



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0007928

(43) 공개일자 2015년01월21일

(51) 국제특허분류(Int. Cl.)

G06F 17/00 (2006.01)

(21) 출원번호 10-2014-0029214

(22) 출원일자 2014년03월12일

심사청구일자 2014년03월12일

(30) 우선권주장

1020130082108 2013년07월12일 대한민국(KR)

(71) 출원인

경희대학교 산학협력단

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

(72) 발명자

이영구

대전 서구 청사서로 11, 106동 302호 (월평동, 무지개아파트)

김건희

경기 안양시 만안구 박달로479번길 31, 5동 206호 (박달동, 신한아파트)

(뒷면에 계속)

(74) 대리인

서재승

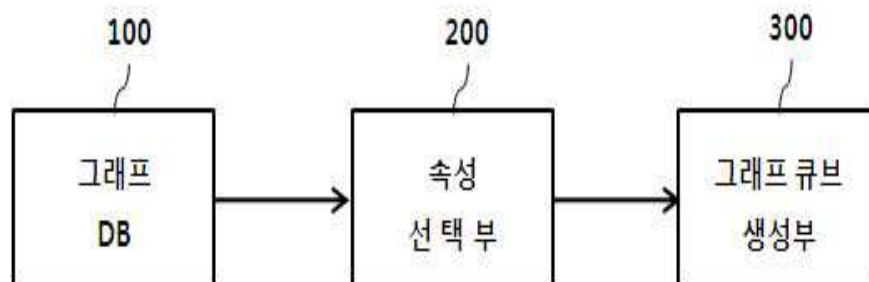
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 온라인 분석 처리를 위한 그래프 큐브의 생성 방법

(57) 요약

본 발명은 온라인 분석 처리를 위한 그래프 큐브의 생성 방법에 관한 것으로, 보다 구체적으로 노드가 가지는 속성의 발현 빈도뿐만 아니라 노드가 이웃 노드와 가지는 구조적 관계를 고려하여 그래프 큐브를 생성하는 방법에 관한 것이다.

대표도 - 도2



(72) 발명자

키파야트 올라 칸

경기도 용인시 기흥구 덕영대로 1732 경희대학교
국제캠퍼스

박기성

경기 부천시 원미구 원미로 123-4, E동 4호 (원미
동, 백송연립)

이 발명을 지원한 국가연구개발사업

과제고유번호 2012047478

부처명 교육과학기술부

연구관리전문기관 한국연구재단

연구사업명 중견연구자지원사업

연구과제명 행위 데이터베이스에서 행위패턴 모델링 및 마이닝 기술 연구

기 여 율 1/1

주관기관 경희대학교 산학협력단

연구기간 2012.09.01 ~ 2015.08.31

특허청구의 범위

청구항 1

다양한 속성을 가지는 각 노드가 서로 연결되어 있는 그래프에서 속성 조합으로 이루어진 그래프 큐브를 생성하는 방법에 있어서,

속성별로 해당 속성을 가지는 속성 노드와 상기 속성 노드에 연결되어 있는 이웃 노드 사이의 구조적 연결 정도를 나타내는 속성 구조 연결값을 계산하는 단계;

상기 이웃 노드끼리의 구조적 연결 정도를 나타내는 이웃 구조 연결값을 계산하는 단계;

상기 속성 구조 연결값과 상기 이웃 구조 연결값으로부터 상기 해당 속성에 대한 선택 중요도를 계산하는 단계; 및

상기 선택 중요도와 선택 임계값을 비교하여 상기 선택 임계값을 초과하는 해당 속성을 상기 속성 조합을 구성하는 속성으로 선택하는 단계를 포함하는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 2

제 1 항에 있어서,

상기 속성 구조 연결값은 상기 속성 노드와 상기 속성 노드에 연결되어 있는 이웃 노드 사이의 에지 수에 기초하여 계산되는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 3

제 2 항에 있어서,

상기 속성 구조 연결값은 상기 해당 속성을 가지는 속성 노드와 상기 이웃 노드 사이의 에지 수의 총합, 및 상기 그래프에서 상기 해당 속성을 가지는 속성 노드의 전체 수의 비율로부터 계산되는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 4

제 3 항에 있어서,

상기 속성 구조 연결값은 아래의 수학식(1)로 표현되는 밀집도(density, DS)에 의해 계산되며,

[수학식 1]

$$DS = \frac{E_T}{N_T}$$

여기서 E_T 는 상기 해당 속성을 가지는 속성 노드와 상기 속성 노드에 연결되어 있는 이웃 노드 사이의 에지 수의 총합이며, 상기 N_T 는 상기 그래프에서 상기 해당 속성을 가지는 속성 노드의 전체 수인 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 5

제 1 항에 있어서,

상기 이웃 구조 연결값은 상기 속성 노드에 연결되어 있는 상기 이웃 노드끼리의 단위 이웃 구조 연결값의 합으로 계산되는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 6

제 5 항에 있어서, 상기 단위 이웃 구조 연결값은

상기 이웃 노드끼리 연결되어 있는 에지의 총수와 상기 이웃 노드들이 모두 연결된 경우의 에지 총수의 비율로 계산되는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 7

제 6 항에 있어서,

상기 단위 이웃 구조 연결값(UCC)은 아래의 수학적식(2)에 의해 계산되며,

[수학적식 2]

$$UCC = \frac{NE_T}{NE_P}$$

여기서 NE_T 는 상기 이웃 노드끼리 연결되어 있는 에지의 총수이며, NE_P 는 상기 이웃 노드들이 모두 연결된 경우의 에지 총수를 의미하는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 8

제 1 항 내지 제 7 항 중 어느 한 항에 있어서, 상기 선택 중요도는

상기 속성 구조 연결값과 상기 이웃 구조 연결값의 합으로부터 계산되는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 9

제 8 항에 있어서, 상기 그래프 큐브의 생성 방법은

동일 속성 그룹에 속해 있는 해당 속성의 선택 중요도를 합하여 상기 동일 속성 그룹의 그룹 선택 중요도를 계산하는 단계를 더 포함하며,

상기 그룹 선택 중요도와 상기 선택 임계값을 비교하여 상기 선택 임계값을 초과하는 속성 그룹의 해당 속성을 상기 속성 조합을 구성하는 속성으로 선택하는 것을 특징으로 하는 그래프 큐브의 생성 방법.

