



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0066219
(43) 공개일자 2016년06월10일

(51) 국제특허분류(Int. Cl.)

G06F 17/30 (2006.01)

(21) 출원번호 10-2014-0170280

(22) 출원일자 2014년12월02일

심사청구일자 2014년12월02일

(71) 출원인

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 14 항

(54) 발명의 명칭 스트림 환경의 감쇄 윈도우 및 근접 프루닝 기반 최신 빈발 패턴 유지 방법, 이를 수행하는 최신 빈발 패턴 유지 서버 및 이를 저장하는 기록매체

(57) 요약

최신 빈발 패턴 유지 방법은 제1 배치의 크기를 기초로 트랜잭션을 수집하는 단계, 상기 트랜잭션이 상기 제1 배치의 크기만큼 수집된 경우에는 적어도 하나의 제1 빈발 항목을 마이닝하고 상기 적어도 하나의 제1 빈발 항목 각각을 노드로 하여 트리를 생성하는 단계, 제2 배치의 크기를 기초로 트랜잭션을 수집하는 단계 및 상기 트랜잭션이 상기 제2 배치의 크기만큼 수집된 경우에는 적어도 하나의 제2 빈발 항목을 마이닝하고 상기 적어도 하나의 제2 빈발 항목 각각을 노드로 하여 상기 트리에 추가하는 단계를 포함한다.

대표도 - 도1

100



(72) 발명자

이강인

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

김동규

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

이 발명을 지원한 국가연구개발사업

과제고유번호 NIPA-2014-H0502-14-3008

부처명 미래창조과학부

연구관리전문기관 정보통신산업진흥원

연구사업명 IT/SW 창의연구과정(대학주도형 기술개발)

연구과제명 비정형 빅데이터 분석을 통한 스트림 근접 패턴 마이닝 기반의 실시간 이슈 검색 기술 개

발

기 여 율 1/1

주관기관 세종대학교 산학협력단

연구기간 2014.06.01 ~ 2015.05.31

명세서

청구범위

청구항 1

- (a) 제1 배치(batch)의 크기를 기초로 트랜잭션을 수집하는 단계;
- (b) 상기 트랜잭션이 상기 제1 배치의 크기만큼 수집된 경우에는 적어도 하나의 제1 빈발 항목을 마이닝하고 상기 적어도 하나의 제1 빈발 항목 각각을 노드로 하여 트리를 생성하는 단계;
- (c) 제2 배치의 크기를 기초로 트랜잭션을 수집하는 단계; 및
- (d) 상기 트랜잭션이 상기 제2 배치의 크기만큼 수집된 경우에는 적어도 하나의 제2 빈발 항목을 마이닝하고 상기 적어도 하나의 제2 빈발 항목 각각을 노드로 하여 상기 트리에 추가하는 단계를 포함하는 최신 빈발 패턴 유지 방법.

청구항 2

제1항에 있어서, 상기 (d) 단계는

상기 제2 빈발 항목 중에서 상기 트리에 저장되어 있는 제1 빈발 항목과 동일한 빈발 항목이 있는 경우에는 해당 제1 빈발 항목이 포함되어 있는 노드의 빈도수를 감소 시킨 후 증가시키는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 3

제2항에 있어서, 상기 (d) 단계는

상기 노드의 배치 정보를 갱신하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 4

제1항에 있어서, 상기 (d) 단계는

상기 트리에 저장되어 있는 제1 빈발 항목 중에서 상기 제2 빈발 항목과 중복되지 않는 빈발 항목이 있는 경우에는 해당 제1 빈발 항목이 저장되어 있는 노드의 빈도수를 감소시키는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 5

제1항에 있어서,

(e) 상기 트리의 노드를 순차적으로 탐색하면서 해당 노드의 빈도수가 현재까지 처리한 트랜잭션의 개수 및 사용자 단말에 의하여 기 설정된 빈도수 감소 팩터, 최소 빈도수 임계치 및 신뢰율을 기초로 산출된 근접 최소 빈도수보다 작은 경우에는 해당 노드의 빈도수를 0으로 하거나 또는 해당 노드를 상기 트리에서 삭제하는 단계를 더 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 6

제5항에 있어서, 상기 (e) 단계는

상기 해당 노드가 상기 트리의 단말노드에 해당하는 경우에는 상기 해당 노드를 삭제하고 상기 삭제된 노드로부터 루트 노드까지 탐색하면서 빈도수가 0에 해당하는 노드를 삭제하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 7

제1항에 있어서,

(f) 사용자로부터 최신 빈발 항목에 대한 출력을 요청 받는 경우에는 상기 트리에 저장되어 있는 빈발 항목을 탐색하여 제공하는 단계를 더 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 8

제7항에 있어서, 상기 (f) 단계는

상기 트리의 루트 노드로부터 순차적으로 각각의 노드를 탐색하면서 해당 노드의 빈도수가 현재까지 처리한 트랜잭션의 개수 및 사용자 단말에 의하여 기 설정된 빈도수 감소 팩터 및 최소 빈도수 임계치를 기초로 산출된 최소 빈도수 보다 큰 경우에는 상기 해당 노드로부터 상기 루트 노드까지의 경로에 포함되는 노드의 빈발 항목을 추출하는 단계를 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 9

제1항에 있어서, 상기 트리는

빈발 항목 및 빈발 항목이 저장되어 있는 노드를 연결하기 위한 링크를 포함하는 헤더리스트를 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 10

제1항에 있어서, 상기 노드는

빈발 항목 및 해당 빈발 항목에 대한 빈도수 및 배치 정보를 포함하고 해당 노드의 부모 노드 및 자식 노드에 접근할 수 있는 포인터를 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 방법.

청구항 11

제1 배치의 크기 또는 제2 배치의 크기를 기초로 트랜잭션을 수집하는 트랜잭션 수집부;

상기 트랜잭션이 상기 제1 배치의 크기만큼 수집된 경우에는 적어도 하나의 제1 빈발 항목을 마이닝하고 상기 적어도 하나의 제1 빈발 항목 각각을 노드로 하여 트리를 생성하는 트리 생성부; 및

상기 트랜잭션이 상기 제2 배치의 크기만큼 수집된 경우에는 적어도 하나의 제2 빈발 항목을 마이닝하고 상기 적어도 하나의 제2 빈발 항목 각각을 노드로 하여 상기 트리에 추가하는 빈발 항목 추가부를 포함하는 최신 빈발 패턴 유지 서버.

