



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0103453
(43) 공개일자 2023년07월07일

(51) 국제특허분류(Int. Cl.)
G06F 11/36 (2006.01) G06F 21/64 (2013.01)
(52) CPC특허분류
G06F 11/3604 (2013.01)
G06F 11/3664 (2013.01)
(21) 출원번호 10-2021-0194357
(22) 출원일자 2021년12월31일
심사청구일자 없음

(71) 출원인
삼성에스디에스 주식회사
서울특별시 송파구 올림픽로35길 125 (신천동)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
지윤찬
서울특별시 송파구 올림픽로35길 125 (신천동, 삼성SDS West Campus)
금영준
서울특별시 송파구 올림픽로35길 125 (신천동, 삼성SDS West Campus)
(뒷면에 계속)
(74) 대리인
특허법인가산

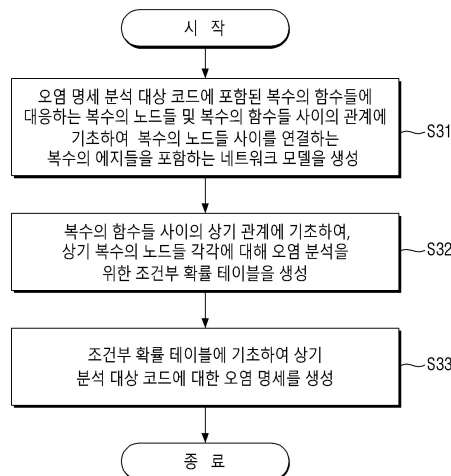
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 프로그램 코드의 오염 명세 생성 방법, 그리고 이를 구현하기 위한 장치

(57) 요약

본 발명의 일 실시예에 따른 컴퓨팅 장치에 의해 수행되는 방법은, 오염 명세 분석 대상 코드에 포함된 복수의 함수들에 대응하는 복수의 노드들 및 상기 복수의 함수들 사이의 관계에 기초하여 상기 복수의 노드들 사이를 연결하는 복수의 에지들(edge)을 포함하는 네트워크 모델을 생성하는 단계, 상기 복수의 함수들 사이의 상기 관계에 기초하여, 상기 복수의 노드들 각각에 대해 오염 분석을 위한 조건부 확률 테이블(CPT: Conditional Probability Table)을 생성하는 단계, 및 상기 조건부 확률 테이블에 기초하여 상기 분석 대상 코드에 대한 오염 명세(taint specification)를 생성하는 단계를 포함한다.

대표도 - 도3



(52) CPC특허분류

G06F 11/3684 (2013.01)

G06F 11/3688 (2013.01)

G06F 11/3692 (2013.01)

G06F 21/64 (2013.01)

(72) 발명자

오학주

서울특별시 강남구 압구정로 151, 122동 1002호(압구정동, 현대아파트)

이정수

서울특별시 동대문구 왕산로 86, 102동 2219호(용두동)

명세서

청구범위

청구항 1

컴퓨팅 장치에 의해 수행되는 방법에 있어서,

오염 명세 분석 대상 코드에 포함된 복수의 함수들에 대응하는 복수의 노드들 및 상기 복수의 함수들 사이의 관계에 기초하여 상기 복수의 노드들 사이를 연결하는 복수의 에지들(edge)을 포함하는 네트워크 모델을 생성하는 단계;

상기 복수의 함수들 사이의 상기 관계에 기초하여, 상기 복수의 노드들 각각에 대해 오염 분석을 위한 조건부 확률 테이블(CPT: Conditional Probability Table)을 생성하는 단계; 및

상기 조건부 확률 테이블에 기초하여 상기 분석 대상 코드에 대한 오염 명세(taint specification)를 생성하는 단계를 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 2

제1항에 있어서,

상기 조건부 확률 테이블을 생성하는 단계는,

상기 조건부 확률 테이블에 초기 값을 부여하는 단계;

상기 복수의 노드들 중 제1 노드의 테인트 분류를 나타내는 외부 제공 정보를 이용하여 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 단계; 및

상기 업데이트된 제1 노드의 테인트 분류 각각에 해당하는 확률 값이 사전 설정된 조건을 만족한다는 판정에 기초하여, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값의 업데이트를 종료하는 단계를 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 3

제2항에 있어서,

상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 단계는,

상기 제1 노드를 선정하는 단계;

상기 제1 노드에 대한 상기 외부 제공 정보에 기초하여, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 갱신하는 단계; 및

상기 제1 노드에 직접 연결된 제2 노드의 테인트 분류 각각에 해당하는 확률 값을 갱신하는 단계를 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 4

제3항에 있어서,

상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 단계는,

상기 제2 노드를 통해 상기 제1 노드에 간접적으로 연결된 제3 노드의 테인트 분류 각각에 해당하는 확률 값을 갱신하는 단계

를 더 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 5

제3항에 있어서,

상기 제1 노드를 선정하는 단계는,

상기 복수의 노드들 중 에지의 수가 최대인 노드를 상기 제1 노드로 선정하는 단계를 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 6

제1항에 있어서,

상기 복수의 엣지들은,

상호간 데이터를 주고받는 함수들에 각각 대응되는 노드들을 연결하는 제1 타입의 에지를 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 7

제1항에 있어서,

상기 복수의 엣지들은,

유사도가 높은 함수들에 각각 대응되는 노드들을 연결하는 제2 타입의 에지를 포함하고,

상기 유사도는 함수의 형태적 특성 및 의미적 특성 중 적어도 하나에 기초하여 결정되는,

프로그램 코드의 오염 명세 생성 방법.

청구항 8

제1항에 있어서,

상기 네트워크 모델은, 베이지안(Bayesian) 네트워크 모델인,

프로그램 코드의 오염 명세 생성 방법.

청구항 9

제1항에 있어서,

상기 조건부 확률 테이블을 생성하는 단계는,

상기 복수의 함수들에 대응되는 각각의 노드에 대하여, 각 노드와 연관된 조건 별로, 각 노드가 제1 테인트 분류에 해당할 제1 확률 값, 제2 테인트 분류에 해당할 제2 확률 값, 그리고 제3 테인트 분류에 해당할 제3 확률 값을 각각 설정하는 단계를 포함하고,

상기 제1 테인트 분류는 테인트 소스(taint source)이고, 상기 제2 테인트 분류는 테인트 싱크(taint sink)이며, 제3 테인트 분류는 테인트 소스 및 테인트 싱크 중 어느 것도 아닌 것에 해당하며,

상기 제1 확률 값 내지 상기 제3 확률 값의 총합은 1인,

프로그램 코드의 오염 명세 생성 방법.

청구항 10

제1항에 있어서,

상기 조건부 확률 테이블을 생성하는 단계는,

상기 복수의 함수들에 대응되는 각각의 노드에 대하여, 각 노드와 연관된 조건 별로, 각 노드가 제1 테인트 분류에 해당할 제1 확률 값, 제2 테인트 분류에 해당할 제2 확률 값, 제3 테인트 분류에 해당할 제3 확률 값, 그

리고 제4 테인트 분류에 해당할 제4 확률 값을 각각 설정하는 단계를 포함하고,

상기 제1 테인트 분류는 테인트 소스이고, 상기 제2 테인트 분류는 테인트 싱크이며, 상기 제4 테인트 분류는 테인트 새니타이저(taint sanitizer)이고, 상기 제3 테인트 분류는 테인트 소스, 테인트 싱크, 테인트 새니타이저 중 어느 것도 아닌 것에 해당하며,

상기 제1 확률 값 내지 상기 제4 확률 값의 총합은 1인,

프로그램 코드의 오염 명세 생성 방법.

청구항 11

제9항에 있어서,

상기 조건부 확률 테이블을 생성하는 단계는,

상기 네트워크 모델의 제2 노드의 자식 노드인 제3 노드가 상기 제2 노드와 동일한 테인트 분류를 가질 조건부 확률 값을, 상기 제3 노드가 상기 제2 노드와 상이한 테인트 분류를 가질 조건부 확률 값보다 높게 설정하는 단계를 포함하고,

상기 제3 노드는 상기 제2 노드와 유사성(similarity)을 가지는 노드인,

프로그램 코드의 오염 명세 생성 방법.

청구항 12

제9항에 있어서,

상기 조건부 확률 테이블을 생성하는 단계는,

상기 네트워크 모델의 제2 노드가 상기 제1 테인트 분류일 경우, 상기 제2 노드의 자식 노드인 제4 노드가 상기 제1 테인트 분류일 조건부 확률 값을, 상기 제4 노드가 상기 제2 테인트 분류 또는 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계; 및

상기 제2 노드가 상기 제2 테인트 분류일 경우, 상기 제4 노드가 상기 제1 테인트 분류 또는 상기 제2 테인트 분류일 조건부 확률 값을 상기 제4 노드가 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계

를 포함하고,

상기 제4 노드는 상기 제2 노드와 데이터를 교환하는 관계를 가지는 노드인,

프로그램 코드의 오염 명세 생성 방법.

청구항 13

제10항에 있어서,

상기 조건부 확률 테이블을 생성하는 단계는,

상기 네트워크 모델의 제2 노드가 상기 제1 테인트 분류일 경우, 상기 제2 노드의 자식 노드인 제4 노드가 상기 제1 테인트 분류일 조건부 확률 값을, 상기 제4 노드가 상기 제2 테인트 분류 내지 상기 제4 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계;

상기 제2 노드가 상기 제2 테인트 분류일 경우, 상기 제4 노드가 상기 제1 테인트 분류 또는 상기 제2 테인트 분류 또는 상기 제4 테인트 분류일 조건부 확률 값을 상기 제4 노드가 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계; 및

상기 제2 노드가 상기 제4 테인트 분류일 경우, 상기 제4 노드가 상기 제1 테인트 분류 또는 상기 제4 테인트 분류일 조건부 확률 값을 상기 제4 노드가 상기 제2 테인트 분류 또는 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계

를 포함하고,

상기 제4 노드는 상기 제2 노드와 데이터를 교환하는 관계를 가지는 노드인,

프로그램 코드의 오염 명세 생성 방법.

청구항 14

제1항에 있어서,

상기 복수의 노드들의 개수가 기준치 이상인 경우 상기 생성된 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계;

상기 네트워크 모델의 복수의 노드들 각각에 대해 생성된 조건부 확률 테이블을 이용하여 상기 제1 네트워크 모델 및 상기 제2 네트워크 모델에 포함된 각 노드 별 조건부 확률 테이블의 초기 확률 값을 설정하는 단계; 및

상기 제1 네트워크 모델 및 상기 제2 네트워크 모델에 포함된 각 노드의 테인트 분류를 나타내는 외부 제공 정보를 이용하여 상기 각 노드에 대해 설정된 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 단계를 더 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 15

제14항에 있어서,

상기 복수의 노드들의 개수가 기준치 이상인 경우 상기 생성된 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계는,

상기 네트워크 모델에서 삭제될 경우 상기 네트워크 모델에서 고립되는 노드의 개수가 최소가 되도록 하는 제1 에지를 결정하는 단계; 및

상기 제1 에지를 삭제하여 상기 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계를 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 16

제14항에 있어서,

상기 복수의 노드의 개수가 기준치 이상인 경우 상기 생성된 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계는,

상기 제1 네트워크 모델 및 제2 네트워크 모델에 속하게 될 노드의 개수의 차이가 기준치 이하가 되도록 분할하는 단계를 포함하는,

프로그램 코드의 오염 명세 생성 방법.