청구항 10

다양한 속성을 가지는 각 노드가 서로 연결되어 있는 그래프에서 속성 조합으로 이루어진 그래프 큐브를 생성하는 장치에 있어서,

속성별로 해당 속성을 가지는 속성 노드와 상기 속성 노드에 연결되어 있는 이웃 노드 사이의 구조적 연결 정도를 나타내는 속성 구조 연결값을 계산하는 속성 구조 연결값 계산부;

상기 이웃 노드끼리의 구조적 연결 정도를 나타내는 이웃 구조 연결값을 계산하는 이웃 구조 연결값 계산부;

상기 속성 구조 연결값과 상기 이웃 구조 연결값으로부터 상기 해당 속성에 대한 선택 중요도를 계산하는 선택 중요도 계산부; 및

상기 선택 중요도와 선택 임계값을 비교하여 상기 선택 임계값을 초과하는 해당 속성을 상기 속성 조합을 구성하는 속성으로 선택하는 선택부를 포함하는 것을 특징으로 하는 그래프 큐브의 생성 장치.

청구항 11

제 10 항에 있어서, 상기 속성 구조 연결값 계산부는

상기 해당 속성을 가지는 속성 노드와 상기 이웃 노드 사이의 에지 수의 총합을 계산하는 제1 에지 계산부; 및

상기 해당 속성을 가지는 속성 노드와 상기 이웃 노드 사이의 에지 수의 총합 및, 상기 그래프에서 상기 해당 속성을 가지는 속성 노드의 전체 수의 비율로 상기 속성 노드의 속성 구조 연결값을 계산하는 제1 연결 계산부를 포함하는 것을 특징으로 하는 그래프 큐브의 생성 장치.

청구항 12

제 10 항에 있어서, 상기 이웃 구조 연결값 계산부는

상기 이웃 노드끼리 연결되어 있는 에지의 총 수를 계산하는 제2 에지 계산부;

상기 이웃 노드끼리 연결되어 있는 에지의 총 수 및 상기 이웃 노드들이 모두 연결된 경우의 에지 총수의 비율로부터 상기 속성 노드에 연결되어 있는 상기 이웃 노드끼리의 단위 이웃 구조 연결값을 계산하는 제2 연결 계산부; 및

속성별 상기 속성 노드의 단위 이웃 구조 연결값을 합하여 상기 해당 속성의 이웃 구조 연결값을 계산하는 제3 연결 계산부를 포함하는 것을 특징으로 하는 그래프 큐브의 생성 장치.

명세서

기술분야

[0001] 본 발명은 온라인 분석 처리를 위한 그래프 큐브의 생성 방법에 관한 것으로, 보다 구체적으로 노드가 가지는 속성의 발현 빈도뿐만 아니라 노드가 이웃 노드와 가지는 구조적 관계를 고려하여 그래프 큐브를 생성하는 방법에 관한 것이다.

배경기술

[0002] 소셜 네트워크, 화학 구성물 등은 속성을 가지는 노드들이 서로 연결 관계를 가지고 형성되는데, 이러한 노드와 노드 사이의 관계를 그래프로 표현할 수 있다. 최근에는 유비쿼터스 환경 및 이를 이용한 다양한 소셜 네트워크 서비스의 경우를 예로 들면, 소셜 네트워크 서비스에서 사용자를 노드로 판단하고 사용자가 가지는 성별, 나이, 직업 등을 노드가 가지는 다양한 속성들로 판단하는 경우 사용자 수와 사용자가 가지는 속성의 수는 기하급수적으로 증가하고 되며, 이와 함께 이러한 노드를 표현하는 그래프의 크기도 기하급수적으로 증가하고 있다. 이러한 대량의 그래프 정보로부터 사용자가 원하는 속성에 따라 그래프를 정리하고 그로부터 필요한 정보를 얻기 위한 빅 데이터 기술이 제안되고 있다.

[0003] 온라인 분석 처리(Online Analytical Processing, OLAP)는 의사결정 지원 시스템 가운데 대표적인 예로, 사용자가 동일한 데이터를 여러 기준을 이용하는 다양한 방식으로 바라보면서 다차원 데이터 분석을 할 수 있도록 도와준다.

[0004] 온라인 분석 처리는 1993년 에드거 F. 커드에 의해 처음 제안된 것으로, 그는 온라인 분석 처리를 사용자가 다차원 정보에 직접 접근하여 대화 형태로 정보를 분석하고 의사결정에 활용하는 과정이라고 정의하였다. 온라인 분석 처리 기술은 기업들에게 단순한 거래처리를 넘어선 정보의 활용 가능성을 보여주었고, 이를 계기로 적극적인 데이터의 활용을 통한 의사결정의 중요성이 강조되었다.

[0005] 온라인 분석 처리의 특징은 최종 사용자가 다차원 정보에 직접 접근하여 대화식으로 정보를 분석하고 의사결정에 활용하는 것으로, 사용자는 온라인 상에서 직접 데이터에 접근하며, 대화식으로 정보를 분석하므로 사용자가 기업의 전반적인 상황을 이해할 수 있게 하고 의사결정을 지원한다.