청구항 12

제11항에 있어서,

상기 트리의 노드를 순차적으로 탐색하면서 해당 노드의 빈도수가 현재까지 처리한 트랜잭션의 개수 및 사용자 단말에 의하여 기 설정된 빈도수 감소 팩터, 최소 빈도수 임계치 및 신뢰율을 기초로 산출된 근접 최소 빈도수

보다 작은 경우에는 해당 노드의 빈도수를 0으로 하거나 또는 해당 노드를 상기 트리에서 삭제하는 빈발 항목 관리부를 더 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 서버.

청구항 13

제11항에 있어서,

사용자로부터 최신 빈발 항목에 대한 출력을 요청 받는 경우에는 상기 트리에 저장되어 있는 빈발 항목을 탐색하여 출력하는 빈발 항목 제공부를 더 포함하는 것을 특징으로 하는 최신 빈발 패턴 유지 서버.

청구항 14

제1 배치의 크기를 기초로 트랜잭션을 수집하는 기능;

상기 트랜잭션이 상기 제1 배치의 크기만큼 수집된 경우에는 적어도 하나의 제1 빈발 항목을 마이닝하고 상기 적어도 하나의 제1 빈발 항목 각각을 노드로 하여 트리를 생성하는 기능;

제2 배치의 크기를 기초로 트랜잭션을 수집하는 기능; 및

상기 트랜잭션이 상기 제2 배치의 크기만큼 수집된 경우에는 적어도 하나의 제2 빈발 항목을 마이닝하고 상기 적어도 하나의 제2 빈발 항목 각각을 노드로 하여 상기 트리에 추가하는 기능을 포함하는 최신 빈발 패턴 유지 방법에 관한 컴퓨터 프로그램을 기록한 기록매체.

발명의 설명

기술 분야

[0001] 본 발명은 최신 빈발 패턴 유지 기술에 관한 것으로서, 특정 배치 크기를 기초로 트랜잭션을 수집하고 빈발 항목을 마이닝 하여 트리에 삽입하고 비 빈발 항목에 대하여는 해당 노드를 삭제하는 것을 통하여 트리에 빈발 항목만을 유지할 수 있는 최신 빈발 패턴 유지 방법, 이를 수행하는 최신 빈발 패턴 유지 서버 및 이를 저장하는 기록매체에 관한 것이다.

배경 기술

[0002] 일반적으로, 트랜잭션 정보를 트리에 저장하여 유지하며 사용자로부터 빈발 항목 추출에 대한 요청을 받은 경우에, 트랜잭션 정보가 저장되어 있는 트리를 마이닝 하여 빈발 항목을 검출한다. 또한, 새로운 트랜잭션이 도착하면 이를 트리에 반영하며, 빈발 항목 추출에 대한 요청을 받을 때마다 모든 트랜잭션 정보에 대하여 마이닝 연산을 수행한다. 즉, 새로운 트랜잭션이 도착할 때마다 이를 처리하여 트리를 갱신하고, 빈발 항목 추출에 대한 요청을 받을 때마다 트리에 대한 마이닝 연산이 요구 된다.

[0003] 한국등록특허 제10-0899144호는 데이터 스트림으로부터 최대 빈발항목집합을 탐색하는 방법 및 장치에 관한 것으로서, 이미 발생한 트랜잭션들에 포함된 항목집합들 및 항목 집합들의 출현 빈도수가 전위 트리를 통하여 관리되고, 이를 이용하여 노드를 방문하면서 해당하는 항목 집합이 빈발한지 여부를 검사함으로써 최대 빈발 항목 집합을 찾는 기술이다.

[0004] 한국등록특허 제10-1105363호는 룬 트랜잭션 데이터 스트림을 위한 빈발항목집합 탐색 방법에 관한 것으로서, 발생하는 트랜잭션을 분할하여 복수 개의 분할 트랜잭션을 생성하고, 복수개의 분할 트랜잭션 각각을 복수 개의 제1 계층 전위 트리를 이용하여 마이닝하여 빈발 항목 집합을 생성하고, 생성된 빈발 항목 집합을 압축 및 병합하고 제2 계층 전위 트리를 이용하여 마이닝 하는 기술이다.

선행기술문헌

특허문헌

- [0005] (특허문헌 0001) 한국등록특허 제10-0899144호
(특허문헌 0002) 한국등록특허 제10-1105363호

발명의 내용

해결하려는 과제

- [0006] 본 발명의 일 실시예는 하나의 트리를 이용하여 최신 빈발 항목만을 유지 및 관리할 수 있는 최신 빈발 패턴 유지 방법을 제공하고자 한다.
- [0007] 본 발명의 일 실시예는 특정 배치의 크기를 기초로 트랜잭션을 수집한 다음에 빈발 항목을 마이닝할 수 있는 최신 빈발 패턴 유지 방법을 제공하고자 한다.
- [0008] 본 발명의 일 실시예는 빈발 항목을 노드로 하여 트리에 추가하고 비 빈발 항목은 해당 트리에서 삭제할 수 있는 최신 빈발 패턴 유지 방법을 제공하고자 한다.
- [0009] 본 발명의 일 실시예는 트리에 저장되어 있는 빈발 항목을 탐색하여 사용자에게 최신 빈발 항목의 집합을 제공할 수 있는 최신 빈발 패턴 유지 방법을 제공하고자 한다.

