



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0102930

(43) 공개일자 2013년09월23일

(51) 국제특허분류(Int. Cl.)

G06F 17/30 (2006.01) *G06F 17/40* (2006.01)

(21) 출원번호 10-2012-0024135

(22) 출원일자 2012년03월08일

심사청구일자 2012년03월08일

(71) 출원인

경희대학교 산학협력단

경기도 용인시 기흥구 덕영대로 1732, 국제캠퍼스
내 (서천동, 경희대학교)

(72) 발명자

이영구

경기도 용인시 기흥구 서천동 1번지 경희대학교
전자정보관 314호

한용구

경기도 용인시 기흥구 서천동 경희대학교 전자정보관 349호

박기성

경기도 용인시 기흥구 서천동 경희대학교 전자정보관 349호

(74) 대리인

서재승

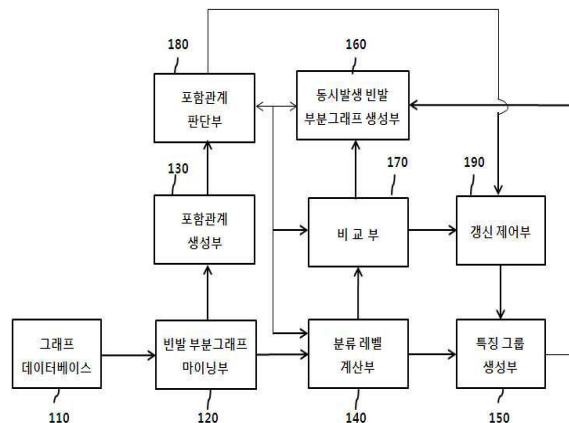
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 동시발생 빈발 부분그래프의 상한 분류값을 고려한 특징 그룹 생성 방법

(57) 요약

본 발명은 그래프 데이터베이스에서 그래프를 분류하기 위한 특징 그룹을 생성하는 방법에 관한 것으로, 보다 구체적으로 그래프 데이터베이스에 저장되어 있는 레이블(labeled) 그래프와 언레이블(unlabeled) 그래프에서 높은 분류 레벨을 가지는 특징 그룹을 생성하며, 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 추가한 추가 동시발생 빈발 부분그래프의 분류 레벨은 동시발생 빈발 부분그래프의 상한 분류값보다 작거나 같음에 기초하여 기생성된 특징 그룹의 최소 분류 레벨보다 작은 상한 분류값을 가지는 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 조합하여 추가 동시발생 빈발 부분그래프를 생성하는 것을 중지함으로써 동시발생 빈발 부분그래프의 조합에 소요되는 연산량을 줄일 수 있는 특징 그룹 생성 방법에 관한 것이다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 NIPA-2012-C1090-1221-0003

부처명 지식경제부

연구사업명 대학 IT 연구센터(ITRC)

연구과제명 동서신의학 u-라이프케어 기술 연구

주관기관 경희대학교 산학협력단

연구기간 2006.11.01 ~ 2014.12.31

특허청구의 범위

청구항 1

컴퓨터에서 그래프 데이터베이스에 저장되어 있는 그래프에 기초하여 특징 그룹을 생성하는 방법에 있어서,
 상기 그래프 데이터베이스에서 패턴 확장되는 개별 빈발 부분그래프의 분류값 또는 상한값에 기초하여 상위 분류레벨을 가지는 개별 빈발 부분그래프로 특징 그룹을 생성하는 단계; 및
 상기 특징 그룹을 구성하는 상기 개별 빈발 부분그래프의 조합으로 이루어진 동시발생 빈발 부분그래프의 분류 상한값과 상기 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨을 비교하는 단계를 포함하며,
 상기 비교 결과에 기초하여 상기 동시발생 빈발 부분그래프의 분류 상한값이 상기 최소 분류레벨보다 작은 경우
 상기 특징 집합에서 상기 동시발생 빈발 부분그래프를 구성하는 개별 빈발 부분그래프를 제외한 나머지 개별 빈발 부분그래프를 상기 동시발생 빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단하는 것을 특징으로 하는 특징 그룹 생성 방법.

청구항 2

제 1 항에 있어서, 상기 그래프 데이터베이스는
 클래스 레이블이 결정된 레이블 그래프(labeled graph)와 클래스 레이블이 결정되지 않은 언레이블 그래프(unlabeled graph)를 구비하는 것을 특징으로 하는 특징 그룹 생성 방법.

청구항 3

제 2 항에 있어서, 상기 특징 그룹 생성 방법은
 그래프 데이터베이스로부터 최소 지지도(MinSup) 이상으로 출현하는 개별 빈발 부분그래프를 추출하는 단계;
 상기 추출한 개별 빈발 부분그래프의 DFS 코드를 생성하는 단계;
 상기 생성한 개별 빈발 부분그래프의 DFS 코드와 상기 추출한 개별 빈발 부분그래프의 정규 연산 코드를 비교하여 상기 추출한 개별 빈발 부분그래프가 최소 DFS 코드를 가지는 정규 그래프인지 판단하는 단계;
 상기 추출한 개별 빈발 부분그래프가 정규 그래프인 경우, 상기 추출한 개별 빈발 부분그래프를 부모 노드로 패턴 확장하여 생성되는 자식 노드의 개별 빈발 부분그래프 사이의 포함관계 정보를 생성하는 단계를 포함하는 것을 특징으로 하는 특징 그룹 생성 방법.

청구항 4

제 3 항에 있어서, 상기 특징 그룹 생성 방법은
 상기 비교 결과에 기초하여 상기 동시발생 빈발 부분그래프의 분류 상한값이 상기 최소 분류레벨보다 큰 경우,
 상기 동시발생 빈발 부분그래프를 구성하는 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프의 포함관계를 판단하는 단계를 더 포함하며,
 상기 제1 개별 빈발 부분그래프와 상기 제2 개별 빈발 부분그래프가 서로 포함 관계인 경우, 상기 특징 집합에서 상기 동시발생 빈발 부분그래프를 구성하는 개별 빈발 부분그래프를 제외한 나머지 개별 빈발 부분그래프를 상기 동시발생 빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단하는 것을 특징으로 하는 특징 그룹 생성 방법.

청구항 5

제 4 항에 있어서, 상기 특징 그룹 생성 방법은
 상기 제1 개별 빈발 부분그래프와 상기 제2 개별 빈발 부분그래프가 서로 불포함 관계인 경우, 상기 동시발생 빈발 부분그래프의 분류값을 계산하는 단계;
 상기 동시발생 빈발 부분그래프의 분류값과 상기 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨

을 비교하는 단계; 및

상기 동시발생 빈발 부분그래프의 분류값이 상기 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨보다 큰 경우, 상기 최소 분류레벨을 가지는 개별 빈발 부분그래프를 상기 특징 그룹에서 삭제하고 상기 동시발생 빈발 부분그래프를 상기 특징 그룹에 추가하여 상기 특징 그룹을 갱신하는 단계를 포함하는 것을 특징으로 하는 특징 그룹 생성 방법.