[0006]

[0007] 종래 그래프 정보를 온라인 분석 처리에 이용하기 위하여 그래프를 구성하는 노드 속성에 따라 그래프를 마이닝(mining)하기 위한 기법으로 그래프 큐브(graph cube)가 제안되었다. 그래프 큐브는 사용자가 요구하는 노드 속성에 따라 다차원적으로 그래프를 분류 정리하기 위한 기술로 아이스버그(Iceberg)로 알려진 그래프 큐빙 기술이 제안되었다. 도 1과 도 2를 참고로 종래 아이스버그에 대해 보다 구체적으로 살펴본다.

[0008] 도 1(a)에 도시되어 있는 바와 같이 그래프는 다수의 노드로 구성되어 있으며, 각 노드는 연결관계에 따라 에지로 연결되어 있다. 그리고 도 1(b)에 도시되어 있는 바와 같이 그래프를 구성하는 각 노드는 서로 상이한 다양

한 속성(A, B, C)을 가진다.

[0009] 사용자가 다차원적으로 온라인 분석 처리를 위해 1차원, 2차원, ... 등의 다차원으로 노드의 속성 조합으로 이루어진 그래프 큐브를 작성하는데, 모든 노드 속성을 고려하여 그래프 큐브를 작성하는 경우 방대한 양의 그래프 큐브가 생성되어, 그래프 큐브를 생성하거나 생성한 그래프 큐브로부터 필요한 정보를 분석하는데 오래 시간과 연산량이 필요하게 된다. 여기서 n 차원은 조합 속성 수를 의미한다.

[0010] 따라서 종래 아이스버그 그래프 큐빙 기술에서는 노드에서 각 속성이 발현하는 빈도를 카운트하고 최소 빈발 지지도 이상의 빈발 지지도를 가지지 않는 속성은 그래프 큐브를 생성하는데 제외하여 불필요한 속성이 포함되어 그래프 큐브가 생성되는 것을 방지한다. 예를 들어 도 1(b)에서 최소 빈발 지지도가 3으로 설정되어 있는 경우 C2의 출현 빈발도는 2이므로 C2 속성은 그래프 큐브를 생성하는데 이용되지 않는다.

[0011] 그러나 종래 아이스버그 그래프 큐빙 기술에서는 단순히 노드에서 각 속성이 출현하는 빈도만을 카운트하여 그래프 큐브를 생성하는데 사용되는 속성을 선택하며, 따라서 비록 출현 빈도는 작아도 이웃 노드와 높은 관계를 가지는 노드의 속성에 대해서는 고려하지 못한다는 문제점을 가진다.

발명의 내용

해결하려는 과제

[0012] 따라서 속성의 출현 빈도, 즉 빈발 지지도를 고려함과 동시에 속성 노드가 이웃 노드와 가지는 관계도 고려하여 그래프 큐브를 생성하는 것이 필요하다.

[0013] 본 발명은 위에서 언급한 그래프 큐브의 생성 방법이 가지는 문제점을 해결하기 위한 것으로 본 발명이 이루고자 하는 목적은 노드가 가지는 속성의 출현 빈도뿐만 아니라 이웃 노드와 가지는 구조적 관계 정도를 고려하여 그래프 큐브를 생성하는 방법을 제공하는 것이다.

[0014] 또한 본 발명이 이루고자 하는 목적은 속성 노드와 속성 노드에 연결되어 있는 이웃 노드와의 구조적 연결 관계뿐만 아니라 이웃 노드끼리의 구조적 연결 관계도 고려하여 그래프 큐브를 생성하는 방법을 제공하는 것이다.

과제의 해결 수단

[0015] 본 발명의 목적을 달성하기 위하여 본 발명에 따른 그래프 큐브의 생성 방법은 속성별로 해당 속성을 가지는 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 구조적 연결 정도를 나타내는 속성 구조 연결값을 계산하는 단계와, 이웃 노드끼리의 구조적 연결 정도를 나타내는 이웃 구조 연결값을 계산하는 단계와, 속성 구조 연결값과 이웃 구조 연결값으로부터 해당 속성에 대한 선택 중요도를 계산하는 단계와, 선택 중요도와 선택 임계값을 비교하여 선택 임계값을 초과하는 해당 속성을 속성 조합을 구성하는 속성으로 선택하는 단계를 포함하는 것을 특징으로 한다.

[0016] 여기서 속성 구조 연결값은 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 에지 수에 기초하여 계산되는 것을 특징으로 한다.

[0017] 바람직하게, 속성 구조 연결값은 해당 속성을 가지는 속성 노드와 이웃 노드 사이의 에지 수의 총합, 및 그래프에서 해당 속성을 가지는 속성 노드의 전체 수의 비율로부터 계산되는 것을 특징으로 하며, 보다 바람직하게 속성 구조 연결값은 아래의 수학식으로 표현되는 밀집도(density, DS)에 의해 계산되며,

[0018] [수학식]

$$DS = \frac{E_T}{N_T}$$

[0019]

[0020] 여기서 E_T 는 해당 속성을 가지는 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 에지 수의 총합이며, N_T 는 상기 그래프에서 해당 속성을 가지는 속성 노드의 전체 수인 것을 특징으로 한다.

[0021] 여기서 이웃 구조 연결값은 속성 노드에 연결되어 있는 이웃 노드끼리의 단위 이웃 구조 연결값의 합으로 계산되는 것을 특징으로 한다.

