(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0062297
(43) 공개일자 2017년06월07일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) *G06F 17/00* (2006.01)
(52) CPC특허분류
G06F 17/30327 (2013.01)
G06F 17/00 (2013.01)
(21) 출원번호 10-2015-0167936
(22) 출원일자 2015년11월27일
심사청구일자 2015년11월27일

(71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
윤은일
서울특별시 성북구 종암로5길 7 (종암동)
김동규
경기도 김포시 전원로 28, 114동 1304호 (장기동, 전원마을월드1단지아파트)
(뒷면에 계속)
(74) 대리인
정부연

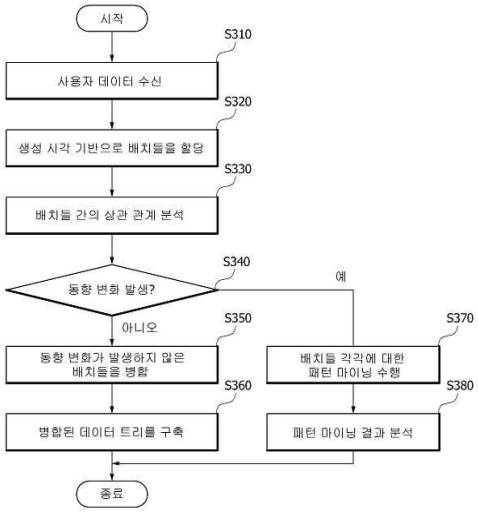
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **슬라이딩 윈도우 기반의 패턴 마이닝 방법, 이를 수행하는 장치 및 이를 저장하는 기록매체**

(57) 요 약

슬라이딩 윈도우 기반의 패턴 마이닝 방법은 (a) 트랜잭션 정보 및 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 상기 초기 사용자 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성하는 단계, (b) 추가 사용자 데이터를 수신하고 상기 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 상기 제1 및 제2 배치들의 사용자 특성에 대한 상관 관계를 분석하는 단계 및 (c) 상기 상관 관계 분석의 결과를 기초로 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 단계를 포함한다. 따라서, 슬라이딩 윈도우 기반의 패턴 마이닝 방법은 배치들 각각의 사용자 특성에 대한 상관 관계 분석을 통해 배치들 간의 동향 변화 발생 여부를 판단하고, 동향 변화가 발생하지 않으면 배치들의 데이터 트리를 병합할 수 있다.

대 표 도 - 도3



(52) CPC특허분류
G06F 17/30324 (2013.01)
G06F 17/30339 (2013.01)
G06F 17/30539 (2013.01)

(72) 발명자
양홍모
충청북도 청주시 상당구 수영로 350 (금천동, 장자마을1단지부영아파트) 104동 302동
이강인
충청북도 청주시 흥덕구 모충로12번길 4 (개신동, 송학아파트)103동 905호
신동춘
경기도 안산시 상록구 안산천동로1길 9, 19동 301호(월피동, 한양아파트)

이 발명을 지원한 국가연구개발사업
과제고유번호 1425094939
부처명 중소기업청
연구관리전문기관 한국산학연협회
연구사업명 산학연협력기술개발
연구과제명 데이터 마이닝 기법을 이용한 빅데이터 내 동향 분석 기술 개발
기 여 율 1/1
주관기관 세종대학교산학협력단
연구기간 2015.06.01 ~ 2016.05.31

명 세 서

청구범위

청구항 1

(a) 트랜잭션 정보 및 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 상기 초기 사용자 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성하는 단계;

(b) 추가 사용자 데이터를 수신하고 상기 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 상기 제1 및 제2 배치들의 사용자 특성에 대한 상관 관계를 분석하는 단계; 및

(c) 상기 상관 관계 분석의 결과를 기초로 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 단계를 포함하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 2

제1항에 있어서,

(d) 상기 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 상기 제1 및 제2 배치들을 병합하여, 병합된 데이터 트리를 구축하는 단계를 더 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 3

제1항에 있어서,

(e) 상기 제1 및 제2 배치들 간의 동향 변화가 발생하면 상기 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하여 상기 패턴 마이닝의 결과를 분석하는 단계를 더 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 4

제1항에 있어서, 상기 (b) 단계는

각각의 원소가 가지는 상관 계수를 통해 사용자 특성들 간의 상관 관계를 나타내는 상관 행렬을 생성하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 5

제4항에 있어서, 상기 (b) 단계는

적어도 하나의 사용자 특성 정보를 포함하는 특성 행렬을 생성하는 단계;

상기 특성 행렬을 정규화하여 단위 행렬을 생성하고, 상기 단위 행렬을 전치하여 전치 행렬을 생성하는 단계; 및

상기 단위 행렬 및 상기 전치 행렬의 행렬 곱셈을 통해 상관 행렬을 도출하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 6

제5항에 있어서, 상기 정규화는

상기 특성 행렬 내의 원소들의 평균 값 및 표준 편차 값을 연산하여 상기 단위 행렬의 각 행의 합이 0에 해당하도록 변환하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 7

제1항에 있어서, 상기 (c) 단계는

상기 제1 및 제2 배치들 각각의 상관 행렬을 비교하여 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 8

제7항에 있어서, 상기 (c) 단계는

상기 상관 행렬의 상관 계수를 비교하여 상기 상관 계수의 변화량이 기 설정된 최소 변화량을 초과하면 해당 상관 계수에 대한 동향 변화의 발생을 판단하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 9

제8항에 있어서, 상기 (c) 단계는

상기 최소 변화량을 초과하는 상관 계수들의 개수가 기 설정된 최소 임계치를 초과하면 상기 제1 및 제2 배치들 간의 동향 변화의 발생을 판단하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 10

제8항에 있어서, 상기 상관 계수들 각각의 최소 변화량은

상기 상관 계수들 각각의 평균 변화량 및 특성을 기초로 서로 다르게 설정되는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 11

제9항에 있어서, 상기 최소 임계치는

상기 제1 및 제2 배치들의 총 사용자 수, 총 사용자 특성 수를 기초로 설정되는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 12

제8항에 있어서, 상기 상관 계수는

그 값이 1에 가까울수록 해당 사용자들 간의 관련도가 높은 것을 나타내는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 13

제2항에 있어서, 상기 (d) 단계는

상기 병합된 제1 및 제2 배치들의 사용자 데이터를 기초로 사용자 아이디 정보, 사용자 특성 정보 및 사용자 링크 정보를 포함하는 헤더 테이블을 생성하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 14

제13항에 있어서, 상기 (d) 단계는

상기 병합된 제1 및 제2 배치들의 사용자 데이터의 적어도 하나의 키워드를 기초로 데이터 트리 방문을 수행하고 데이터 트리 경로를 생성하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 15

제14항에 있어서, 상기 (d) 단계는

상기 데이터 트리 경로 상에 있는 테일 노드에 사용자 아이디 정보, 생성 시각 정보, 사용자 특성 정보를 추가적으로 연관시키는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 16

제14항에 있어서, 상기 (d) 단계는

상기 헤더 테이블의 사용자 링크들 각각으로부터 해당 사용자에 관한 테일 노드를 순차적으로 연결하는 단계를 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법.

청구항 17

트랜잭션 정보 및 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 상기 초기 사용자 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성하는 배치 생성부;

추가 사용자 데이터를 수신하고 상기 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 상기 제1 및 제2 배치들의 사용자 특성에 대한 상관 관계를 분석하는 상관 관계 분석부; 및

상기 상관 관계 분석의 결과를 기초로 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 동향 변화 판단부를 포함하는 슬라이딩 윈도우 기반의 패턴 마이닝 장치.

청구항 18

제17항에 있어서,

상기 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 상기 제1 및 제2 배치들을 병합하여, 병합된 데이터 트리를 구축하는 데이터 트리 병합부를 더 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 장치.

청구항 19

제17항에 있어서,

상기 제1 및 제2 배치들 간의 동향 변화가 발생하면 상기 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하

여 상기 패턴 마이닝의 결과를 분석하는 배치별 패턴 분석부를 더 포함하는 것을 특징으로 하는 슬라이딩 윈도우 기반의 패턴 마이닝 장치.