청구항 6

제 3 항에 있어서, 상기 개별 빈발 부분그래프의 특징 그룹을 생성하는 단계는

상기 그래프 데이터베이스에서 최소 지지도를 가지는 정규 코드의 개별 빈발 부분그래프를 추출하는 단계;

상기 추출한 개별 빈발 부분그래프의 분류값을 계산하는 단계;

상기 계산한 분류값에 기초하여 순차적으로 높은 상위 분류값을 가지는 임계 수의 개별 빈발 부분그래프를 선택하고, 상기 선택한 임계 수의 개별 빈발 부분그래프 중 가장 작은 분류값을 가지는 개별 빈발 부분그래프보다 높은 분류 상한값을 가지는 개별 빈발 부분그래프를 선택하는 단계; 및

상기 선택한 개별 빈발 부분그래프로 특징 그룹을 생성하는 단계를 포함하는 것을 특징으로 하는 특징 그룹의 생성 방법.

청구항 7

제 6 항에 있어서, 상기 특징 그룹의 생성 방법은

상기 생성한 개별 빈발 부분그래프의 특징 그룹을 구성하는 개별 빈발 부분그래프의 분류 상한값의 내림 순서로 배열하는 단계를 더 포함하며,

상기 동시발생 빈발 부분그래프는 상기 개별 빈발 부분그래프의 특징 그룹에서 내림 순서로 높은 분류 상한값을 가지는 개별 빈발 부분그래프를 조합하여 생성되는 것을 특징으로 하는 특징 그룹의 생성 방법.

청구항 8

제 2 항 내지 제 7 항 중 어느 한 항에 있어서, 상기 개별 빈발 부분그래프의 분류값 또는 상기 동시발생 빈발 부분그래프의 분류값은

상기 그래프 데이터베이스에서 서로 다른 클래스 레이블을 가지는 레이블 그래프를 다른 영역으로 분류하는 양의 분류값, 상기 그래프 데이터베이스에서 서로 다른 클래스 레이블을 가지는 레이블 그래프를 동일 영역으로 분류하는 음의 분류값 및 상기 그래프 데이터베이스에서 클래스 레이블을 가지지 않는 언레이블 그래프를 다른 영역에 균등하게 분류하는 양의 분류값의 총합으로 계산되는 것을 특징으로 하는 특징 그룹의 생성 방법.

청구항 9

제 2 항 내지 제 7 항에 중 어느 한 항에 있어서, 상기 개별 빈발 부분그래프의 상한 분류값 또는 상기 동시발생 빈발 부분그래프의 분류 상한값은

상기 그래프 데이터베이스에서 서로 다른 클래스 레이블을 가지는 레이블 그래프를 다른 영역으로 분류하는 양의 분류값과 상기 그래프 데이터베이스에서 클래스 레이블을 가지지 않는 언레이블 그래프를 다른 영역에 균등하게 분류하는 양의 분류값의 합으로 계산되는 것을 특징으로 하는 특징 그룹의 생성 방법.

청구항 10

그래프 데이터베이스에 저장되어 있는 그래프에 기초하여 특징 그룹을 생성하는 장치에 있어서,

상기 그래프 데이터베이스에서 패턴 확장되는 개별 빈발 부분그래프를 추출하는 빈발 부분그래프 마이닝부;

상기 추출한 개별 빈발 부분그래프의 분류값 또는 분류 상한값의 분류 레벨을 계산하는 분류 레벨 계산부;

상기 계산한 분류 레벨에 기초하여 상위 분류 레벨을 가지는 개별 빈발 부분그래프로 특징 그룹을 생성하는 특징 그룹 생성부;

상기 특징 그룹을 구성하는 상기 개별 빈발 부분그래프의 조합으로 동시발생 빈발 부분그래프를 생성하는 동시

발생 빈발 부분그래프 생성부; 및

상기 동시발생 빈발 부분그래프의 분류 상한값과 상기 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류 레벨을 비교하는 비교부를 포함하며,

상기 동시발생 빈발 부분그래프 생성부는 상기 비교부의 비교 결과에 기초하여 상기 동시발생 빈발 부분그래프의 분류 상한값이 상기 최소 분류레벨보다 작은 경우 상기 특징 그룹에서 상기 동시발생 빈발 부분그래프를 제외한 나머지 개별 빈발 부분그래프를 상기 동시발생 빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단 제어하는 것을 특징으로 하는 특징 그룹 생성 장치.

청구항 11

제 10 항에 있어서, 상기 특징 그룹 생성 장치는

상기 비교부의 비교 결과에 기초하여 상기 동시발생 빈발 부분그래프의 분류 상한값이 상기 최소 분류레벨보다 큰 경우, 상기 동시발생 빈발 부분그래프를 구성하는 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프의 포함관계를 판단하는 포함관계 판단부를 더 포함하며,

상기 제1 개별 빈발 부분그래프와 상기 제2 개별 빈발 부분그래프가 서로 포함 관계인 경우, 상기 동시발생 빈발 부분그래프 생성부는 상기 특징 그룹에서 상기 동시발생 빈발 부분그래프를 제외한 나머지 개별 빈발 부분그래프를 상기 동시발생 빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단하는 것을 특징으로 하는 특징 그룹 생성 장치.

청구항 12

제 11 항에 있어서, 상기 특징 그룹 생성 장치는

상기 비교부의 비교 결과에 기초하여 상기 동시발생 빈발 부분그래프의 분류 상한값이 상기 최소 분류레벨보다 크고, 상기 포함관계 판단부의 판단 결과에 기초하여 상기 제1 개별 빈발 부분그래프와 상기 제2 개별 빈발 부분그래프가 서로 불포함 관계인 경우,

상기 최소 분류레벨을 가지는 개별 빈발 부분그래프를 상기 특징 그룹에서 삭제하고 상기 동시발생 빈발 부분그래프를 상기 특징 그룹에 추가하여 상기 특징 그룹을 갱신하는 갱신 제어부를 더 포함하는 것을 특징으로 하는 특징 그룹 생성 장치.

명세서

기술분야

[0001] 본 발명은 그래프 데이터베이스에서 분류력이 높은 특징 집합을 선택하기 위한 방법에 관한 것으로, 보다 구체적으로 그래프 데이터베이스에 저장되어 있는 레이블(labeled) 그래프와 레이블 되지 않은 그래프(이하 언레이블 그래프라 언급)에서 높은 분류력을 가지는 특징 집합을 선택하며, 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 추가한 추가 동시발생 빈발 부분그래프의 분류력은 동시발생 빈발 부분그래프의 상한 분류력보다 작거나 같음에 기초하여 기생성된 특징 그룹의 최소 분류력보다 작은 상한 분류력을 가지는 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 조합하여 추가 동시발생 빈발 부분그래프를 생성하는 것을 중지함으로써 동시발생 빈발 부분그래프의 조합에 소요되는 연산량을 줄일 수 있는 특징 그룹 생성 방법에 관한 것이다.

배경기술

[0002] 그래프 분류는 화합물, XML, 웹 문서, 소셜 네트워크 등과 같은 그래프 구조를 갖는 데이터를 분류하는 기술이다. 그래프 분류를 위해 구조적 정보를 갖는 빈발 부분그래프들이 특징으로 널리 사용되고 있다. 여기서 빈발 부분그래프란 그래프 데이터베이스에서 최소 지지도 이상으로 발생하는 부분그래프를 의미하는데, 빈발 부분그래프는 그래프의 고유한 특성을 나타내어 그래프의 분류, 군집화, 인덱싱 등에서 사용된다. 빈발 부분그래프는 그래프 데이터베이스에 따라 수만에서 수십만 개가 생성되기 때문에 그래프 데이터베이스의 분류에 도움이 되는 빈발 부분그래프로 이루어진 특징 그룹을 선택해야 한다.