[0022] 바람직하게 단위 이웃 구조 연결값은 이웃 노드끼리 연결되어 있는 에지의 총수와 이웃 노드들이 모두 연결된 경우의 에지 총수의 비율로 계산되는 것을 특징으로 하며, 보다 바람직하게 단위 이웃 구조 연결값(UCC)은 아래의 수학적식에 의해 계산되며,

[0023] [수학적식]

$$UCC = \frac{NE_T}{NE_P}$$

[0024]

[0025] 여기서 NE_T 는 이웃 노드끼리 연결되어 있는 에지의 총수이며, NE_P 는 이웃 노드들이 모두 연결된 경우의 에지 총수를 의미하는 것을 특징으로 한다.

[0026] 여기서 선택 중요도는 속성 구조 연결값과 이웃 구조 연결값의 합으로부터 계산되는 것을 특징으로 한다.

[0027] 본 발명의 일 실시예에 따른 그래프 큐브의 생성 방법은 동일 속성 그룹에 속해 있는 해당 속성의 선택 중요도를 합하여 동일 속성 그룹의 그룹 선택 중요도를 계산하는 단계를 더 포함하며, 그룹 선택 중요도와 선택 임계값을 비교하여 선택 임계값을 초과하는 속성 그룹의 해당 속성을 속성 조합을 구성하는 속성으로 선택하는 것을 특징으로 한다.

[0028] 한편, 본 발명의 일 실시예에 따른 그래프 큐브의 생성 장치는 속성별로 해당 속성을 가지는 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 연결 정도를 나타내는 속성 구조 연결값을 계산하는 속성 구조 연결값 계산부와, 이웃 노드끼리의 연결 정도를 나타내는 이웃 구조 연결값을 계산하는 이웃 구조 연결값 계산부와, 속성 구조 연결값과 이웃 구조 연결값으로부터 해당 속성에 대한 선택 중요도를 계산하는 선택 중요도 계산부와, 선택 중요도와 선택 임계값을 비교하여 선택 임계값을 초과하는 해당 속성을 속성 조합을 구성하는 속성으로 선택하는 선택부를 포함하는 것을 특징으로 한다.

[0029] 여기서 속성 구조 연결값 계산부는 해당 속성을 가지는 속성 노드와 이웃 노드 사이의 에지 수의 총합을 계산하는 제1 에지 계산부와, 해당 속성을 가지는 속성 노드와 이웃 노드 사이의 에지 수의 총합 및, 그래프에서 해당 속성을 가지는 속성 노드의 전체 수의 비율로 속성 노드의 속성 구조 연결값을 계산하는 제1 연결 계산부를 포함하는 것을 특징으로 한다.

[0030] 여기서 이웃 구조 연결값 계산부는 이웃 노드끼리 연결되어 있는 에지의 총 수를 계산하는 제2 에지 계산부와, 이웃 노드끼리 연결되어 있는 에지의 총 수 및 이웃 노드들이 모두 연결된 경우의 에지 총수의 비율로부터 속성 노드에 연결되어 있는 이웃 노드끼리의 단위 이웃 구조 연결값을 계산하는 제2 연결 계산부와, 속성별 속성 노드의 단위 이웃 구조 연결값을 합하여 해당 속성의 이웃 구조 연결값을 계산하는 제3 연결 계산부를 포함하는 것을 특징으로 한다.

발명의 효과

[0031] 본 발명에 따른 그래프 큐브의 생성 방법은 노드가 가지는 속성의 출현 빈도뿐만 아니라 이웃 노드와 가지는 구조적 관계 정도를 고려하여 그래프 큐브를 생성함으로써, 출현 빈도가 작아도 이웃 노드와 구조적으로 긴밀한 연결 관계를 가지는 노드의 속성도 고려하여 그래프 큐브를 생성할 수 있다.

[0032] 또한 본 발명에 따른 그래프 큐브의 생성 방법은 속성 노드와 연결되어 있는 이웃 노드와의 구조적 연결 관계뿐만 아니라 속성 노드의 이웃 노드끼리의 구조적 연결 관계도 고려하여 그래프 큐브를 생성할 수 있다.

도면의 간단한 설명

[0033] 도 1은 종래 아이스버그의 그래프 큐빙 기법에 따른 속성을 선택하는 방법을 설명하기 위한 도면이다.

도 2는 본 발명에 따른 그래프 큐브의 생성 장치를 설명하기 위한 기능 블록도이다.

도 3은 본 발명에 따른 속성 선택부를 보다 구체적으로 설명하기 위한 기능 블록도이다.

도 4는 본 발명에 따른 그래프 큐브 생성 장치의 속성 구조 연결값 계산부를 설명하기 위한 기능 블록도이다.

도 5는 본 발명에 따른 그래프 큐브 생성 장치의 이웃 구조 연결값 계산부를 설명하기 위한 기능 블록도이다.

도 6은 본 발명에 따른 그래프 큐브의 생성 방법을 설명하기 위한 흐름도이다.

도 7은 속성 노드의 에지를 계산하는 방법의 일 예를 설명하기 위한 도면이다.

도 8은 속성 노드의 에지에 기초하여 계산된 밀집도의 일 예를 설명하기 위한 도면이다.

도 9는 단위 이웃 구조 연결값과 이웃 구조 연결값을 계산하는 방법의 일 예를 설명하기 위한 도면이다.

도 10은 선택 중요도를 계산하는 방법의 일 예를 설명하기 위한 도면이다.

도 11은 선택한 속성 그룹을 구성하는 속성들의 역색인 포맷의 일 예를 도시하고 있다

도 12는 역색인 포맷을 이용하여 생성되는 그래프 큐브의 일 예를 도시하고 있다.

