



(51) 국제특허분류:

G06F 17/30 (2006.01)

(21) 국제출원번호:

PCT/KR2017/012962

(22) 국제출원일:

2017년 11월 18일 (18.11.2017)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2017-0131551 2017년 10월 11일 (11.10.2017) KR

(71) 출원인: 서울대학교 산학협력단 (SEOUL NATIONAL

UNIVERSITY R&DB FOUNDATION) [KR/KR]; 08826
서울시 관악구 관악로 1 (신림동), Seoul (KR).

(72) 발명자: 강유 (KANG, U); 08826 서울시 관악구 관악로 1, 서울대학교 301동 502호, Seoul (KR). 윤민지 (YOON, Minji); 08826 서울시 관악구 관악로 1, 서울대학교 301동 519호, Seoul (KR). 정진홍 (JUNG, Jinhong); 08826 서울시 관악구 관악로 1, 서울대학교 301동 519호, Seoul (KR).

(74) 대리인: 특허법인 아이스퀘어 (ISQUARE PATENT & LAW FIRM); 06174 서울시 강남구 영동대로82길 11, 5층(대치동, 도원빌딩), Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

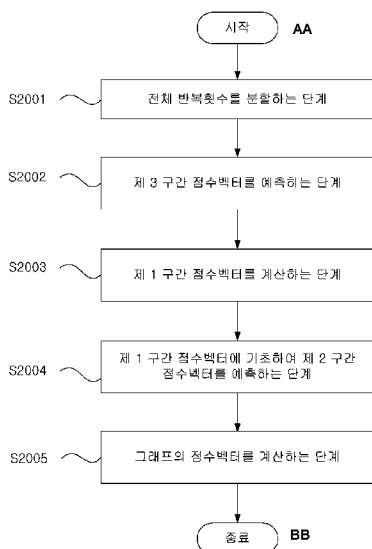
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

규칙 4.17에 의한 선언서:

— 신규성을 해치지 아니하는 개시 또는 신규성 상실의 예외에 관한 선언 (규칙 4.17(v))

(54) Title: METHOD AND DEVICE FOR EFFICIENTLY CALCULATING SIMILARITY BETWEEN NODES FOR LARGE SCALE GRAPH

(54) 발명의 명칭: 대용량 그래프를 위한 효율적 노드간 유사도 계산 방법 및 장치



S2001 ... Step of dividing total number of iterations
 S2002 ... Step of predicting third section score vector
 S2003 ... Step of calculating first section score vector
 S2004 ... Step of predicting second section score vector on basis of first section score vector
 S2005 ... Step of calculating score vector of graph
 AA ... Start
 BB ... End

(57) Abstract: The present invention relates to a method and device for calculating similarity between nodes in a large scale graph. According to a first aspect of the present invention, a similarity calculation device for calculating similarity between nodes in a large scale graph may comprise: a section dividing unit for dividing, into at least one section, the total number of iterations of an operation for calculating a score vector having, as an element, a score for each of at least one node constituting the graph; and an operation execution unit for calculating a score vector by executing an operation corresponding to each divided section.

(57) 요약서: 본 발명은 대용량 그래프에서 노드간 유사도 계산 방법 및 장치에 관한 것이다. 본 발명의 제1 측면에 따르면, 유사도계산장치는, 본 발명의 제 1 측면에 따르면, 대용량 그래프에서 노드간의 유사도를 계산하는 장치에 있어서, 상기 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할하는 구간분할부 및 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산하는 연산수행부를 포함할 수 있다.

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 대용량 그래프를 위한 효율적 노드간 유사도 계산 방법 및 장치

기술분야

- [1] 본 발명은 대용량 그래프를 위한 효율적 노드간 유사도 계산 방법 및 장치에 관한 것으로, 보다 구체적으로는 대용량 그래프에서 RWR (Random Walk with Restart) 기법을 이용한 노드간 유사도를 빠르고 정확하게 계산하는 유사도 계산 방법 및 장치에 관한 것이다.

배경기술

- [2] 데이터 마이닝(data mining)이란 많은 양의 데이터로부터 유용한 정보를 추출하는 것을 의미한다. 그리고, 이러한 데이터 마이닝을 그래프에 적용하는 것을 그래프 마이닝이라고 하며, 특히 그래프에 포함된 노드들의 우선순위를 결정하는 기술을 그래프 랭킹(graph ranking)이라고 한다.
- [3] 최근 SNS 이용률이 증가하고, 인터넷상에 저장되는 웹 문서의 양이 증가하면서, SNS 이용자간의 연결 관계 또는 웹 문서들간의 연결 관계를 나타내는 그래프에서 노드들의 유사도를 파악하는 것이 중요해지고 있다.
- [4] 그래프 랭킹 수행 방법을 제시하는 문헌으로서, 한국공개특허 제2011-0131094호가 있다. 본 문헌에는 정보 네트워크로부터 하나 이상의 노드들 및 간선들을 수집하고, 수집된 노드들에 랜덤 워크를 수행하여 얻은 시퀀스들을 데이터 베이스에 저장하고, 데이터 베이스를 마이닝하여 정보 네트워크에 존재하는 패턴을 파악하는 내용이 개시되었다.
- [5] 한편, 전술한 배경기술은 발명자가 본 발명의 도출을 위해 보유하고 있었거나, 본 발명의 도출 과정에서 습득한 기술 정보로서, 반드시 본 발명의 출원 전에 일반 공중에게 공개된 공지기술이라 할 수는 없다.

발명의 상세한 설명

기술적 과제

- [6] 본 발명의 일 실시예는 대용량 그래프를 위한 효율적 노드간 유사도 계산 방법 및 장치를 제공하는 데에 그 목적이 있다.
- [7] 본 발명의 일 실시예는 노드간 유사도 계산을 위해 점수벡터를 계산하기 위해 전체 반복횟수만큼의 연산을 수행하지 않고도 점수벡터를 예측하는 유사도 계산 방법 및 장치를 제공하는 데에 그 목적이 있다.
- [8] 본 발명의 일 실시예는 반복 행렬 벡터 곱 형태인 CPI(Cumulative Power Iteration)를 이용하여 노드간의 유사도를 계산하는 유사도 계산 방법 및 장치를 제공하는 데에 그 목적이 있다.