청구항 20

트랜잭션 정보 및 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 상기 초기 사용자 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성하는 기능;

추가 사용자 데이터를 수신하고 상기 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 상기 제1 및 제2 배치들의 사용자 특성에 대한 상관 관계를 분석하는 기능;

상기 상관 관계 분석의 결과를 기초로 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 기능;

상기 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 상기 제1 및 제2 배치들을 병합하여, 병합된 데이터 트리를 구축하는 기능; 및

상기 제1 및 제2 배치들 간의 동향 변화가 발생하면 상기 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하여 상기 패턴 마이닝의 결과를 분석하는 기능을 수행하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법을 구현하기 위한 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체.

발명의 설명

기 술 분 야

[0001] 본 발명은 슬라이딩 윈도우 기반의 패턴 마이닝 기술에 관한 것으로, 보다 상세하게는 배치들 각각의 사용자 특성에 대한 상관 관계 분석을 통해 배치들 간의 동향 변화 발생 여부를 판단하고, 동향 변화가 발생하면 배치들의 데이터 트리를 병합하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법, 이를 수행하는 장치 및 이를 저장하는 기록매체에 관한 것이다.

배 경 기 술

[0002] 패턴 마이닝 기술은 지능적인 서비스를 필요로 하는 환경에서 유용하게 사용된다. 패턴 마이닝 기술은 수집된 데이터로부터 정보를 추출하기 위해 데이터를 분석하여 패턴을 추출하고, 추출된 패턴을 기초로 서비스를 제공한다. 종래의 패턴 마이닝 기술은 시간의 흐름에 따라 변화하는 데이터를 분석하는 방법이 제공된다. 일반적으로 데이터 마이닝의 대상이 되는 데이터 집합에서는 응용 도메인에 나타나는 모든 단위 정보들을 단위항목(또는 키워드)으로 정의하고 응용 도메인에서 의미적으로 서로 함께 발생하는 단위정보들의 모임을 트랜잭션이라 정의한다. 트랜잭션은 의미적인 동시성을 갖는 단위항목들의 정보를 가지며 데이터 마이닝의 분석 대상이 되는 데이터 집합은 해당 응용 도메인에서 발생된 트랜잭션들의 집합으로 정의된다.

[0003] 한편, 종래 기술의 패턴 마이닝 기술은 스트림 데이터를 효율적으로 처리하기 위하여 슬라이딩 윈도우 기술을 사용하여 데이터를 생성 시각 기준으로 복수의 배치들로 분리하고, 배치들 각각에 대한 데이터 트리를 통해 패턴을 분석하는 방법을 제시하였으나, 이러한 기술은 데이터의 생성 시각만을 고려하고 지속적으로 생성되는 데이터의 동향 변화를 고려하지 않지 않기 때문에 데이터의 특성 변화를 효율적으로 판단할 수 없다.

[0004] 한국등록특허 제10-1093521호는 연속적으로 발생하는 데이터의 패턴 분석 방법에 관한 것으로, 연속적으로 발생하는 데이터를 단위 블록으로 분할하는 단계, 분할된 단위 블록-상기 단위 블록은 N번째 분할된 단위 블록임-의 데이터로부터 패턴을 추출하는 단계 및 추출된 패턴을 패턴 모델 PMN-1과 결합하여 새로운 패턴 모델 PMN을 생성하는 단계를 개시함으로써, 연속적으로 발생하는 데이터를 효과적으로 분석할 수 있는 방법에 대하여 개시한다.

[0005] 한국등록특허 제10-0896528호는 데이터 스트림으로부터 연관규칙을 생성하는 방법 및 데이터 마이닝 시스템에 관한 것으로, 현재 발생한 트랜잭션에 포함된 항목집합 및 상기 항목집합의 출현 빈도수에 따라서 상기 항목집합에 대응하는 노드가 가지고 있는 정보를 갱신하거나 새로운 노드를 추가하는 단계, 전위트리의 각 노드에 해당하는 항목집합에 대하여 지지도를 최소 지지도와 비교하여 빈발항목집합들을 선택하는 단계 및 선택된 빈발항목집합들에 해당하는 노드들 중 전부 또는 일부의 노드들을 방문하면서 상기 방문된 노드들 각각이 가지고 있는

정보를 이용하여 연관규칙을 생성하는 단계를 개시함으로써 지속적으로 발생되는 트랜잭션들로 구성되는 비한정적 데이터 집합인 데이터스트림으로부터 효율적으로 연관규칙을 생성하는 기술을 개시한다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국등록특허 제10-1093521호 (2011. 12. 07 등록)

(특허문헌 0002) 한국등록특허 제10-0896528호 (2003. 04. 29 등록)

발명의 내용

해결하려는 과제

[0007] 본 발명의 일 실시예는 배치들 각각의 사용자 특성에 대한 상관 관계 분석을 통해 배치들 간의 동향 변화 발생 여부를 판단하고, 동향 변화가 발생하면 배치들의 데이터 트리를 병합하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법 및 장치를 제공하고자 한다.

[0008] 본 발명의 일 실시예는 배치들 간의 동향 변화가 발생하지 않으면 배치들 각각에 대한 패턴 마이닝을 수행하여 패턴 마이닝의 결과를 분석하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법 및 장치를 제공하고자 한다.

[0009] 본 발명의 일 실시예는 사용자 특성 정보를 기초로 배치들 각각에 대한 상관 행렬을 도출하고, 상관 행렬의 상관 계수 변화량을 통해 동향 변화의 발생 여부를 판단하는 슬라이딩 윈도우 기반의 패턴 마이닝 방법 및 장치를 제공하고자 한다.

과제의 해결 수단

[0010] 실시예들 중에서, 슬라이딩 윈도우 기반의 패턴 마이닝 방법은 (a) 트랜잭션 정보 및 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 상기 초기 사용자 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성하는 단계, (b) 추가 사용자 데이터를 수신하고 상기 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 상기 제1 및 제2 배치들의 사용자 특성에 대한 상관 관계를 분석하는 단계 및 (c) 상기 상관 관계 분석의 결과를 기초로 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 단계를 포함 한다.

[0011] 상기 슬라이딩 윈도우 기반의 패턴 마이닝 방법은 (d) 상기 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 상기 제1 및 제2 배치들을 병합하여, 병합된 데이터 트리를 구축하는 단계를 더 포함할 수 있다. 상기 슬라이딩 윈도우 기반의 패턴 마이닝 방법은 (e) 상기 제1 및 제2 배치들 간의 동향 변화가 발생하면 상기 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하여 상기 패턴 마이닝의 결과를 분석하는 단계를 더 포함할 수 있다.

[0012] 상기 (b) 단계는 각각의 원소가 가지는 상관 계수를 통해 사용자 특성들 간의 상관 관계를 나타내는 상관 행렬을 생성하는 단계를 포함할 수 있다. 상기 (b) 단계는 적어도 하나의 사용자 특성 정보를 포함하는 특성 행렬을 생성하는 단계, 상기 특성 행렬을 정규화하여 단위 행렬을 생성하고, 상기 단위 행렬을 전치하여 전치 행렬을 생성하는 단계 및 상기 단위 행렬 및 상기 전치 행렬의 행렬 곱셈을 통해 상관 행렬을 도출하는 단계를 포함할 수 있다. 일 실시예에서, 상기 정규화는 상기 특성 행렬 내의 원소들의 평균 값 및 표준 편차 값을 연산하여 상기 단위 행렬의 각 행의 합이 0에 해당하도록 변환할 수 있다.

[0013] 상기 (c) 단계는 상기 제1 및 제2 배치들 각각의 상관 행렬을 비교하여 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단할 수 있다. 상기 (c) 단계는 상기 상관 행렬의 상관 계수를 비교하여 상기 상관 계수의 변화량이 기 설정된 최소 변화량을 초과하면 해당 상관 계수에 대한 동향 변화의 발생을 판단하는 단계를 포함할 수 있다. 상기 (c) 단계는 상기 최소 변화량을 초과하는 상관 계수들의 개수가 기 설정된 최소 임계치를 초과하면 상기 제1 및 제2 배치들 간의 동향 변화의 발생을 판단하는 단계를 포함할 수 있다.