발명을 실시하기 위한 구체적인 내용

[0034] 이하 첨부한 도면을 참고로 본 발명에 따른 그래프 큐브의 생성 방법에 대해 보다 구체적으로 살펴본다.

[0035] 도 2는 본 발명에 따른 그래프 큐브의 생성 장치를 설명하기 위한 기능 블록도이다.

[0036] 도 2를 참고로 보다 구체적으로 살펴보면, 속성 선택부(200)는 그래프 데이터베이스(100)에 저장되어 있는 그래프에 기초하여 그래프를 구성하는 노드의 속성들의 선택 중요도를 노드 속성의 출현 빈도뿐만 아니라 노드가 연결되어 있는 이웃 노드와의 구조적 관계를 고려하여 계산하고, 선택 중요도에 기초하여 선택 중요도가 선택 임계값보다 높은 선택 중요도를 가지는 속성을 선택한다.

[0037] 그래프 큐브 생성부(300)는 선택한 노드의 속성들에 기초하여 선택한 노드의 속성 조합으로 이루어진 그래프 큐브를 생성한다. 그래프 큐브 생성부(300)는 1개의 속성으로 이루어진 1차원 그래프 큐브, 2개의 속성 조합으로 이루어진 2차원 그래프 큐브, 3개의 속성 조합으로 이루어진 3차원 그래프 큐브를 순차적으로 생성한다.

[0038] 최소 빈발 지지도를 낮게 설정하는 경우 출현 빈도가 낮은 노드 속어도 선택되어 이로부터 방대한 양의 그래프 큐브가 생성되며, 최소 빈발 지지도를 높게 설정하는 경우 출현 빈도가 높은 노드 속성만이 선택되어 이로부터 적은 수의 그래프 큐브만이 생성된다. 그래프 큐브가 많은 경우 이를 생성하거나 저장하거나 처리하는데 많은 시간과 비용이 필요하게 되며, 그래프 큐브가 적은 경우 필요한 정보를 정확하게 얻어내기 곤란하다. 따라서 생성되는 그래프 큐브의 양에 기초하여 최소 빈발 지지도의 크기를 적절하게 설정하여야 하는데, 이 경우 최소 빈발 지지도만을 고려하여 노드 속성을 선택하는 경우 출현 빈도는 최소 빈발 지지도보다 낮지만 노드와 이웃 노드 사이에 중요한 구조적 관계를 가지는 노드 속성은 고려되지 않게 된다.

[0039] 그러나 본원발명의 속성 선택부(200)는 그래프를 구성하는 노드가 갖는 노드 속성의 출현 빈도뿐만 아니라 노드가 연결되어 있는 이웃 노드와의 구조적 연결 관계 및 이웃 노드끼리의 구조적 연결 관계를 고려하여 선택 중요도를 계산하며, 선택한 선택 중요도에 기초하여 노드 속성을 선택한다.

[0040] 도 3은 본 발명에 따른 속성 선택부를 보다 구체적으로 설명하기 위한 기능 블록도이다.

[0041] 도 3을 참고로 보다 구체적으로 살펴보면, 속성 구조 연결값 계산부(210)는 그래프 데이터베이스(100)에 저장되어 있는 그래프 노드의 속성별로 해당 속성을 가지는 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 구조적 연결 정도를 나타내는 속성 구조 연결값을 계산하며, 이웃 구조 연결값 계산부(220)는 속성 노드에 연결되어 있는 이웃 노드끼리의 구조적 연결 정도를 나타내는 이웃 구조 연결값을 계산한다. 여기서 이웃 구조 연결값 계산부(220)는 해당 속성을 가지는 각 속성 노드에 연결되어 있는 이웃 노드끼리의 단위 속성 구조 연결값을 계산하고, 해당 속성을 가지는 각 속성 노드의 단위 이웃 구조 연결값을 합하여 해당 속성의 이웃 구조 연결값을 계산한다.

[0042] 선택 중요도 계산부(230)는 속성 구조 연결값 계산부(210)에서 계산한 해당 속성의 속성 구조 연결값과 이웃 구조 연결값 계산부(220)에서 계산한 이웃 구조 연결값에 기초하여 해당 속성의 선택 중요도를 계산하며, 선택부(240)는 해당 속성의 선택 중요도와 선택 임계값을 비교하여 해당 속성의 선택 중요도가 선택 임계값을 초과하는 경우 해당 속성을 그래프 큐브를 생성하는데 이용하는 속성으로 선택한다. 그러나 해당 속성의 선택 중요도가 선택 임계값을 초과하지 못하는 경우 해당 속성을 제외하여 해당 속성이 그래프 큐브를 생성하는데 이용되지

않도록 한다.

[0043] 도 4는 본 발명에 따른 그래프 큐브 생성 장치의 속성 구조 연결값 계산부를 설명하기 위한 기능 블록도이며, 도 5는 본 발명에 따른 그래프 큐브 생성 장치의 이웃 구조 연결값 계산부를 설명하기 위한 기능 블록도이다.

[0044] 먼저, 도 4를 참고로 속성 구조 연결값 계산부를 보다 구체적으로 살펴보면, 제1 에지 계산부(211)는 해당 속성을 가지는 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 에지 수의 총합을 계산한다. 제1 연결 계산부(213)는 제1 에지 계산부(211)에서 계산한 해당 속성을 가지는 속성 노드와 이웃 노드 사이의 에지 수의 총합 및, 그래프에서 해당 속성을 가지는 속성 노드의 전체 수의 비율로 속성 노드의 속성 구조 연결값을 계산한다. 바람직하게, 제1 연결 계산부(213)는 아래의 수학식(1)과 같이 해당 속성의 속성 구조 연결값을 밀집도(DS)로 계산하는데,

[0045] [수학식 1]

$$DS = \frac{E_T}{N_T}$$

[0046]

[0047] 여기서 E_T 는 해당 속성을 가지는 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 에지 수의 총합이며, N_T 는 그래프에서 해당 속성을 가지는 속성 노드의 전체 수를 의미한다.