과제 해결 수단

- [9] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1

측면에 따르면, 대용량 그래프에서 노드간의 유사도를 계산하는 장치에 있어서, 상기 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할하는 구간분할부 및 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산하는 연산수행부를 포함할 수 있다.

[10] 본 발명의 제2 측면에 따르면, 유사도계산방법은, 유사도계산장치가 대용량 그래프에서 노드간의 유사도를 계산하는 방법에 있어서, 상기 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할하는 단계 및 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산하는 단계를 포함할 수 있다.

[11] 본 발명의 제3 측면에 따르면, 유사도계산방법을 수행하기 위한 컴퓨터 프로그램으로서, 유사도계산방법은 상기 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할하는 단계 및 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산하는 단계를 포함할 수 있다.

[12] 본 발명의 제 4 측면에 따르면, 유사도계산방법을 수행하는 프로그램이 기록된 컴퓨터 판독가능한 기록매체로서, 유사도계산방법은 상기 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할하는 단계 및 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산하는 단계를 포함할 수 있다.

발명의 효과

[13] 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 본 발명의 일 실시예는 노드간 유사도 계산 방법 및 장치를 제시할 수 있다.

[14] 또한, 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 노드간 유사도 계산을 위해 점수벡터를 계산하기 위한 연산의 전체 반복횟수를 분할하여 각 구간에 대한 점수벡터를 예측함으로써 계산량을 줄이는 유사도 계산 방법 및 장치를 제시할 수 있다.

[15] 그리고 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 반복 행렬 벡터 곱을 계산할 수 있는 CPI(Cumulative Power Iteration)를 이용하여 노드간 유사도 계산 방법 및 장치를 제시할 수 있다.

[16] 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[17] 도 1은 본 발명의 일 실시예에 따른 노드간 유사도계산장치를 도시한

도면이다.

- [18] 도 2는 본 발명의 일 실시예에 따른 노드간 유사도계산방법을 설명하기 위한 순서도이다.

발명의 실시를 위한 최선의 형태

- [19] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [20] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [21] 이하 첨부된 도면을 참고하여 본 발명을 상세히 설명하기로 한다.
- [22] 본 발명의 실시예들을 자세히 설명하기에 앞서, 우선 대용량 그래프에서 노드간의 유사도서 이용되는 예측값을 계산하는 페이지랭크(Page Rank)에 대한 설명을 한다.
- [23] '페이지랭크'는 그래프에서 노드의 중요도를 측정하는데 사용되는 알고리즘으로 노드로 진입하는 간선의 수가 많으면 중요 노드로 예측되어 페이지랭크 점수가 높고, 노드로 진입하는 간선의 수가 적으면 중요도가 낮은 노드로 예측되어 페이지랭크 점수가 낮다.
- [24] 모든 노드에 대한 페이지랭크 점수인 p 는 아래와 같은 수식으로 정의된다.
- [25] [수학식 1]
- [26]
$$\mathbf{p} = (1 - c)\tilde{\mathbf{A}}^T \mathbf{p} + \frac{c}{n} \mathbf{1}$$
- [27] $\tilde{\mathbf{A}}$: 열을 기준으로 정규화된 인접 행렬, c : 재시작 확률
- [28] 이러한 페이지랭크에 의해 제공되는 거시적 관점에서의 노드 중요성은 각 노드의 선호도를 반영하지 못한다.
- [29] 반면, 'RWR(Random Walk with Restart)'은 그래프에서 주어진 시드노드 s 에 대한 각 노드의 관련성을 측정할 수 있다. 그래프에서 간선을 따라 이동하고, 노드 s 에서 때때로 재시작을 하는 랜덤서퍼를 가정하면, 각 단계에서 랜덤서퍼는 $1-c$ 의 확률로 간선을 이동하거나 또는 c 의 확률로 시드노드로 점프한다.
- [30] 여기서 RWR 점수 벡터 r 에 대한 반복 수식은 아래와 같다.
- [31] [수학식2]

[32] $\mathbf{r} = (1 - c)\tilde{\mathbf{A}}^T \mathbf{r} + c\mathbf{q}$

[33] $\mathbf{q}$: 시드노드 s 의 인덱스를 1로 하고 나머지를 0으로 한 시드벡터

[34] 페이지랭크에서 1은 시드벡터의 역할로 RWR과의 차이점이다. 즉, 페이지랭크에서 랜덤워크는 그래프상의 어느 노드에서 동일한 확률로 재시작할 수 있는 반면, RWR에서 랜덤워크는 기 설정된 시드노드에서만 재시작할 수 있다.

[35] 그리고 ‘CPI(Cumulative Power Iteration)’는 RWR의 문제점을 반복 행렬-벡터 곱 형태(iterative matrix-vector multiplication form)에서 그래프 전체로의 점수 전파로 설명한다.

[36] 점수 $\mathbf{x}^{(0)}$ 은 초기 시작 시 시드노드로부터 생성된다. 각 단계에서 점수는 현재 노드의 진출 간선(out-edges)의 개수에 따라 균등하게 나뉘어 ‘1-c’만큼 감소되어 전파된다. 예를 들어, 진출간선의 수가 n_v 인 노드 v 의 점수 x_v 는 이웃 노드 v 로 $\frac{1}{n_v}(1 - c)x_v$ 만큼 전파된다.

[37] 행렬 벡터 곱 형태에서 $\mathbf{x}^{(i)}$ 에서 점수는 진출 간선과 연결된 이웃 노드들로 전파되고, 전파된 점수는 $\mathbf{x}^{(i+1)}$ 에 저장된다.

[38] CPI는 중간 점수 벡터 $\mathbf{x}^{(i)}$ 를 적층한 최종결과인 점수벡터 $\mathbf{r}_{\text{CPI}}$ 는 아래의 수식과 같다.

[39] [수학식 3]

[40] $\mathbf{x}^{(0)} = c\mathbf{q}$

$$\mathbf{x}^{(i)} = (1 - c)\tilde{\mathbf{A}}^T \mathbf{x}^{(i-1)} = c \left((1 - c)\tilde{\mathbf{A}}^T \right)^i \mathbf{q}$$

$$\mathbf{r}_{\text{CPI}} = \sum_{i=0}^{\infty} \mathbf{x}^{(i)} = c \sum_{i=0}^{\infty} \left((1 - c)\tilde{\mathbf{A}}^T \right)^i \mathbf{q}$$

[41] 이하에서는 상술된 페이지랭킹과 CPI를 이용하여 대용량 그래프에서 노드간의 유사도를 측정하는 유사도측정장치(100)에 대해 도면을 참조하여 설명한다.

