 있다.

[0061] 일 실시예에서, 빈발 패턴 추출부는 생성된 조건적 트리 자료 구조가 단일 경로로 구성되는 경우에는 단일 경로에 포함되는 아이테를 기초로 후보 빈발 패턴을 생성하고 후보 빈발 패턴의 실제 가중화 지지도가 임계값을 만족하는 경우에는 후보 빈발 패턴을 가중화 최대 빈발 패턴으로 추출할 수 있다. 보다 구체적으로, 빈발 패턴 추출부는 후보 빈발 패턴의 최대성을 판별하고 패턴을 구성하는 아이테의 가중치 평균과 지지도를 곱한 실제 가중화 지지도를 산출하고, 실제 가중화 지지도가 임계값 이상의 값을 가지는 경우에는 해당 패턴을 가중화 최대 빈발 패턴으로 추출할 수 있다.

[0062] 일 실시예에서, 실제 가중화 지지도가 임계값 이상의 값을 갖지 않는 경우에는 빈발 패턴 추출부는 후보 빈발 패턴의 모든 요소에 대한 하위 집합을 추출하여 각 하위집합에 대한 유효성을 검증하여 유효한 가중화 최대 빈발 패턴을 선별적으로 추출할 수 있다. 일 실시예에서, 추출이 완료된 가중화 최대 빈발 패턴은 사용자에게 반환될 수 있다.

[0063] 제어부(250)는 스캔부(210), 슬라이딩 윈도우 삽입부(220), 트리 자료 구조 갱신부(230) 및 트리 자료 구조 재구축부(240)의 동작 및 데이터의 흐름을 제어한다.

[0064] 도 3은 도 1의 슬라이딩 윈도우 기반의 최신 빈발 패턴 관리 시스템에서 수행되는 최신 빈발 패턴 관리 방법에 관한 흐름도이다.

[0065] 스캔부(210)는 트랜잭션이 수집되었는지 여부를 스캔한다(단계 S301). 일 실시예에서, 스캔부(210)는 트랜잭션이 사용자에게 의하여 기 설정된 특정 배치 크기만큼 수집되었는지 여부를 스캔하고, 트랜잭션이 수집되지 않은 경우에는 수집될때까지 대기할 수 있다.

[0066] 슬라이딩 윈도우 삽입부(220)는 특정 배치 크기만큼 수집된 트랜잭션을 슬라이딩 윈도우에 삽입한다(단계 S302). 일 실시예에서, 슬라이딩 윈도우 삽입부(220)는 슬라이딩 윈도우에 트랜잭션을 삽입할 수 있는지 여부를 확인할 수 있다. 슬라이딩 윈도우 삽입부(220)는 슬라이딩 윈도우가 이미 가득 찬 상태에 대하여 트랜잭션을 삽입할 수 없는 경우에는, 슬라이딩 윈도우에 포함되어 있는 배치 중에서 가장 오래된 배치를 삭제하고 새로운 트랜잭션을 삽입할 수 있다.

[0067] 트리 자료 구조 갱신부(230) 슬라이딩 윈도우 삽입부(220)를 통하여 삽입된 트랜잭션을 기초로 트리 자료 구조를 갱신한다(단계 S303). 일 실시예에서, 기 구축된 트리 자료 구조가 있는 경우에는 해당 트리 자료 구조에 새로운 트랜잭션에 포함되어 있는 아이테를 삽입하고, 기 구축된 트리 자료 구조가 없는 경우에는 해당 슬라이딩 윈도우에 포함되어 있는 아이테를 기초로 새로운 트리 자료 구조를 생성할 수 있다.

[0068] 트리 자료 구조 재구축부(240) 트리 자료 구조 갱신부(303)를 통하여 갱신 되거나 또는 생성된 트리 자료 구조를 기초로 트리 자료 구조를 재구축한다(단계 S304). 보다 구체적으로, 트리 자료 구조 재구축부(240)는 트리 자료 구조에 포함되어 있는 아이테이 지지도를 기초로 내림차순으로 정렬되도록 아이테를 정렬할 수 있다.

[0069] 도 4는 도 3의 최신 빈발 패턴 관리 방법에 관한 보다 구체적인 흐름도이다.

[0070] 스캔부(210)는 트랜잭션의 수집 여부를 스캔하고(단계 S401), 트랜잭션이 특정 배치 크기만큼 수집되었는지 여부를 확인한다(단계 S402). 일 실시예에서, 트랜잭션이 특정 배치 크기만큼 수집되지 않은 경우에는 단계 S401로 돌아간다.