[0048] 다음으로, 도 5를 참고로 이웃 구조 연결값 계산부를 보다 구체적으로 살펴보면, 제2 에지 계산부(221)는 이웃 노드끼리 연결되어 있는 에지의 총 수를 계산하며, 제2 연결 계산부(223)는 제2 에지 계산부(221)에서 계산한 이웃 노드끼리 연결되어 있는 에지의 총 수 및 이웃 노드들이 모두 연결된 경우의 에지 총수의 비율로부터 해당 속성을 가지는 각 속성 노드에 연결되어 있는 이웃 노드끼리의 단위 이웃 구조 연결값(UCC)을 계산한다.

[0049] 제3 연결 계산부(225)는 해당 속성을 가지는 모든 속성 노드의 단위 이웃 구조 연결값을 합하여 해당 속성의 이웃 구조 연결값을 계산한다. 바람직하게, 단위 이웃 구조 연결값(UCC)은 아래의 수학식(2)에 의해 계산되며,

[0050] [수학식 2]

$$UCC = \frac{NE_T}{NE_P}$$

[0051]

[0052] 여기서 NE_T 는 이웃 노드끼리 연결되어 있는 에지의 총수이며, NE_P 는 이웃 노드들이 모두 연결된 경우의 에지 총수를 의미한다.

[0053] 도 6은 본 발명에 따른 그래프 큐브의 생성 방법을 설명하기 위한 흐름도이다.

[0054] 도 6을 참고로 보다 구체적으로 살펴보면, 그래프 데이터베이스에 저장된 그래프를 구성하는 노드가 갖는 속성의 출현 빈도, 및 해당 속성을 가지는 속성 노드와 속성 노드에 연결되어 있는 이웃 노드의 구조적 연결 관계를 고려하여 해당 속성의 속성 구조 연결값을 계산한다(S110). 여기서 속성 구조 연결값은 속성 노드와 속성 노드에 연결되어 있는 이웃 노드 사이의 에지 수에 기초하여 계산되는데, 바람직하게 속성 구조 연결값은 해당 속성을 가지는 속성 노드와 이웃 노드 사이의 에지 수의 총합, 및 그래프에서 해당 속성을 가지는 속성 노드의 전체 수의 비율로부터 계산된다.

[0055] 도 7과 도 8을 참고로 해당 속성의 속성 구조 연결값을 계산하는 일 예를 살펴보면, 먼저 도 7(a)에 도시되어 있는 그래프에 대해 그래프를 구성하는 노드의 속성값은 도 7(b)와 같다. 도 7(c)와 같이 해당 속성별로 해당 속성을 가지는 속성 노드에 연결된 에지의 수를 계산한다. 예를 들어 해당 속성 A1을 가지는 속성 노드는 1, 3, 4이며, 그래프를 구성하는 노드들 중 해당 속성 A1을 가지는 속성 노드 1, 3, 4에 연결된 에지의 수는 속성

노드 1이 3개, 속성 노드 3이 2개, 속성 노드 4가 2개로 총 7개이다.

[0056] 한편, 도 7에 도시되어 있는 그래프의 해당 속성별로 해당 속성을 가지는 속성 노드와 이웃 노드 사이의 에지 수의 총합, 및 그래프에서 해당 속성을 가지는 속성 노드의 전체 수의 비율로부터 계산된 속성 구조 연결값으로 밀집도는 도 8과 같다. 예를 들어, 해당 속성 A1에 대해 살펴보면, 해당 속성 A1을 가지는 속성 노드는 1, 3, 4로 모두 3개이며 해당 속성 A1을 가지는 속성 노드 1, 3, 4가 가지는 총 에지의 수는 7개로 해당 속성 A1의 밀집도는 2.33이다. 이와 같이 모든 속성별로 해당 속성의 속성 구조 연결값을 계산한다.

[0057] 다시 도 6을 참고로 살펴보면, 각 속성 노드에 연결되어 있는 이웃 노드끼리의 구조적 연결 정도를 나타내는 이웃 구조 연결값을 계산한다(S120). 여기서 이웃 구조 연결값은 속성별로 속성 노드에 연결되어 있는 이웃 노드끼리의 단위 이웃 구조 연결값의 합으로 계산되는데, 단위 이웃 구조 연결값은 속성 노드에 연결되어 있는 이웃 노드를 기준으로 이웃 노드에 연결되어 있는 에지의 총수와 이웃 노드들이 모두 연결된 경우의 에지 총수의 비율로 계산된다.