[42] 도 1은 본 발명의 일 실시예에 따른 유사도측정장치(100)의 구성을 나타내는 블록도이다. 도 1을 참조하면, 본 발명의 일 실시예에 따른 유사도측정장치(100)는 구간분할부(110)와 연산수행부(120)를 포함할 수 있다.

[43] 우선, 구간분할부(110)는 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할할 수 있다.

[44] [수학식 4]

[45] $\mathbf{r}_{\text{CPI}}$

$$\begin{aligned} &= \mathbf{r}_{\text{family}} + \mathbf{r}_{\text{neighbor}} + \mathbf{r}_{\text{stranger}} \\ &= \underbrace{\mathbf{x}^{(0)} + \dots + \mathbf{x}^{(S-1)}}_{\text{family part}} + \underbrace{\mathbf{x}^{(S)} + \dots + \mathbf{x}^{(T-1)}}_{\text{neighbor part}} + \underbrace{\mathbf{x}^{(T)} + \dots}_{\text{stranger part}} \end{aligned}$$

[46] S: $\mathbf{r}_{\text{neighbor}}$ 의 시작하는 차수, T: $\mathbf{r}_{\text{stranger}}$ 의 시작하는 차수

[47] 이때, 제 1 구간 점수벡터 $\mathbf{r}_{\text{family}} = \mathbf{x}^{(0)} + \dots + \mathbf{x}^{(S-1)}$ 는 CPI연산에서 0회에서 S-1회까지의 반복구간에서 시드노드로부터 인접 노드로의 점수의 전파를 나타낸다.

[48] 그리고 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}} = \mathbf{x}^{(S)} + \dots + \mathbf{x}^{(T-1)}$ 는 네이버 파트를 따른 점수의 전파를 나타내고, 나머지 반복 구간인 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}} = \mathbf{x}^{(T)} + \dots$. S는 T회에서 끝까지 점수 전파를 나타낸다.

[49] 즉, 구간분할부(110)는 점수 벡터를 계산하는 연산인 CPI의 반복과정을 수학적 4와 같이 3개의 구간으로 분할하기 위해 S와 T를 설정할 수 있다.

[50] 이때 S와 T는 정확도와 계산시간을 고려하여 설정될 수 있다. 그리고 이러한 분할을 통해 정확한 RWR점수벡터는 직접 계산되는 제 1 구간 점수벡터 $\mathbf{r}_{\text{family}}$ 와 예측되는 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}}$ 및 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}}$ 를 통해 계산될 수 있다.

[51] 한편, 연산수행부(120)는 분할된 구간별로 각 구간에 대응되는 점수벡터의 연산을 나누어 수행할 수 있다.

[52] 우선, 연산수행부(120)는 점수의 분배가 시드노드의 위치보다 분배되는 간선에 영향을 더 많이 받는 관측을 근거로 페이지랭크를 이용하여 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}}$ 를 예측하는 스트레인저예측단계(stranger approximation phase)를 수행할 수 있다.

[53] 그리고 연산수행부(120)는 정확하게 계산되는 제 1 구간 점수벡터 $\mathbf{r}_{\text{family}}$ 를 기초로 실제 세상을 구현한 그래프의 블록 단위 구조의 이점을 이용하여 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}}$ 를 예측하는 네이버예측단계(neighbor approximation phase)를 수행할 수 있다.

[54] 이후, 연산수행부(120)는 계산된 제 1 구간 점수벡터 $\mathbf{r}_{\text{family}}$ 와 예측된 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}}$ 및 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}}$ 를 기초로 RWR점수벡터 $\mathbf{r}_{\text{TPA}}$ 를 계산할 수 있다.

[55] 이하에서는 연산수행부(120)가 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}}$ 와 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}}$ 를 예측하는 각 단계를 구체적으로 살펴본다.

[56] 우선, 연산수행부(120)는 스트레인저예측단계에서 페이지랭크를 이용하여 제 3 구간 점수벡터를 예측할 수 있다.

[57] 페이지랭크 점수벡터인 $\mathbf{p}_{\text{CPI}}$ 는 아래의 수식과 같다.

[58] [수학식 5]

$$\begin{aligned}
[59] \quad \mathbf{p}_{\text{CPI}} &= \mathbf{p}_{\text{family}} + \mathbf{p}_{\text{neighbor}} + \mathbf{p}_{\text{stranger}} \\
&= \underbrace{\mathbf{x}'^{(0)} + \dots + \mathbf{x}'^{(S-1)}}_{\text{family part}} + \underbrace{\mathbf{x}'^{(S)} + \dots + \mathbf{x}'^{(T-1)}}_{\text{neighbor part}} + \underbrace{\mathbf{x}'^{(T)} + \dots}_{\text{stranger part}}
\end{aligned}$$

$$[60] \quad \mathbf{x}'^{(i)} = (1 - c)\tilde{\mathbf{A}}^\top \mathbf{x}'^{(i-1)}, \mathbf{x}'^{(0)} = \frac{c}{n} \mathbf{1}$$

[61] 여기서 RWR점수벡터 $\mathbf{r}_{\text{CPI}}$ 와 페이지랭크 점수벡터 $\mathbf{p}_{\text{CPI}}$ 의 차이는 시드벡터이다. 그러므로 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}}$ 는 페이지랭크상의 제 3 구간 점수벡터 $\mathbf{p}_{\text{stranger}}$ 와 유사하여 아래와 같다.

[62] [수학식 6]

$$[63] \quad \tilde{\mathbf{r}}_{\text{stranger}} = \mathbf{p}_{\text{stranger}}$$

[64] 이를 증명하면, 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}}$ 와 예측된 제 3 구간 점수벡터 $\tilde{\mathbf{r}}_{\text{stranger}}$ 는 아래와 같다.