과제의 해결 수단

- [0010] 실시예들 중에서, 최신 빈발 패턴 유지 방법은 (a) 제1 배치의 크기를 기초로 트랜잭션을 수집하는 단계, (b) 상기 트랜잭션이 상기 제1 배치의 크기만큼 수집된 경우에는 적어도 하나의 제1 빈발 항목을 마이닝하고 상기 적어도 하나의 제1 빈발 항목 각각을 노드로 하여 트리를 생성하는 단계, (c) 제2 배치의 크기를 기초로 트랜잭션을 수집하는 단계 및 (d) 상기 트랜잭션이 상기 제2 배치의 크기만큼 수집된 경우에는 적어도 하나의 제2 빈발 항목을 마이닝하고 상기 적어도 하나의 제2 빈발 항목 각각을 노드로 하여 상기 트리에 추가하는 단계를 포함한다.
- [0011] 일 실시예에서, 상기 (d) 단계는 상기 제2 빈발 항목 중에서 상기 트리에 저장되어 있는 제1 빈발 항목과 동일한 빈발 항목이 있는 경우에는 해당 제1 빈발 항목이 포함되어 있는 노드의 빈도수를 증가시키는 단계를 포함할 수 있다.
- [0012] 일 실시예에서, 상기 (d) 단계는 상기 노드의 배치 정보를 갱신하는 단계를 포함할 수 있다.
- [0013] 일 실시예에서, 상기 (d) 단계는 상기 트리에 저장되어 있는 제1 빈발 항목 중에서 상기 제2 빈발 항목과 중복되지 않는 빈발 항목이 있는 경우에는 해당 제1 빈발 항목이 저장되어 있는 노드의 빈도수를 감소시키는 단계를 포함할 수 있다.
- [0014] 일 실시예에서, (e) 상기 트리의 노드를 순차적으로 탐색하면서 해당 노드의 빈도수가 현재까지 처리한 트랜잭션의 개수 및 사용자 단말에 의하여 기 설정된 빈도수 감소 팩터, 최소 빈도수 임계치 및 신뢰율을 기초로 산출된 근접 최소 빈도수보다 작은 경우에는 해당 노드의 빈도수를 0으로 하거나 또는 해당 노드를 상기 트리에서 삭제하는 단계를 더 포함할 수 있다.
- [0015] 일 실시예에서, 상기 (e) 단계는 상기 해당 노드가 상기 트리의 단말노드에 해당하는 경우에는 상기 해당 노드를 삭제하고 상기 삭제된 노드로부터 루트 노드까지 탐색하면서 빈도수가 0에 해당하는 노드를 삭제하는 단계를 포함할 수 있다.
- [0016] 일 실시예에서, (f) 사용자로부터 최신 빈발 항목에 대한 출력을 요청 받는 경우에는 상기 트리에 저장되어 있는 빈발 항목을 탐색하여 출력하는 단계를 더 포함할 수 있다.
- [0017] 일 실시예에서, 상기 (f) 단계는 상기 트리의 루트 노드로부터 순차적으로 각각의 노드를 탐색하면서 해당 노드의 빈도수가 현재까지 처리한 트랜잭션의 개수 및 사용자 단말에 의하여 기 설정된 빈도수 감소 팩터 및 최소 빈도수 임계치를 기초로 산출된 최소 빈도수 보다 큰 경우에는 상기 해당 노드로부터 상기 루트 노드까지의 경

로에 포함되는 노드의 빈발 항목을 추출하는 단계를 포함할 수 있다.

- [0018] 일 실시예에서, 상기 트리는 빈발 항목 및 빈발 항목이 저장되어 있는 노드를 연결하기 위한 링크를 포함하는 헤더리스트를 포함할 수 있다.
- [0019] 일 실시예에서, 상기 노드는 빈발 항목 및 해당 빈발 항목에 대한 빈도수 및 배치 정보를 포함하고 해당 노드의 부모 노드 및 자식 노드에 접근할 수 있는 포인터를 포함할 수 있다.
- [0020] 실시예들 중에서, 최신 빈발 패턴 유지 서버는 제1 배치의 크기 또는 제2 배치의 크기를 기초로 트랜잭션을 수집하는 트랜잭션 수집부, 상기 트랜잭션이 상기 제1 배치의 크기만큼 수집된 경우에는 적어도 하나의 제1 빈발 항목을 마이닝하고 상기 적어도 하나의 제1 빈발 항목 각각을 노드로 하여 트리를 생성하는 트리 생성부 및 상기 트랜잭션이 상기 제2 배치의 크기만큼 수집된 경우에는 적어도 하나의 제2 빈발 항목을 마이닝하고 상기 적어도 하나의 제2 빈발 항목 각각을 노드로 하여 상기 트리에 추가하는 빈발 항목 추가부를 포함한다.
- [0021] 일 실시예에서, 상기 트리의 노드를 순차적으로 탐색하면서 해당 노드의 빈도수가 현재까지 처리한 트랜잭션의 개수 및 사용자 단말에 의하여 기 설정된 빈도수 감소 팩터, 최소 빈도수 임계치 및 신뢰율을 기초로 산출된 근접 최소 빈도수보다 작은 경우에는 해당 노드의 빈도수를 0으로 하거나 또는 해당 노드를 상기 트리에서 삭제하는 빈발 항목 관리부를 더 포함할 수 있다.
- [0022] 일 실시예에서, 사용자로부터 최신 빈발 항목에 대한 출력을 요청 받는 경우에는 상기 트리에 저장되어 있는 빈발 항목을 탐색하여 출력하는 빈발 항목 제공부를 더 포함할 수 있다.
- [0023] 실시예들 중에서, 최신 빈발 패턴 유지 방법에 관한 컴퓨터 프로그램을 기록한 기록매체는 제1 배치의 크기를 기초로 트랜잭션을 수집하는 기능, 상기 트랜잭션이 상기 제1 배치의 크기만큼 수집된 경우에는 적어도 하나의 제1 빈발 항목을 마이닝하고 상기 적어도 하나의 제1 빈발 항목 각각을 노드로 하여 트리를 생성하는 기능, 제2 배치의 크기를 기초로 트랜잭션을 수집하는 기능 및 상기 트랜잭션이 상기 제2 배치의 크기만큼 수집된 경우에는 적어도 하나의 제2 빈발 항목을 마이닝하고 상기 적어도 하나의 제2 빈발 항목 각각을 노드로 하여 상기 트리에 추가하는 기능을 포함하는 최신 빈발 패턴 유지 방법에 관한 컴퓨터 프로그램을 기록한다.

발명의 효과

- [0024] 본 발명의 일 실시예에 따른 최신 빈발 패턴 유지 방법은 하나의 트리를 이용하여 최신 빈발 항목만을 유지 및 관리할 수 있다.
- [0025] 본 발명의 일 실시예에 따른 최신 빈발 패턴 유지 방법은 특정 배치의 크기를 기초로 트랜잭션을 수집한 다음에 빈발 항목을 마이닝할 수 있다.
- [0026] 본 발명의 일 실시예에 따른 최신 빈발 패턴 유지 방법은 빈발 항목을 노드로 하여 트리에 추가하고 비 빈발 항목은 해당 트리에서 삭제할 수 있다.
- [0027] 본 발명의 일 실시예에 따른 최신 빈발 패턴 유지 방법은 트리에 저장되어 있는 빈발 항목을 탐색하여 사용자에게 최신 빈발 항목의 집합을 제공할 수 있다.