- [0071] 일 실시예에서, 트랜잭션이 특정 배치 크기만큼 수집된 경우에는 슬라이딩 윈도우 삽입부(220)는 트랜잭션을 슬라이딩 윈도우에 삽입할 수 있는지 여부를 확인한다(단계 S403). 일 실시예에서, 슬라이딩 윈도우에 트랜잭션을 삽입할 수 있는 경우에는 슬라이딩 윈도우 삽입부(220)는 슬라이딩 윈도우에 트랜잭션을 삽입하고(단계 S404), 슬라이딩 윈도우에 트랜잭션을 삽입할 수 없는 경우에는 슬라이딩 윈도우 삽입부(220)는 슬라이딩 윈도우에서 가장 오래된 배치를 삭제하고(단계 S405) 트랜잭션을 삽입한다(단계 S404).
- [0072] 트리 자료 구조 갱신부(230)는 슬라이딩 윈도우에 포함되어 있는 아이템을 기초로 트리 자료 구조를 갱신하기 위하여 기 구축된 트리가 있는지 여부를 확인한다(단계 S406). 일 실시예에서, 기 구축된 트리가 있는 경우에는 트리 자료 구조 갱신부(230)는 해당 트리를 갱신하고(단계 S407), 기 구축된 트리가 없는 경우에는 슬라이딩 윈도우에 포함되어 있는 아이템을 기초로 새로운 트리를 생성한다(단계 S408).
- [0073] 트리 자료 구조 재구축부(240)는 트리 자료 구조 갱신부(230)를 통하여 갱신된 트리 자료 구조를 기초로 아이템이 지도도에 대한 내림차순으로 정렬되어 있는지 여부를 확인한다(단계 S409). 일 실시예에서, 내림차순으로 정렬되어 있지 않은 경우에는 트리 자료 구조 재구축부(240)는 아이템을 지도도에 내림차순으로 정렬하여 트리 자료 구조를 재구축한다(단계 S411).
- [0074] 빈발 패턴 추출부는 트리 자료 구조 재구축부(240)를 통하여 재구축이 완료된 트리 자료 구조를 기초로 사용로부터 최대 빈발 패턴 추출 요청이 있는 경우에는 최대 빈발 패턴을 추출하여 사용자에게 반환한다. 최대 빈발 패턴 추출 방법은, 이하, 도 5를 참조하여 보다 상세하게 설명한다.
- [0075] 도 5는 최대 빈발 패턴 추출 방법에 관한 보다 구체적인 흐름도이다.
- [0076] 빈발 패턴 추출부는 트리 자료 구조에서 프리픽스를 선정한다(단계 S501). 일 실시예에서, 빈발 패턴 추출부의 트리 자료 구조의 테이블에 포함되어 있는 아이템 각각을 프리픽스로 선정할 수 있다. 여기에서, 프리픽스가 선정되는 순서는 테이블의 가장 아래에 존재하는 아이템을 시작으로 상향식 방식을 따르고, 이하의 과정들은 각각의 프리픽스에 대하여 수행된다.
- [0077] 빈발 패턴 추출부는 선정된 프리픽스를 기초로 해당 프리픽스와 노드 링크로 연결된 아이템에 대한 잠재적 가중화 지도도를 산출하여 내림차순으로 정렬한다(단계 S502). 보다 구체적으로, 빈발 패턴 추출부는 프리픽스와 연관된 모든 아이템의 잠재적 가중화 지도도를 산출하고 산출된 잠재적 가중화 지도도 중 사용자에게 의하여 설정된 임계값 이상의 값을 가지는 아이템에 대하여 지도도를 기초로 내림차순 정렬 순서를 계산한다. 여기에서, 내림차순 정렬 순서는 현재 탐색된 아이템의 특징에 따라 이정의 지도도를 기준으로 한 내림차순 정렬 순서와 달라질 수 있다.
- [0078] 빈발 패턴 추출부는 임계값 이상의 값을 가지는 아이템을 지도도 내림차순으로 정렬한 다음 트리를 재탐색하면서 현재의 트리 자료 구조 중 일부 아이템을 포함하는 조건적 트리 자료 구조를 생성한다(단계 S503).
- [0079] 빈발 패턴 추출부는 생성된 조건적 트리 자료 구조가 단일 경로에 해당하는지 여부를 확인한다(단계 S504). 일 실시예에서, 조건적 트리 자료 구조가 단일 경로에 해당하지 않는 경우에는 빈발 패턴 추출부는 조건적 트리에서 프리픽스를 선정하고 조건적 트리를 재귀적으로 생성한다(단계 S506). 보다 구체적으로, 빈발 패턴 추출부는 조건적 트리에 대한 테이블에 포함되어 있는 아이템 각각을 기존의 프리픽스에 추가시켜 확장시키고 확장된 프리픽스를 기초로 조건적 트리를 재귀적으로 생성한 다음 단계 S504로 돌아가 조건적 트리가 단일 경로인지를 다시 확인할 수 있다.
- [0080] 일 실시예에서, 조건적 트리 자료 구조가 단일 경로에 해당하는 경우에는 빈발 패턴 추출부는 최대 빈발 후보 패턴을 생성한다(단계 S505). 보다 구체적으로, 빈발 패턴 추출부는 현재의 프리픽스와 해당 트리를 구성하는 아이템들을 모두 조합하여 최대 빈발 후보 패턴을 생성할 수 있다.
- [0081] 빈발 패턴 추출부는 최대 빈발 후보 패턴을 기초로 가중화 최대 빈발 패턴을 추출한다(단계 S507). 보다 구체적으로, 빈발 패턴 추출부는 최대 빈발 후보 패턴에 대한 최대성 및 실제 가중화 지도도 값을 기초로 가중화 최대 빈발 패턴을 추출할 수 있다.
- [0082] 일 실시예에서, 모든 프리픽스에 대하여 전 단계를 수행하여 가중화 최대 빈발 패턴 추출이 완료되면 해당 최대 빈발 패턴은 사용자에게 제공될 수 있다.