청구항 17

컴퓨터로 하여금 제1 항 내지 제16 항 중 어느 한 항의 방법을 수행하도록 하는 컴퓨터 프로그램이 저장된, 컴퓨터 판독 가능한 비일시적 기록 매체.

청구항 18

오염 명세 생성 장치에 있어서,

하나 이상의 프로세서;

외부장치와 통신하는 통신 인터페이스;

상기 프로세서에 의하여 수행되는 컴퓨터 프로그램을 로드(load)하는 메모리; 및

상기 컴퓨터 프로그램을 저장하는 스토리지를 포함하되,

상기 컴퓨터 프로그램은,

오염 명세 분석 대상 코드에 포함된 복수의 함수들에 대응하는 복수의 노드들 및 상기 복수의 함수들 사이의 관계에 기초하여 상기 복수의 노드들 사이를 연결하는 복수의 에지들(edge)을 포함하는 네트워크 모델을 생성하는 동작,

상기 복수의 함수들 사이의 상기 관계에 기초하여, 상기 복수의 노드들 각각에 대해 오염 분석을 위한 조건부 확률 테이블(CPT: Conditional Probability Table)을 생성하는 동작, 및

상기 조건부 확률 테이블에 기초하여 상기 분석 대상 코드에 대한 오염 명세(taint specification)를 생성하는 동작을 수행하기 위한 인스트럭션을 포함하는,

오염 명세 생성 장치.

청구항 19

제18항에 있어서,

상기 조건부 확률 테이블을 생성하는 동작은,

상기 조건부 확률 테이블에 초기 값을 부여하는 동작,

상기 복수의 노드들 중 제1 노드의 테인트 분류를 나타내는 외부 제공 정보를 이용하여 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 동작, 및

상기 업데이트된 제1 노드의 테인트 분류 각각에 해당하는 확률 값이 사전 설정된 조건을 만족한다는 판정에 기초하여, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값의 업데이트를 종료하는 동작을 포함하는,

오염 명세 생성 장치.

청구항 20

제19항에 있어서,

상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 동작은,

상기 제1 노드를 선정하는 동작,

상기 제1 노드에 대한 상기 외부 제공 정보에 기초하여, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 갱신하는 동작, 및

상기 제1 노드에 직접 연결된 제2 노드의 테인트 분류 각각에 해당하는 확률 값을 갱신하는 동작을 포함하는,

오염 명세 생성 장치.

발명의 설명

기술 분야

[0001] 본 발명은 프로그램 코드의 오염 명세 생성 방법, 그리고 이를 구현하기 위한 장치에 관한 것으로서, 보다 자세하게는, 프로그램 코드에 포함된 함수들에 대한 오염 명세를 생성하기 위한 프로그램 코드의 오염 명세 생성 방법, 그리고 이를 구현하기 위한 장치에 관한 것이다.

배경 기술

[0002] 오염 분석(taint analysis)은 프로그램 코드에 포함되어 있는 함수들 중 외부 데이터를 유입하는 함수로부터 취약점을 발생시키는 함수에 이르기까지 데이터의 전파 경로를 추적하는 기법이다.

[0003] 오염 분석을 위해서는, 프로그램 코드에 포함되어 있는 각각의 함수가 외부 데이터를 유입하는 근원에 해당하는 함수(taint source)인지, 취약점을 발생시킬 수 있는 함수인지(taint sink), 혹은 데이터에 포함된 위협을 제거하는 함수인지(taint sanitizer)를 사전에 알 수 있어야 하는데, 이러한 사전 정보를 담은 명세를 오염 명세

(taint specification)라고 한다.

- [0005] 종래에는 오염 분석을 위해 필요한 오염 명세를 확보하기 위해 프로그램의 개발자 또는 사용자가 직접 분석하는 수동 방식을 주로 사용해 왔고, 이러한 방식은 확장성과 품질면에서 명확한 한계가 존재한다.
- [0006] 최근에는 이러한 단점을 해결하기 위해 머신 러닝 알고리즘 기반으로 오염 명세를 추론하는 방법이 연구되고 있다. 그러나, 이러한 방법 또한 추론 결과에 영향을 주는 기초 지식이 사전에 완전한 형태로 준비되어야 한다는 한계가 존재하고, 사용자가 최적의 기초 지식을 구하기 위한 과정을 반복해야 하므로 그 과정에서 많은 시행 착오가 발생할 수 있다.
- [0007] 따라서, 오염 명세를 추론하기 위해 필요한 기초 지식을 사전에 준비하는 과정을 최소화하면서도, 추가되는 새로운 지식을 용이하게 반영할 수 있는 기술이 요구된다.

발명의 내용

해결하려는 과제

- [0008] 본 발명이 해결하고자 하는 기술적 과제는, 프로그램 코드에 포함되어 있는 함수들 각각이 테인트 소스(taint source), 테인트 싱크(taint sink), 및 테인트 새니타이저(taint sanitizer) 중 어떠한 타입의 함수인지에 관한 정보를 포함하는 오염 명세를 제공할 수 있는 프로그램 코드의 오염 명세 생성 방법, 그리고 이를 구현하기 위한 장치에 관한 것이다.
- [0009] 본 발명이 해결하고자 하는 다른 기술적 과제는, 프로그램 코드에 포함되어 있는 함수들의 테인트 분류와 관련된 사전 정보가 마련되어 있지 않은 경우에도, 함수들 간의 관계에 기초하여 생성한 네트워크 모델을 이용하여 오염 명세를 제공할 수 있는 프로그램 코드의 오염 명세 생성 방법, 그리고 이를 구현하기 위한 장치에 관한 것이다.
- [0010] 본 발명이 해결하고자 하는 또 다른 기술적 과제는, 프로그램 코드에 포함되어 있는 함수들 각각에 대한 오염 명세를 생성 시, 새롭게 추가되는 각 함수에 관한 정보를 실시간으로 반영하여 최적의 오염 명세를 생성할 수 있는 프로그램 코드의 오염 명세 생성 방법, 그리고 이를 구현하기 위한 장치에 관한 것이다.
- [0011] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 개시의 기술분야에서의 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0012] 상기 기술적 과제를 해결하기 위한, 본 발명의 일 실시예에 따른 컴퓨팅 장치에 의해 수행되는 방법은, 오염 명세 분석 대상 코드에 포함된 복수의 함수들에 대응하는 복수의 노드들 및 상기 복수의 함수들 사이의 관계에 기초하여 상기 복수의 노드들 사이를 연결하는 복수의 에지들(edge)을 포함하는 네트워크 모델을 생성하는 단계, 상기 복수의 함수들 사이의 상기 관계에 기초하여, 상기 복수의 노드들 각각에 대해 오염 분석을 위한 조건부 확률 테이블(CPT: Conditional Probability Table)을 생성하는 단계, 및 상기 조건부 확률 테이블에 기초하여 상기 분석 대상 코드에 대한 오염 명세(taint specification)를 생성하는 단계를 포함한다.
- [0013] 일 실시예로서, 상기 조건부 확률 테이블을 생성하는 단계는, 상기 조건부 확률 테이블에 초기 값을 부여하는 단계, 상기 복수의 노드들 중 제1 노드의 테인트 분류를 나타내는 외부 제공 정보를 이용하여 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 단계, 및 상기 업데이트된 제1 노드의 테인트 분류 각각에 해당하는 확률 값이 사전 설정된 조건을 만족한다는 판정에 기초하여, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값의 업데이트를 종료하는 단계를 포함할 수 있다.
- [0014] 일 실시예로서, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 단계는, 상기 제1 노드를 선정하는 단계, 상기 제1 노드에 대한 상기 외부 제공 정보에 기초하여, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 갱신하는 단계, 및 상기 조건부 확률 테이블 내의 값들 중 상기 제1 노드에 직접 연결된 제2 노드에 대응되는 값을 갱신하는 단계를 포함할 수 있다.
- [0015] 일 실시예로서, 상기 제1 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 단계는, 상기 제2 노드를 통해 상기 제1 노드에 간접적으로 연결된 제3 노드의 테인트 분류 각각에 해당하는 확률 값을 갱신하는 단계

를 더 포함할 수 있다.