도면의 간단한 설명

- [0028] 도 1은 본 발명의 일 실시예에 따른 최신 빈발 패턴 유지 시스템에 대한 블록도이다.
- 도 2는 도 1의 최신 빈발 패턴 유지 서버에 대한 블록도이다.
- 도 3은 본 발명의 일 실시예에 따른 최신 빈발 패턴 유지 방법에 대한 흐름도이다.
- 도 4는 빈발 항목을 트리에 추가하는 방법에 대한 보다 구체적인 흐름도이다.
- 도 5는 비 빈발 항목을 트리에서 관리하는 방법에 대한 보다 구체적인 흐름도이다.
- 도 6은 빈발 항목을 유지하는 트리에 대한 예시이다.

발명을 실시하기 위한 구체적인 내용

- [0029] 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.
- [0030] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0031] "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.
- [0032] 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.
- [0033] 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.
- [0034] 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.
- [0035] 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.
- [0036] 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 최신 빈발 패턴 유지 시스템에 대한 블록도이다.
- [0038] 도 1을 참조하면, 최신 빈발 패턴 유지 시스템(100)은 사용자 단말(110), 최신 빈발 패턴 유지 서버(120)를 포함한다. 여기에서, 사용자 단말(110)을 네트워크(130)를 통하여 최신 빈발 패턴 유지 서버(120)와 연결된다.
- [0039] 사용자 단말(110)은 최신 빈발 패턴 유지 서버(120)와 연결되어 최신 빈발 패턴 유지 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 유지 서버(120)에 최신 빈발 항목들로 트리를 유지하기 위하여 요구되는 특정 기준 값을 설정하는데 필요한 특정 정보를 전송할 수 있고 최신 빈발 항목의 출력을 요청할 수 있으며 최신 빈발 패턴 유지 서버(120)로부터 최신 빈발 항목에 대한 정보를 수신할 수 있다.
- [0040] 최신 빈발 패턴 유지 서버(120)는 사용자 단말(110)과 연결되어 특정 작업을 실행할 수 있는 컴퓨팅 장치에 해당할 수 있다. 일 실시예에서, 최신 빈발 패턴 유지 서버(120)는 사용자 단말(110)로부터 수신한 특정 정보를 기초로 최신 빈발 항목을 트리 형태로 유지 및 관리 할 수 있고, 트리에서 빈발 항목을 탐색하여 사용자 단말

(110)에 제공할 수 있다.

- [0041] 도 2는 도 1의 최신 빈발 패턴 유지 서버에 대한 블록도이다.
- [0042] 도 2를 참조하면, 최신 빈발 패턴 유지 서버(120)는 트랜잭션 수집부(210), 트리 생성부(220), 빈발 항목 추가부(230), 빈발 항목 관리부(240) 및 제어부(250)를 포함한다.
- [0043] 트랜잭션 수집부(210)는 특정 배치(batch)의 크기를 기초로 해당 크기만큼 트랜잭션을 수집한다. 일 실시예에서, 트랜잭션은 하나의 정보 문서에 해당할 수 있고, 예를 들어, 트랜잭션은 하나의 웹페이지(Webpage)에 해당할 수 있다. 또한, 배치의 크기는 사용자 단말(110)에 의하여 기 설정될 수 있다.
- [0044] 트리 생성부(220)는 특정 배치 크기만큼 수집된 트랜잭션으로부터 마이닝된 빈발 항목 각각을 노드로 하는 트리를 생성한다. 일 실시예에서, 도 6을 참조하면, 트리는 헤더리스트를 포함할 수 있고, 여기에서, 헤더리스트는 빈발 항목 및 빈발 항목이 저장되어 있는 노드를 연결하기 위한 링크를 포함할 수 있다.
- [0045] 일 실시예에서, 도 6을 참조하면, 트리에 포함되는 노드 각각은 빈발 항목 및 해당 빈발 항목에 대한 빈도수 및 배치 정보를 포함할 수 있고, 해당 노드의 부모 노드 및 자식 노드에 접근할 수 있는 포인터 및 동일한 빈발 항목을 가지는 노드를 연결하기 위한 링크를 포함할 수 있다. 여기에서, 빈도수는 해당 빈발 항목이 특정 배치 내에서 얼마나 자주 나타나는지에 대한 수치이며 배치 정보는 해당 빈발 항목이 마이닝된 배치가 몇번째 배치인지에 대한 정보에 해당한다. 예를 들어, 빈발 항목 A는 제3 배치의 크기를 기초로 수집된 트랜잭션에서 마이닝되고, 제3 배치 내에서 빈발 항목 A가 총 45회 나타나는 경우에는, 빈발 항목 A에 대한 {빈도수, 배치 정보}는 {45, 3}이 될 수 있다.
- [0046] 일 실시예에서, 트리 생성부(220)는 첫번째 배치에 해당하는 제1 배치의 크기를 기초로 수집된 트랜잭션으로부터 마이닝된 제1 빈발 항목을 기초로 하여 트리를 생성할 수 있다. 즉, 트리 생성부(220)는 최신 빈발 패턴 유지 과정에서 트리를 한번만 생성한다. 일 실시예에서, 트리 생성부(220)는 제1 빈발 항목 각각을 노드로 하는 트리를 생성할 수 있고, 각각의 노드는 해당 빈발 항목에 대한 빈도수 및 배치 정보를 포함할 수 있다.
- [0047] 일 실시예에서, 빈발 항목은 트리 이외의 다른 형태의 데이터 구조를 통하여 유지 및 관리 될 수 있고, 예를 들어, 빈발 항목은 리스트 형태의 데이터 구조를 이용하여 유지 및 관리 될 수 있다.
- [0048] 빈발 항목 추가부(230)는 트리 생성부(220)를 통하여 제1 빈발 항목에 대한 트리가 생성된 경우에는 제2 빈발 항목 각각을 노드로 하여 트리 생성부(220)를 통하여 생성된 트리에 추가한다. 여기에서, 제2 빈발 항목은 두번째 배치에 해당하는 제2 배치의 크기를 기초로 수집된 트랜잭션으로부터 마이닝된 빈발 항목에 해당한다.
- [0049] 일 실시예에서, 빈발 항목 추가부(230)는 제2 빈발 항목 중에서 트리에 저장되어 있는 제1 빈발 항목과 동일한 빈발 항목이 있는 경우에는 해당 제1 빈발 항목이 포함되어 있는 노드의 빈도수를 감쇄시킨 후, 제2 빈발 항목의 빈도수만큼 증가시키고 배치 정보를 갱신할 수 있다. 예를 들어, 트리에 저장되어 있는 특정 제1 빈발 항목의 노드에 {빈도수, 배치 정보}가 {45, 1}로 저장되어 있고, 특정 제1 빈발 항목과 동일한 특정 제2 빈발 항목의 제2 배치 내에서의 빈도수가 20 인 경우에는 특정 제1 빈발 항목의 노드의 {빈도수, 배치 정보}에서 빈도수가 {44(=45-1), 1}로 감쇄된 후 {64(=44+20), 2}로 갱신될 수 있다. 여기에서, 배치 정보는 특정 제1 빈발 항목이 나타난 가장 최근의 배치로서 현재 처리되고 있는 배치에 해당할 수 있다.
- [0050] 일 실시예에서, 빈발 항목 추가부(230)는 트리에 저장되어 있는 제1 빈발 항목 중에서 제2 빈발 항목과 중복되지 않는 빈발 항목이 있는 경우에는 해당 제1 빈발 항목이 저장되어 있는 노드의 빈도수를 감소시킬 수 있다. 예를 들어, 특정 제1 빈발 항목에 해당하는 빈발 항목 B가 트리에 저장되어 있고 빈발 항목 B에 대한 {빈도수, 배치 정보}가 {15, 1}에 해당하고, 제2 빈발 항목 중에서 빈발 항목 B와 동일한 빈발 항목이 없는 경우에는 빈발 항목 B에 대한 {빈도수, 배치 정보}를 {14, 1}로 수정할 수 있다.
- [0051] 빈발 항목 관리부(240)는 트리의 노드를 순차적으로 탐색하면서 해당 노드의 빈도수가 근접 최소 빈도수보다 작은 경우에는 해당 노드의 빈도수를 0으로 하거나 또는 해당 노드를 트리에서 삭제한다. 여기에서, 근접 최소 빈도수와 비교되는 노드의 빈도수는 기 저장되어 있는 빈도수에 현재 처리 중인 배치 정보 및 해당 노드에 마지막으로 접근한 배치 정보 사이의 차이만큼 λ 를 곱하여 재 산출된 빈도수에 해당하고, λ 는 사용자 단말(110)로부터 입력 받은 빈도수 감소 팩터에 해당한다.
- [0052] 일 실시예에서, 근접 최소 빈도수는 현재까지 처리한 트랜잭션의 개수 및 사용자 단말(110)에 의하여 기 설정