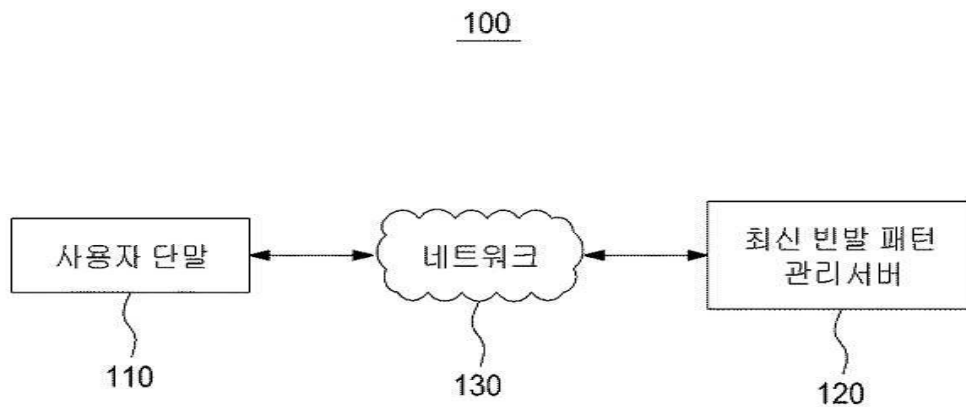
[0083] 상기에서는 본 출원의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 출원을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

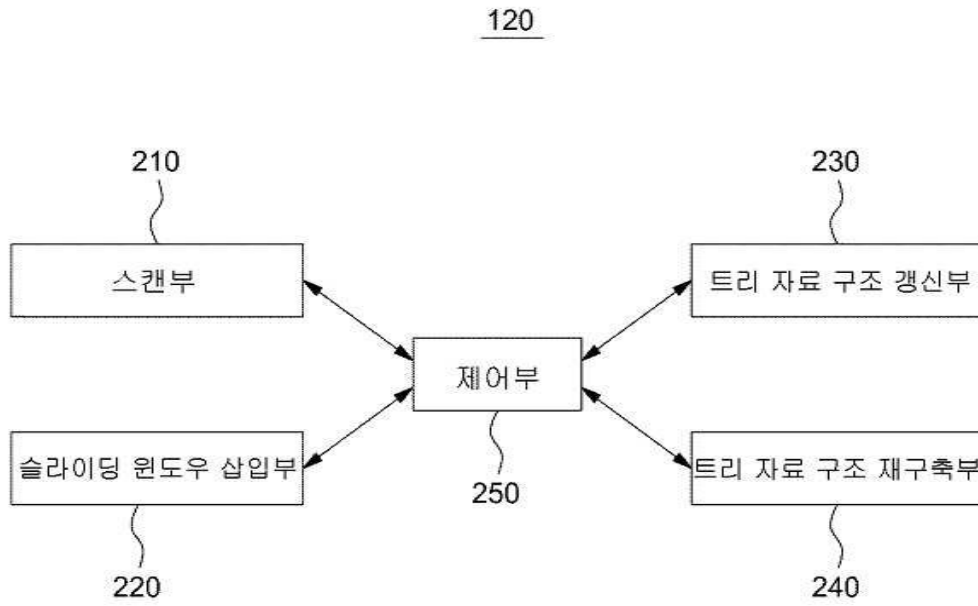
[0084] 100: 최신 빈발 패턴 관리 시스템
 110: 사용자 단말
 120: 최신 빈발 패턴 관리 서버
 130: 네트워크
 210: 스캔부
 220: 슬라이딩 윈도우 삼입부
 230: 트리 자료 구조 갱신부
 240: 트리 자료 구조 재구축부