[65] [수학식 7]

$$\begin{aligned}
[66] \quad \mathbf{r}_{\text{stranger}} &= \mathbf{x}^{(T)} + \mathbf{x}^{(T+1)} + \dots \\
\tilde{\mathbf{r}}_{\text{stranger}} &= \mathbf{x}'^{(T)} + \mathbf{x}'^{(T+1)} + \dots
\end{aligned}$$

[67] 그리고 $\|\mathbf{r}_{\text{stranger}} - \tilde{\mathbf{r}}_{\text{stranger}}\|_1$ 는 아래와 같이 바운드된다.

[68] [수학식 8]

$$\begin{aligned}
[69] \quad \|\mathbf{r}_{\text{stranger}} - \tilde{\mathbf{r}}_{\text{stranger}}\|_1 &= \|(\mathbf{x}^{(T)} + \dots) - (\mathbf{x}'^{(T)} + \dots)\|_1 \\
&\leq \sum_{i=T}^{\infty} \|\mathbf{x}^{(i)} - \mathbf{x}'^{(i)}\|_1
\end{aligned}$$

[70] CPI에서 i회의 점수벡터 $\mathbf{x}^{(i)}$ 와 $\mathbf{x}'^{(i)}$ 는 아래와 같다.

[71] [수학식 9]

$$\begin{aligned}
[72] \quad \mathbf{x}^{(i)} &= (1 - c)\tilde{\mathbf{A}}^\top \mathbf{x}^{(i-1)} = c(1 - c)^i (\tilde{\mathbf{A}}^\top)^i \mathbf{q} \\
\mathbf{x}'^{(i)} &= (1 - c)\tilde{\mathbf{A}}^\top \mathbf{x}'^{(i-1)} = c(1 - c)^i (\tilde{\mathbf{A}}^\top)^i \mathbf{b}
\end{aligned}$$

[73] q: s 번째 단위벡터, $\mathbf{b} = \frac{1}{n} \mathbf{1}$

[74] $(\tilde{\mathbf{A}}^\top)^i$ 가 $\left[\mathbf{c}_1^{(i)}, \dots, \mathbf{c}_n^{(i)} \right]$ 에 의해 표현된다고 가정하면, $\mathbf{x}^{(i)} - \mathbf{x}'^{(i)}$ 는 아래와 같이 표현된다.

[75] [수학식 10]

$$\begin{aligned}
[76] \quad \mathbf{x}^{(i)} - \mathbf{x}'^{(i)} &= c(1-c)^i (\tilde{\mathbf{A}}^\top)^i (\mathbf{q} - \mathbf{b}) \\
&= c(1-c)^i (\tilde{\mathbf{A}}^\top)^i \left(-\frac{1}{n}, -\frac{1}{n}, \dots, \underbrace{\frac{n-1}{n}}_{s\text{-th entry}}, -\frac{1}{n}, \dots \right)^\top \\
&= c(1-c)^i \left(-\frac{1}{n} \mathbf{c}_1^{(i)} \dots + \frac{n-1}{n} \mathbf{c}_s^{(i)} \dots - \frac{1}{n} \mathbf{c}_n^{(i)} \right) \\
&= \frac{c(1-c)^i}{n} \sum_{j \neq s} (\mathbf{c}_s^{(i)} - \mathbf{c}_j^{(i)})
\end{aligned}$$

[77] $\mathbf{c}_j^{(i)}$: 행렬 $(\tilde{\mathbf{A}}^\top)^i$ 의 j 번째 열

[78] 그러면, $\|\mathbf{r}_{\text{stranger}} - \tilde{\mathbf{r}}_{\text{stranger}}\|_1$ 는 아래와 같이 바운드(bounded)된다.

[79] [수학식 11]

$$\begin{aligned}
[80] \quad \|\mathbf{x}^{(i)} - \mathbf{x}'^{(i)}\|_1 &= \frac{c(1-c)^i}{n} \left\| \sum_{j \neq s} (\mathbf{c}_s^{(i)} - \mathbf{c}_j^{(i)}) \right\|_1 \\
&\leq \frac{c(1-c)^i}{n} \sum_{j \neq s} \|\mathbf{c}_s^{(i)} - \mathbf{c}_j^{(i)}\|_1 \\
&\leq \frac{c(1-c)^i}{n} \times 2(n-1) \leq 2c(1-c)^i
\end{aligned}$$

[81] 두번째 부등식에서 $\|\mathbf{c}_j^{(i)}\|_1 = 1$ 과 $\|\mathbf{c}_s^{(i)} - \mathbf{c}_j^{(i)}\|_1 \leq \|\mathbf{c}_s^{(i)}\|_1 + \|\mathbf{c}_j^{(i)}\|_1 = 2$ 를 이용하고, $\tilde{\mathbf{A}}^\top$ 와 $(\tilde{\mathbf{A}}^\top)^i$ 는 열 확률행렬이므로 $\|\mathbf{r}_{\text{stranger}} - \tilde{\mathbf{r}}_{\text{stranger}}\|_1$ 는 아래와 같이 바운드된다.

[82] [수학식 12]

$$\begin{aligned}
[83] \quad \|\mathbf{r}_{\text{stranger}} - \tilde{\mathbf{r}}_{\text{stranger}}\|_1 &\leq \sum_{i=T}^{\infty} \|\mathbf{x}^{(i)} - \mathbf{x}'^{(i)}\|_1 \\
&\leq \sum_{i=T}^{\infty} 2c(1-c)^i = 2(1-c)^T
\end{aligned}$$

[84] 그리고 연산수행부(120)는 네이버예측단계에서 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}}$ 를 예측할 수 있다.

[85] 즉, 연산수행부(120)는 할당된 시드노드를 기준으로 제 1 구간 점수벡터 $\mathbf{r}_{\text{family}}$ 를 계산할 수 있고, 아래와 같이 스케일링된 제 1 구간 점수벡터 $\mathbf{r}_{\text{family}}$ 에 의해 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}}$ 를 예측할 수 있다.

[86] [수학식 13]

$$[87] \quad \tilde{\mathbf{r}}_{\text{neighbor}} = \frac{\|\mathbf{r}_{\text{neighbor}}\|_1}{\|\mathbf{r}_{\text{family}}\|_1} \mathbf{r}_{\text{family}} = \frac{(1-c)^S - (1-c)^T}{1 - (1-c)^S} \mathbf{r}_{\text{family}}$$

[88] c: 재시작확률

[89] 이를 증명하면, 제 1 구간 점수벡터 와 제 2 구간 점수벡터 는 아래의 수식과 같다.