된 빈도수 감소 팩터, 최소 빈도수 임계치 및 신뢰율을 기초로 산출될 수 있다. 보다 구체적으로, 근접 최소 빈도수는 아래의 [수식 1]을 통하여 산출될 수 있다.

$$\mu = T_x \times \lambda \times \text{minsup} - \sqrt{\frac{2 \ln(\frac{2}{\delta})}{T_x \times \lambda}}$$

[0053] [수식 1]

[0054] μ : 근접 최소 빈도수

[0055] T_x : 현재까지 처리한 트랜잭션 수

[0056] λ : 빈도수 감소 팩터

[0057] minsup: 최소 빈도수 임계치 ($0 < \text{minsup} \leq 1$)

[0058] δ : 신뢰율

[0059] 여기에서, $T_x \times \lambda$ ($= T_x'$)는 전체 트랜잭션의 개수이고, $T_x' \times \text{minsup}$ ($= \xi$)는 최소 빈도수이고, $\sqrt{\frac{2 \ln(\frac{2}{\delta})}{T_x}} (= \varepsilon)$ 는 오차값으로서, 즉, 근접 최소 빈도수는 최소 빈도수에서 오차를 뺀으로써 계산될 수 있다. 또한, 최소 빈도수(minsup)는 체르노프 바운드에 의하여 결정될 수 있다.

[0060] 일 실시예에서, 빈발 항목 관리부(240)는 근접 최소 빈도수보다 작은 빈도수를 가지는 노드가 트리의 단말 노드에 해당하는 경우에는 해당 노드를 삭제하고 삭제된 노드로부터 루트 노드까지 탐색하면서 빈도수가 0에 해당하는 노드를 삭제할 수 있다. 또한, 빈발 항목 관리부(240)는 삭제된 노드로부터 루트 노드까지 탐색하면서 빈도수가 0이 아닌 노드가 나타나면 해당 과정을 중단할 수 있다.

[0061] 일 실시예에서, 빈발 항목 관리부(240)는 트리의 메모리 효율성을 높이기 위하여 체르노프 바운드(chernoff bound)를 응용하여 사용자 단말(110)에 의하여 설정된 신뢰율(δ)을 벗어나지 않는 범위 내에서 트리의 크기를 최대한 줄일 수 있다.

[0062] 일 실시예에서, 최신 빈발 패턴 유지 서버(120)는 빈발 항목 제공부를 더 포함할 수 있고, 빈발 항목 제공부는 사용자 단말(110)로부터 최신 빈발 항목에 대한 출력을 요청 받는 경우에는 트리에 저장되어 있는 빈발 항목을 탐색하여 사용자 단말(110)에 제공할 수 있다.

[0063] 일 실시예에서, 빈발 항목 제공부는 트리의 루트 노드로부터 깊이 우선 탐색 방법을 이용하여 순차적으로 각각의 노드를 탐색하면서 해당 노드의 빈도수가 최소 빈도수 보다 큰 경우에는 해당 노드로부터 루트 노드까지의 경로에 포함되는 노드의 빈발 항목을 추출할 수 있다. 여기에서, 최소 빈도수는 사용자 단말(110)에 의하여 기 설정된 빈도수 감소 팩터(λ) 및 최소 빈도수 임계치(minsup)를 기초로 산출될 수 있고, [수식 1]에서 설명한 최소 빈도수(ξ)와 동일하다.

[0064] 일 실시예에서, 빈발 항목 제공부는 트랜잭션 수집부(210), 트리 생성부(220), 빈발 항목 추가부(230) 또는 빈발 항목 관리부(240)를 통하여 특정 작업이 수행중인 경우에는 해당 작업이 종료될 때까지 대기하고, 해당 작업이 종료된 후에 빈발 항목을 추출할 수 있다.