- [0016] 일 실시예로서, 상기 제1 노드를 선정하는 단계는, 상기 복수의 노드들 중 에지의 수가 최대인 노드를 상기 제1 노드로 선정하는 단계를 포함할 수 있다.
- [0017] 일 실시예로서, 상기 복수의 엣지들은, 상호간 데이터를 주고받는 함수들에 각각 대응되는 노드들을 연결하는 제1 타입의 에지를 포함할 수 있다.
- [0018] 일 실시예로서, 상기 복수의 엣지들은, 유사도가 높은 함수들에 각각 대응되는 노드들을 연결하는 제2 타입의 에지를 포함하고, 상기 유사도는 함수의 형태적 특성 및 의미적 특성 중 적어도 하나에 기초하여 결정될 수 있다.
- [0019] 일 실시예로서, 상기 네트워크 모델은, 베이지안(Bayesian) 네트워크 모델일 수 있다.
- [0020] 일 실시예로서, 상기 조건부 확률 테이블을 생성하는 단계는, 상기 복수의 함수들에 대응되는 각각의 노드에 대하여, 각 노드와 연관된 조건 별로, 각 노드가 제1 테인트 분류에 해당할 제1 확률 값, 제2 테인트 분류에 해당할 제2 확률 값, 그리고 제3 테인트 분류에 해당할 제3 확률 값을 각각 설정하는 단계를 포함하고, 상기 제1 테인트 분류는 테인트 소스(taint source)이고, 상기 제2 테인트 분류는 테인트 싱크(taint sink)이며, 제3 테인트 분류는 테인트 소스 및 테인트 싱크 중 어느 것도 아닌 것에 해당하며, 상기 제1 확률 값 내지 상기 제3 확률 값의 총합은 1일 수 있다.
- [0021] 일 실시예로서, 상기 조건부 확률 테이블을 생성하는 단계는, 상기 복수의 함수들에 대응되는 각각의 노드에 대하여, 각 노드와 연관된 조건 별로, 각 노드가 제1 테인트 분류에 해당할 제1 확률 값, 제2 테인트 분류에 해당할 제2 확률 값, 제3 테인트 분류에 해당할 제3 확률 값, 그리고 제4 테인트 분류에 해당할 제4 확률 값을 각각 설정하는 단계를 포함하고, 상기 제1 테인트 분류는 테인트 소스이고, 상기 제2 테인트 분류는 테인트 싱크이며, 상기 제4 테인트 분류는 테인트 새니타이저(taint sanitizer)이고, 상기 제3 테인트 분류는 테인트 소스, 테인트 싱크, 테인트 새니타이저 중 어느 것도 아닌 것에 해당하며, 상기 제1 확률 값 내지 상기 제4 확률 값의 총합은 1일 수 있다.
- [0022] 일 실시예로서, 상기 조건부 확률 테이블을 생성하는 단계는, 상기 네트워크 모델의 제2 노드의 자식 노드인 제3 노드가 상기 제2 노드와 동일한 테인트 분류를 가질 조건부 확률 값을, 상기 제3 노드가 상기 제2 노드와 상이한 테인트 분류를 가질 조건부 확률 값보다 높게 설정하는 단계를 포함하고, 상기 제3 노드는 상기 제2 노드와 유사성(similarity)을 가지는 노드일 수 있다.
- [0023] 일 실시예로서, 상기 조건부 확률 테이블을 생성하는 단계는, 상기 네트워크 모델의 제2 노드가 상기 제1 테인트 분류일 경우, 상기 제2 노드의 자식 노드인 제4 노드가 상기 제1 테인트 분류일 조건부 확률 값을, 상기 제4 노드가 상기 제2 테인트 분류 또는 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계, 및 상기 제2 노드가 상기 제2 테인트 분류일 경우, 상기 제4 노드가 상기 제1 테인트 분류 또는 상기 제2 테인트 분류일 조건부 확률 값을 상기 제4 노드가 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계를 포함하고, 상기 제4 노드는 상기 제2 노드와 데이터를 교환하는 관계를 가지는 노드일 수 있다.
- [0024] 일 실시예로서, 상기 조건부 확률 테이블을 생성하는 단계는, 상기 네트워크 모델의 제2 노드가 상기 제1 테인트 분류일 경우, 상기 제2 노드의 자식 노드인 제4 노드가 상기 제1 테인트 분류일 조건부 확률 값을, 상기 제4 노드가 상기 제2 테인트 분류 내지 상기 제4 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계, 상기 제2 노드가 상기 제2 테인트 분류일 경우, 상기 제4 노드가 상기 제1 테인트 분류 또는 상기 제2 테인트 분류 또는 상기 제4 테인트 분류일 조건부 확률 값을 상기 제4 노드가 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계, 및 상기 제2 노드가 상기 제4 테인트 분류일 경우, 상기 제4 노드가 상기 제1 테인트 분류 또는 상기 제4 테인트 분류일 조건부 확률 값을 상기 제4 노드가 상기 제2 테인트 분류 또는 상기 제3 테인트 분류일 조건부 확률 값보다 낮게 설정하는 단계를 포함하고, 상기 제4 노드는 상기 제2 노드와 데이터를 교환하는 관계를 가지는 노드일 수 있다.
- [0025] 일 실시예로서, 상기 복수의 노드들의 개수가 기준치 이상인 경우 상기 생성된 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계, 상기 네트워크 모델의 복수의 노드들 각각에 대해 생성된 조건부 확률 테이블을 이용하여 상기 제1 네트워크 모델 및 상기 제2 네트워크 모델에 포함된 각 노드 별 조건부 확률 테이블의 초기 확률 값을 설정하는 단계, 및 상기 제1 네트워크 모델 및 상기 제2 네트워크 모델에 포함된 각 노드의 테인트 분류를 나타내는 외부 제공 정보를 이용하여 상기 각 노드에 대해 설정된 테인트 분류 각각에 해

당하는 확률 값을 업데이트하는 단계를 더 포함할 수 있다.

- [0026] 일 실시예로서, 상기 복수의 노드들의 개수가 기준치 이상인 경우 상기 생성된 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계는, 상기 네트워크 모델에서 삭제될 경우 상기 네트워크 모델에서 고립되는 노드의 개수가 최소가 되도록 하는 제1 에지를 결정하는 단계, 및 상기 제1 에지를 삭제하여 상기 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계를 포함할 수 있다.
- [0027] 일 실시예로서, 상기 복수의 노드의 개수가 기준치 이상인 경우 상기 생성된 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 단계는, 상기 제1 네트워크 모델 및 제2 네트워크 모델에 속하게 될 노드의 개수의 차이가 기준치 이하가 되도록 분할하는 단계를 포함할 수 있다.
- [0028] 상기 기술적 과제를 해결하기 위한, 본 발명의 일 실시예에 따른 컴퓨터 판독 가능한 비일시적 기록 매체는, 컴퓨터로 하여금 상기 방법을 수행하도록 하는 컴퓨터 프로그램을 저장할 수 있다.
- [0029] 상기 기술적 과제를 해결하기 위한, 본 발명의 일 실시예에 따른 오염 명세 생성 장치는, 하나 이상의 프로세서, 외부장치와 통신하는 통신 인터페이스, 상기 프로세서에 의하여 수행되는 컴퓨터 프로그램을 로드(load)하는 메모리, 및 상기 컴퓨터 프로그램을 저장하는 스토리지를 포함하되, 상기 컴퓨터 프로그램은, 오염 명세 분석 대상 코드에 포함된 복수의 함수들에 대응하는 복수의 노드들 및 상기 복수의 함수들 사이의 관계에 기초하여 상기 복수의 노드들 사이를 연결하는 복수의 에지들(edge)을 포함하는 네트워크 모델을 생성하는 동작, 상기 복수의 함수들 사이의 상기 관계에 기초하여, 상기 복수의 노드들 각각에 대해 오염 분석을 위한 조건부 확률 테이블(CPT: Conditional Probability Table)을 생성하는 동작, 및 상기 조건부 확률 테이블에 기초하여 상기 분석 대상 코드에 대한 오염 명세(taint specification)를 생성하는 동작을 수행하기 위한 인스트럭션을 포함한다.

도면의 간단한 설명

- [0030] 도 1은 본 발명의 몇몇 실시예들을 설명하기 위한 개념도이다.
- 도 2는 본 발명의 일 실시예에 따른 오염 명세 생성 장치의 구성을 도시한 블록도이다.
- 도 3은 본 발명의 다른 실시예에 따른 프로그램 코드의 오염 명세 생성 방법을 설명하기 위한 순서도이다.
- 도 4는 프로그램 코드에 대한 오염 분석에 관하여 설명하기 위한 도면이다.
- 도 5는 본 발명의 몇몇 실시예에 따라, 예시적인 프로그램 코드를 대상으로 오염 분석을 수행한 결과로서 획득될 수 있는 오염 명세의 예이다.
- 도 6은 본 발명의 몇몇 실시예에 따른 프로그램 코드의 오염 명세 추론 프로세스를 도시한 예이다.
- 도 7은 본 발명의 몇몇 실시예에 따른 네트워크 모델을 구성하는 과정을 도시한 예이다.
- 도 8은 본 발명의 몇몇 실시예에 따른 오염 명세를 추론하는 과정을 도시한 예이다.
- 도 9는 본 발명의 몇몇 실시예에서, 두 노드 간 유사도 에지가 연결된 경우에 CPT의 초기 값을 부여하는 예시적인 방법을 설명하기 위한 도면이다.
- 도 10은 본 발명의 몇몇 실시예에서, 세 노드 간 두 개의 유사도 에지가 연결된 경우에 CPT의 초기 값을 부여하는 예시적인 방법을 설명하기 위한 도면이다.
- 도 11은 본 발명의 몇몇 실시예에서, 두 노드 간 데이터 흐름 에지가 연결된 경우의 CPT의 초기 값을 부여하는 예시적인 방법을 설명하기 위한 도면이다.
- 도 12는 본 발명의 몇몇 실시예에서, 세 노드 간 두 개의 데이터 흐름 에지가 연결된 경우에 CPT의 초기 값을 부여하는 예시적인 방법을 설명하기 위한 도면이다.
- 도 13은 본 발명의 몇몇 실시예에서, 유사도 에지와 데이터 흐름 에지가 조합된 경우에 CPT의 초기 값을 부여하는 예시적인 방법을 설명하기 위한 도면이다.
- 도 14는 본 발명의 몇몇 실시예에 따라, 서로 연결된 복수의 노드들에 대한 CPT의 초기 값을 부여하는 예시적인 방법을 설명하기 위한 도면이다.
- 도 15는 본 발명의 몇몇 실시예에 따른 각 노드에 대한 CPT를 기반으로 초기 확률을 할당하는 예이다.

도 16 내지 도 18은 본 발명의 몇몇 실시예에 따라, 특정 노드의 오염 분석 결과에 관한 외부 정보를 이용하여 인접 노드들의 오염 분류에 관한 확률 값을 업데이트하는 예시적인 방법을 설명하는 도면이다.

도 19는 본 발명의 몇몇 실시예에 따른 네트워크 모델을 분할하는 예이다.

도 20은 본 발명의 몇몇 실시예에 따른 네트워크 모델 분할 시 오염 명세를 추론하는 과정을 도시한 예이다.

도 21은 본 발명의 몇몇 실시예에 따른 방법들을 구현할 수 있는 예시적인 컴퓨팅 장치의 하드웨어 구성도이다.

발명을 실시하기 위한 구체적인 내용

- [0031] 이하, 첨부된 도면을 참조하여 본 개시의 바람직한 실시 예들을 상세히 설명한다. 본 개시의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 개시의 기술적 사상은 이하의 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 이하의 실시예들은 본 개시의 기술적 사상을 완전하도록 하고, 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 본 개시의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 개시의 기술적 사상은 청구항의 범주에 의해 정의될 뿐이다.
- [0032] 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 개시를 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 개시의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.
- [0033] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있다. 또 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다. 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 개시를 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다.
- [0034] 또한, 본 개시의 구성 요소를 설명하는 데 있어서, 제1, 제2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 어떤 구성 요소가 다른 구성요소에 "연결", "결합" 또는 "접속"된다고 기재된 경우, 그 구성 요소는 그 다른 구성요소에 직접적으로 연결되거나 또는 접속될 수 있지만, 각 구성 요소 사이에 또 다른 구성 요소가 "연결", "결합" 또는 "접속"될 수도 있다고 이해되어야 할 것이다.
- [0035] 명세서에서 사용되는 "포함한다 (comprises)" 및/또는 "포함하는 (comprising)"은 언급된 구성 요소, 단계, 동작 및/또는 소자는 하나 이상의 다른 구성 요소, 단계, 동작 및/또는 소자의 존재 또는 추가를 배제하지 않는다.
- [0036] 이하, 본 개시의 몇몇 실시예들에 대하여 첨부된 도면에 따라 상세하게 설명한다.
- [0037] 도 1은 본 발명의 몇몇 실시예들을 설명하기 위한 개념도이다. 도 1을 참조하면, 본 발명의 실시예에 따른 오염 명세 생성 장치(1)는 프로그램 코드의 오염 분석을 위해 필요한 오염 명세를 생성하고, 그 결과를 출력할 수 있는 인터페이스를 제공한다.
- [0038] 먼저, 오염 명세 생성 장치(1)는 사용자 단말(미도시)로부터 분석 대상 코드(2)를 입력 받고, 입력된 분석 대상 코드(2)를 이용하여 네트워크 모델(3)을 생성한다. 이 때, 네트워크 모델(3)은 베이저안 네트워크(Bayesian Network) 알고리즘에 기반한 모델일 수 있고, 그 외에 확률 이론에 기초한 다른 종류의 그래프 모델이 사용될 수도 있다.
- [0039] 오염 명세 생성 장치(1)는 분석 대상 코드(2)에 포함되어 있는 복수의 함수들 각각에 대응하는 복수의 노드들(node)과, 복수의 노드들 사이를 연결하는 복수의 에지들(edge)을 포함하는 네트워크 모델(3)을 생성한다. 이 때, 복수의 에지들은 복수의 함수들 간 관계를 분석한 결과에 기초하여 생성될 수 있다. 예로서, 복수의 함수들 간 관계가 상호간 데이터를 주고 받는 관계인지, 또는 형태 또는 의미상 유사한 관계인지에 따라 다른 타입의 에지가 생성될 수 있다.
- [0040] 일 실시예로서, 오염 명세 생성 장치(1)는 네트워크 모델(3)에 포함된 각 노드 별로 해당 함수가 테인트 소스(taint source), 테인트 싱크(taint sink), 및 테인트 새니타이저(taint sanitizer)가 될 각각의 확률 값을 설

정할 수 있다. 이 때, 각 노드 별로 설정되는 확률 값은 베이지안(Bayesian) 이론에 기초한 조건부 확률 값으로, 연결되어 있는 노드들의 확률 값에 의해 영향을 받을 수 있다.