[0003] 종래 그래프 분류를 위한 다양한 특징 그룹 선택에 대한 연구가 수행되고 있다. 종래 특징 그룹 선택 방법 중

하나인 도약 탐색(leap search) 방식은 빈발 부분그래프 마이닝 단계에서 G-test 점수가 낮은 특징들에 대한 탐색공간을 줄임으로써 빠른 수행시간과 높은 분류 정확도를 가지는 특징 그룹을 선택할 수 있다. 한편 다른 종래 특징 그룹 선택 방법인 모델 기반 탐색 트리(model based search tree)는 인포메이션 게인(information gain)이 높은 특징의 포함여부로 그래프 DB를 분할하는 과정을 반복하여 높은 분류 성능을 가지는 특징 그룹을 선택할 수 있다. 이 밖에도 PCC(pearson correlation coefficient)[3], CBA(classification based on associations) 등과 같은 특징 그룹 선택 기법들이 그래프 분류 연구에서 사용되었다.

[0004] 앞서 설명한 종래 특징 그룹 선택 기법은 클래스 레이블이 결정된 레이블 그래프를 훈련용 그래프로 사용하는 감독기반 특징 평가 방식이다. 그러나, 그래프 분류의 많은 애플리케이션에서는 훈련용 그래프로 사용할 충분한 양의 레이블 그래프를 얻기 어렵다. 왜냐하면, 그래프 데이터의 클래스 레이블을 결정하는 것은 도메인 전문가의 분석과 많은 시간이 요구되는 고비용의 과정이기 때문이다. 예를 들어, 화합물에 대한 항암 활성 정보 클래스 레이블을 결정하기 위해서는, 전문가가 대상 화합물이 포함된 약물을 다양한 피실험자들에게 복용시키고 분석하는 과정을 반복해야 하기 때문에 원하는 결과를 수집하기까지 어려울 뿐만 아니라 오랜 시간과 비용이 소요된다.

[0005] 충분한 양의 레이블 그래프를 요구하는 종래의 감독기반 특징 그룹 선택 방식에서 특징 그룹을 선택하는 특징 평가 함수는 충분한 양의 레이블 그래프로 훈련시키지 않으면, 특징들과 클래스의 상관관계를 계산할 수 없어 특징 선택 성능이 저하되며 따라서 충분한 양의 레이블 데이터를 요구한다. 그러나 레이블 그래프를 획득하는 것은 고비용으로 종래 감독기반 특징 그룹 선택 방식에서는 한정된 수의 레이블 그래프만을 이용하기 때문에 높은 분류 성능을 가지는 특징 그룹을 선택하기 곤란하다는 문제점을 가진다.

[0006] 종래 감독기반 특징 선택 방식이 가지는 문제점을 해결하기 위하여, 소량의 클래스 레이블이 결정된 레이블 그래프와 대량의 클래스 레이블이 결정되지 않은 그래프를 함께 이용한 부분 감독기반 특징 선택 방식이 제안되었다. 그래프는 미리 정해진 특징이 없기 때문에, 빈발 부분그래프를 마이닝하여 빈발 부분그래프 특징을 먼저 생성한다. 종래 부분감독 특징 선택 방식에서는 생성한 빈발 부분그래프 특징에 기초하여 그래프 데이터베이스에 저장되어 있는 레이블 그래프와 레이블 되지 않은 그래프의 분류 특성을 평가한다. 종래 부분 감독기반 특징 그룹 선택 방식은 소량의 레이블 그래프들로 구성된 데이터베이스에서 종래 감독기반 특징 선택 방식의 문제점을 해결하였다.

[0007] 그러나 위에서 설명한 종래 감독기반 특징 그룹 선택 방식과 부분 감독기반 특징 그룹 선택 방식은 모두 개별 빈발 부분그래프만의 분류력을 평가하는 것으로, 개별 빈발 부분그래프의 조합으로 생성되는 동시발생 빈발 부분그래프의 분류력에 대해서는 평가하지 않았다.

[0008] 분류력이 낮은 개별 빈발 부분그래프들의 동시 발생을 특징으로 사용하면 더 높은 분류력을 가질 수 있다는 것을 도 1을 참고로 설명한다. 도 1은 개별 빈발 부분그래프 특징보다 동시에 발생하는 빈발 부분그래프들을 특징으로 사용할 때의 분류력이 더 높을 수 있음을 보여주는 예시이다. 원으로 표현된 그래프는 개별 빈발 부분그래프(A-B)만을 가지고 있는 그래프들을, 삼각형으로 표현된 그래프는 개별 빈발 부분그래프(C-D)만을 가지고 있는 그래프들을, 사각형으로 표현된 그래프는 개별 빈발 부분그래프(A-B, C-D)를 모두 가지고 있는 그래프들을 나타낸다.

[0009] 개별 빈발 부분그래프(A-B)는 포지티브 레이블 그래프와 네거티브 레이블 그래프를 개별 빈발 부분그래프(A-B)로 나누어진 영역의 동일한 영역에 분포시키고 언레이블 그래프도 고르게 분포시키지 못하므로 분류력이 낮다. 한편, 개별 빈발 부분그래프(C-D)는 포지티브 레이블 그래프와 네거티브 레이블 그래프를 개별 빈발 부분그래프(C-D)로 나누어진 영역의 서로 다른 영역에 분포시키지만 언레이블 그래프를 고르게 분포시키지 못하므로 분류력이 높지 않다. 한편, 개별 빈발 부분그래프(A-B, C-D)의 조합으로 이루어진 동시발생 빈발 부분그래프의 경우 포지티브 레이블 그래프와 네거티브 레이블 그래프를 서로 다른 영역에 분포시키며 언레이블 그래프를 고르게 분포시키므로 높은 분류력을 가짐을 알 수 있다.

발명의 내용

해결하려는 과제

[0010] 동시발생 빈발 부분그래프는 빈발 부분그래프의 조합으로부터 생성되는데, 최소 지지도가 낮을수록 많은 수의 빈발 부분그래프가 생성되며 생성된 빈발 부분그래프의 조합으로부터 동시발생 빈발 부분그래프를 생성하기 위해서는 지수 연산량이 필요하므로 동시발생 빈발 부분그래프의 생성은 고비용이 소요되는 작업이다. 더욱이 레

이들 그래프와 언레이블 그래프를 모두 사용하여 높은 분류력을 가지는 동시발생 빈발 부분그래프 조합을 생성하는 경우에는 더욱 많은 빈발 부분그래프를 생성하고 이로부터 생성되는 동시발생 빈발 부분그래프의 생성에는 더욱 많은 연산량을 필요로 한다.

[0011] 따라서 본 발명이 이루고자 하는 목적은 레이블 그래프와 언레이블 그래프를 이용하여 특징 그룹을 선택하는 부분 감독기반의 특징 그룹 선택 방식에서 동시발생 빈발 부분그래프의 탐색 공간을 줄여 동시발생 빈발 부분그래프의 생성에 소요되는 연산량을 줄인 특징 그룹 선택 방법을 제공하는 것이다.

[0012] 본 발명이 이루고자 하는 다른 목적은 분류값이 높은 개별 빈발 부분그래프뿐만 아니라 동시발생 빈발 부분그래프로 특징 그룹을 선택하는 방법을 제공하는 것이다.