[0065] 제어부(250)는 최신 빈발 패턴 유지 서버(120)는 트랜잭션 수집부(210), 트리 생성부(220), 빈발 항목 추가부(230) 및 빈발 항목 관리부(240)의 동작 및 데이터의 흐름을 제어한다.

[0066] 도 3은 본 발명의 일 실시예에 따른 최신 빈발 패턴 유지 방법에 대한 흐름도이다.

[0067] 트랜잭션 수집부(210)는 제1 배치의 크기를 기초로 해당 크기만큼 트랜잭션을 수집하고(단계 S301), 트리 생성부(220)는 수집된 트랜잭션에서 제1 빈발 항목을 마이닝하여 트리를 생성한다(단계 S302). 일 실시예에서, 트리 생성부(220)는 제1 빈발 항목 각각을 노드로 하는 트리를 생성할 수 있고, 각각의 노드는 해당 빈발 항목에 대한 빈도수 및 배치 정보를 포함할 수 있다.

- [0068] 트랜잭션 수집부(210)는 제2 배치의 크기를 기초로 해당 크기만큼 트랜잭션을 수집하고(단계 S302), 빈발 항목 추가부(230)는 제2 배치의 크기를 기초로 수집된 트랜잭션에서 제2 빈발 항목을 마이닝 하여 트리에 추가한다(단계 S304). 일 실시예에서, 빈발 항목 추가부(230)는 제2 빈발 항목 중에서 트리에 저장되어 있는 제1 빈발 항목과 동일한 빈발 항목이 있는지 여부를 기초로 새로운 노드를 생성하여 빈발 항목을 트리에 추가하거나 또는 트리에 기 저장되어 있는 제1 빈발 항목의 빈도수 및 배치 정보를 갱신할 수 있다.
- [0069] 도 4는 빈발 항목을 트리에 추가하는 방법에 대한 보다 구체적인 흐름도이다.
- [0070] 빈발 항목 추가부(230)는 특정 배치의 크기만큼 수집된 트랜잭션에서 빈발 항목을 마이닝하고 사용자 단말(110)에 의하여 기 설정된 기준에 따라 빈발 항목을 정렬하고(단계 S401), 정렬된 순서대로 트리에 해당 빈발 항목을 삽입한다.
- [0071] 빈발 항목 추가부(230)는 특정 빈발 항목을 트리에 삽입하기 전에 해당 특정 빈발 항목이 트리에 삽입하려는 정렬된 빈발 항목 중에서 마지막 빈발 항목인지 여부를 확인한다(단계 S402).
- [0072] 일 실시예에서, 특정 빈발 항목이 마지막 빈발 항목에 해당하는 경우에는 빈발 항목 추가부(230)는 이전 빈발 항목이 추가된 노드의 자식 노드에 해당 빈발 항목이 있는지 여부를 확인한다(단계 S403). 일 실시예에서, 자식 노드에 해당 빈발 항목이 있는 경우에는 자식 노드의 빈도수 및 배치 정보를 갱신한다(단계 S404). 여기에서, 빈도수는 기 저장되어 있는 빈도수가 감쇄된 후 해당 빈발 항목의 빈도수를 더한 값으로 갱신될 수 있고, 배치 정보는 해당 빈발 항목에 대한 배치 정보로 갱신될 수 있다. 다른 일 실시예에서, 자식 노드에 해당 빈발 항목이 없는 경우에는 해당 빈발 항목에 대한 노드를 새로 생성하여 이전 빈발 항목이 추가된 노드의 자식 노드로 삽입한다(단계 S405).
- [0073] 일 실시예에서, 특정 빈발 항목이 마지막 빈발 항목에 해당하지 않는 경우에는 빈발 항목 추가부(230)는 이전 빈발 항목이 추가된 노드의 자식 노드에 해당 빈발 항목이 있는지 여부를 확인한다(단계 S406). 여기에서, 특정 빈발 항목이 정렬된 빈발 항목 중에서 처음으로 트리에 추가되는 빈발 항목에 해당하는 경우에는 루트 노드의 자식 노드에 해당 빈발 항목이 있는지 여부를 확인할 수 있다.
- [0074] 일 실시예에서, 자식 노드에 해당 빈발 항목이 있는 경우에는 해당 자식 노드를 다음 빈발 항목에 대하여 현재 노드로 설정하고(단계 S407), 해당 자식 노드의 빈도수 및 배치 정보를 갱신한다. 여기에서, 빈도수 및 배치 정보의 갱신은 단계 S404에서 수행된 것과 동일한 방법으로 갱신 될 수 있다. 보다 구체적으로, 빈발 항목 추가부(230)는 현재 위치한 노드에서 해당 자식 노드로 이동하여, 다음 빈발 항목을 추가할 때 해당 자식 노드를 기준으로 해당 자식 노드의 자식 노드에 빈발 항목이 있는지 여부를 확인한다. 즉, 해당 자식 노드를 다음 빈발 항목에 대하여는 현재 노드로 설정한다. 다른 일 실시예에서, 자식 노드에 해당 빈발 항목이 없는 경우에는 단계 S405와 동일한 단계 S408을 수행한다.
- [0075] 도 5는 비 빈발 항목을 트리에서 관리하는 방법에 대한 보다 구체적인 흐름도이다.
- [0076] 빈발 항목 관리부(240)는 트리의 각 노드들을 깊이 우선 탐색 방법으로 탐색하고(단계 S501), 각 노드의 빈도수를 재산출한다(단계 S502). 일 실시예에서, 각 노드의 빈도수는 현재 처리 중인 배치 정보 및 해당 노드에 마지막으로 접근한 배치 정보간의 차이만큼 사용자 단말(110)에 의하여 기 설정된 빈도수 감소 팩터(λ)를 기 저장된 빈도수에 곱함으로써 재 산출될 수 있다.
- [0077] 빈발 항목 관리부(240)는 각 노드를 탐색하면서 재산출된 빈도수와 특정 기준을 비교하여 재산출된 빈도수가 특정 기준보다 작은지 여부를 확인한다(단계 S503). 일 실시예에서, 특정 기준은 근접 최소 빈도수(μ)에 해당할 수 있다.
- [0078] 일 실시예에서, 빈발 항목 관리부(240)는 재산출된 빈도수가 특정 기준보다 작은 노드에 대하여 해당 노드가 트리에서 내부노드에 해당하는지 여부를 확인한다(단계 S504). 여기에서, 내부노드는 단말노드가 아닌 노드에 해당한다.
- [0079] 일 실시예에서, 해당 노드가 내부노드에 해당하는 경우에는, 빈발 항목 관리부(240)는 해당 노드의 빈도수를 0으로 설정할 수 있다(단계 S505). 다른 일 실시예에서, 해당 노드가 내부노드에 해당하지 않는 경우에는, 빈발 항목 관리부(240)는 해당 노드를 트리에서 삭제할 수 있다(단계 S506). 또한, 빈발 항목 관리부(240)는 삭제된