- [0041] 오염 명세 생성 장치(1)는 네트워크 모델(3)에 포함된 노드에 대응하는 특정 함수에 대한 질의를 외부로 요청할 수 있고, 요청된 질의에 응답하여 외부로부터 제공되는 외부 제공 정보(4)를 이용하여 특정 함수에 대응하는 노드에 대해 신뢰 전파 알고리즘에 기초하여 계산된 테인트 분류 각각에 해당하는 확률 값을 실시간으로 업데이트할 수 있다. 여기서, 테인트 분류는 테인트 소스, 테인트 싱크, 테인트 새니타이저, 및 테인트 새니타이저 중 어느 것도 아닌 것을 포함할 수 있다.
- [0042] 외부 제공 정보(4)는 해당 함수가 테인트 소스(taint source), 테인트 싱크(taint sink), 및 테인트 새니타이저(taint sanitizer) 중 어느 타입의 함수인지에 대한 기초 지식을 포함할 수 있다. 외부 제공 정보(4)는 사용자 단말(미도시)로부터 입력되거나, 외부 장치에 저장되어 있는 정보가 자동으로 제공될 수도 있다.
- [0043] 또한, 오염 명세 생성 장치(1)는 신뢰 전파(belief propagation) 알고리즘에 기반하여 특정 함수에 대응하는 노드와 직접 또는 간접적으로 연결되어 있는 인접 노드에 대한 테인트 분류 각각에 해당하는 확률 값을 업데이트할 수 있다.
- [0044] 이에 따라, 오염 명세 생성 장치(1)는 최종적으로 업데이트 된 확률 값에 기초하여 분석 대상 코드(2)에 대한 오염 명세(5)를 생성하고, 이를 화면에 출력하거나 외부 장치로 제공할 수 있다.
- [0045] 상기와 같이, 본 발명의 실시예에 따른 오염 명세 생성 장치(1)에 의하면, 프로그램 코드에 포함되어 있는 함수들 간의 관계에 기초하여 생성한 네트워크 모델을 이용하여 프로그램 코드에 대한 오염 명세를 제공할 수 있다. 이에 따라, 각 함수들의 테인트 분류와 관련된 사전 정보가 마련되어 있지 않은 경우에도, 각 함수에 대해 추가되는 정보를 실시간으로 반영함에 의해 정확도가 높은 최적의 오염 명세를 제공할 수 있다.
- [0046] 도 2는 본 발명의 일 실시예에 따른 오염 명세 생성 장치의 구성을 도시한 블록도이다. 본 발명의 실시예에 따른 오염 명세 생성 장치(1)는 통신부(16), 저장부(17), 및 프로세서(11)를 포함하고, 통신부(16)를 통해 외부 장치(10)와 연결될 수 있다.
- [0047] 오염 명세 생성 장치(1)는 프로그램 코드의 오염 명세 생성을 위한 일련의 동작을 수행할 수 있는 인터페이스를 제공하는 장치로서, 예컨대 서버 또는 PC와 같은 컴퓨팅 장치로 구현될 수 있다.
- [0048] 통신부(16)는 외부 장치(10)와 유선 또는 무선 통신방식을 이용하여 통신한다. 통신부(16)는 이더넷(Ethernet) 등과 같은 유선 통신방식으로 사용자 단말(10)과 통신하거나, 와이파이(Wi-Fi) 또는 블루투스(Bluetooth) 등과 같은 무선 통신방식으로 사용자 단말(10)과 통신할 수 있다. 통신부(16)가 통신하는 방식은 이에 한정되지 않고, 다른 통신방식을 이용하여 통신할 수도 있다.
- [0049] 저장부(17)는 프로그램 코드에 포함된 복수의 함수들 간의 관계에 기초하여 생성되는 네트워크 모델에 관한 정보와, 네트워크 모델에 포함된 각각의 노드에 대해 설정되는 조건부 확률 테이블(CPT: Conditional Probability Table)에 관한 정보, 및 조건부 확률 테이블에 기초하여 생성되는 오염 명세에 관한 정보를 저장할 수 있다.
- [0050] 프로세서(11)는 모델 생성 모듈(12), CPT 생성 모듈(13), 및 오염 명세 생성 모듈(14)을 포함하고, 이 외에 네트워크 모델의 분할과 관련된 추가적인 모듈을 더 포함할 수도 있다.
- [0051] 모델 생성 모듈(12)은 프로그램 코드에 포함된 복수의 함수들을 분석하여 복수의 노드들 및 복수의 노드들 사이를 연결하는 복수의 에지들을 포함하는 네트워크 모델, 예컨대 베이지안 네트워크 모델을 생성한다. 이 때, 복수의 노드들은 복수의 함수들 각각에 대응하여 생성되고, 복수의 에지들은 복수의 함수들 간 관계에 기초하여 생성된다. 예로서, 복수의 에지들은 복수의 함수들 간 데이터를 주고 받는 관계를 나타내는 데이터 흐름 에지(data flow edge)와, 복수의 함수들 간 형태 또는 의미상 유사한 관계를 나타내는 유사도 에지(similarity edge)로 구분될 수 있다.
- [0052] CPT 생성 모듈(13)은 모델 생성 모듈(12)에서 생성된 네트워크 모델 상의 각 노드에 대해 조건부 확률 테이블(CPT)을 생성한다. 이 때, 각 노드에 대해 생성되는 조건부 확률 테이블은, 각 노드에 대응하는 해당 함수의 테인트 분류 별 확률 값을 포함한다. 여기서, 테인트 분류는, 테인트 소스(taint source), 테인트 싱크(taint sink), 및 테인트 새니타이저(taint sanitizer)의 세 가지 타입을 포함하고, 세 가지 타입 중 어느 것도 아닌 경우를 더 포함할 수 있다.
- [0053] 테인트 소스는, 프로그램 코드 중 외부에서 유입되는 값을 반환하는 함수로, 예컨대 HTTP request를 읽어 반환

하는 함수나, 사용자 입력을 읽어서 반환하는 함수 등이 이에 해당될 수 있다. 테인트 싱크는, 프로그램 코드 중 조작된 값이 파라미터로 입력될 경우 보안 문제를 일으킬 수 있는 함수로, 예컨대 SQL 구분 실행문이나 시스템 명령어 실행문 등이 해당될 수 있다. 테인트 새니타이저는, 데이터에 포함된 의도하지 않은 내용을 제거하는 함수로, 데이터를 변형 또는 정제하는 과정에서 입력한 값을 무력화시키는 함수가 이에 해당될 수 있다.

[0054] 일 실시예로서, CPT 생성 모듈(13)은 각 노드의 테인트 분류를 나타내는 외부 제공 정보를 이용하여 각 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트할 수 있다. 즉, 각 노드가 테인트 소스인지, 테인트 싱크인지, 및 테인트 새니타이저인지를 알 수 있는 기초 지식이 사용자 입력 또는 외부 장치(10)로부터 제공되면, 신뢰 전파 알고리즘에 기초하여 외부로부터 제공된 기초 지식을 이용하여 각 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트할 수 있다. 이에 따라, 기초 지식이 제공된 노드 및 그 노드와 직접 또는 간접적으로 연결된 다른 노드들의 테인트 분류 각각에 해당하는 확률 값이 업데이트될 수 있다.

[0055] 또한, CPT 생성 모듈(13)은 사전 설정된 조건을 만족하게 되면 각 노드의 테인트 분류 각각에 해당하는 확률 값의 업데이트를 중단함에 의해, 업데이트 과정을 완료할 수 있다. 예로서, 각 노드에 대한 외부 제공 정보에 기초하여 각 노드의 테인트 분류 각각에 해당하는 확률 값의 업데이트가 반복적으로 수행되는 과정에서, 모든 노드들의 테인트 분류에 따른 각 확률 값들의 변동치가 기준치 이하가 되면 업데이트를 종료하고, 최종적으로 모든 노드의 테인트 분류 각각에 해당하는 확률 값을 획득할 수 있다.

[0056] 오염 명세 생성 모듈(14)은 CPT 생성 모듈(13)에서 생성된 최종적인 조건부 확률 테이블을 이용하여 각 함수에 대한 테인트 분류 정보를 포함하는 오염 명세를 생성할 수 있다. 오염 명세 생성 모듈(14)이 생성하는 오염 명세의 형태는 도 5를 참조하여 확인할 수 있다.

[0057] 도 5에 도시된 예에서, 오염 명세(50)는 프로그램 코드에 포함되어 있는 각 함수(51)가 테인트 소스, 테인트 싱크, 및 테인트 새니타이저 중 어느 타입의 함수인지, 즉 테인트 분류를 나타내는 라벨 정보(52)를 포함할 수 있다. 또한, 오염 명세(50)는 테인트 소스로 분류된 함수에 대한 반환값 정보와, 테인트 싱크로 분류된 함수에 대한 파라미터 정보를 포함할 수 있다.

[0058] 이와 같이 생성된 오염 명세(50)는 프로그램 코드의 오염 분석을 위해 사용될 수 있다. 예로서, 도 4에 도시된 예에서, 프로그램 코드(40)에 대해 생성된 오염 명세(50)에 있어, 'theCookie.getvalue()'는 테인트 소스로 분류되고, 'batchUpdate()'는 테인트 싱크로 분류될 수 있다. 이 때, 프로그램 코드(40) 내에서 오염된 값인 'theCookie'(41)가 'HTTP request'를 통해 유입되면, 'theCookie'(41)는 테인트 소스인 'theCookie.getvalue()'를 통해 'param'(42)로 전파(propagation)되고, 다시 'bar'(43)와 'sql'(44)를 거쳐 테인트 싱크인 'batchUpdate(sql)')(45)로 전파될 수 있다. 이에 따라, 오염된 값인 'theCookie'(41)가 'batchUpdate(sql)')(45)에 유입되어 보안 문제가 발생할 수 있으므로, 즉시 전파를 중단하고 경고를 발생시킬 수 있다.

[0059] 도 3은 본 발명의 다른 실시예에 따른 프로그램 코드의 오염 명세 생성 방법을 설명하기 위한 순서도이다.