[0014] 일 실시예에서, 상기 상관 계수들 각각의 최소 변화량은 상기 상관 계수들 각각의 평균 변화량 및 특성을 기초로 서로 다르게 설정될 수 있다. 일 실시예에서, 상기 최소 임계치는 상기 제1 및 제2 배치들의 총 사용자 수, 총 사용자 특성 수를 기초로 설정될 수 있다. 일 실시예에서, 상기 상관 계수는 그 값이 1에 가까울수록 해당 사용자들 간의 관련도가 높은 것을 나타낼 수 있다.

[0015] 상기 (d) 단계는 상기 병합된 제1 및 제2 배치들의 사용자 데이터를 기초로 사용자 아이디 정보, 사용자 특성 정보 및 사용자 링크 정보를 포함하는 헤더 테이블을 생성하는 단계를 포함할 수 있다. 상기 (d) 단계는 상기 병합된 제1 및 제2 배치들의 사용자 데이터의 적어도 하나의 키워드를 기초로 데이터 트리 방문을 수행하고 데이터 트리 경로를 생성하는 단계를 포함할 수 있다. 상기 (d) 단계는 상기 데이터 트리 경로 상에 있는 테일 노드에 사용자 아이디 정보, 생성 시각 정보, 사용자 특성 정보를 추가적으로 연관시킬 수 있다. 상기 (d) 단계는 상기 헤더 테이블의 사용자 링크들 각각으로부터 해당 사용자에 관한 테일 노드를 순차적으로 연결할 수 있다.

[0016] 실시예들 중에서, 슬라이딩 윈도우 기반의 패턴 마이닝 장치는 트랜잭션 정보 및 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 상기 초기 사용자 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성하는 배치 생성부, 추가 사용자 데이터를 수신하고 상기 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 상기 제1 및 제2 배치들의 사용자 특성에 대한 상관 관계를 분석하는 상관 관계 분석부 및 상기 상관 관계 분석의 결과를 기초로 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 동향 변화 판단부를 포함한다.

[0017] 상기 슬라이딩 윈도우 기반의 패턴 마이닝 장치는 상기 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 상기 제1 및 제2 배치들을 병합하여, 병합된 데이터 트리를 구축하는 데이터 트리 병합부를 더 포함할 수 있다. 상기 슬라이딩 윈도우 기반의 패턴 마이닝 장치는 상기 제1 및 제2 배치들 간의 동향 변화가 발생하면 상기 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하여 상기 패턴 마이닝의 결과를 분석하는 배치별 패턴 분석부를 더 포함할 수 있다.

[0018] 실시예들 중에서, 슬라이딩 윈도우 기반의 패턴 마이닝 방법을 구현하기 위한 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체는 트랜잭션 정보 및 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 상기 초기 사용자 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성하는 기능, 추가 사용자 데이터를 수신하고 상기 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 상기 제1 및 제2 배치들의 사용자 특성에 대한 상관 관계를 분석하는 기능, 상기 상관 관계 분석의 결과를 기초로 상기 제1 및 제2 배치들 간의 동향 변화 여부를 판단하는 기능, 상기 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 상기 제1 및 제2 배치들을 병합하여, 병합된 데이터 트리를 구축하는 기능 및 상기 제1 및 제2 배치들 간의 동향 변화가 발생하면 상기 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하여 상기 패턴 마이닝의 결과를 분석하는 기능을 수행한다.

발명의 효과

[0019] 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.

[0020] 본 발명의 일 실시예에 따른 슬라이딩 윈도우 기반의 패턴 마이닝 방법 및 장치는 배치들 각각의 사용자 특성에 대한 상관 관계 분석을 통해 배치들 간의 동향 변화 발생 여부를 판단하고, 동향 변화가 발생하지 않으면 배치들의 데이터 트리를 병합할 수 있다.

[0021] 본 발명의 일 실시예에 따른 슬라이딩 윈도우 기반의 패턴 마이닝 방법 및 장치는 배치들 간의 동향 변화가 발생하면 배치들 각각에 대한 패턴 마이닝을 수행하여 패턴 마이닝의 결과를 분석할 수 있다.

[0022] 본 발명의 일 실시예에 따른 슬라이딩 윈도우 기반의 패턴 마이닝 방법 및 장치는 사용자 특성 정보를 기초로 배치들 각각에 대한 상관 행렬을 도출하고, 상관 행렬의 상관 계수 변화량을 통해 동향 변화의 발생 여부를 판단할 수 있다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 일 실시예에 따른 슬라이딩 윈도우 기반의 패턴 마이닝 시스템을 설명하는 블록도이다.

도 2는 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치를 나타내는 블록도이다.

도 3은 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 수행되는 슬라이딩 윈도우 기반의 패턴 마이닝 과정을 설명하는 순서도이다.

도 4는 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 생성되는 헤더 테이블 및 데이터 트리를 나타내는 도면이다.

도 5는 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 수행되는 상관 관계의 분석 과정을 설명하는 순서도이다.

도 6은 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 수행되는 상관 행렬의 생성 과정을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0024] 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.

[0025] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.

[0026] "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.

[0027] 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.

[0028] 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.

[0029] 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.

[0030] 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.

[0031] 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.

[0032] 도 1은 본 발명의 일 실시예에 따른 슬라이딩 윈도우 기반의 패턴 마이닝 시스템을 설명하는 블록도이다.

[0033] 도 1을 참조하면, 슬라이딩 윈도우 기반의 패턴 마이닝 시스템(100)은 사용자 단말기(110), 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120) 및 데이터베이스(130)를 포함한다.

[0034] 사용자 단말기(110)는 사용자의 제어에 의해 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)에 접속하여 패턴의 마이닝을 요청할 수 있다. 사용자 단말기(110)는 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)에서 제공된 마이닝 결과 정보 또는 데이터 그룹의 특성 정보를 화면에 디스플레이 할 수 있다. 사용자 단말기(110)는 모바일 단말기, 태블릿 PC, 랩탑 PC 또는 데스크탑 PC에 해당할 수 있다.

[0035] 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 사용자 단말기(110)의 요청에 따라 데이터베이스(130)에 저장된 데이터에 대한 중요 패턴 마이닝을 수행할 수 있다. 보다 구체적으로, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 사용자 데이터들의 생성 시각을 기초로 사용자 데이터들을 배치(Batch)별로 분류할 수 있다. 즉, 배치는 생성 시각에 따라 분류된 사용자 데이터들의 집합에 해당할 수 있다. 결과적으로, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 데이터 스트림(또는 복수의 사용자 데이터들)을 수신하여 생성 시각을 기초로 배치별로 분류할 수 있다.

[0036] 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 배치들 각각의 사용자 특성에 대한 상관 관계 분석을 통해 배치들 간의 동향 변화 발생 여부를 판단할 수 있다. 일 실시예에서, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 배치들 간의 동향 변화가 발생하면 배치들의 데이터 트리를 병합할 수 있고, 동향 변화가 발생하지 않으면 배치들 각각에 대한 패턴 마이닝을 수행할 수 있다.

[0037] 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 분류된 배치들 각각에 대하여 헤더 테이블 및 데이터 트리를 생성할 수 있다. 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 데이터 트리 경로의 사용자 링크를 기초로 해당 경로에 대한 특성 값을 추출하고 데이터 트리 경로의 대표 특성 값을 도출하여, 패턴 마이닝의 결과를 분석할 수 있다. 즉, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 마이닝 결과 정보 또는 데이터 그룹의 특성 정보를 사용자 단말기(110)에 제공할 수 있다.

[0038] 도 2는 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치를 나타내는 블록도이다.

[0039] 도 2를 참조하면, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 배치 생성부(210), 상관 관계 분석부(220), 동향 변화 판단부(230), 데이터 트리 병합부(240), 배치별 패턴 분석부(250) 및 제어부(260)를 포함할 수 있다.