과제의 해결 수단

[0013] 본 발명의 목적을 달성하기 위하여 본 발명에 따른 특징 그룹 선택 방법은 그래프 데이터베이스에서 개별 빈발 부분그래프의 분류값 또는 상한값에 기초하여 상위 분류레벨을 가지는 개별 빈발 부분그래프로 특징 그룹을 생성하는 단계와, 특징 그룹을 구성하는 개별 빈발 부분그래프의 조합으로 이루어진 동시발생 빈발 부분그래프의 분류 상한값과 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨을 비교하는 단계를 포함하며, 비교 결과에 기초하여 동시발생 빈발 부분그래프의 분류 상한값이 최소 분류레벨보다 작은 경우 특징 집합에서 동시발생 빈발 부분그래프를 제외한 나머지 개별 빈발 부분그래프를 동시발생 빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단하는 것을 특징으로 한다.

[0014] 본 발명의 일 실시예에 따른 특징 그룹 선택 장치는 그래프 데이터베이스에서 패턴 확장되는 개별 빈발 부분그래프를 추출하는 빈발 부분그래프 마이닝부와, 추출한 개별 빈발 부분그래프의 분류값 또는 분류 상한값의 분류레벨을 계산하는 분류 레벨 계산부와, 계산한 분류 레벨에 기초하여 상위 분류 레벨을 가지는 개별 빈발 부분그래프로 특징 그룹을 생성하는 특징 그룹 생성부와, 특징 그룹을 구성하는 개별 빈발 부분그래프의 조합으로 동시발생 빈발 부분그래프를 생성하는 동시발생 빈발 부분그래프 생성부와, 동시발생 빈발 부분그래프의 분류 상한값과 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨을 비교하는 비교부를 포함하며, 동시발생 빈발 부분그래프 생성부는 비교부의 비교 결과에 기초하여 동시발생 빈발 부분그래프의 분류 상한값이 최소 분류레벨보다 작은 경우 특징 그룹에서 동시발생 빈발 부분그래프를 제외한 나머지 개별 빈발 부분그래프를 동시발생 빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단 제어하는 것을 특징으로 한다.

발명의 효과

[0015] 본 발명에 따른 특징 그룹 선택 방법은 종래 특징 그룹 선택 방법과 비교하여 다음과 같은 다양한 효과를 가진다.

[0016] 첫째, 본 발명에 따른 특징 그룹 선택 방법은 소량의 레이블 그래프와 더불어 언레이블 그래프를 이용하여 분류특징을 선택함으로써, 소량의 레이블 그래프로부터 선택한 분류 특징보다 높은 분류력을 가지는 분류 특징을 선택할 수 있다.

[0017] 둘째, 본 발명에 따른 특징 그룹 선택 방법은 분류값이 높은 상위 개별 빈발 부분그래프만을 이용하여 동시발생 빈발 부분그래프를 생성함으로써, 동시발생 빈발 부분 그래프의 생성에 소요되는 연산량을 줄일 수 있다.

[0018] 셋째, 본 발명에 따른 특징 그룹 선택 방법은 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 추가한 추가 동시발생 빈발 부분그래프의 분류 레벨은 동시발생 빈발 부분그래프의 상한 분류값보다 작거나 같음에 기초하여 기생성된 특징 그룹의 최소 분류 레벨보다 작은 상한 분류값을 가지는 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 조합하여 추가 동시발생 빈발 부분그래프를 생성하는 것을 중지함으로써 동시발생 빈발 부분그래프의 조합에 소요되는 연산량을 줄일 수 있다.

도면의 간단한 설명

[0019] 도 1은 분류력이 낮은 개별 빈발 부분그래프들의 동시 발생을 특징으로 사용하면 더 높은 분류력을 가질 수 있다는 것을 설명하기 위한 도면이다.

도 2는 본 발명의 일 실시예에 따른 특징 그룹 선택 장치를 설명하기 위한 기능블록도이다.

도 3은 본 발명에 따른 특징 그룹 선택 방법을 설명하기 위한 흐름도이다.

도 4는 본 발명에 따른 빈발 부분그래프 마이닝 방법의 흐름도를 도시하고 있다.

도 5는 본 발명에 따른 빈발 부분그래프 마이닝 방법의 흐름도를 도시하고 있다.

도 6은 개별 빈발 부분그래프와 개별 빈발 부분그래프 조합으로 이루어진 동시발생 빈발 부분그래프의 분류레벨 관계를 도시하고 있다.

도 7은 본 발명에 따른 특징 그룹을 갱신하는 방법(A)의 일 예를 설명하는 흐름도이다.

도 8은 본 발명에 따른 특징 그룹을 갱신하는 방법(A)의 다른 예를 설명하는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하 첨부한 도면을 참고로 본 발명에 따른 특징 그룹 선택 방법 및 그 장치에 대해 보다 구체적으로 설명한다.

[0021] 도 2는 본 발명의 일 실시예에 따른 특징 그룹 선택 장치를 설명하기 위한 기능블록도이다.

[0022] 도 2를 참고로 보다 구체적으로 살펴보면, 그래프 데이터베이스(110)에는 클래스 레이블이 결정된 레이블(labeled) 그래프와 클래스 레이블이 결정되지 않은 언레이블(unlabeled) 그래프가 저장되어 있다. 여기서 클래스 레이블이란 특정 대상에 대한 그래프의 활성 또는 비활성의 성질을 나타내는 정보로, 예를 들어 암의 치료 물질의 경우 암이라는 특정 대상에서 항암 성질을 가지는 치료물질의 화학 성분 그래프는 활성 클래스로 레이블되며, 항암 성질을 가지지 않는 치료물질의 화학 성분 그래프는 비활성 클래스로 레이블된다. 한편, 항암 성질을 가지는지 결정되지 않은 치료물질의 화학 성분 그래프는 클래스가 결정되지 않은 언레이블 그래프이다.

[0023] 빈발 부분그래프 마이닝부(120)는 그래프 데이터베이스부(110)에 저장되어 있는 레이블 그래프와 언레이블 그래프에서 최소 지지도 이상으로 출현하는 개별 빈발 부분그래프를 추출하고, 추출한 개별 빈발 부분그래프의 DFS 코드와 최소 DFS 코드를 가지는 정규 연산 코드를 비교하여 추출한 개별 빈발 부분그래프의 DFS 코드가 최소 DFS 코드인 경우 추출한 개별 빈발 부분그래프를 깊이우선탍색 방식으로 패턴 확장하여 레이블 그래프와 언레이블 그래프에 포함되어 있는 모든 개별 빈발 부분그래프를 추출한다. 한편, 포함관계 생성부(130)는 빈발 부분그래프 마이닝 과정에서 패턴 확장되는 개별 빈발 부분그래프 사이의 포함관계를 판단하고, 판단한 개별 빈발 부분그래프 사이의 포함관계를 나타내는 포함관계 정보를 생성한다. 제1 빈발 부분그래프에서 패턴 확장되어 생성되는 제2 빈발 부분그래프는 제1 빈발 부분그래프를 포함하는 것으로 판단된다.