[0060] 본 발명의 실시예에 따른 프로그램 코드의 오염 명세 생성 방법은 컴퓨팅 장치(100)에 의하여 실행될 수 있고, 예컨대 오염 명세 생성 장치(1)에 의해 실행될 수 있다. 본 실시예에 따른 방법을 실행하는 상기 컴퓨팅 장치(100)는 응용 프로그램 실행 환경을 구비한 컴퓨팅 장치일 수 있다. 본 발명의 실시예에 따른 방법에 포함되는 일부 동작의 수행 주체에 대한 기재가 생략될 수 있으며, 그러한 경우 그 주체는 상기 컴퓨팅 장치(100)임을 유의한다.

[0061] 도 3을 참조하면, 먼저, 동작 S31에서, 오염 명세 분석 대상 코드에 포함된 복수의 함수들에 대응하는 복수의 노드들 및 복수의 함수들 사이의 관계에 기초하여 복수의 노드들 사이를 연결하는 복수의 에지들을 포함하는 네트워크 모델이 생성된다. 여기서, 네트워크 모델은 베이지안 네트워크 모델일 수 있다.

[0062] 일 실시예로서, 복수의 에지들은, 상호간 데이터를 주고받는 함수들에 각각 대응되는 노드들을 연결하는 데이터 흐름 에지(data flow edge)를 포함할 수 있다. 또한, 복수의 에지들은, 유사도가 높은 함수들에 각각 대응되는 노드들을 연결하는 유사도 에지(similarity edge)를 포함할 수 있다. 여기서, 함수들 간 유사도는 함수의 형태적 특성 및 의미적 특성 중 적어도 하나에 기초하여 결정될 수 있다. 함수들 간 유사도 결정을 위해 고려하는 형태적 특성 및 의미적 특성은 아래와 같은 특성들을 포함할 수 있다.

[0063] 1) 형태적 특성

[0064] 1-1) 파라미터/리턴값 존재 여부

- [0065] 1-2) 메소드명/클래스명에 동일 단어의 존재 여부
- [0066] 1-3) 파라미터에 사용한 타입을 리턴 타입에 사용하는지 여부
- [0067] 1-4) 동일한 리턴 타입을 사용하는지 여부
- [0068] 1-5) 메소드명에 사용된 리턴 타입을 사용하는지 여부
- [0069] 2) 의미적 특성
- [0070] 2-1) 적어도 하나의 공통 함수를 호출하는지 여부
- [0071] 2-2) 호출하는 함수 중 파라미터가 없는 호출 존재 여부
- [0072] 2-3) 입력 파라미터 존재 여부
- [0073] 2-4) 호출하는 함수로 데이터 플로우를 전이하는지 여부
- [0074] 2-5) 입력 파라미터와 출력 파라미터를 모두 가지는지 여부
- [0075] 2-6) 넘겨받은 파라미터 중 메소드 안에서 소멸하는 파라미터 존재 여부
- [0076] 2-7) 재정의 하는 변수의 존재 여부

- [0078] 상기와 같은 특성들에 기초하여 임의의 두 함수가 형태 또는 의미상 유사한 지를 판단할 수 있고, 각 특성에 부여되는 점수를 합산하여 두 함수의 유사도 점수를 산출할 수 있다. 이 때, 각 특성에 부여되는 점수는 중요도에 따라 상이하게 부여될 수 있다. 예로서, 흔하게 만족되는 특성에는 낮은 점수가 부여되고, 흔하지 않게 만족되는 특성에는 높은 점수가 부여될 수 있다.
- [0079] 이를 통해 산출되는 두 함수의 유사도 점수가 기준치를 초과하면, 두 함수는 서로 유사한 관계가 있다고 결정되고, 네트워크 모델 상에서 두 함수를 연결하는 유사도 에지가 생성할 수 있다.
- [0080] 다음으로, 동작 S32에서, 복수의 함수들 사이의 관계에 기초하여, 복수의 노드들 각각에 대해 오염 분석을 위한 조건부 확률 테이블(CPT)이 생성된다.
- [0081] 일 실시예로서, 동작 S32는, 조건부 확률 테이블에 초기 값이 부여되는 동작, 복수의 노드들 중 특정 노드의 테인트 분류를 나타내는 외부 제공 정보를 이용하여 특정 노드의 테인트 분류 각각에 해당하는 확률 값이 업데이트되는 동작, 및 업데이트된 특정 노드의 테인트 분류 각각에 해당하는 확률 값의 업데이트를 종료하는 동작을 포함할 수 있다. 이 때, 복수의 노드들 중 외부 제공 정보를 요청할 특정 노드를 선택 시, 확률 값에 영향을 줄 수 있는 인접 노드의 수가 많은 노드를 선택하는 것이 바람직하다. 예로서, 연결된 에지의 수가 최대인 노드를 선택하고, 선택된 노드가 어떤 오염 타입을 갖는 함수인지에 관한 정보를 사용자 단말 또는 외부 장치로 요청할 수 있다.
- [0082] 여기서, 테인트 분류는, 분석 대상 코드 내에서 각 함수가 어떠한 오염 타입으로 분류되는 지에 관한 것으로, 예컨대 테인트 소스, 테인트 싱크, 및 테인트 세니타이저의 세 가지 타입을 포함하고, 세 가지 타입 중 어느 것도 아닌 경우를 더 포함할 수 있다.
- [0083] 일 실시예로서, 특정 노드의 테인트 분류 각각에 해당하는 확률 값이 업데이트되는 동작은, 복수의 노드들 중 특정 노드가 선정되는 동작, 특정 노드에 대한 외부 제공 정보에 기초하여, 특정 노드의 테인트 분류 각각에 해당하는 확률 값이 갱신되는 동작, 및 특정 노드에 직접 연결된 인접 노드의 테인트 분류 각각에 해당하는 확률 값이 갱신되는 동작을 포함할 수 있다. 이 때, 특정 노드의 선정 시, 복수의 노드들 중 에지의 수가 최대인 노드가 특정 노드로 선정될 수 있다. 이 때, 특정 노드에 대한 외부 제공 정보는, 선정된 특정 노드에 대응하는 함수가 어떤 테인트 분류에 해당하는 함수인지에 관한 기초 지식으로서, 사용자 단말 또는 외부 장치로 요청하여 얻을 수 있다.
- [0084] 또한, 특정 노드의 테인트 분류 각각에 해당하는 확률 값이 업데이트되는 동작은, 특정 노드와 직접 연결된 노드를 통해 특정 노드에 간접적으로 연결된 노드의 테인트 분류 각각에 해당하는 확률 값이 갱신되는 동작을 더 포함할 수 있다.

- [0085] 마지막으로, 동작 S33에서, 조건부 확률 테이블에 기초하여 분석 대상 코드에 대한 오염 명세가 생성된다. 이 때, 생성되는 오염 명세는 분석 대상 코드에 포함되어 있는 각 함수가 테인트 소스, 테인트 싱크, 테인트 새니타이저, 및 어느 것도 아닌 경우 중 어느 타입의 함수인지에 관한 정보를 포함할 수 있다.
- [0086] 상기와 같은 본 발명의 실시예에 따른 방법에 의하면, 프로그램 코드에 포함되어 있는 함수들 간의 관계에 기초하여 생성한 네트워크 모델을 이용하여 오염 명세를 제공할 수 있다. 또한, 새롭게 추가되는 각 함수에 관한 정보를 실시간으로 반영하여 최적의 오염 명세를 생성할 수 있다.
- [0087] 도 6은 본 발명의 몇몇 실시예에 따른 프로그램 코드의 오염 명세 추론 프로세스를 도시한 예이다. 도 6을 참조하면, 먼저 동작 S51에서, 분석 대상 코드가 입력되고, 동작 S52에서, 입력된 분석 대상 코드에 포함된 복수의 함수들에 대한 관계 분석을 통해 네트워크 모델이 생성된다.
- [0088] 이 때, 동작 S52는, 도 7에 도시된 바와 같이, 정적 분석 및 노드를 구성하는 동작(S521)과, 데이터 흐름 분석 및 함수들의 유사도를 분석하는 동작(S5221, S5222), 및 노드들 간 데이터 흐름 에지 및 유사도 에지를 연결하는 동작(S5231, S5232)을 포함할 수 있다.
- [0089] 동작 S521에서는, 분석 대상 코드에 포함된 복수의 함수들을 추출하여 각 함수에 대응하는 노드를 생성하고, 함수들 간 관계를 분석한다. 동작 S5221과 동작 S5222에서는, 함수들 간 관계 분석을 통해 함수들이 상호간 데이터를 주고받는 관계인지, 형태와 의미상 유사한 관계인지를 분석한다. 이에 따라, 동작 S5231과 동작 S5232에서는, 분석 결과에 따른 데이터 흐름 에지 및 유사도 에지를 연결한다.
- [0090] 동작 S52가 수행된 이후, 네트워크 모델이 분할되는 동작 S53이 선택적으로 수행될 수 있다. 동작 S53에서, 복수의 노드들의 개수가 기준치 이상인 경우, 동작 S52에서 생성된 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할할 수 있다.
- [0091] 일 실시예로서, 동작 S53은, 네트워크 모델에서 삭제될 경우 네트워크 모델에서 고립되는 노드의 개수가 최소가 되도록 하는 제1 에지를 결정하는 동작, 및 제1 에지를 삭제하여 네트워크 모델을 제1 네트워크 모델 및 제2 네트워크 모델로 분할하는 동작을 포함할 수 있다. 또한, 동작 S53은, 제1 네트워크 모델 및 제2 네트워크 모델에 속하게 될 노드의 개수의 차이가 기준치 이하가 되도록 분할하는 동작을 포함할 수 있다.
- [0092] 네트워크 모델의 분할과 관련된 구체적인 내용은 이후 도 19 및 도 20에 대한 설명 부분에서 좀더 자세히 설명하도록 한다.
- [0093] 다음으로, 동작 S54에서, 노드 별로 조건부 확률 테이블(CPT)의 초기 확률 값이 설정된다. 이 때, 각 노드에 대해, 해당 함수가 테인트 소스, 테인트 싱크, 및 테인트 새니타이저가 될 각각의 확률 값(P.src, P.sink, P.san)을 설정하게 된다. 조건부 확률 테이블(CPT)의 초기 확률 값은, 테인트 소스, 테인트 싱크, 및 테인트 새니타이저 중 어느 것에도 속하지 않는 경우의 확률 값(P.non)을 추가적으로 더 포함할 수 있다. 이 때, 각각의 확률 값(P.src, P.sink, P.san, P.non)의 합은 1이 되어야 한다. 일 실시예로서, 조건부 확률 테이블(CPT)의 초기 확률 값은, 테인트 새니타이저가 될 확률 값(P.san)을 제외하고 설정될 수도 있다.
- [0094] 마지막으로, 동작 S55에서, 신뢰전과 알고리즘에 의해 각 노드에 대한 오염 명세의 추론이 수행된다.
- [0095] 동작 S55는, 도 8에 도시된 바와 같이, 질의할 노드를 선정하는 동작 S551, 선정된 노드의 오염 명세를 질의하고 답변 받는 동작 S552, 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 동작 S553, 인접 노드의 테인트 분류 각각에 해당하는 확률 값을 업데이트하는 동작 S554, 및 업데이트 종료 조건을 만족하는지 여부를 판단하는 동작 S555를 포함할 수 있다. 여기서, 테인트 분류 각각에 해당하는 확률 값은, 해당 노드가 테인트 소스, 테인트 싱크, 테인트 새니타이저, 및 이들 중 어느 것도 아닌 것이 될 각각의 확률 값(P.src, P.sink, P.san, P.non)을 의미한다.
- [0096] 동작 S551에서는, 복수의 노드들 중 인접 노드에 영향을 많이 줄 수 있는 노드를 질의할 노드로 선정할 수 있다. 예로서, 복수의 노드들 중 에지의 수가 최대인 노드가 선정될 수 있다. 동작 S552에서는, 선정된 노드의 테인트 분류와 관련된 기초 지식을 요청하는 질의를 사용자 단말 또는 외부 장치로 전송하고, 그에 대한 답변을 제공받을 수 있다. 즉, 선정된 노드가 테인트 소스, 테인트 싱크, 및 테인트 새니타이저 중 어느 타입인지에 관한 정보를 사용자 단말로부터 입력 받거나, 외부 장치에 저장되어 있는 정보를 이용하여 제공받을 수 있다.
- [0097] 동작 S553에서는, 선정된 노드에 대한 질의 요청에 대해, 사용자 단말 또는 외부 장치로부터 제공된 기초 지식을 이용하여 해당 노드의 테인트 분류 각각에 해당하는 확률 값(P.src, P.sink, P.san, P.non)을 업데이트할 수 있다. 예로서, 질의 요청에 대해, 해당 노드가 테인트 소스라는 기초 지식(P.src=1)이 제공되면, 그 기초 지식