[0040] 배치 생성부(210)는 트랜잭션 정보 및 사용자 특성 정보를 포함하는 초기 사용자 데이터를 수신하고 초기 데이터를 기초로 제1 데이터 트리를 구축하여 생성 시각 기반의 제1 배치를 생성할 수 있다. 여기에서, 초기 사용자 데이터는 제1 배치 생성을 위하여 데이터 스트림으로부터 생성 시각을 기초로 분류된 데이터 집합에 해당할 수 있다. 또한, 초기 사용자 데이터는 복수의 사용자 데이터들 중 어느 하나에 해당하고, 데이터 생성 시각에 의하여 한정되지 않는다.

[0041] 배치 생성부(210)는 사용자 아이디, 사용자 특성 정보, 사용자 링크를 포함하는 헤더 테이블을 생성할 수 있다. 여기에서, 사용자 특성 정보는 데이터 생성 시각, 데이터 길이, 데이터 특성(예를 들어, SNS 상에서 시간당 글 등록 개수, 시간당 댓글 개수), 데이터 작성 횟수, 적어도 하나의 키워드 정보를 포함할 수 있다. 또한, 데이터 생성 시각은 사용자 데이터가 생성되어 데이터베이스(130)에 저장된 시간에 해당할 수 있다.

[0042] 상관 관계 분석부(220)는 추가 사용자 데이터를 수신하고 추가 사용자 데이터를 기초로 제2 데이터 트리를 구축하여 생성 시각 기반의 제2 배치를 생성하고, 제1 및 제2 배치들 각각의 사용자 특성에 대한 상관 관계를 분석할 수 있다. 여기에서, 추가 사용자 데이터는 제2 배치 생성을 위하여 데이터 스트림으로부터 생성 시각을 기초로 분류된 데이터 집합에 해당할 수 있다. 또한, 추가 사용자 데이터는 복수의 사용자 데이터들 중 어느 하나에 해당하고, 초기 사용자 데이터의 데이터 생성 시각에 의하여 한정되지 않는다. 즉, 제1 및 제2 배치들 각각은 서로 다른 생성 시각 범위 내에서 생성된 사용자 데이터들을 기초로 생성될 수 있다.

[0043] 상관 관계 분석부(220)는 적어도 하나의 사용자 특성 정보를 기초로 상관 행렬을 생성하여 제1 및 제2 배치들 각각의 사용자 특성에 대한 상관 관계를 분석할 수 있다. 보다 구체적으로, 상관 관계 분석부(220)는 제1 및 제2 배치들 각각에 대하여 적어도 하나의 사용자 특성 정보를 포함하는 특성 행렬을 생성할 수 있다. 상관 관계

분석부(220)는 특성 행렬을 정규화하여 단위 행렬을 생성하고, 단위 행렬을 전치하여 전치 행렬을 생성할 수 있다. 여기에서, 정규화는 특성 행렬 내의 원소들의 평균 값 및 표준 편차 값을 연산하여 단위 행렬의 각 행의 값이 0에 해당하도록 변환할 수 있다. 즉, 상관 관계 분석부(220)는 사용자 특성들 간의 상관 관계를 효율적으로 도출하기 위하여 정규화를 수행할 수 있다. 상관 관계 분석부(220)는 단위 행렬 및 전치 행렬의 행렬 곱셈을 통해 상관 행렬을 도출할 수 있다. 상관 관계 분석부(220)는 제1 및 제2 배치들 각각의 상관 행렬을 도출하여 제1 및 제2 배치들 각각의 사용자 특성들 간의 상관 관계를 분석할 수 있다. 일 실시예에서, 상관 행렬은 각각의 원소가 가지는 상관 계수를 통해 사용자 특성들 간의 상관 관계를 나타낼 수 있다.

[0044] 동향 변화 판단부(230)는 상관 관계 분석의 결과를 기초로 제1 및 제2 배치들 간의 동향 변화 여부를 판단할 수 있다. 보다 구체적으로, 동향 변화 판단부(230)는 제1 및 제2 배치들 각각의 상관 행렬을 비교하여 제1 및 제2 배치들 간의 동향 변화 여부를 판단할 수 있다. 동향 변화 판단부(230)는 상관 행렬들 각각의 대응되는 상관 계수를 비교함으로써 상관 행렬을 비교할 수 있다. 여기에서, 상관 계수는 사용자 특성들 간의 관련 정도를 나타낼 수 있고, -1 내지 1의 값을 가질 수 있다. 일 실시예에서, 상관 계수는 그 값이 1에 가까울수록 해당 사용자들 간의 관련도가 높은 것을 나타낼 수 있으나, 반드시 이에 한정되는 것은 아니다.

[0045] 일 실시예에서, 동향 변화 판단부(230)는 상관 행렬의 상관 계수를 비교하여 상관 계수의 변화량이 기 설정된 최소 변화량을 초과하면 해당 상관 계수에 대한 동향 변화의 발생을 판단할 수 있다. 예를 들어, 상관 계수의 값은 양 또는 음의 값에서 부호가 변화하거나, 관련도가 높거나 낮은 상태에서 반대의 상태로 변화할 수 있다. 여기에서, 상관 계수들 각각의 최소 변화량은 상관 계수들 각각의 특성 및 상관 계수들 각각의 평균 변화량을 기초로 서로 다르게 설정될 수 있다. 즉, 상관 행렬의 상관 계수들 중 일부는 동향 변화가 발생할 수 있고, 나머지 일부는 동향 변화가 발생하지 않을 수 있다.

[0046] 일 실시예에서, 동향 변화 판단부(230)는 최소 변화량을 초과하는 상관 계수들의 개수가 기 설정된 최소 임계치를 초과하면 제1 및 제2 배치들 간의 동향 변화의 발생을 판단할 수 있다. 여기에서, 최소 임계치는 제1 및 제2 배치들의 총 사용자 수, 총 사용자 특성 수를 기초로 설정될 수 있다. 즉, 동향 변화 판단부(230)는 최소 변화량을 초과하는 상관 계수들의 개수가 최소 임계치를 초과하면 제1 및 제2 배치들 간에 유의미한 동향 변화가 발생한 것을 판단할 수 있고, 판단 결과를 데이터 트리 병합부(240) 또는 배치별 패턴 분석부(250)에 제공할 수 있다.

[0047] 데이터 트리 병합부(240)는 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 제1 및 제2 배치들을 병합하여, 병합된 데이터 트리를 구축할 수 있다. 보다 구체적으로, 데이터 트리 병합부(240)는 병합된 제1 및 제2 배치들의 사용자 데이터를 기초로 헤더 테이블을 생성하고, 데이터 트리를 생성할 수 있다. 이하, 헤더 테이블 및 데이터 트리를 생성하는 과정에 대하여 도 3 및 도 4에서 상세히 설명한다.

[0048] 배치별 패턴 분석부(250)는 제1 및 제2 배치들 간의 동향 변화가 발생하면 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하여 패턴 마이닝의 결과를 분석할 수 있다. 보다 구체적으로, 배치별 패턴 분석부(250)는 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행하여 제1 및 제2 배치들 각각의 데이터 트리에 대한 특성 값을 추출할 수 있다. 결과적으로, 배치별 패턴 분석부(250)는 제1 및 제2 배치들 각각에 대한 대표 특성 값을 도출하여 패턴 마이닝 결과의 정보 또는 데이터 그룹(또는 배치)의 특성 정보를 사용자 단말기(110)에 제공할 수 있다.

[0049] 제어부(260)는 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)의 전체적인 동작을 제어하고, 배치 생성부(210), 상관 관계 분석부(220), 동향 변화 판단부(230), 데이터 트리 병합부(240) 및 배치별 패턴 분석부(250) 간의 제어 흐름 또는 데이터 흐름을 제어할 수 있다.

[0050] 도 3은 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 수행되는 슬라이딩 윈도우 기반의 패턴 마이닝 과정을 설명하는 순서도이고, 도 4는 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 생성되는 헤더 테이블 및 데이터 트리를 나타내는 도면이다.