[0024] 분류 레벨 계산부(140)는 추출한 개별 빈발 부분그래프의 분류 레벨을 계산하며, 특징 그룹 생성부(150)는 계산한 개별 빈발 부분그래프의 분류 레벨에 기초하여 특징 그룹을 구성하는 개별 빈발 부분그래프를 선택한다. 분류 레벨 계산부(140)는 개별 빈발 부분그래프의 분류값 또는 분류 상한값을 계산하며, 특징 그룹 생성부(150)는 추출한 개별 빈발 부분그래프 중 순차적으로 분류값이 높은 임계 수의 개별 빈발 부분그래프를 후보 특징 그룹으로 선택하고, 후보 특징 그룹을 구성하는 개별 빈발 부분그래프 중 가장 낮은 분류값을 가지는 개별 빈발 부분그래프의 분류값보다 높은 분류 상한값을 가지는 개별 빈발 부분그래프를 선택한다. 특징 그룹 생성부(150)는 후보 특징 그룹과 후보 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류값보다 높은 분류 상한값을 가지는 개별 빈발 부분그래프로 특징 그룹을 생성한다.

[0025] 동시발생 빈발 부분그래프 생성부(160)는 특징 그룹에 구비되어 있는 개별 빈발 부분그래프의 상한값에 따라 특징 그룹의 개별 빈발 부분그래프를 내림차순으로 정렬하고, 특징 그룹의 개별 빈발 부분그래프를 정렬 순서에 따라 조합하여 동시발생 빈발 부분그래프를 생성한다.

[0026] 비교부(170)는 동시발생 빈발 부분그래프의 분류 상한값과 생성한 특징 그룹의 개별 빈발 부분그래프의 최소 분류레벨을 비교한다. 동시발생 빈발 부분그래프 생성부(160)는 비교부(170)의 비교 결과에 기초하여 특징 그룹에서 동시발생 빈발 부분그래프를 구성하는 빈발 부분그래프 이외의 나머지 빈발 부분그래프를 동시발생 빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프를 생성할지 판단한다.

[0027] 한편, 동시발생 빈발 부분그래프의 분류 상한값이 특징 그룹의 개별 빈발 부분그래프의 최소 분류레벨보다 큰 경우, 포함관계 판단부(180)는 동시발생 빈발 부분그래프를 구성하는 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프의 포함관계 정보에 기초하여 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프 사이의 포함관계를 판단한다. 동시발생 빈발 부분그래프 생성부(160)는 포함관계 판단부(170)의 판단 결과에 기초하여 특징 그룹에서 동시발생 빈발 부분그래프를 구성하는 빈발 부분그래프 이외의 나머지 빈발 부분그래프를 동시발생

빈발 부분그래프에 추가하여 새로운 동시발생 빈발 부분그래프를 생성할지 판단한다.

[0028] 갱신 제어부(190)는 비교부(170)의 비교 결과와 포함관계 판단부(180)의 판단 결과에 기초하여 동시발생 부분 그래프를 특징 그룹에 갱신할지 결정한다.

[0029] 도 3은 본 발명에 따른 특징 그룹 선택 방법을 설명하기 위한 흐름도이다.

[0030] 도 3을 참고로 보다 구체적으로 살펴보면, 그래프 데이터베이스에 저장되어 있는 레이블 그래프와 언레이블 그래프를 빈발 부분그래프 마이닝하여 개별 빈발 부분그래프를 추출한다(S10). 빈발 부분그래프 마이닝 방법의 흐름도를 도시하고 있는 도 4를 참고로 보다 구체적으로 살펴보면, 복잡한 구조를 가지는 그래프들이 대용량으로 저장되어 있는 그래프 데이터베이스로부터 최소 지지도(MinSup) 이상으로 출현하는 개별 빈발 부분그래프를 추출한다(S11). 개별 빈발 부분그래프(FG)은 아래의 수학적식(1)과 같이 정의된다.

[0031] [수학적식 1]

$$FG = \frac{|g \subseteq G|}{|GD|} > MinSup$$

[0032] 여기서 $|g \subseteq G|$ 는 부분그래프(g)를 가지는 그래프(G)의 개수이며 $|GD|$ 는 그래프 데이터베이스에 존재하는 그래프의 개수이다.

[0034] 추출한 개별 빈발 부분그래프의 DFS 코드를 생성한다(S12). 추출한 개별 빈발 부분그래프의 DFS 코드는 추출한 개별 빈발 부분그래프를 구성하는 각 에지에 대해 각 에지의 확장 순서로 깊이 우선 탐색(DFS) 방식으로 생성되는데, 개별 빈발 부분그래프를 구성하는 각 에지의 DFS 코드는 에지를 구성하는 2개 버텍스(vertex) 식별자, 2개 버텍스 레이블 및 2개 버텍스를 연결하는 에지 레이블로 구성된다. 여기서 개별 빈발 부분그래프의 DFS 코드는 버텍스 식별자의 순서로, 즉 깊이 우선 탐색 방식으로 에지를 확장하며 확장한 에지에 대해 각각 DFS 코드를 생성한다. 개별 빈발 부분그래프를 구성하는 각 에지의 DFS 코드는 아래의 수학적식(2)과 같이 표현된다.

[0035] [수학적식 2]

[0036] DFS 코드={i, j, l(v_i), l(e(v_i, v_j)), l(v_j)}

[0037] 여기서 i, j는 각각 에지(e)를 구성하는 2개 버텍스(vertex) 식별자, l(v_i)와 l(v_j)는 각각 2개 버텍스 레이블, l(e(v_i, v_j))는 2개 버텍스를 연결하는 에지 레이블이다.

[0038] 생성한 개별 빈발 부분그래프의 DFS 코드를 정규화 연산하여 추출한 빈발 부분그래프가 최소 DFS 코드를 가지는 정규 그래프인지 판단한다(S13). 판단 결과 추출한 개별 빈발 부분그래프가 정규 그래프인 경우 추출한 개별 빈발 부분그래프에서 에지를 1개씩 확장하여 추출한 개별 빈발 부분그래프로부터 생성되는 새로운 부분그래프를 생성한다(S14). 생성한 새로운 부분그래프에 대해 앞서 설명한 S11 단계 내지 S14 단계를 반복 수행하여 그래프 데이터베이스에 존재하는 최소 지지도를 만족하는 모든 개별 빈발 부분그래프를 탐색한다. 이때 정규그래프인 개별 빈발 부분그래프와 개별 빈발 부분그래프로부터 확장된, 정규그래프인 개별 빈발 부분그래프는 서로 부모 노드와 자식 노드의 관계를 가지며 자식 노드인 개별 빈발 부분그래프는 부모 노드인 개별 빈발 부분그래프를 포함한다. 즉, 빈발 부분그래프 마이닝 과정에서 생성되는 개별 빈발 부분그래프는 부모 노드와 자식 노드의 관계를 가지는 트리로 생성될 수 있으며, 부모 노드와 자식 노드의 관계로부터 동시발생 빈발 부분그래프를 구성하는 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프의 포함관계를 판단할 수 있다. 이러한 부모 노드와 자식 노드의 포함관계 정보를 생성한다(S15).

[0039] 바람직하게, 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프의 포함관계 정보는 아래의 수학적식(3)의 포함관계 행렬로 표현된다.

[0040] [수학적식 3]

[0041] I=[I_{ij}]

[0042] 여기서 I_{ij}는 i번째 빈발 부분그래프가 j번째 빈발 부분그래프를 포함하는 경우 1의 값을 가지며, i번째 빈발 부분그래프가 j번째 빈발 부분그래프를 포함하지 않는 경우 0의 값을 가진다.