(P.src=1)을 반영하여 해당 노드의 테인트 분류 각각에 해당하는 확률 값(P.src, P.sink, P.san, P.non)을 업데이트할 수 있다. 이 때, 해당 노드 뿐 아니라, 해당 노드에 직접 또는 간접적으로 연결되어 있는 인접 노드들의 테인트 분류 각각에 해당하는 확률 값도 업데이트될 수 있다(동작 S554). 동작 S555에서는, 사전에 설정된 업데이트 종료 조건이 만족되면, 업데이트 동작이 중단되고, 이에 따라 노드의 테인트 분류 각각에 해당하는 최종 확률 값(P.src, P.sink, P.san, P.non)이 결정될 수 있다. 예로서, 업데이트 종료 조건은, 업데이트 동작이 반복적으로 수행되는 과정에서 노드의 확률 값의 변동치가 기준치 이하인지 여부를 판단하는 것일 수 있다. 즉, 외부로부터 제공되는 기초 지식에 의해 노드의 테인트 분류 각각에 해당하는 확률 값이 계속해서 업데이트되는 과정에서 이전 업데이트 시점의 확률 값과 현재 업데이트 시점의 확률 값의 차이가 크지 않다면 업데이트를 종료할 수 있다. 이에 따라, 최종 업데이트된 테인트 분류 각각에 해당하는 확률 값을 이용하여 해당 노드가 테인트 소스, 테인트 싱크, 테인트 새니타이저, 및 아무 것도 아닌 경우 중 어느 타입에 해당하는지를 결정하고, 이로부터 각 노드에 대한 오염 명세를 생성할 수 있다.

[0098] 도 9는 본 발명의 몇몇 실시예에 따른 두 노드 간 유사도 에지가 연결된 경우의 CPT를 구성하는 예이다. 도 9를 참조하면, 오염 명세 생성 장치(1)는, 두 개의 노드(81, 82)가 유사도 에지(83)에 의해 연결된 경우의 CPT(85)를 생성할 수 있다.

[0099] 도시된 예에서, 부모 노드(f)(81)와 유사도 에지(83)에 의해 연결된 자식 노드(g)(82)에 대한 CPT(85)를 생성 시, 자식 노드(g)(82)가 부모 노드(f)(81)와 동일한 테인트 분류를 가질 조건부 확률 값은, 자식 노드(g)(82)가 부모 노드(f)(81)와 상이한 테인트 분류를 가질 조건부 확률 값보다 높게 설정될 수 있다. 이 때, CPT(85)를 생성하기 위해 사전에 확률 할당 규칙이 설정된 테이블(84)이 사용될 수 있다.

[0100] 도 10은 본 발명의 몇몇 실시예에 따른 세 노드 간 두 개의 유사도 에지가 연결된 경우의 CPT를 구성하는 예이다. 도 10을 참조하면, 오염 명세 생성 장치(1)는 세 개의 노드(91, 92, 94)가 두 개의 유사도 에지(93, 95)에 의해 연결된 경우의 CPT를 생성할 수 있다.

[0101] 도시된 예에서, 두 개의 부모 노드(f,g)(91, 94)와 유사도 에지(93, 95)에 의해 연결된 자식 노드(h)(92)에 대한 CPT를 생성 시, 자식 노드(h)(92)가 부모 노드(f,g)(91, 94)와 동일한 테인트 분류를 가질 조건부 확률 값은, 자식 노드(h)(92)가 부모 노드(f,g)(91, 94)와 상이한 테인트 분류를 가질 조건부 확률 값보다 높게 설정될 수 있다. 이 때, CPT를 생성하기 위해 사전에 확률 할당 규칙이 설정된 테이블(96)이 사용될 수 있다.

[0102] 도 11은 본 발명의 몇몇 실시예에 따른 두 노드 간 데이터 흐름 에지가 연결된 경우의 CPT를 구성하는 예이다. 도 11을 참조하면, 오염 명세 생성 장치(1)는, 두 개의 노드(1001, 1002)가 데이터 흐름 에지(1003)에 의해 연결된 경우의 CPT(1005)를 생성할 수 있다. 이 때, CPT(1005)를 생성하기 위해 이하에서 설명하는 확률 할당 규칙이 설정된 테이블(1004)이 사용될 수 있다.

[0103] 도시된 예에서, 부모 노드(f)(1001)와 데이터 흐름 에지(1003)에 의해 연결된 자식 노드(g)(1002)에 대한 CPT(1005)를 생성 시, 부모 노드(f)(1001)가 테인트 소스이고 자식 노드(g)(1002)가 테인트 싱크 또는 테인트 새니타이저 또는 아무것도 아닌 경우의 조건부 확률 값은, 부모 노드(f)(1001)가 테인트 소스이고 자식 노드(g)(1002)가 테인트 소스인 경우의 조건부 확률 값보다 높게 설정될 수 있다.

[0104] 또한, 부모 노드(f)(1001)가 테인트 싱크이고 자식 노드(g)(1002)가 테인트 소스 또는 테인트 싱크 또는 테인트 새니타이저인 경우의 조건부 확률 값은, 부모 노드(f)(1001)가 테인트 싱크이고 자식 노드(g)(1002)가 아무 것도 아닌 경우의 조건부 확률 값보다 낮게 설정될 수 있다.

[0105] 또한, 부모 노드(f)(1001)가 테인트 새니타이저이고 자식 노드(g)(1002)가 테인트 싱크 또는 아무 것도 아닌 경우의 조건부 확률 값은, 부모 노드(f)(1001)가 테인트 새니타이저이고 자식 노드(g)(1002)가 테인트 소스 또는 테인트 새니타이저인 경우의 조건부 확률 값보다 높게 설정될 수 있다.

[0106] 도 12는 본 발명의 몇몇 실시예에 따른 세 노드 간 두 개의 데이터 흐름 에지가 연결된 경우의 CPT를 구성하는 예이다. 도 12를 참조하면, 오염 명세 생성 장치(1)는 세 개의 노드(111, 112, 114)가 두 개의 데이터 흐름 에지(113, 115)에 의해 연결된 경우의 CPT(117)를 생성할 수 있다. 이 경우, 앞서 도 11에서 설명된 확률 할당 규칙이 세 개의 노드에 대해 동일하게 적용된 테이블(116)이 사용될 수 있다.

[0107] 도 13은 본 발명의 몇몇 실시예에 따른 유사도 에지와 데이터 흐름 에지가 조합된 경우의 CPT를 구성하는 예이다. 도 13을 참조하면, 오염 명세 생성 장치(1)는 세 개의 노드(121, 122, 124)가 데이터 흐름 에지(123)와 유사도 에지(125)에 의해 연결된 경우의 CPT(127)를 생성할 수 있다. 이 경우, 앞서 도 12에서 세 개의 노드가 두 개의 데이터 흐름 에지에 의해 연결된 경우에 생성된 CPT(117)의 확률 할당 규칙과, 도 9에서 두 개의 노드가

유사도 예지에 의해 연결된 경우에 생성된 CPT(85)의 확률 할당 규칙을 결합하여 CPT(127)를 생성할 수 있다.