[0051] 도 3 및 도 4를 참조하면, 배치 생성부(210)는 데이터베이스(130)에 저장된 사용자 데이터(410)를 수신할 수 있다(단계 S310). 여기에서, 사용자 데이터(410)는 트랜잭션 정보 및 사용자 특성 정보를 포함할 수 있다. 예를 들어, 트랜잭션 정보는 적어도 하나의 키워드 정보를 포함할 수 있고, 특성 정보는 사용자 아이디 정보, 생성 시각 정보 및 사용자 특성 정보를 포함할 수 있다.

[0052] 배치 생성부(210)는 데이터 스트림(또는 복수의 사용자 데이터들)을 수신하여 생성 시각 기반으로 배치들을 할당할 수 있다(단계 S320). 예를 들어, 동향 변화 판단부(230)는 제1 및 제2 배치들의 상관 관계를 판단할 수 있

고, 제1 및 제2 배치들의 상관 관계가 인정되면(또는 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면) 제1 및 제2 배치들을 병합할 수 있고, 상관 관계가 인정되지 않으면(또는 제1 및 제2 배치들 간의 동향 변화가 발생하면) 제1 및 제2 배치들 각각에 대하여 패턴 마이닝을 수행할 수 있다. 일 실시예에서, 배치 생성부(210) 및 상관 관계 분석부(220)는 데이터베이스(130)에 저장된 사용자 데이터를 스캔하여 모든 사용자 데이터의 처리를 완료할 때까지 배치들을 할당할 수 있다. 결과적으로, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 사용자 데이터의 생성 시각 및 사용자 특성의 상관 관계를 기초로 배치들을 할당하여, 시간의 흐름에 따른 사용자 데이터의 동향 변화 결과 및 마이닝 분석 결과를 제공할 수 있다.

[0053] 상관 관계 분석부(220)는 제1 및 제2 배치들 각각의 사용자 특성에 대한 상관 관계를 분석할 수 있다(단계 S330). 동향 변화 판단부(230)는 상관 관계 분석의 결과를 기초로 제1 및 제2 배치들 간의 동향 변화 여부를 판단할 수 있다(단계 S340).

[0054] 일 실시예에서, 데이터 트리 병합부(240)는 제1 및 제2 배치들 간의 동향 변화가 발생하지 않으면 제1 및 제2 배치들을 병합할 수 있다(단계 S350). 데이터 트리 병합부(240)는 병합된 제1 및 제2 배치들의 사용자 데이터를 기초로 헤더 테이블(420)을 생성하고, 데이터 트리(430)를 생성할 수 있다(단계 S360). 보다 구체적으로, 데이터 트리 병합부(240)는 병합된 제1 및 제2 배치들의 사용자 데이터를 기초로 사용자 아이디 정보, 사용자 특성 정보 및 사용자 링크 정보를 포함하는 헤더 테이블(420)을 생성할 수 있다. 데이터 트리 병합부(240)는 병합된 제1 및 제2 배치들의 사용자 데이터의 적어도 하나의 키워드를 기초로 데이터 트리 방문을 수행하고 데이터 트리(430)의 경로를 생성할 수 있다. 데이터 트리 병합부(240)는 데이터 트리 경로 상에 있는 테일 노드(434)에 사용자 아이디 정보, 생성 시각 정보 및 사용자 특성 정보를 추가적으로 연관시킬 수 있다. 또한, 데이터 트리 병합부(240)는 헤더 테이블(420)의 사용자 링크들 각각으로부터 해당 사용자에 관한 테일 노드(434)를 순차적으로 연결할 수 있다.

[0055] 일 실시예에서, 사용자 데이터(410)는 복수의 트랜잭션 데이터들(T)을 포함할 수 있다. 예를 들어, 제1 트랜잭션 데이터(T1)은 {1,3}의 트랜잭션 정보와 {1, 3/1}의 특성 정보를 포함할 수 있다. 데이터 트리 병합부(240)는 트랜잭션 정보(또는 키워드 {1,3})을 기초로 데이터 트리 방문을 수행하여 데이터 트리(430)의 경로를 생성할 수 있다. 데이터 트리 병합부(240)는 데이터 트리(430)의 경로 상에 있는 트리 종단 노드에 특성 정보 {1, 3/1}를 테일 노드(434)로서 저장할 수 있다.

[0056] 일 실시예에서, 배치별 패턴 분석부(250)는 제1 및 제2 배치들 간의 동향 변화가 발생하면 제1 및 제2 배치들 각각에 대한 패턴 마이닝을 수행할 수 있다(단계 S370). 배치별 패턴 분석부(250)는 제1 및 제2 배치들 각각의 데이터 트리에 대한 특성 값을 추출하여 대표 특성 값을 도출함으로써 패턴 마이닝의 결과를 분석할 수 있다(단계 S380). 결과적으로, 결과적으로, 배치별 패턴 분석부(250)는 패턴 마이닝 결과의 정보 또는 데이터 그룹(또는 배치)의 특성 정보를 사용자 단말기(110)에 제공할 수 있다.

[0057] 일 실시예에서, 배치별 패턴 분석부(250)는 복수의 배치들 각각에 대하여 헤더 테이블(420)을 생성하고 데이터 트리(430)를 구축하여 패턴 마이닝을 수행할 수 있다. 헤더 테이블(420)은 사용자 아이디 항목, 사용자 특성 항목 및 링크 항목을 포함할 수 있다. 사용자 아이디 항목은 사용자의 명칭 또는 식별 부호를 통해 복수의 사용자들을 구분할 수 있다. 사용자 특성 항목은 사용자 데이터에 포함된 데이터 생성 시간, 데이터 길이, 데이터 특성(예를 들어, SNS 상에서 시간당 글 등록 개수, 시간당 댓글 개수), 데이터 작성 횟수, 적어도 하나의 키워드 정보를 포함할 수 있다. 링크 항목은 해당 사용자에 관한 테일 노드(434)와 연결되는 사용자 링크를 포함할 수 있다.

[0058] 배치별 패턴 분석부(250)는 적어도 하나의 키워드의 조합으로 대표될 수 있는 사용자 데이터를 수신하여 데이터 트리(430)를 생성할 수 있다. 보다 구체적으로, 배치별 패턴 분석부(250)는 적어도 하나의 키워드를 기초로 데이터 트리 방문을 수행하고 데이터 트리(430)의 경로를 구축할 수 있다. 여기에서, 데이터 트리(430)의 경로는 복수의 키워드들의 조합 및 키워드의 빈도수를 기초로 형성될 수 있다.

[0059] 일 실시예에서, 데이터 트리(430)는 트리 중간 노드 및 트리 종단 노드를 포함할 수 있다. 또한, 데이터 트리(430)는 일반 노드(432) 및 테일 노드(434)를 포함할 수 있다. 즉, 데이터 트리(430)의 트리 중간 노드는 일반 노드(432)와 연관되고, 트리 종단 노드는 테일 노드(434)와 연관될 수 있다.

[0060] 일 실시예에서, 사용자 데이터에 포함된 적어도 하나의 키워드는 일반 노드(432) 또는 테일 노드(434)에 저장될 수 있다. 여기에서, 테일 노드(734)는 키워드 명칭, 키워드 빈도수, 사용자 특성 정보, 생성 시각 정보를 포함할 수 있다. 보다 구체적으로, 일반 노드(432)의 키워드 빈도수는 그와 연결된 적어도 하나의 테일 노드(434)에

저장되는 키워드 빈도수들의 합에 해당할 수 있다. 즉, 데이터 트리 경로들은 적어도 하나의 키워드의 조합으로 대표될 수 있고, 서로 공유하는 키워드를 상위 노드(또는 일반 노드)에 저장할 수 있다. 예를 들어, 일반 노드(432) 및 테일 노드(434)는 [키워드 명칭: 키워드 빈도수]의 형태로 키워드 정보를 저장할 수 있다. 또한, 적어도 하나의 키워드의 조합들({1, 3}, {1, 2, 4}, {2, 3, 4}) 중 키워드 빈도수가 가장 적은 키워드(3:2, 4:2, 4:1)가 테일 노드(434a, 434b, 434c)에 저장될 수 있다.