도면

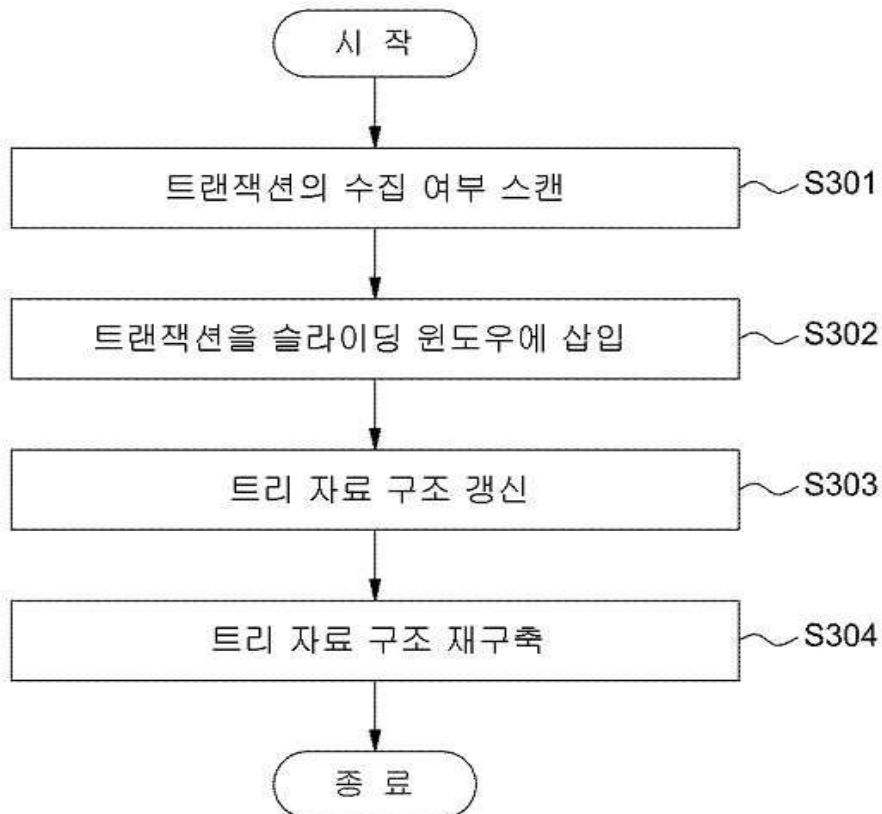
도면1



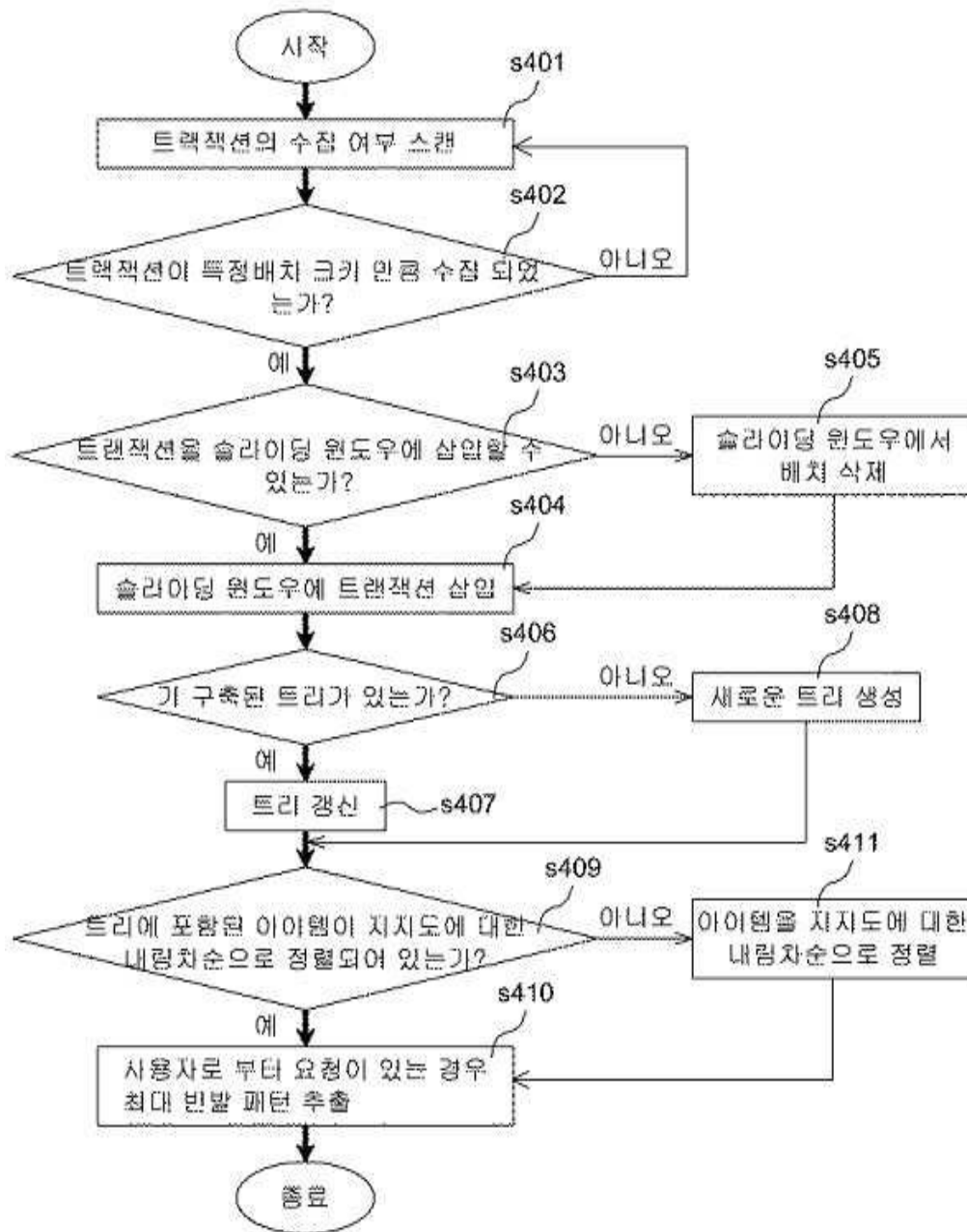