[90] [수학식 14]

$$[91] \quad \mathbf{r}_{\text{family}} = \mathbf{x}^{(0)} + \mathbf{x}^{(1)} + \dots + \mathbf{x}^{(S-1)}$$

$$\mathbf{r}_{\text{neighbor}} = \mathbf{x}^{(S)} + \mathbf{x}^{(S+1)} + \dots + \mathbf{x}^{(T-1)}$$

$$[92] \quad \mathbf{x}^{(i)} = c(1-c)^i (\tilde{\mathbf{A}}^\top)^i \mathbf{q}$$

[93] 그러면, $\|\mathbf{r}_{\text{family}}\|_1$ 와 $\|\mathbf{r}_{\text{neighbor}}\|_1$ 는 아래의 수식과 같다.

[94] [수학식 15]

$$[95] \quad \|\mathbf{r}_{\text{family}}\|_1 = \|\mathbf{x}^{(0)} + \mathbf{x}^{(1)} + \dots + \mathbf{x}^{(S-1)}\|_1 = \sum_{i=0}^{S-1} \|\mathbf{x}^{(i)}\|_1$$

$$\|\mathbf{r}_{\text{neighbor}}\|_1 = \|\mathbf{x}^{(S)} + \mathbf{x}^{(S+1)} + \dots + \mathbf{x}^{(T-1)}\|_1 = \sum_{i=S}^{T-1} \|\mathbf{x}^{(i)}\|_1$$

[96] 여기서 $\mathbf{x}^{(i)}$ 는 모든 원소가 음수가 아니고, $\tilde{\mathbf{A}}^\top$ 와 $(\tilde{\mathbf{A}}^\top)^i$ 는 열 확률 행렬이다. 그래서 $\|(\tilde{\mathbf{A}}^\top)^i \mathbf{q}\|_1 = \|\mathbf{q}\|_1 = 1$ 이고

$$\|\mathbf{x}^{(i)}\|_1 = \|c(1-c)^i (\tilde{\mathbf{A}}^\top)^i \mathbf{q}\|_1 = c(1-c)^i \text{ 이다.}$$

[97] 그러면, $\|\mathbf{r}_{\text{family}}\|_1$ 와 $\|\mathbf{r}_{\text{neighbor}}\|_1$ 는 아래의 수식과 같다.

[98] [수학식 16]

$$[99] \quad \|\mathbf{r}_{\text{family}}\|_1 = \sum_{i=0}^{S-1} c(1-c)^i = 1 - (1-c)^S$$

$$\|\mathbf{r}_{\text{neighbor}}\|_1 = \sum_{i=S}^{T-1} c(1-c)^i = (1-c)^S - (1-c)^T$$

[100] 이후, 연산수행부(120)는 계산된 제 1 구간 점수벡터 $\mathbf{r}_{\text{family}}$ 와 예측된 제 2 구간 점수벡터 $\mathbf{r}_{\text{neighbor}}$ 및 제 3 구간 점수벡터 $\mathbf{r}_{\text{stranger}}$ 를 기초로 RWR점수벡터 $\mathbf{r}_{\text{TPA}}$ 를 계산할 수 있다.

[101] 이하에서는 상술한 바와 같은 유사도계산장치(100)를 이용한 유사도계산방법을 설명한다. 도 2 는 본 발명의 일 실시예에 따른 유사도계산방법을 설명하기 위한 순서도이다.

- [102] 도 2에 도시된 실시예에 따른 유사도계산방법은 도 2에 도시된 유사도계산장치(100)에서 시계열적으로 처리되는 단계들을 포함한다. 따라서, 이하에서 생략된 내용이라고 하더라도 도 2에 도시된 유사도계산장치(100)에 관하여 이상에서 기술한 내용은 도 2에 도시된 실시예에 따른 유사도계산방법에 적용될 수도 있다.
- [103] 우선, 유사도계산장치(100)는 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할할 수 있다(S2001).
- [104] 예를 들어, 유사도계산장치(100)는 그래프의 점수벡터를 계산하기 위한 CIP 연산이 반복되는 횟수를 기준으로 전체 반복횟수를 제 1 구간 내지 제 3 구간으로 나눌 수 있다.
- [105] 이를 위해 유사도계산장치(100)는 구간을 나누는 기준차수를 설정할 수 있다. 예를 들어, 유사도계산장치(100)는 제 1 구간과 제 2 구간을 나누는 회수인 기준차수인 S와 제 2 구간과 제 3 구간을 나누는 기준차수인 T를 사용자로부터 획득할 수 있고, 획득된 S와 T에 기초하여 구간을 나눌 수 있다.
- [106] 이때, S의 값이 커질수록 계산시간은 증가하나 정확도가 높아지는 반면, T가 작아질수록 에러율이 증가한다.
- [107] 이후, 유사도계산장치(100)는 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산할 수 있다.
- [108] 우선, 유사도계산장치(100)는 각 노드에서 진출하는 간선의 수에 따라 결정되는 노드의 중요도에 기초하여 제 3 구간 점수벡터를 예측하여 제 3 구간 점수벡터를 계산할 수 있다(S2002).
- [109] 예를 들어, 유사도계산장치(100)는 페이지랭크를 이용하여 T차수부터 마지막까지 페이지랭크 점수벡터를 계산할 수 있고, 계산된 페이지랭크 점수벡터를 예측된 제 3 구간 점수벡터로 이용할 수 있다.
- [110] 이후, 유사도계산장치(100)는 제 1 구간 점수벡터를 계산할 수 있고(S2003), 계산된 제 1 구간 점수벡터를 스케일링하여 제 2 구간 점수벡터를 예측할 수 있다(S2004).
- [111] 예를 들어, 유사도계산장치(100)는 제 1 구간인 0차수부터 S-1까지 RWR 점수벡터를 계산할 수 있고, 계산된 제 1 구간 점수벡터를 제시작확률 c에 의한 값에 기초하여 스케일링함으로써 제 2 구간 점수벡터를 예측할 수 있다.
- [112] 그리고 유사도계산장치(100)는 계산된 제 1 구간 점수벡터와 예측된 제 2 구간 점수벡터 및 제 3 구간 점수벡터에 기초하여 대용량 그래프의 점수벡터를 계산할 수 있다(S2005).
- [113] 본 실시예에서 사용되는 '~부'라는 용어는 소프트웨어 또는 FPGA(field programmable gate array) 또는 ASIC와 같은 하드웨어 구성요소를 의미하며, '~부'는 어떤 역할들을 수행한다. 그렇지만 '~부'는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '~부'는 어드레싱할 수 있는 저장 매체에 있도록 구성될

수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '~부'는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램특히 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들, 및 변수들을 포함한다.