[0058] 도 9(a)는 그래프를 구성하는 각 노드의 단위 이웃 구조 연결값의 일 예를 도시하고 있으며, 도 9(b)는 각 해당 속성별 이웃 구조 연결값의 일 예를 도시하고 있다. 도 7과 도 9를 참고로 해당 속성의 이웃 구조 연결값을 계산하는 일 예를 살펴보면, 그래프를 구성하는 노드 1은 노드 2, 노드 3, 노드 4에 연결되어 있으며 따라서 노드 1의 이웃 노드는 노드 2, 노드 3, 노드 4이다. 노드 1의 이웃 노드인 노드 2, 3, 4를 기준으로 이웃 노드 3과 이웃 노드 4는 서로 연결되어 있어 이웃 노드 2, 3, 4의 에지 총수는 1개이며, 이웃 노드 2, 3, 4가 모두 연결된 경우의 에지의 총수는 3개이다. 따라서 노드 1의 단위 이웃 구조 연결값은 $1/3=0.33$ 으로 계산된다. 다만 속성 노드에 연결되어 있는 이웃 노드의 수가 1개인 경우 단위 이웃 구조 연결값은 1로 계산된다. 이와 같이 그래프를 구성하는 모든 노드에 대해 단위 이웃 구조 연결값을 계산한다. 해당 속성을 가지는 속성 노드별로 단위 이웃 구조 연결값을 합하여 이웃 구조 연결값을 가지는데, 예를 들어 해당 속성 A1에 대한 이웃 구조 연결값은 해당 속성 A1을 가지는 속성 노드 1, 3, 4의 단위 이웃 구조 연결값인 0.33, 1, 1을 합하여 2.33으로 계산된다.

[0059] 다시 도 6을 참고로 살펴보면, 속성 구조 연결값과 이웃 구조 연결값으로부터 해당 속성에 대한 선택 중요도를 계산하고(S130), 계산한 선택 중요도가 선택 임계값보다 큰지를 판단한다(S140). 여기서 선택 중요도(SC)는 속성 구조 연결값(A)과 이웃 구조 연결값(B)의 합으로부터 계산되는데, 바람직하게 아래의 수학적식(3)과 같이 계산된다.

[0060] [수학적식 3]

$$SC = \alpha \times A + \beta \times B$$

[0061] 여기서 α 는 속성 구조 연결값에 대한 가중치이며 β 는 이웃 구조 연결값에 대한 가중치이다. 본 발명이 적용되는 분야에 따라 속성 구조 연결값에 대한 가중치 또는 이웃 구조 연결값에 대한 가중치를 조절하여 속성 구조 연결값 또는 이웃 구조 연결값의 가중비율을 조절할 수 있다. 도 10(a)는 속성 구조 연결값에 대한 가중치와 이웃 구조 연결값에 대한 가중치가 각각 1인 경우 선택 중요도를 계산한 일 예를 도시하고 있다.

[0063] 해당 속성의 선택 중요도가 선택 임계값보다 큰 경우 해당 속성을 그래프 큐브를 생성하는데 이용되는 속성으로 선택하고(S150), 해당 속성의 선택 중요도가 선택 임계값보다 작은 경우 해당 속성을 그래프 큐브를 생성하는데 이용되는 속성에서 제외한다(S160). 선택한 속성을 이용하여 선택한 속성의 조합으로 이루어지는 그래프 큐브를 생성한다. 예를 들어, 선택 임계값이 3.5로 설정되어 있는 경우 A2, B1의 해당 속성은 그래프 큐브를 생성하는데 이용되는 속성에서 제외된다.

[0064] 더욱 바람직하게, 본 발명이 적용되는 분야에 따라 동일 속성 그룹에 속해 있는 해당 속성의 선택 중요도를 합하여 동일 속성 그룹의 그룹 선택 중요도를 계산하며, 그룹 선택 중요도와 선택 임계값을 비교하여 선택 임계값을 초과하는 속성 그룹의 해당 속성을 속성 조합을 구성하는 속성으로 선택할 수 있다. 도 10(b)는 속성 구조 연결값에 대한 가중치와 이웃 구조 연결값에 대한 가중치가 각각 1인 경우 선택 중요도를 계산한 일 예를 도시하고 있다. 예를 들어, 속성 A1과 A2는 속성 그룹 A에 속하며, 속성 B1과 B2는 속성 그룹 B에 속하며, 속성 C1과 C2는 속성 그룹 C에 속하며, 선택 임계값이 6.5로 설정되어 있는 경우, 속성 그룹A와 속성 그룹C는 선택되고

여기서 노드 리스트는 해당 속성을 가지는 노드를 의미하며, 에지 리스트는 각 속성 노드와 연결되어 있는 이웃 노드를 의미한다. 이러한 노드 리스트와 에지 리스트를 이용하여 속성 노드의 연결 관계를 고려하여 해당 속성을 노드로 그래프 큐브를 생성한다.

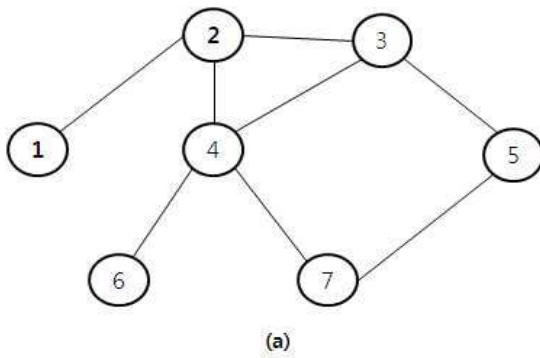
상기 컴퓨터로 읽을 수 있는 기록 매체는 전기 또는 자기식 저장 매체(예를 들어, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장 매체를 포함한다.