도면2



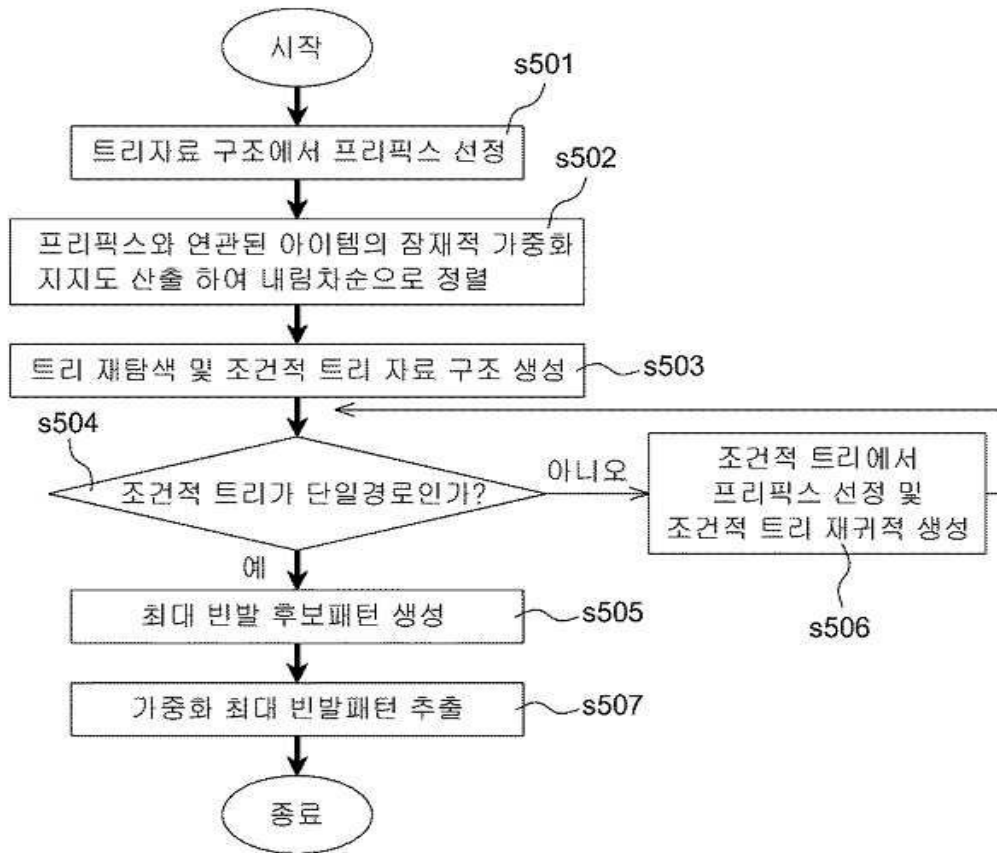
도면3



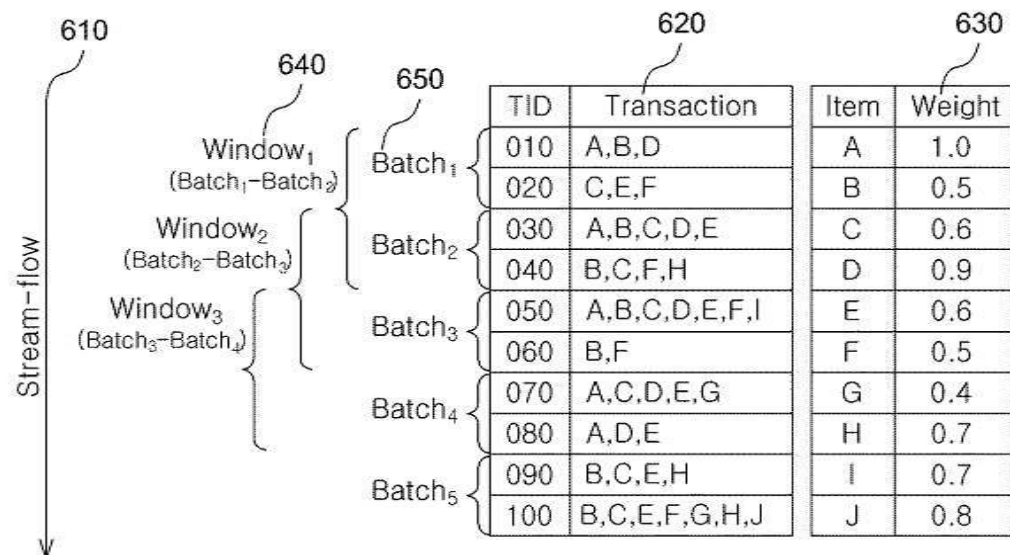