[114] 구성요소들과 '~부'들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '~부'들로 결합되거나 추가적인 구성요소들과 '~부'들로부터 분리될 수 있다.

[115] 뿐만 아니라, 구성요소들 및 '~부'들은 디바이스 또는 보안 멀티미디어카드 내의 하나 또는 그 이상의 CPU 들을 재생시키도록 구현될 수도 있다.

[116] 또한 본 발명의 일실시에에 따르는 유사도계산방법은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 컴퓨터 프로그램(또는 컴퓨터 프로그램 상품)으로 구현될 수도 있다. 컴퓨터 프로그램은 프로세서에 의해 처리되는 프로그래밍 가능한 기계 명령어를 포함하고, 고레벨 프로그래밍 언어(High-level Programming Language), 객체 지향 프로그래밍 언어(Object-oriented Programming Language), 어셈블리 언어 또는 기계 언어 등으로 구현될 수 있다. 또한 컴퓨터 프로그램은 유형의 컴퓨터 판독가능 기록매체(예를 들어, 메모리, 하드디스크, 자기/광학 매체 또는 SSD(Solid-State Drive) 등)에 기록될 수 있다.

[117] 따라서 본 발명의 일실시에에 따르는 유사도계산방법은 상술한 바와 같은 컴퓨터 프로그램이 컴퓨팅 장치에 의해 실행됨으로써 구현될 수 있다. 컴퓨팅 장치는 프로세서와, 메모리와, 저장 장치와, 메모리 및 고속 확장포트에 접속하고 있는 고속 인터페이스와, 저속 버스와 저장 장치에 접속하고 있는 저속 인터페이스 중 적어도 일부를 포함할 수 있다. 이러한 성분들 각각은 다양한 버스를 이용하여 서로 접속되어 있으며, 공통 머더보드에 탑재되거나 다른 적절한 방식으로 장착될 수 있다.

[118] 여기서 프로세서는 컴퓨팅 장치 내에서 명령어를 처리할 수 있는데, 이런 명령어로는, 예컨대 고속 인터페이스에 접속된 디스플레이처럼 외부 입력, 출력 장치상에 GUI(Graphic User Interface)를 제공하기 위한 그래픽 정보를 표시하기 위해 메모리나 저장 장치에 저장된 명령어를 들 수 있다. 다른 실시예로서, 다수의 프로세서 및(또는) 다수의 버스가 적절히 다수의 메모리 및 메모리 형태와 함께 이용될 수 있다. 또한 프로세서는 독립적인 다수의 아날로그 및(또는) 디지털 프로세서를 포함하는 칩들이 이루는 칩셋으로 구현될 수 있다.

[119] 또한 메모리는 컴퓨팅 장치 내에서 정보를 저장한다. 일례로, 메모리는 휘발성 메모리 유닛 또는 그들의 집합으로 구성될 수 있다. 다른 예로, 메모리는 비휘발성 메모리 유닛 또는 그들의 집합으로 구성될 수 있다. 또한 메모리는 예컨대, 자기 혹은 광 디스크와 같이 다른 형태의 컴퓨터 판독 가능한 매체일 수도 있다.

[120] 그리고 저장장치는 컴퓨팅 장치에게 대용량의 저장공간을 제공할 수 있다.

저장 장치는 컴퓨터 판독 가능한 매체이거나 이런 매체를 포함하는 구성일 수 있으며, 예를 들어 SAN(Storage Area Network) 내의 장치들이나 다른 구성도 포함할 수 있고, 플로피 디스크 장치, 하드 디스크 장치, 광 디스크 장치, 혹은 테이프 장치, 플래시 메모리, 그와 유사한 다른 반도체 메모리 장치 혹은 장치 어레이일 수 있다.

- [121] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [122] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.
- [123]
- [124] 2017년 과제번호 및 사사표기
- [125] 1-1. 과제번호 : 1711055475
- [126] 1-2. 사사 국문표기: "본 연구는 미래창조과학부 및 정보통신기술연구진흥센터의 정보통신방송 연구개발사업의 일환으로 수행하였음" [2013-0-00179, (엑소브레인-3세부)컨텍스트 인지형 Deep-Symbolic 하이브리드 지능 원천 기술 개발 및 언어 지식자원 구축].
- [127] 1-3. 사사 영문표기: "This work was supported by ICT R&D program of MSIP/IITP" [2013-0-00179, Development of Core Technology for Context-aware Deep-Symbolic Hybrid Learning and Construction of Language Resources].
- [128]
- [129] 2-1. 과제번호 : 1711056733
- [130] 2-2. 사사 국문표기 : "이 연구는 2015년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행하였음" (No.R0190-15-2012, 빅데이터 처리 고도화 핵심기술개발 사업 총괄 및 고성능 컴퓨팅 기술을 활용한 성능 가속화 기술 개발).
- [131] 2-3. 사사 영문표기 : "This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT)" (No.R0190-15-2012, High Performance Big Data Analytics Platform Performance Acceleration Technologies Development).
- [132]

산업상 이용가능성

- [133] 본 발명은 대량 그래프에서 노드간의 유사도를 계산하는 방법에 관한 것으로보다 상세하게는 그래프를 구성하는 각 노드의 점수인 점수벡터를 계산하기 위한 반복적인 방법에서 연산을 반복하여 계산하지 않고도 일부 반복구간에 대한 점수벡터를 예측함으로써 점수벡터의 계산량을 줄이는 유사도 계산 방법 및 장치에 대해 산업상 이용가능성이 있다.