- [0043] 개별 빈발 부분그래프의 DFS 코드 중 최소 DFS 코드를 개별 빈발 부분그래프의 정규 코드 형태로 사용하여 개별 빈발 부분그래프를 유일하게 표현한다. 개별 빈발 부분그래프를 정규 코드 형태로 표현하는 것을 정규화 연산(canonized operation)이라 정의하고 정규 코드 형태로 표현되는 개별 빈발 부분그래프를 개별 빈발 부분그래프의 정규 그래프라 언급한다. 정규화 연산을 위하여 개별 빈발 부분그래프의 최소 DFS코드를 찾기 위한 과정은 다음과 같다. 개별 빈발 부분그래프의 에지들을 DFS코드로 표현하고, 각 에지의 DFS코드 중 가장 작게 표현된 값을 탐색 시작 지점으로 설정한다. 탐색 시작 지점부터 깊이우선탐색(DFS) 순서에 따라 버텍스의 식별자를 순차적으로 부여한다. 만약 동일 버텍스가 여러 개의 에지를 가지면, 가장 작은 DFS코드로 표현되는 에지를 선택하여 탐색한다. 만약 가장 작게 표현되는 에지가 여러 개 존재할 경우, 하나의 에지를 선택하여 모두 탐색한 후, 나머지 에지들에 대해서도 재귀적으로 모든 에지를 탐색한다. 탐색하여 얻은 DFS코드 집합 중에서 가장 작게 표현되는 최소 DFS코드를 찾는다.
- [0044] 한편, 판단 결과 추출한 개별 빈발 부분그래프가 최소 DFS 코드를 가지는 정규 그래프가 아닌 경우, 추출한 개별 빈발 부분그래프를 제거하여 개별 빈발 부분그래프가 중복하여 확장되는 것을 방지한다.
- [0045] 다시 도 3을 참고로 살펴보면, 추출한 개별 빈발 부분그래프의 분류값 또는 분류 상한값에 기초하여 상위 분류 레벨을 가지는 개별 빈발 부분그래프로 특징 그룹을 생성한다(S20). 도 5를 참고로 보다 구체적으로 살펴보면, 추출한 개별 빈발 부분그래프의 분류값을 계산한다(S21). 여기서 개별 빈발 부분그래프의 분류값은 그래프 데이터베이스에서 서로 다른 클래스 레이블을 가지는 레이블 그래프를 개별 빈발 부분그래프 특성에 의해 다른 영역으로 분류하는 양의 분류값, 그래프 데이터베이스에서 서로 다른 클래스 레이블을 가지는 레이블 그래프를 개별 빈발 부분그래프 특성에 의해 동일 영역으로 분류하는 음의 분류값 및 그래프 데이터베이스에서 클래스 레이블을 가지지 않는 언레이블 그래프를 개별 빈발 부분그래프 특성에 의해 다른 영역에 균등하게 분류하는 양의 분류값의 총합으로 계산된다.
- [0046] 개별 빈발 부분그래프의 분류값을 계산하여 상위 분류값을 가지는 임계수의 개별 빈발 부분그래프를 선택하여 후보 특성 그룹을 생성한다(S22). 생성한 후보 특성 그룹에서 최소 분류값을 가지는 개별 빈발 부분그래프보다 높은 분류값을 가지는, 그래프 데이터베이스에서 추출되는 개별 빈발 부분그래프로 후보 특성 그룹을 갱신하여 높은 상위 분류값을 가지는 임계 수의 개별 빈발 부분그래프를 선택하고(S23), 선택한 임계 수의 개별 빈발 부분그래프 중 가장 작은 분류값을 가지는 개별 빈발 부분그래프보다 높은 분류 상한값을 가지는 개별 빈발 부분그래프를 선택한다(S24). 선택한 개별 빈발 부분그래프로 특징 그룹을 생성한다(S25).
- [0047] 여기서, 특징 그룹은 임계수의 개별 빈발 부분그래프 이외에 선택한 임계 수의 개별 빈발 부분그래프 중 최소 분류값을 가지는 개별 빈발 부분그래프보다 높은 분류 상한값을 가지는 개별 빈발 부분그래프를 선택하여 포함함으로써, 다른 개별 빈발 부분그래프와 조합에 의해 분류력이 향상될 수 있는 잠재 분류력이 높은 개별 빈발 부분그래프도 특징 그룹에 추가 구성시킨다.
- [0048] 다시 도 3을 참고로 살펴보면, 생성한 특징 그룹을 구성하는 개별 빈발 부분그래프의 분류 상한값의 내림순서로 개별 빈발 부분그래프를 정렬하고, 분류 상한값의 내림순서로 다수의 개별 빈발 부분그래프를 조합하여 동시발생 빈발 부분그래프를 생성한다(S30). 예를 들어, 특징 그룹을 구성하는 개별 빈발 부분그래프(f1, f2, f3, f4, f5)가 존재하는 경우, 개별 빈발 부분그래프의 분류 상한값에 따라 개별 빈발 부분그래프를 내림 순서로 정렬한다. 정렬한 개별 빈발 부분그래프(F1, f2, f3, f4, f5)에서 개별 빈발 부분그래프를 순차적으로 조합하여 동시발생 빈발 부분그래프(F1F2, F1F3, F1F4, F1F5)를 생성한다.
- [0049] 동시발생 빈발 부분그래프의 분류 상한값을 계산하고 계산한 동시발생 빈발 부분그래프의 분류 상한값과 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨을 비교한다(S40). 비교 결과 동시발생 빈발 부분그래프의 분류 상한값이 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨보다 작은 경우 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 추가한 추가 동시발생 빈발 부분그래프의 분류 레벨은 동시발생 빈발 부분그래프의 상한 분류값보다 작거나 같음에 기초하여 동시발생 빈발 부분그래프에 새로운 개별 빈발 부분그래프를 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단한다(S50).
- [0050] 개별 빈발 부분그래프와 개별 빈발 부분그래프 조합으로 이루어진 동시발생 빈발 부분그래프의 분류레벨 관계를

도시하고 있는 도 6을 참고로 살펴보면, 제1 개별 빈발 부분그래프(f1)와 제2 빈발 부분그래프(f2)의 조합으로 이루어진 제1 동시발생 빈발 부분그래프(f1f2)의 분류 상한값이 A_{MAX} 인 경우 동시발생 빈발 부분그래프(f1f2)에 새로운 개별 빈발 부분그래프(f3)를 추가하여 생성되는 제2 동시발생 빈발 부분그래프(f1f2f3)의 분류값은 제1 동시발생 빈발 부분그래프(f1f2)의 분류 상한값을 초과하지 못한다. 따라서 제1 동시발생 빈발 부분그래프(f1f2)의 분류 상한값이 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨을 초과하지 못하는 경우 제1 동시발생 빈발 부분그래프(f1f2)에 새로운 개별 빈발 부분그래프를 추가하여 생성되는 동시발생 빈발 부분그래프의 분류값도 특징 그룹을 구성하는 개별 빈발 부분그래프의 최소 분류레벨을 초과하지 못한다. 따라서 제1 동시발생 빈발 부분그래프(f1f2)에 새로운 개별 빈발 부분그래프를 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단한다.

[0051] 도 7은 특징 그룹을 갱신하는 방법(A)의 일 예를 설명하는 흐름도이다.

[0052] 도 7을 참고로 보다 구체적으로 살펴보면, 생성한 동시발생 빈발 부분그래프의 분류 상한값이 특징 그룹의 최소 분류레벨보다 큰 경우, 생성한 동시발생 빈발 부분그래프를 구성하는 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프의 포함관계를 포함관계 행렬식을 이용하여 판단한다(S111). 판단 결과 제2 개별 빈발 부분그래프가 제1 개별 빈발 부분그래프를 포함하는 경우, 생성한 동시발생 빈발 부분그래프에 특징 그룹의 새로운 개별 빈발 부분그래프를 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단한다(S113).