노드로부터 루트 노드까지 탐색하면서(단계 S507), 각 노드의 빈도수가 0인지 여부를 확인한다(단계 S508). 일 실시예에서, 특정 노드의 빈도수가 0에 해당하는 경우에는 해당 노드를 제거하고(단계 S509), 단계 S507로 이동하여 다음 노드를 탐색한다.

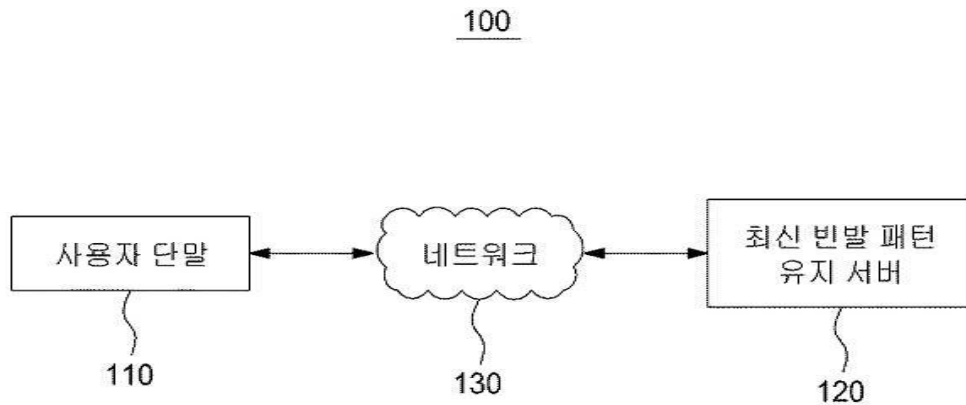
[0080] 상기에서는 본 출원의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 출원을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

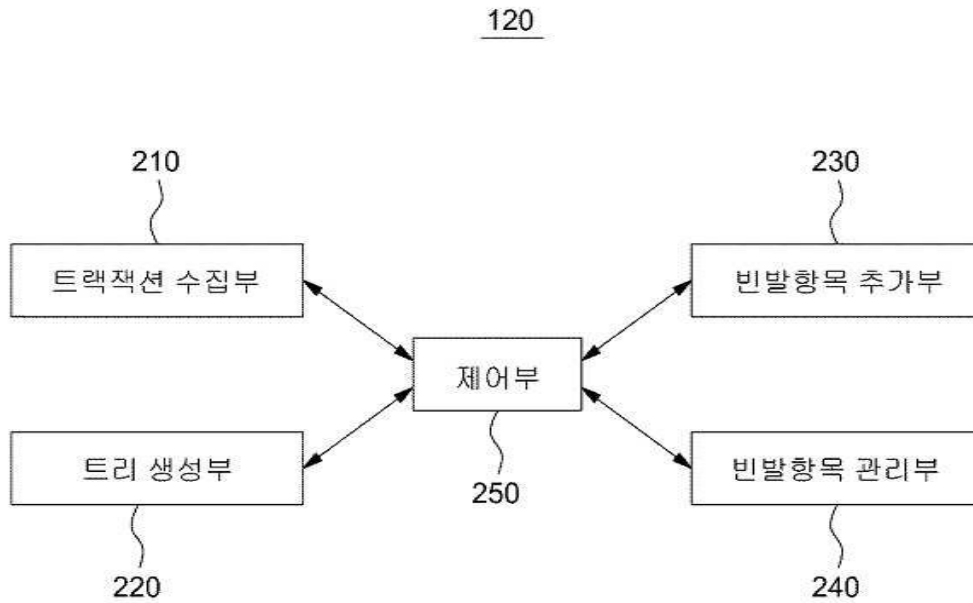
[0081] 100: 최신 빈발 패턴 유지 시스템
110: 사용자 단말
120: 최신 빈발 패턴 유지 서버
130: 네트워크
210: 트랜잭션 수집부
220: 트리 생성부
230: 빈발 항목 추가부
240: 빈발 항목 관리부
250: 제어부