부호의 설명

- 11 -

도면

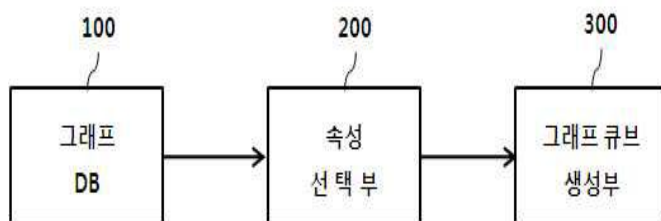
도면1



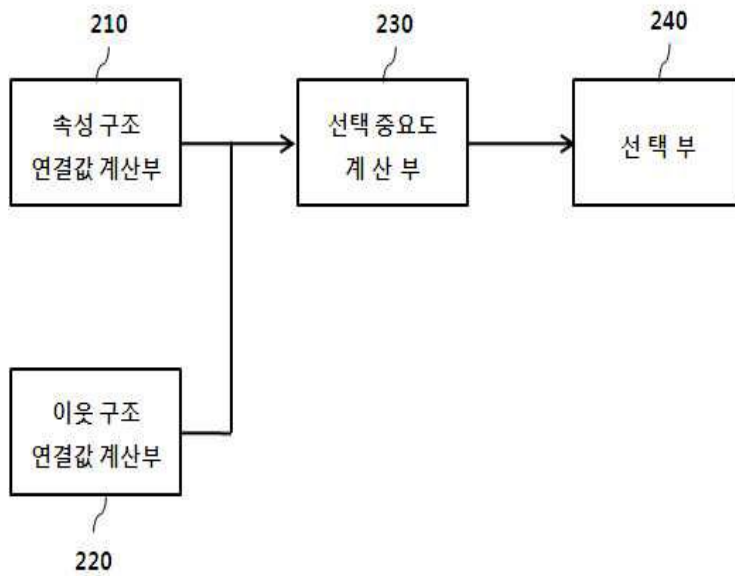
	A	B	C
1	A1	B1	C1
2	A1	B1	C2
3	A2	B2	C1
4	A1	B1	C2
5	A2	B2	C1
6	A2	B1	C1
7	A1	B2	C1

(b)

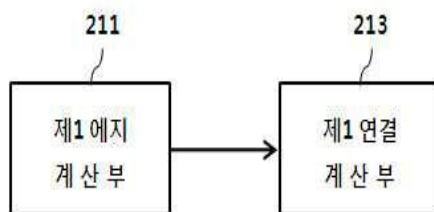
도면2



도면3



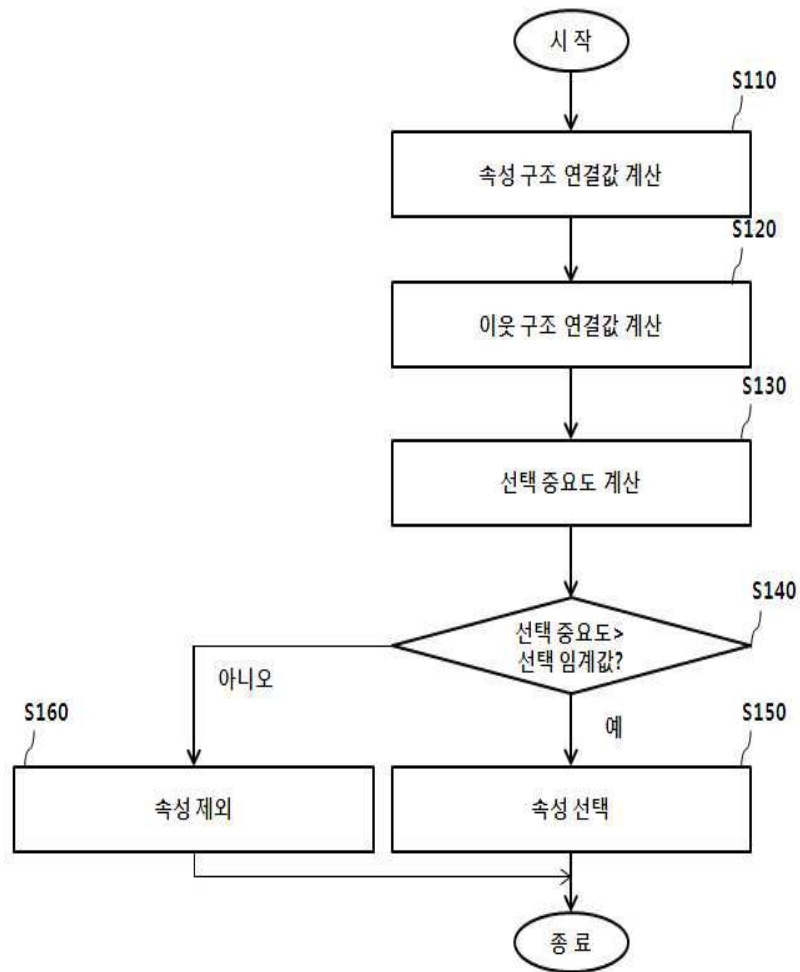
도면4



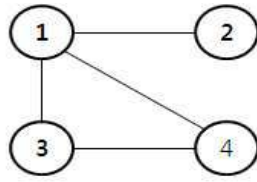
도면5



도면6



도면7



(a)

	A	B	C
1	A1	B1	C1
2	A2	B2	C2
3	A1	B2	C1
4	A1	B1	C2

(b)

속성	에지 수
A1	7
A2	1
B1	5
B2	3
C1	5
C2	3

(b)

도면8

속성	밀집도
A1	2.33
A2	1.0
B1	1.5
B2	1.5
C1	2.5
C2	1.5

도면9

노드	단위 이웃 구조 연결값
1	0.33
2	1
3	1
4	1

(a)

속성	이웃 구조 연결값
A1	2.33
A2	1
B1	1.33
B2	2
C1	1.33
C2	2

(b)

도면10

속성	선택 중요도
A1	4.66
A2	2.0
B1	2.83
B2	3.5
C1	3.83
C2	3.5

(a)

속성	선택 중요도
A	6.66
B	6.33
C	7.33

(b)

도면11

속성	노드 리스트	에지 리스트
A1	1,3,4	{2,3,4},{1,4},{1,3}
A2	2	{1}
C1	1,3	{2,3,4},{1,4}
C2	2,4	{1},{1,3}

도면12