도면4



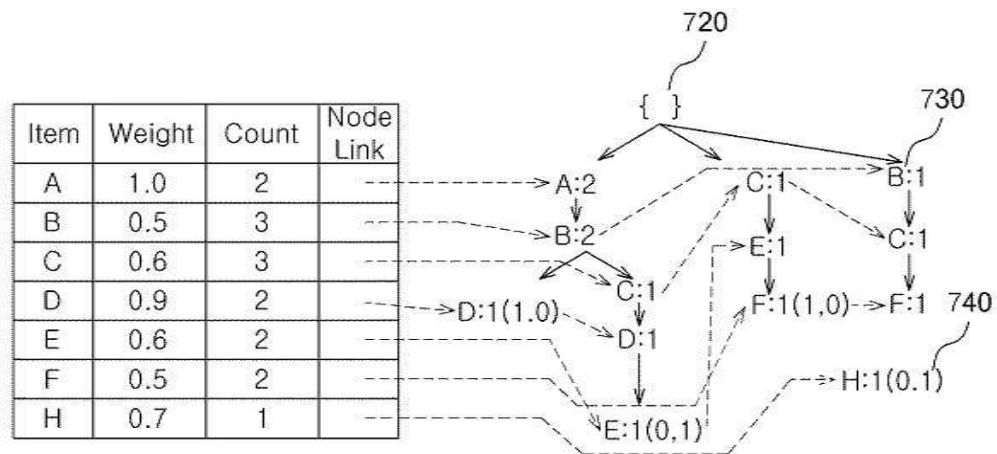
도면5



도면6



도면7a



도면7b