도면

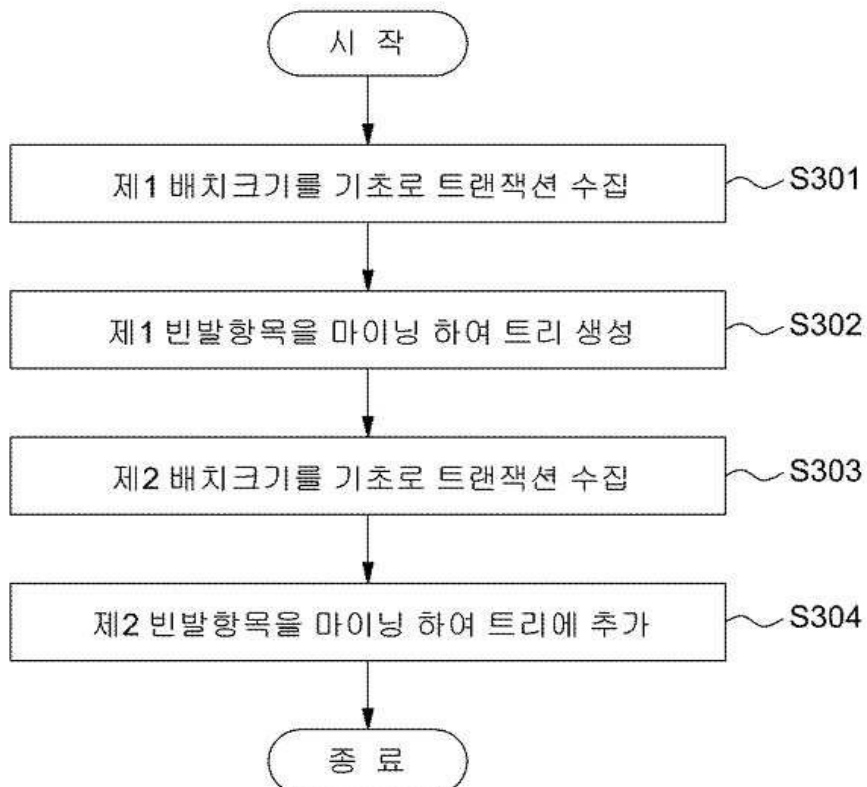
도면1



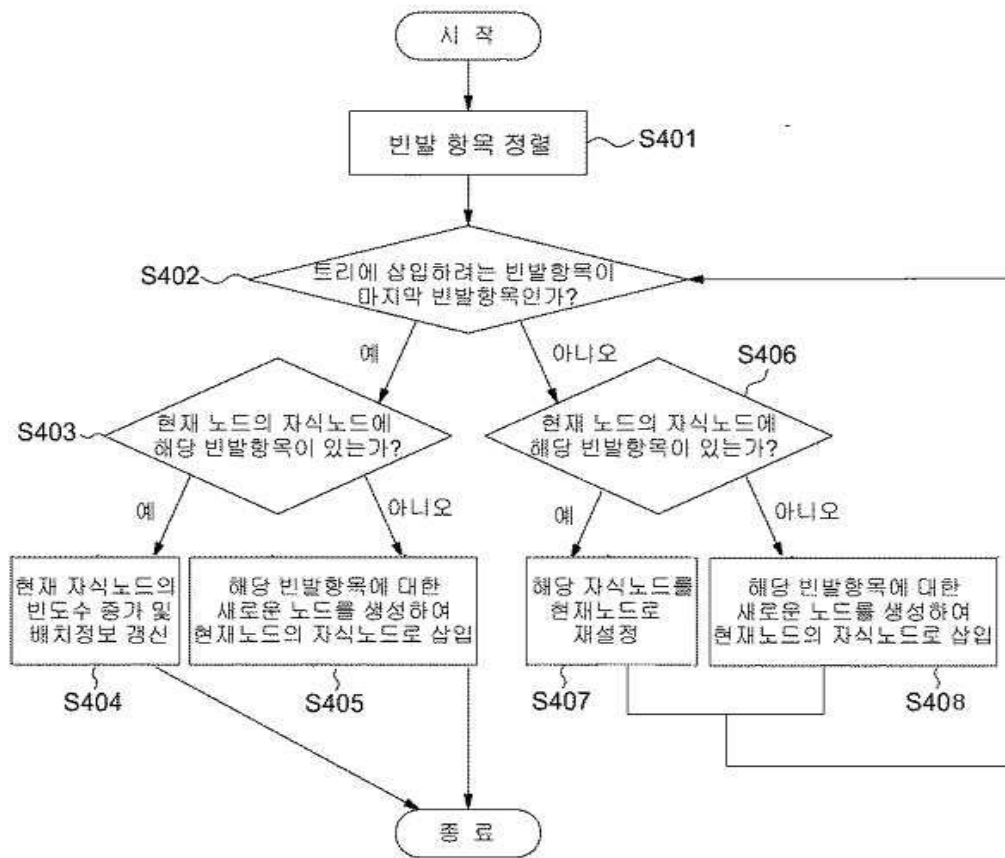
도면2



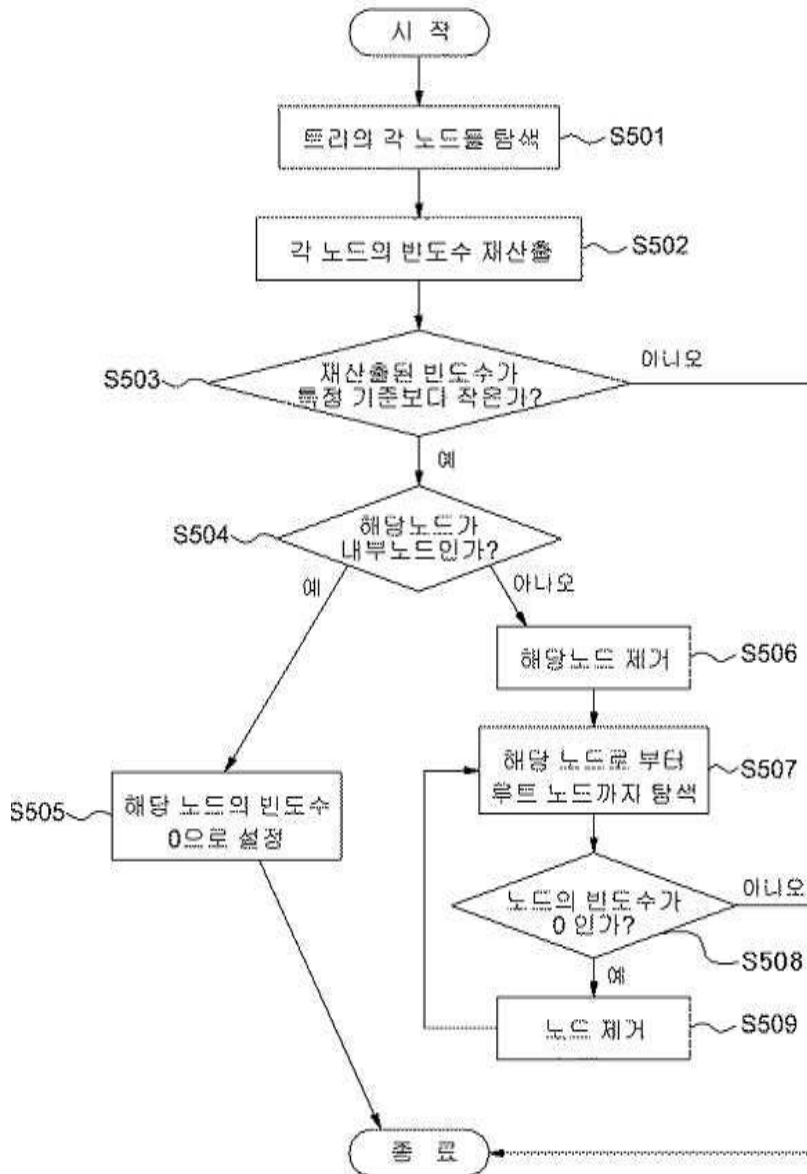
도면3



도면4



도면5



도면6