- [0108] 도 14는 본 발명의 몇몇 실시예에 따른 각 노드에 대한 CPT를 생성하는 예이다. 도 14에 도시된 바와 같이, 오염 명세 장치(1)는 네트워크 모델에 포함된 각 노드(queryForMap, query, map.get, getlogger, printer, logger.info)(131~136)에 대한 CPT(1310~1360)를 생성할 수 있다.
- [0109] 도 15의 예에서, 오염 명세 장치(1)는 도 14에서 생성한 CPT(1310~1360)를 이용하여 각 노드가 테인트 소스인지, 테인트 싱크인지, 테인트 새니타이저인지, 아무 것도 아닌 경우인지에 대한 각각의 확률 값(P.src, P.sink, P.san, P.non)(151~156)을 설정할 수 있다.
- [0110] 도 16의 예에서, 오염 명세 장치(1)는 'queryForMap' 노드(151)가 테인트 소스라는 외부 제공 정보(P.src=1)가 입력되면, 입력된 정보를 이용하여 'queryForMap' 노드(151)와 직접 또는 간접적으로 연결되어 있는 다른 노드들(152~156)의 확률 값을 업데이트 할 수 있다.
- [0111] 또한, 도 17의 예에서, 오염 명세 장치(1)는 'getlogger' 노드(164)가 테인트 소스라는 외부 제공 정보(P.src=1)가 새롭게 입력되면, 입력된 정보를 이용하여 'getlogger' 노드(164)와 직접 또는 간접적으로 연결되어 있는 다른 노드들(162, 163, 165, 166)의 확률 값을 업데이트 할 수 있다.
- [0112] 도 18의 예에서, 오염 명세 장치(1)는 'logger.info' 노드(176)이 테인트 싱크라는 외부 제공 정보(P.sin=1)가 새롭게 입력되면, 입력된 정보를 이용하여 'logger.info' 노드(176)와 직접 또는 간접적으로 연결되어 있는 다른 노드들(171, 172, 173, 175)의 확률 값을 업데이트 할 수 있다.
- [0113] 상기와 같이, 프로그램 코드에 포함되어 있는 함수들 각각에 대한 오염 명세를 생성 시, 새롭게 추가되는 각 함수에 관한 외부 제공 정보를 각 함수의 테인트 분류 각각에 해당하는 확률 값에 실시간으로 반영하여 최적의 오염 명세를 생성할 수 있다.
- [0114] 도 19는 본 발명의 몇몇 실시예에 따른 네트워크 모델을 분할하는 예이다. 도 19를 참조하면, 오염 명세 장치(1)는 복수의 노드들과 복수의 에지들을 포함하는 네트워크 모델(180)을 복수 개의 네트워크 모델(181, 182)로 분할하고, 분할된 각각의 네트워크 모델(181, 182)을 이용하여 오염 명세를 생성할 수 있다.
- [0115] 도시된 예에서, 오염 명세 장치(1)는 네트워크 모델(180)에 포함된 복수의 에지들 중 삭제될 경우 네트워크 모델(180)에서 고립되는 노드의 개수가 최소가 되도록 하는 에지(185)를 선정하고, 선정된 에지(185)를 삭제하여 제1 네트워크 모델(G1)(181)과 제2 네트워크 모델(G2)(182)로 분할할 수 있다. 이 때, 고립되는 노드의 개수가 최소가 되도록 에지(185)를 선정하는 방법은, 예컨대 Kernighan-Lin 이분할 알고리즘에 기반한 것일 수 있다.
- [0116] 또한, 오염 명세 장치(1)는 분할을 통해 제1 네트워크 모델(G1)(181)과 제2 네트워크 모델(G2)(182) 각각에 속하게 될 노드들의 개수의 차이가 기준치 이하가 되도록 네트워크 모델(180)을 분할할 수도 있다.
- [0117] 도 20은 본 발명의 몇몇 실시예에 따른 네트워크 모델 분할 시 오염 명세를 추론하는 과정을 도시한 예이다. 도 20을 참조하면, 먼저 동작 S191에서, 분할된 네트워크 모델 중 추론이 실시되지 않은 네트워크 모델이 선택된다.
- [0118] 동작 S192에서, 분할되기 이전의 네트워크 모델 또는 추론이 실시되었던 다른 분할된 네트워크 모델의 조건부 확률 테이블(CPT)를 이용하여, 동작 S191에서 선택된 분할 네트워크 모델의 각 노드에 대한 CPT의 초기 값이 설정된다. 동작 S193에서는, 각 노드와 직접 또는 간접적으로 연결되어 있는 인접 노드에 대한 테인트 분류 각각에 해당하는 확률 값(P.src, P.sink, P.san, P.non)이 업데이트된다.
- [0119] 다음으로, 동작 S194에서, 분할된 네트워크 모델에 있어 질의할 노드가 선정되고, 동작 S195에서, 선정된 노드의 테인트 분류와 관련된 기초 지식을 요청하는 질의가 사용자 단말 또는 외부 장치로 전송되고, 그에 대한 답변이 제공된다.
- [0120] 동작 S196에서, 동작 S195에서 선정된 노드의 테인트 분류와 관련하여 제공된 답변을 이용하여 해당 노드의 테인트 분류 각각에 해당하는 확률 값(P.src, P.sink, P.san, P.non)이 업데이트 된다. 또한, 동작 S197에서, 해당 노드에 직접 또는 간접적으로 연결되어 있는 인접 노드들의 테인트 분류 각각에 해당하는 확률 값도 업데이트 된다.
- [0121] 동작 S198에서, 해당 노드의 테인트 분류 각각에 해당하는 확률 값(P.src, P.sink, P.san, P.non)의 최대값이 기준치 미만인 노드가 존재하는 경우, 동작 S194 내지 동작 S197이 반복 수행되고, 그렇지 않은 경우 최종적으로 업데이트된 각 노드의 확률 값을 이용하여 오염 명세가 생성된다. 만약 동작 S199에서, 추론 안된 네트워크

모델이 존재한다면, 동작 S191 내지 동작 S198이 반복 수행될 수 있다.

- [0122] 도 21은 본 발명의 몇몇 실시예에 따른 방법들을 구현할 수 있는 예시적인 컴퓨팅 장치의 하드웨어 구성도이다. 도 21에 도시된 바와 같이, 컴퓨팅 장치(100)는 하나 이상의 프로세서(101), 버스(107), 네트워크 인터페이스(102), 프로세서(101)에 의하여 수행되는 컴퓨터 프로그램(105)을 로드(load)하는 메모리(103)와, 컴퓨터 프로그램(105)을 저장하는 스토리지(104)를 포함할 수 있다. 다만, 도 21에는 본 발명의 실시예와 관련 있는 구성요소들만이 도시되어 있다. 따라서, 본 발명이 속한 기술분야의 통상의 기술자라면 도 21에 도시된 구성요소들 외에 다른 범용적인 구성 요소들이 더 포함될 수 있음을 알 수 있다.
- [0123] 프로세서(101)는 컴퓨팅 장치(100)의 각 구성의 전반적인 동작을 제어한다. 프로세서(101)는 CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 또는 본 발명의 기술 분야에 잘 알려진 임의의 형태의 프로세서 중 적어도 하나를 포함하여 구성될 수 있다. 또한, 프로세서(101)는 본 발명의 다양한 실시예들에 따른 방법/동작을 실행하기 위한 적어도 하나의 애플리케이션 또는 프로그램에 대한 연산을 수행할 수 있다. 컴퓨팅 장치(100)는 하나 이상의 프로세서를 구비할 수 있다.
- [0124] 메모리(103)는 각종 데이터, 명령 및/또는 정보를 저장한다. 메모리(103)는 본 발명의 다양한 실시예들에 따른 방법/동작들을 실행하기 위하여 스토리지(104)로부터 하나 이상의 프로그램(105)을 로드(load) 할 수 있다. 예를 들어, 컴퓨터 프로그램(105)이 메모리(103)에 로드 되면, 로직(또는 모듈)이 메모리(103) 상에 구현될 수 있다. 메모리(103)의 예시는 RAM이 될 수 있으나, 이에 한정되는 것은 아니다.
- [0125] 버스(107)는 컴퓨팅 장치(100)의 구성 요소 간 통신 기능을 제공한다. 버스(107)는 주소 버스(Address Bus), 데이터 버스(Data Bus) 및 제어 버스(Control Bus) 등 다양한 형태의 버스로 구현될 수 있다.
- [0126] 네트워크 인터페이스(102)는 컴퓨팅 장치(100)의 유무선 인터넷 통신을 지원한다. 네트워크 인터페이스(102)는 인터넷 통신 외의 다양한 통신 방식을 지원할 수도 있다. 이를 위해, 네트워크 인터페이스(102)는 본 발명의 기술 분야에 잘 알려진 통신 모듈을 포함하여 구성될 수 있다.
- [0127] 스토리지(104)는 하나 이상의 컴퓨터 프로그램(105)을 비임시적으로 저장할 수 있다. 스토리지(104)는 플래시 메모리 등과 같은 비휘발성 메모리, 하드 디스크, 착탈형 디스크, 또는 본 발명이 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터로 읽을 수 있는 기록 매체를 포함하여 구성될 수 있다.
- [0128] 컴퓨터 프로그램(105)은 본 발명의 다양한 실시예들에 따른 방법/동작들이 구현된 하나 이상의 인스트럭션들(instructions)을 포함할 수 있다. 컴퓨터 프로그램(105)이 메모리(103)에 로드 되면, 프로세서(101)는 상기 하나 이상의 인스트럭션들을 실행시킴으로써 본 발명의 다양한 실시예들에 따른 방법/동작들을 수행할 수 있다.
- [0129] 일 실시예로서, 컴퓨터 프로그램(105)은, 오염 명세 분석 대상 코드에 포함된 복수의 함수들에 대응하는 복수의 노드들 및 상기 복수의 함수들 사이의 관계에 기초하여 상기 복수의 노드들 사이를 연결하는 복수의 에지들(edge)을 포함하는 네트워크 모델을 생성하는 동작, 상기 복수의 함수들 사이의 상기 관계에 기초하여, 상기 복수의 노드들 각각에 대해 오염 분석을 위한 조건부 확률 테이블(CPT)을 생성하는 동작, 및 상기 조건부 확률 테이블에 기초하여 상기 분석 대상 코드에 대한 오염 명세를 생성하는 동작을 수행하기 위한 인스트럭션을 포함할 수 있다.
- [0130] 지금까지 도 1 내지 도 21을 참조하여 본 발명의 다양한 실시예들 및 그 실시예들에 따른 효과들을 언급하였다. 본 발명의 기술적 사상에 따른 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.
- [0131] 지금까지 설명된 본 발명의 기술적 사상은 컴퓨터가 읽을 수 있는 매체 상에 컴퓨터가 읽을 수 있는 코드로 구현될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록 매체는, 예를 들어 이동형 기록 매체(CD, DVD, 블루레이 디스크, USB 저장 장치, 이동식 하드 디스크)이거나, 고정식 기록 매체(ROM, RAM, 컴퓨터 구비형 하드 디스크)일 수 있다. 상기 컴퓨터로 읽을 수 있는 기록 매체에 기록된 상기 컴퓨터 프로그램은 인터넷 등의 네트워크를 통하여 다른 컴퓨팅 장치에 전송되어 상기 다른 컴퓨팅 장치에 설치될 수 있고, 이로써 상기 다른 컴퓨팅 장치에서 사용될 수 있다.
- [0132] 이상에서, 본 발명의 실시예를 구성하는 모든 구성 요소들이 하나로 결합되거나 결합되어 동작하는 것으로 설명되었다고 해서, 본 발명의 기술적 사상이 반드시 이러한 실시예에 한정되는 것은 아니다. 즉, 본 발명의 목적 범위 안에서라면, 그 모든 구성요소들이 하나 이상으로 선택적으로 결합하여 동작할 수도 있다.
- [0133] 도면에서 동작들이 특정한 순서로 도시되어 있지만, 반드시 동작들이 도시된 특정한 순서로 또는 순차적 순서로

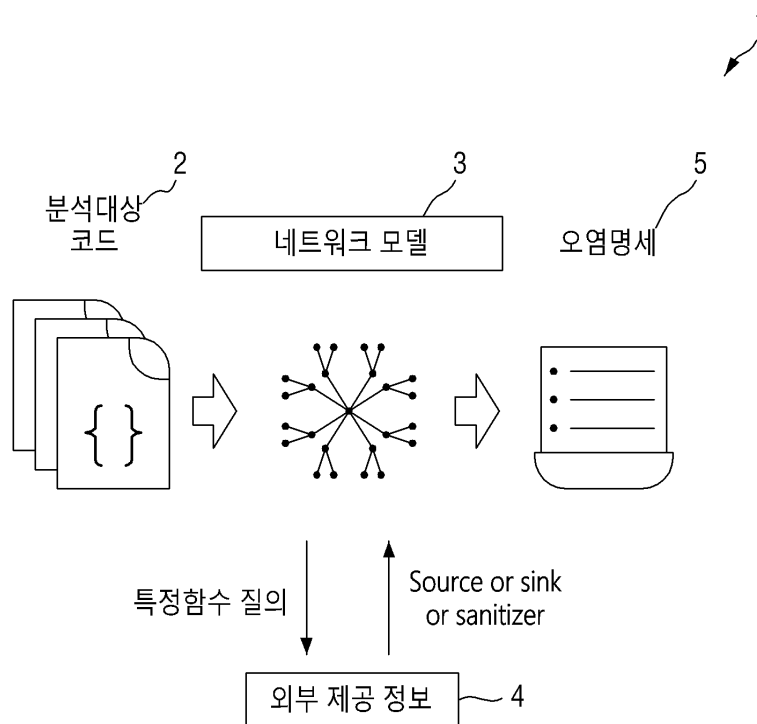
실행되어야만 하거나 또는 모든 도시 된 동작들이 실행되어야만 원하는 결과를 얻을 수 있는 것으로 이해되어서는 안 된다. 특정 상황에서는, 멀티태스킹 및 병렬 처리가 유리할 수도 있다. 더욱이, 위에 설명한 실시예들에서 다양한 구성들의 분리는 그러한 분리가 반드시 필요한 것으로 이해되어서는 안 되고, 설명된 프로그램 컴포넌트들 및 시스템들은 일반적으로 단일 소프트웨어 제품으로 함께 통합되거나 다수의 소프트웨어 제품으로 패키징 될 수 있음을 이해하여야 한다.