[0053] 한편, 판단 결과 제2 개별 빈발 부분그래프가 제1 개별 빈발 부분그래프를 포함하지 않는 경우, 생성한 동시발생 빈발 부분그래프의 분류값을 계산하고, 계산한 동시발생 빈발 부분그래프의 분류값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 큰지 판단한다(S115). 판단결과 생성한 동시발생 빈발 부분그래프의 분류값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 큰 경우 생성한 동시발생 빈발 부분그래프를 특징 그룹에 추가하고 특징 그룹에서 최소 분류레벨을 가지는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프를 삭제하여 특징 그룹을 갱신한다(S117).

[0054] 도 8은 특징 그룹을 갱신하는 방법(A)의 다른 예를 설명하는 흐름도이다.

[0055] 도 8을 참고로 보다 구체적으로 살펴보면, 생성한 동시발생 빈발 부분그래프의 분류 상한값이 특징 그룹의 최소 분류레벨보다 큰 경우 동시발생 빈발 부분그래프를 구성하는 제1 개별 빈발 부분그래프와 제2 개별 빈발 부분그래프의 포함관계를 포함관계 행렬식을 이용하여 판단한다(S121). 판단 결과 제2 개별 빈발 부분그래프가 제1 개별 빈발 부분그래프를 포함하는 경우, 동시발생 빈발 부분그래프에 특징 그룹의 새로운 개별 빈발 부분그래프를 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단한다(S122).

[0056] 한편, 판단 결과 제2 개별 빈발 부분그래프가 제1 개별 빈발 부분그래프를 포함하지 않는 경우, 생성한 동시발생 빈발 부분그래프의 분류 상한값을 계산하고, 계산한 동시발생 빈발 부분그래프의 분류 상한값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 큰지 판단한다(S123). 판단결과 생성한 동시발생 빈발 부분그래프의 분류 상한값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 큰 경우 생성한 동시발생 빈발 부분그래프의 분류값을 계산한다(S124). 생성한 동시발생 빈발 부분그래프의 분류값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 큰지 판단하여(S125), 생성한 동시발생 빈발 부분그래프의 분류값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 큰 경우 생성한 동시발생 빈발 부분그래프를 특징 그룹에 추가하고 특징 그룹에서 최소 분류레벨을 가지는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프를 삭제하여 특징 그룹을 갱신한다(S126). 한편, 생성한 동시발생 빈발 부분그래프의 분류 상한값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 작은 경우 또는 생성한 동시발생 빈발 부분그래프의 분류값이 특징 그룹을 구성하는 개별 빈발 부분그래프 또는 동시발생 빈발 부분그래프의 최소 분류레벨보다 작은 경우 생성한 동시발생 빈발 부분그래프에 특징 그룹의 새로운 개별 빈발 부분그래프를 추가하여 새로운 동시발생 빈발 부분그래프의 생성을 중단한다.

[0057] 한편, 상술한 본 발명의 실시 예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록 매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다.

[0058] 상기 컴퓨터로 읽을 수 있는 기록 매체는 마그네틱 저장 매체(예를 들어, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장 매체를 포함한다.

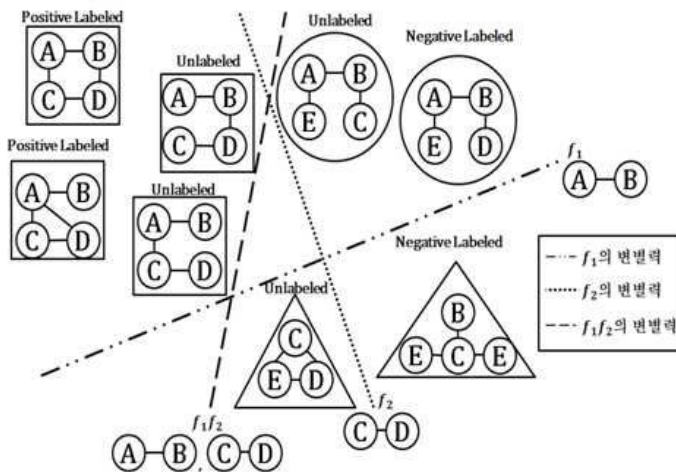
[0059] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다

부호의 설명

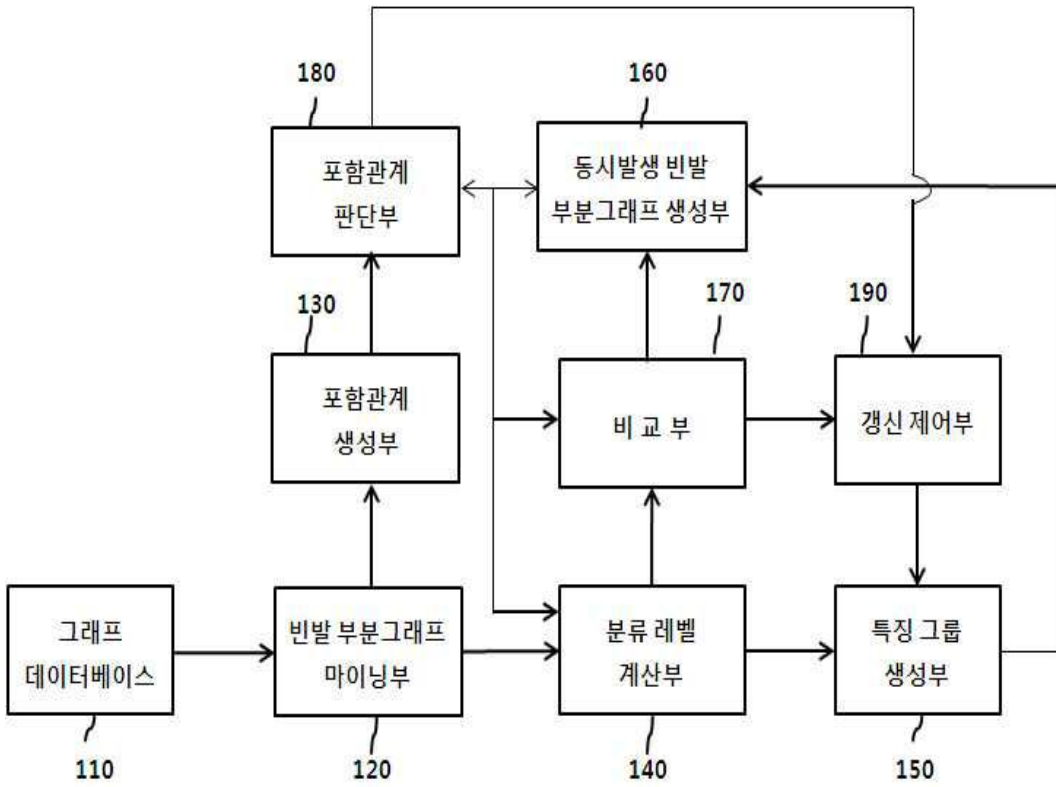
[0060] 110: 그래프 데이터베이스 120: 빈발 부분그래프 마이닝부
130: 포함관계 생성부 140: 분류레벨 계산부
150: 특징 그룹 생성부 160: 동시발생 빈발 부분그래프 생성부
170: 비교부 180: 포함관계 판단부
190: 갱신 제어부