청구범위

- [청구항 1] 대용량 그래프에서 노드간의 유사도를 계산하는 장치에 있어서, 상기 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할하는 구간분할부; 및 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산하는 연산수행부를 포함하는, 유사도계산장치.
- [청구항 2] 제 1 항에 있어서, 상기 구간분할부는, 상기 전체 반복횟수를 상기 적어도 하나의 구간으로 분할하는 기준 차수를 설정하는, 유사도계산장치.
- [청구항 3] 제 2 항에 있어서, 상기 연산수행부는, 상기 기준 차수에 기초하여 분할된 적어도 하나의 구간 중 제 1 구간에 대해 기 설정된 시드노드를 기초로 제 1 구간 점수벡터를 계산하는, 유사도계산장치.
- [청구항 4] 제 3 항에 있어서, 상기 연산수행부는, 상기 제 1 구간 점수벡터를 스케일링하고, 스케일링된 제 1 구간 점수벡터를 기초로 제 2 구간 점수벡터를 예측하는, 유사도계산장치.
- [청구항 5] 제 2 항에 있어서, 상기 연산수행부는, 각 노드에서 진출하는 간선의 수에 따라 결정되는 노드의 중요도에 기초하여 제 3 구간 점수벡터를 예측하는, 유사도계산장치.
- [청구항 6] 유사도계산장치가 대용량 그래프에서 노드간의 유사도를 계산하는 방법에 있어서, 상기 그래프를 구성하는 적어도 하나의 노드 각각에 대한 점수를 원소로 갖는 점수 벡터를 계산하는 연산의 전체 반복횟수를 적어도 하나의 구간으로 분할하는 단계; 및 분할된 구간별로 각 구간에 대응되는 연산을 수행하여 점수벡터를 계산하는 단계를 포함하는, 유사도계산방법.
- [청구항 7] 제 6 항에 있어서, 상기 적어도 하나의 구간으로 분할하는 단계는, 상기 전체 반복횟수를 상기 적어도 하나의 구간으로 분할하는 기준 차수를 설정하는 단계를 포함하는, 유사도계산방법.
- [청구항 8] 제 6 항에 있어서, 상기 점수벡터를 계산하는 단계는,

분할된 적어도 하나의 구간 중 제 1 구간에 대해 기 설정된 시드노드를
기초로 제 1 구간 점수벡터를 계산하는 단계를 포함하는,
유사도계산방법.

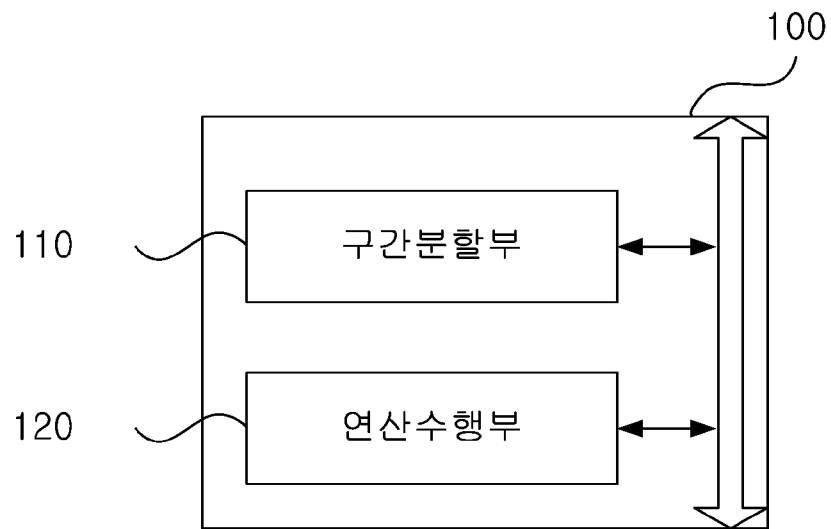
[청구항 9] 제 8 항에 있어서,
상기 점수벡터를 계산하는 단계는,
상기 제 1 구간 점수벡터를 스케일링하여 제 2 구간 점수벡터를 예측하는
단계를 포함하는, 유사도계산방법.

[청구항 10] 제 6 항에 있어서,
상기 점수벡터를 계산하는 단계는,
각 노드에서 진출하는 간선의 수에 따라 결정되는 노드의 중요도에
기초하여 제 3 구간 점수벡터를 예측하는 단계를 포함하는,
유사도계산방법.

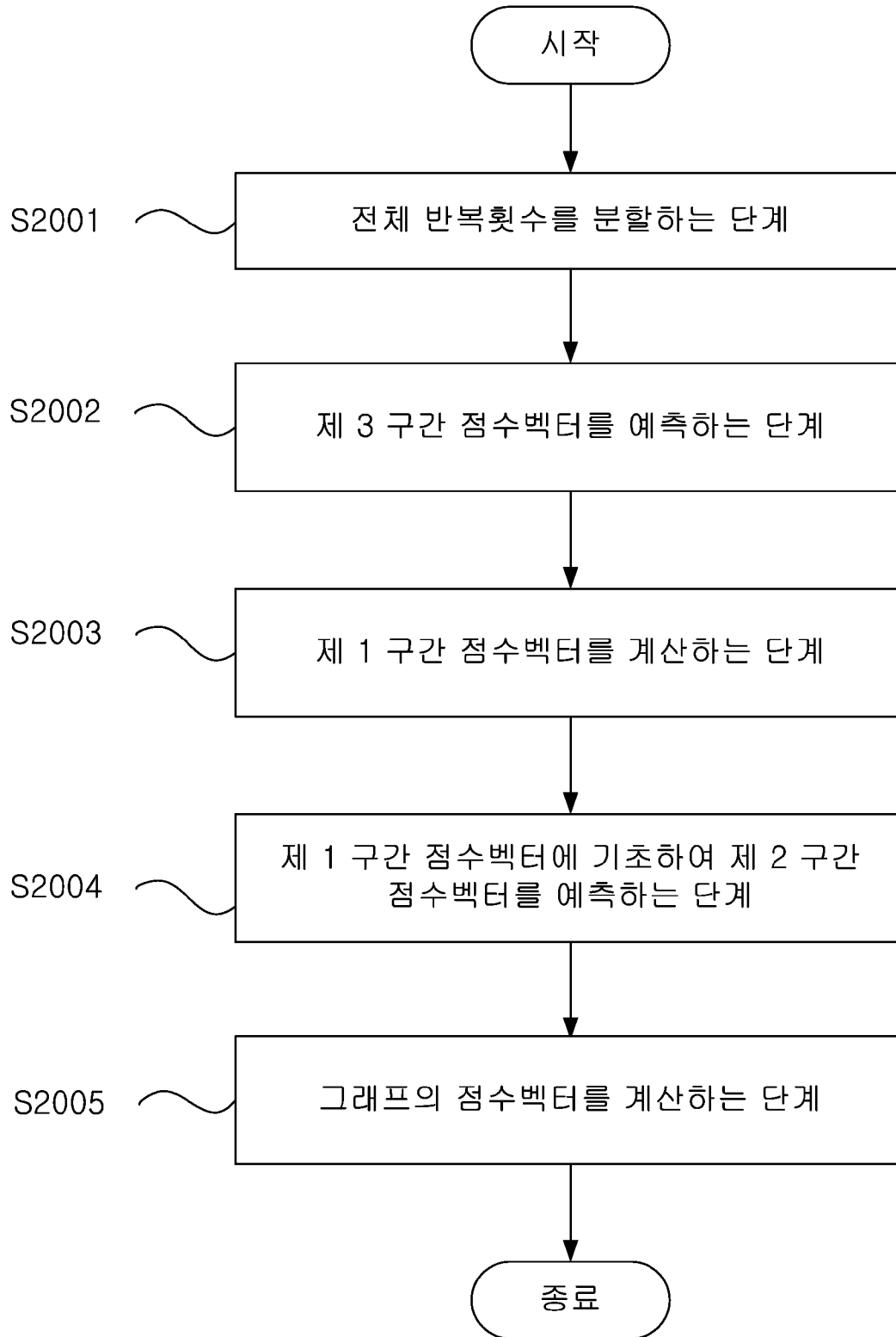
[청구항 11] 유사도계산장치에 의해 수행되며, 제6 항에 기재된 방법을 수행하기 위해
매체에 저장된 컴퓨터 프로그램.