[0061] 배치별 패턴 분석부(250)는 데이터 트리(430)의 경로 상에 있는 트리 종단 노드에 사용자 데이터를 테일 노드(434)로서 추가적으로 연관시키고 해당 사용자 링크를 시작 링크로 해당 사용자에 관한 테일 노드(434)를 순차적으로 연결할 수 있다. 일 실시예에서, 테일 노드(434)는 배치에 대한 특성 정보 및 사용자 링크 정보를 저장할 수 있다. 배치별 패턴 분석부(250)는 테일 노드(434)에 저장되는 트랜잭션 정보 및 특성 정보를 기초로 해당 배치에 대한 특성 값을 추출하고, 배치의 대표 특성 값을 추출할 수 있다.

[0062] 도 5는 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 수행되는 상관 관계의 분석 과정을 설명하는 순서도이고, 도 6은 도 1에 있는 슬라이딩 윈도우 기반의 패턴 마이닝 장치에서 수행되는 상관 행렬의 생성 과정을 나타내는 도면이다.

[0063] 도 5 및 도 6을 참조하면, 상관 관계 분석부(220)는 배치의 사용자 정보를 추출하여(단계 S510), 적어도 하나의 사용자 특성 정보를 포함하는 특성 행렬(610)을 생성할 수 있다(단계 S520). 예를 들어, 특성 행렬(610)은 복수의 사용자들에 대한 사용자 특성들에 대한 정보를 포함할 수 있다. 즉, 해당 사용자에 대한 사용자 특성 정보는 특성 행렬(610)의 원소에 해당할 수 있다.

[0064] 상관 관계 분석부(220)는 특성 행렬(610)을 정규화하여 단위 행렬(620)을 생성할 수 있다(단계 S530). 일 실시예에서, 단위 행렬(620)의 원소들의 값은 특성 행렬(610)의 정규화를 통해 -1 내지 1의 값에 해당할 수 있다. 상관 관계 분석부(220)는 단위 행렬(620)을 전치하여 전치 행렬(630)을 생성할 수 있다(단계 S540). 즉, 단위 행렬(620) 및 전치 행렬(630)은 전치 관계에 해당할 수 있다.

[0065] 상관 관계 분석부(220)는 단위 행렬(620) 및 전치 행렬(630)의 행렬 곱셈을 통해 상관 행렬(640)을 도출할 수 있다(단계 S550). 동향 변화 판단부(230)는 제1 및 제2 배치들 각각의 상관 행렬(640)을 비교할 수 있다(단계 S560). 보다 구체적으로, 동향 변화 판단부(230)는 상관 행렬들(640) 각각의 대응되는 상관 계수를 비교함으로써 상관 행렬을 비교할 수 있다.

[0066] 도 6에서, 상관 행렬(640)이 [0.99, 0.02, 0.98, -0.0; 0.02, 0.99, -0.13, 0.98; 0.98, -0.13, 0.99, -0.20; -0.06, 0.98, -0.20, 0.97]의 상관 계수들을 가지는 경우, 상관 계수의 값은 해당 사용자들 간의 관련도에 해당할 수 있다. 일 실시예에서, 상관 계수는 그 값이 1에 가까울수록 해당 사용자들 간의 관련도가 높은 것을 나타낼 수 있다. 즉, 상관 행렬(640)의 (1,3) 또는 (3,1)의 값이 0.98에 해당하면, 사용자 1 및 사용자 3의 관련도는 높은 것으로 판단될 수 있다. 한편, 상관 행렬(640)의 (1,4) 또는 (4,1)의 값이 -0.06에 해당하면, 사용자 1 및 사용자 4의 관련도는 낮은 것으로 판단될 수 있다.

[0067] 따라서, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 배치들 각각의 사용자 특성에 대한 상관 관계 분석을 통해 배치들 간의 동향 변화 발생 여부를 판단하고, 동향 변화가 발생하지 않으면 배치들의 데이터 트리를 병합하거나 또는 동향 변화가 발생하면 배치들 각각에 대한 패턴 마이닝을 수행하여 패턴 마이닝의 결과를 분석할 수 있다. 또한, 슬라이딩 윈도우 기반의 패턴 마이닝 장치(120)는 사용자 특성 정보를 기초로 배치들 각각에 대한 상관 행렬을 도출하고, 상관 행렬의 상관 계수 변화량을 통해 동향 변화의 발생 여부를 판단할 수 있다.

[0068] 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.