도면

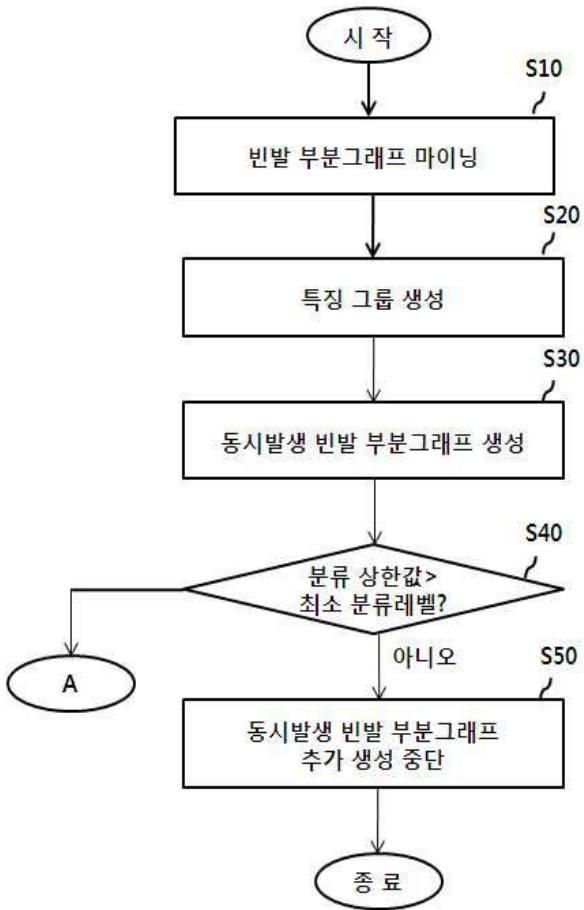
도면1



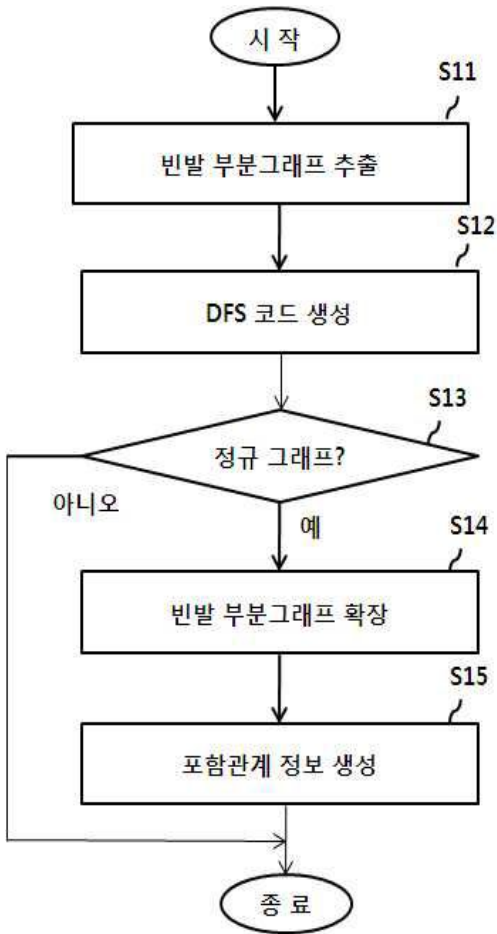
도면2



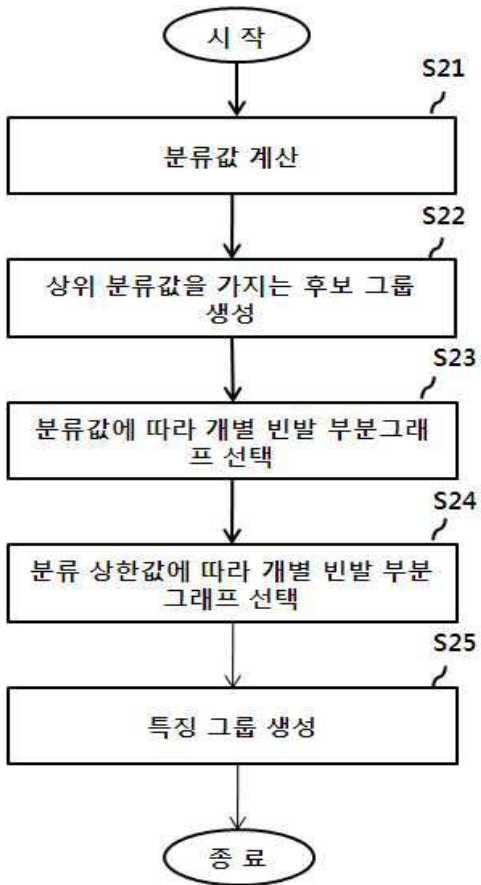
도면3



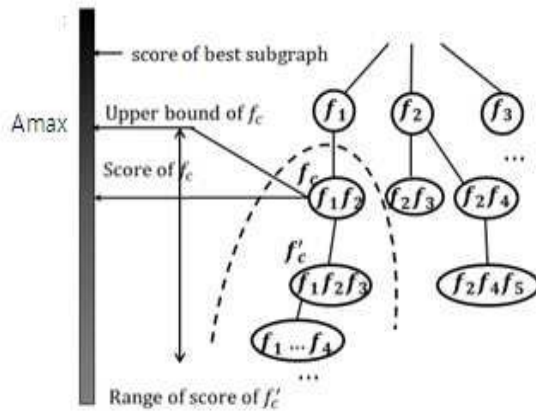
도면4



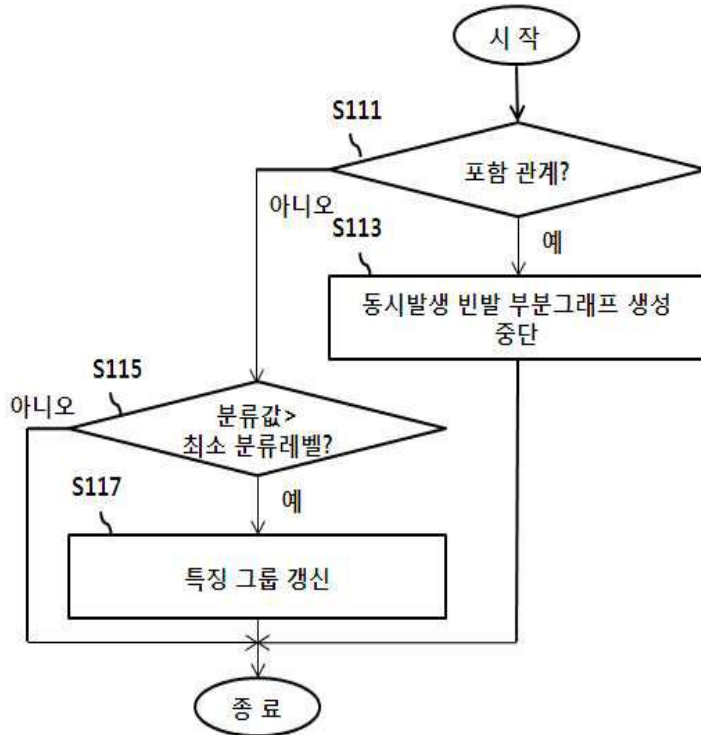
도면5



도면6



도면7



도면8