[청구항 12] 제 6 항에 기재된 방법을 수행하는 프로그램이 기록된 컴퓨터 판독가능한
기록매체.

[도1]



[도2]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2017/012962**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/30(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/30; G06F 7/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models: IPC as above

Japanese Utility models and applications for Utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: graph, node, random walk, restart, similarity

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013-0124488 A1 (KANG, U et al.) 16 May 2013 See claims 1-16.	1-12
A	US 2011-0173189 A1 (SINGH, Ambuj Kumar et al.) 14 July 2011 See claims 1-5.	1-12
A	JP 2012-160015 A (NIPPON TELEGR. & TELEPH. CORP. <NTT>) 23 August 2012 See claims 1-3.	1-12
A	JUNG, Jinhong et al., "Random Walk with Restart on Large Graphs Using Block Elimination", ACM Transactions on Database Systems, Vol. 41, No. 2, Article 12, May 2016. See pages 12:6-12:13.	1-12
A	JUNG, Jinhong et al., "BePI: Fast and Memory-Efficient Method for Billion-Scale Random Walk with Restart", SIGMOD' 17, May 14-19, 2017, Chicago, Illinois, USA See pages 3-8.	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

23 JULY 2018 (23.07.2018)

Date of mailing of the international search report

23 JULY 2018 (23.07.2018)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex Daejeon Building 4, 189, Cheongsu-ro, Seo-gu,
Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2017/012962

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YUN, Minji et al., "TPA: Two Phase Approximation for Random Walk with Restart", arXiv:1708.02574v1, 08 August 2017, Retrieved from <https://arxiv.org/abs/1708.02574> See pages 1-10. *The above document is that in which the applicant declared exceptions to lack of novelty.	1-12

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2017/012962

Patent document cited in search report	Publication date	Patent family member	Publication date
US 2013-0124488 A1	16/05/2013	US 8645339 B2	04/02/2014
US 2011-0173189 A1	14/07/2011	US 2007-0239694 A1	11/10/2007
		US 7933915 B2	26/04/2011
		US 8396884 B2	12/03/2013
		WO 2007-100834 A2	07/09/2007
		WO 2007-100834 A3	31/07/2008
JP 2012-160015 A	23/08/2012	JP 5506711 B2	28/05/2014

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G06F 17/30(2006.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G06F 17/30; G06F 7/00 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 그래프, 노드, 랜덤 워크, 재시작, 유사도		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	US 2013-0124488 A1 (U KANG 등) 2013.05.16 청구항 1-16 참조.	1-12
A	US 2011-0173189 A1 (AMBUJ KUMAR SINGH 등) 2011.07.14 청구항 1-5항 참조.	1-12
A	JP 2012-160015 A (NIPPON TELEGR. & TELEPH. CORP. <NTT>) 2012.08.23 청구항 1-3 참조.	1-12
A	JINHONG JUNG 등, Random Walk with Restart on Large Graphs Using Block Elimination, ACM Transactions on Database Systems, Vol. 41, No. 2, Article 12, May 2016. 페이지 12:6 - 12:13 참조.	1-12
A	JINHONG JUNG 등, BePI: Fast and Memory-Efficient Method for Billion-Scale Random Walk with Restart, SIGMOD' 17, May 14-19, 2017, Chicago, Illinois, USA 페이지 3-8 참조.	1-12
☒ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2018년 07월 23일 (23.07.2018)		국제조사보고서 발송일 2018년 07월 23일 (23.07.2018)
ISA/KR의 명칭 및 우편주소  대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 노지명 전화번호 +82-42-481-8528 

C (계속). 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	MINJI YUN 등, TPA: Two Phase Approximation for Random Walk with Restart, arXiv:1708.02574v1, 08 August 2017 https://arxiv.org/abs/1708.02574에서 검색됨. 페이지 1-10 참조. *위 문헌은 출원인이 신규성 상실의 예외로서 선언한 공지문헌임.	1-12

국제조사보고서
대응특허에 관한 정보

국제출원번호
PCT/KR2017/012962

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 2013-0124488 A1	2013/05/16	US 8645339 B2	2014/02/04
US 2011-0173189 A1	2011/07/14	US 2007-0239694 A1	2007/10/11
		US 7933915 B2	2011/04/26
		US 8396884 B2	2013/03/12
		WO 2007-100834 A2	2007/09/07
		WO 2007-100834 A3	2008/07/31
JP 2012-160015 A	2012/08/23	JP 5506711 B2	2014/05/28