부호의 설명

[0069] 100: 슬라이딩 윈도우 기반의 패턴 마이닝 시스템

110: 사용자 단말기

120: 슬라이딩 윈도우 기반의 패턴 마이닝 장치

130: 데이터 베이스

210: 배치 생성부 220: 상관 관계 분석부

230: 동향 변화 판단부 240: 데이터 트리 병합부

250: 배치별 패턴 분석부 260: 제어부

410: 사용자 데이터 420: 헤더 테이블

430: 데이터 트리 432: 일반 노드

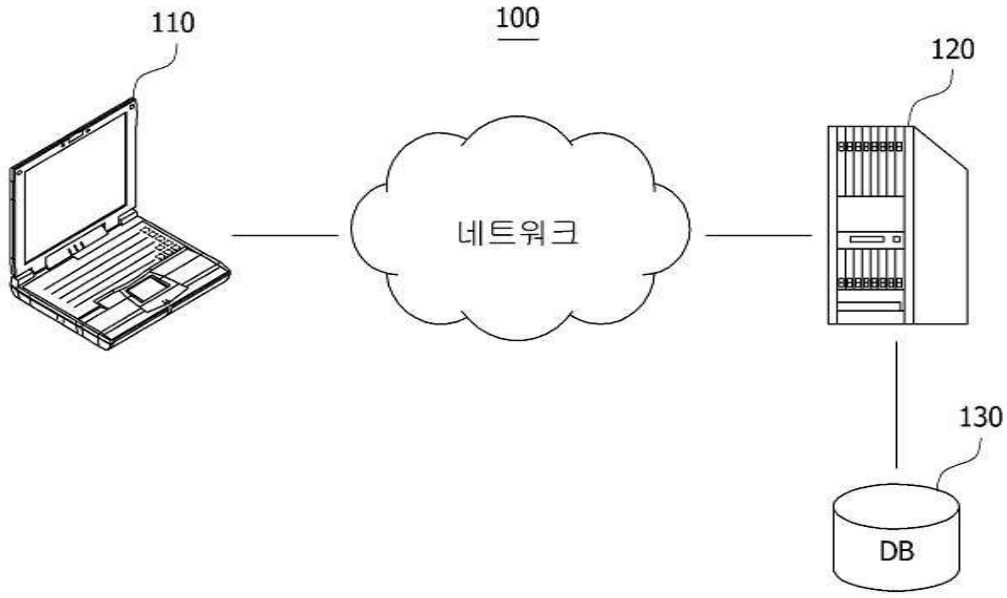
434: 테일 노드 610: 특성 행렬

620: 단위 행렬 630: 전치 행렬

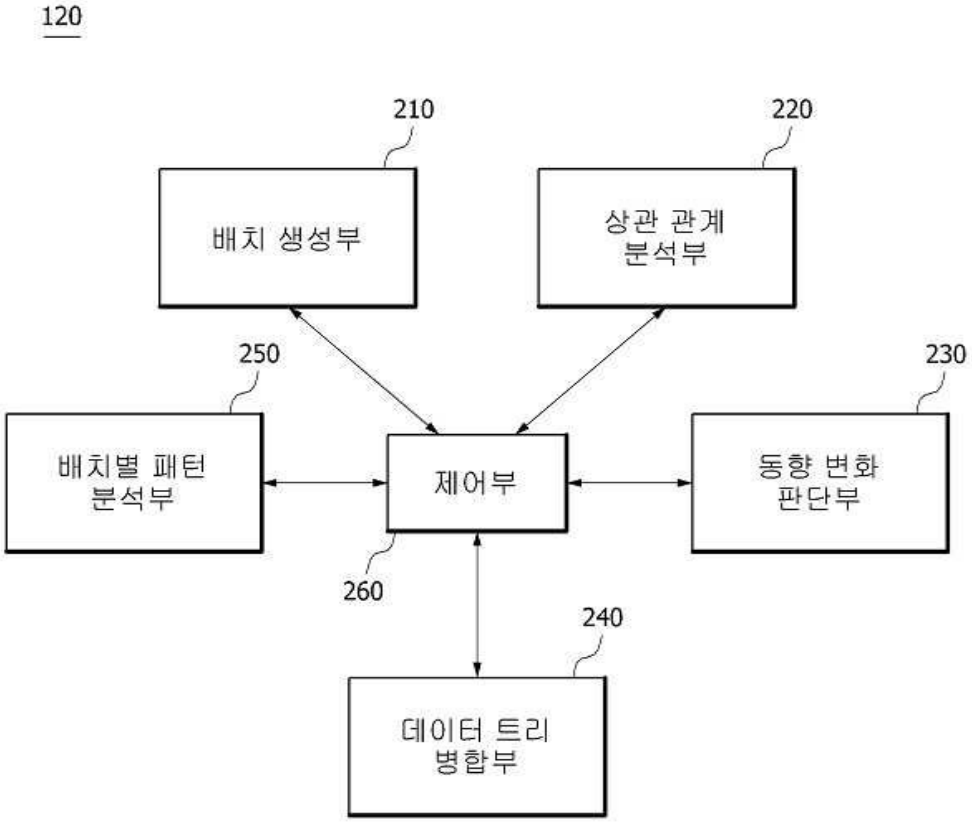
640: 상관 행렬

도면

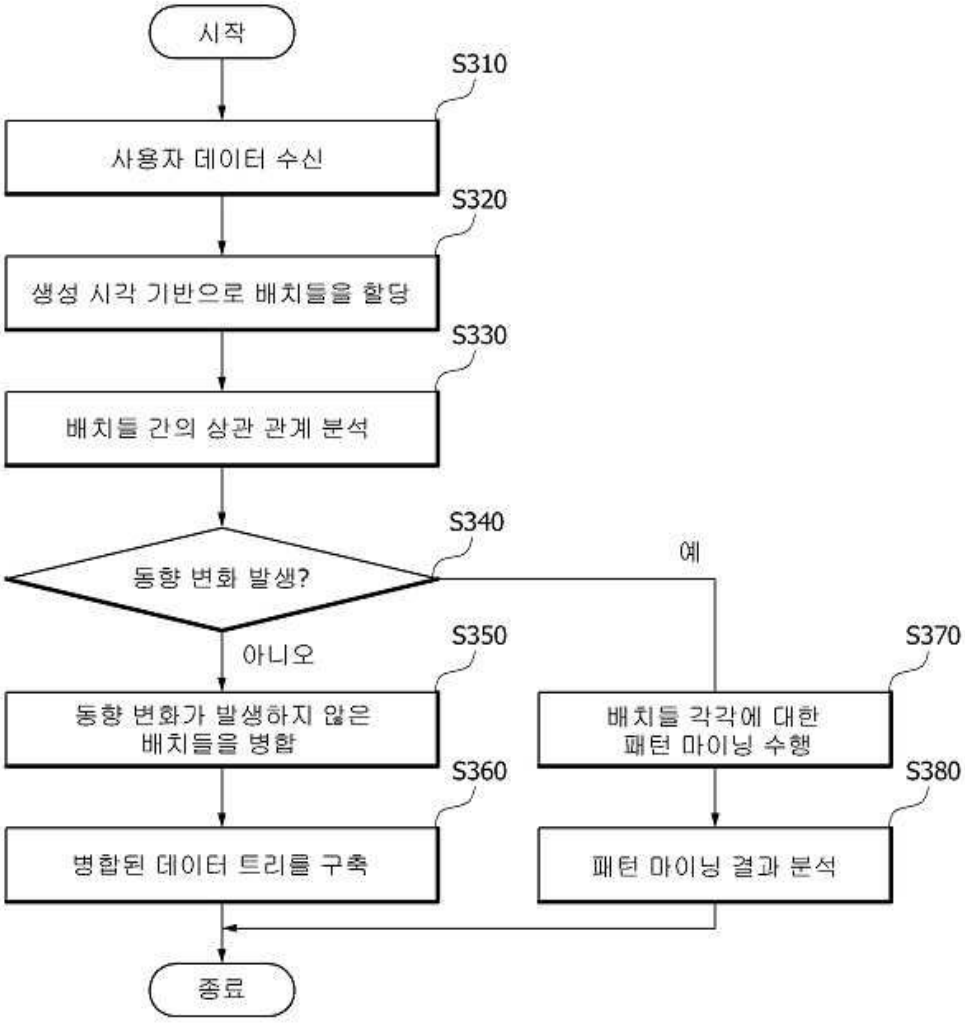
도면1



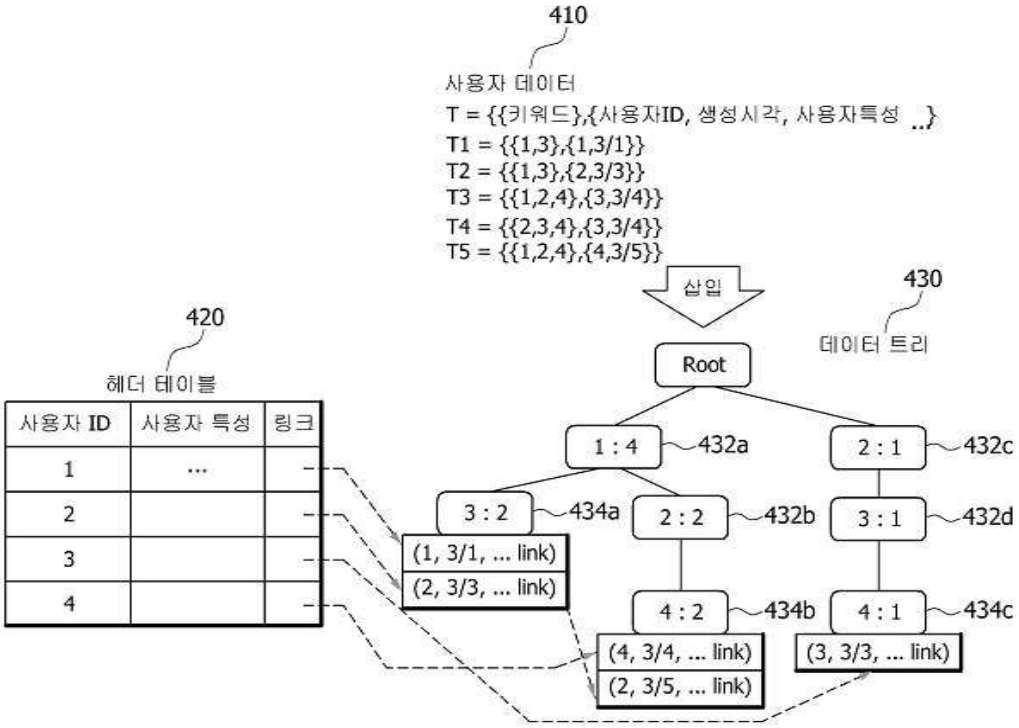
도면2


120
210
배치 생성부
220
상관 관계
분석부
250
배치별 패턴
분석부
제어부
230
동향 변화
판단부
260
240
데이터 트리
병합부

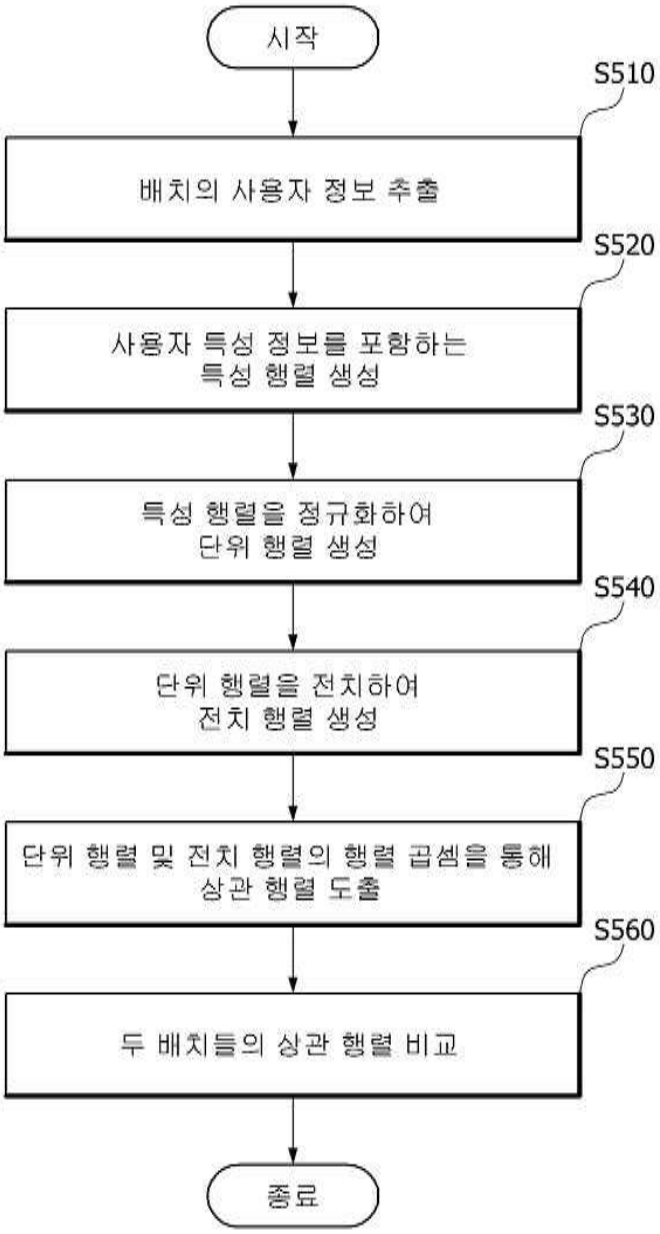
도면3


시작
S310
사용자 데이터 수신
S320
생성 시각 기반으로 배치들을 할당
S330
배치들 간의 상관 관계 분석
S340
동향 변화 발생?
예
아니오
S350
동향 변화가 발생하지 않은 배치들을 병합
S360
병합된 데이터 트리를 구축
S370
배치들 각각에 대한 패턴 마이닝 수행
S380
패턴 마이닝 결과 분석
종료

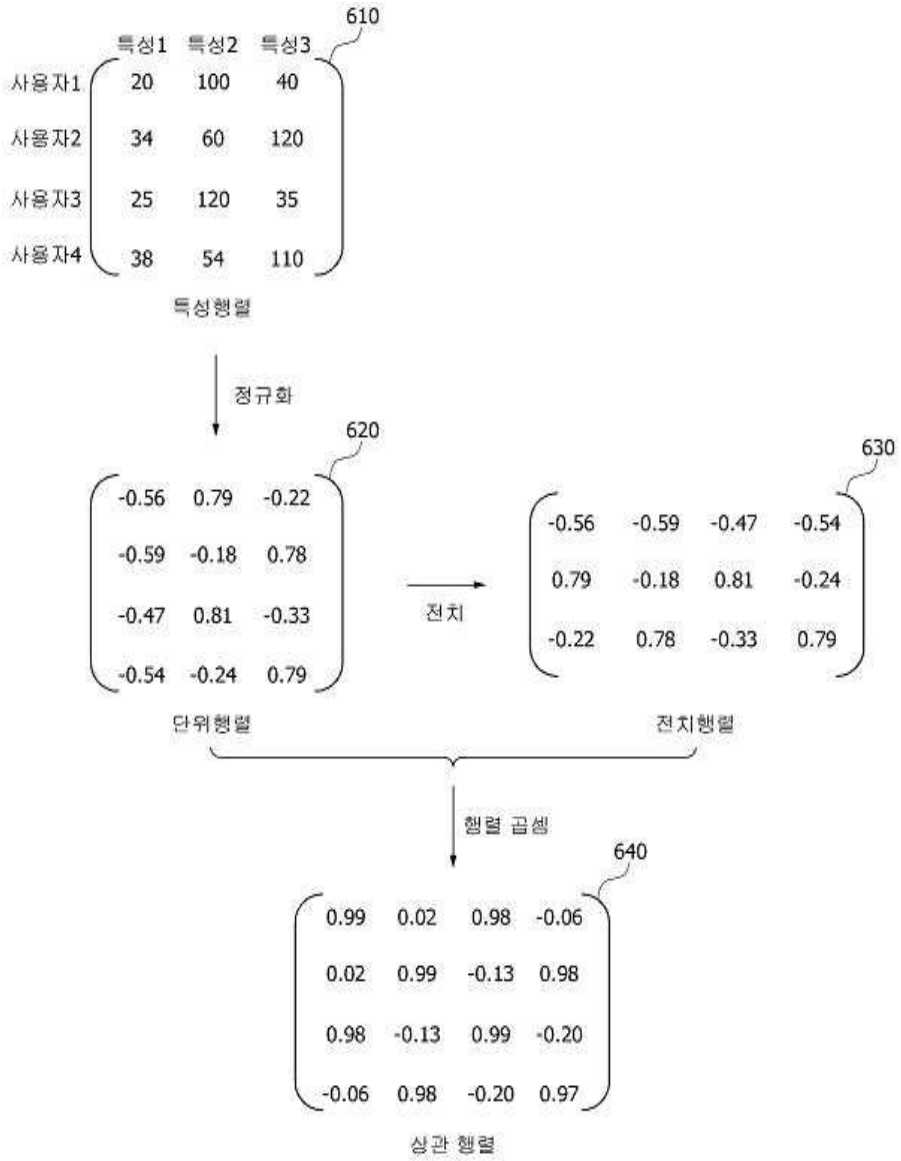
도면4


410
사용자 데이터
T = {{키워드},{사용자ID, 생성시각, 사용자특성 ..}
T1 = {{1,3},{1,3/1}}
T2 = {{1,3},{2,3/3}}
T3 = {{1,2,4},{3,3/4}}
T4 = {{2,3,4},{3,3/4}}
T5 = {{1,2,4},{4,3/5}}
삽입
430
데이터 트리
420
헤더 테이블
사용자 ID
사용자 특성
링크
1
2
3
4
Root
1 : 4
432a
2 : 1
432c
3 : 2
434a
2 : 2
432b
3 : 1
432d
(1, 3/1, ... link)
(2, 3/3, ... link)
4 : 2
434b
4 : 1
434c
(4, 3/4, ... link)
(2, 3/5, ... link)
(3, 3/3, ... link)

도면5


시작
S510
배치의 사용자 정보 추출
S520
사용자 특성 정보를 포함하는
특성 행렬 생성
S530
특성 행렬을 정규화하여
단위 행렬 생성
S540
단위 행렬을 전치하여
전치 행렬 생성
S550
단위 행렬 및 전치 행렬의 행렬 곱셈을 통해
상관 행렬 도출
S560
두 배치들의 상관 행렬 비교
종료

도면6


610
특성1 특성2 특성3
사용자1 20 100 40
사용자2 34 60 120
사용자3 25 120 35
사용자4 38 54 110
특성행렬
정규화
620
-0.56 0.79 -0.22
-0.59 -0.18 0.78
-0.47 0.81 -0.33
-0.54 -0.24 0.79
단위행렬
전치
630
-0.56 -0.59 -0.47 -0.54
0.79 -0.18 0.81 -0.24
-0.22 0.78 -0.33 0.79
전치행렬
행렬 곱셈
640
0.99 0.02 0.98 -0.06
0.02 0.99 -0.13 0.98
0.98 -0.13 0.99 -0.20
-0.06 0.98 -0.20 0.97
상관 행렬