[0134]

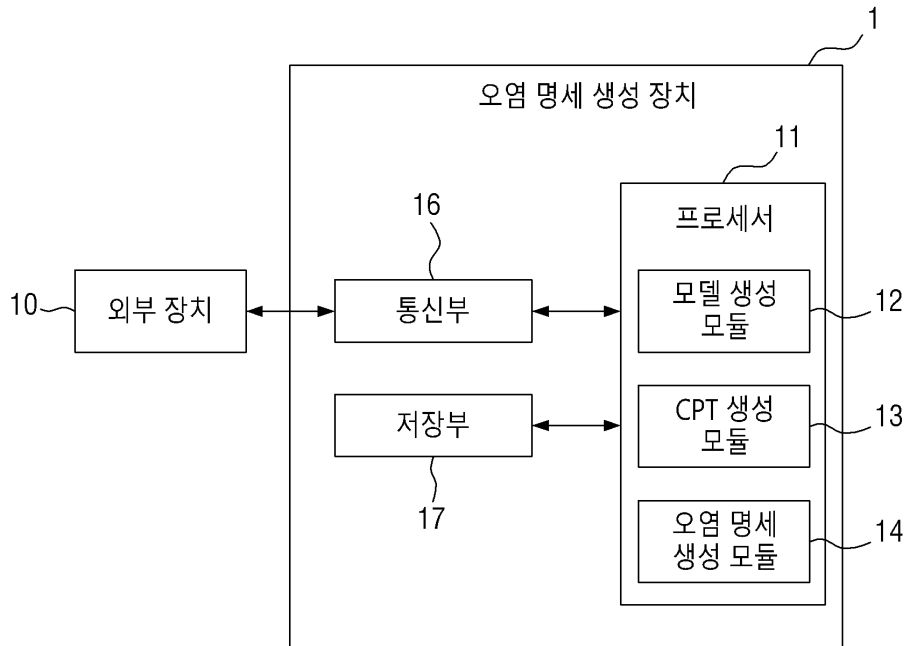
이상 첨부된 도면을 참조하여 본 발명의 실시예들을 설명하였지만, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 본 발명이 다른 구체적인 형태로도 실시될 수 있다는 것을 이해할 수 있다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로 이해해야만 한다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명에 의해 정의되는 기술적 사상의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

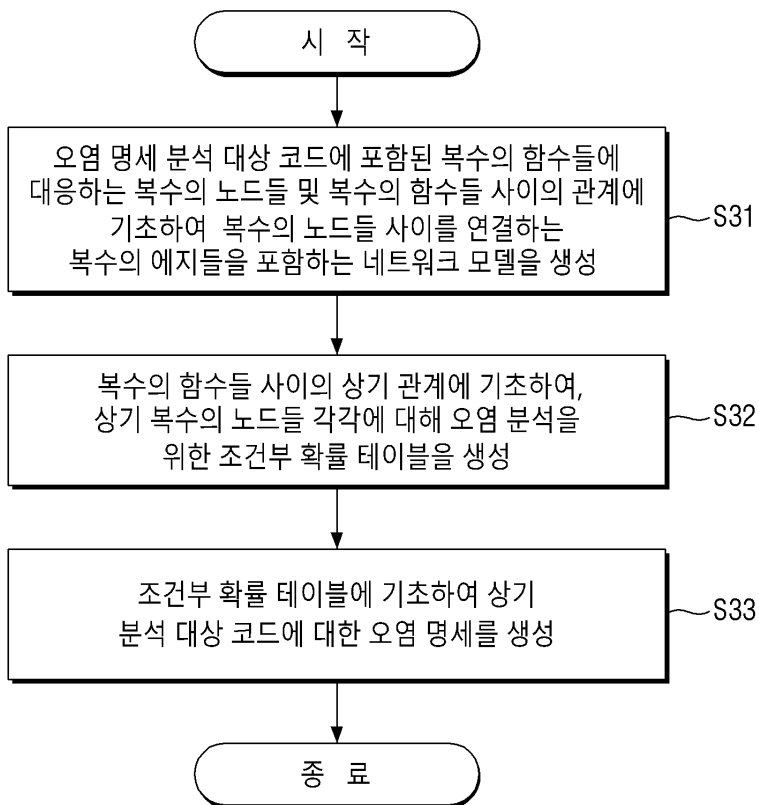
도면1



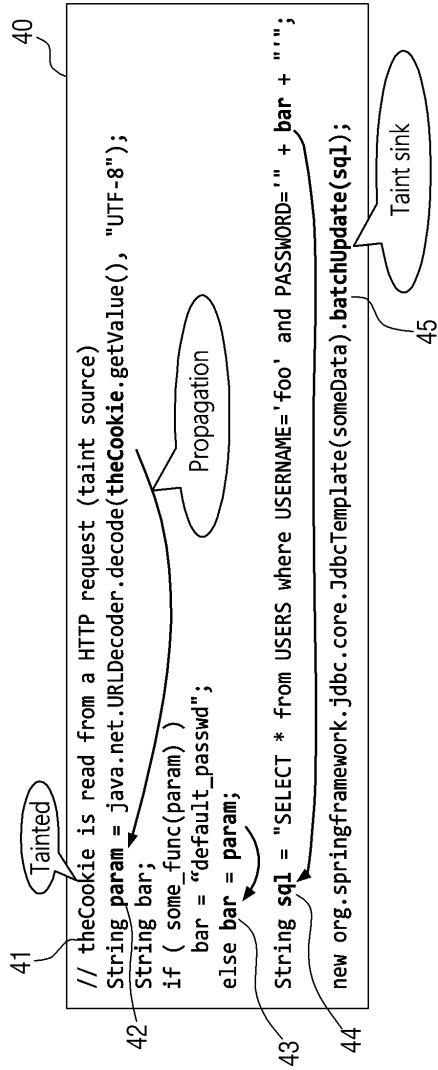
도면2



도면3



도면4



도면5

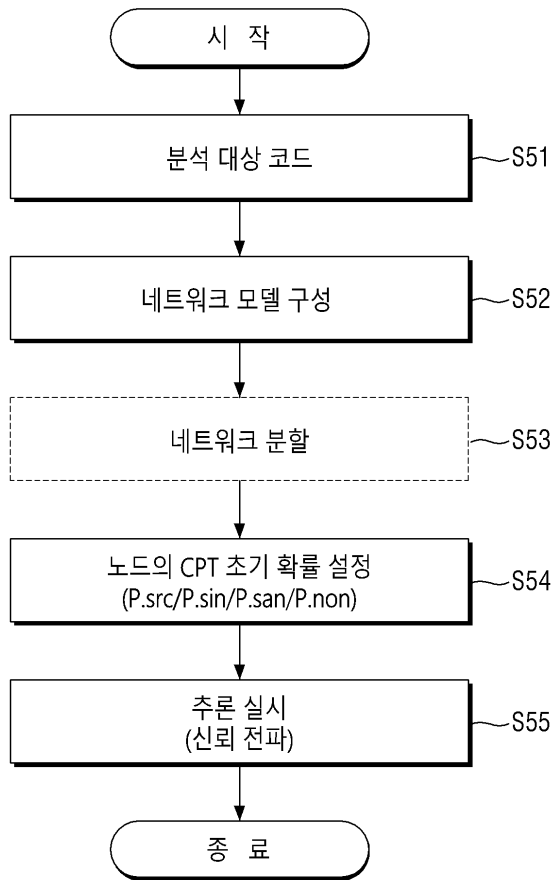
50

51

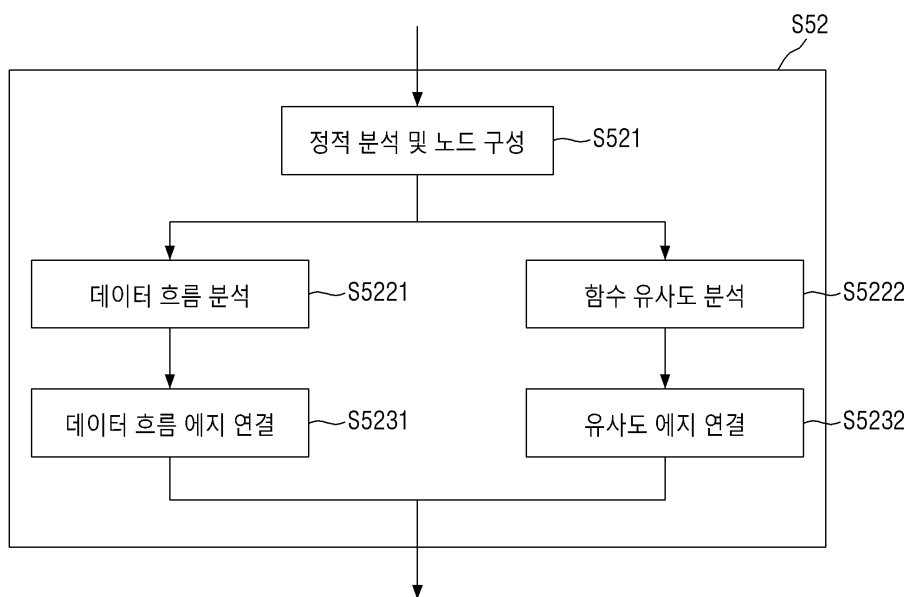
52

Function	Label	Taint	Check
Environment String java.io.BufferedReader.readLine()	Source	retval	-
Map java.lang.System.getenv()	Source	retval	-
String javax.servlet.ServletRequestWrapper.getParameter(String arg0)	Source	retval	-
java.io.File(URL arg0)	Sink	-	arg0
PreparedStatement java.sql.Connection.prepareStatement(String arg0, int[] arg1)	Sink	-	arg0
Process java.lang.Runtime.exec(String[] arg0, String[] arg1)	Sink	-	arg0, arg1
void System.Data.SqlClient.SqlCommand.Prepare()	Sanitizer	-	-
String System.IO.Path.GetExtension(String arg0)	Sanitizer	-	-
String org.apache.commons.lang.StringEscapeUtils.escapeHtml(String arg0)	Sanitizer	-	-

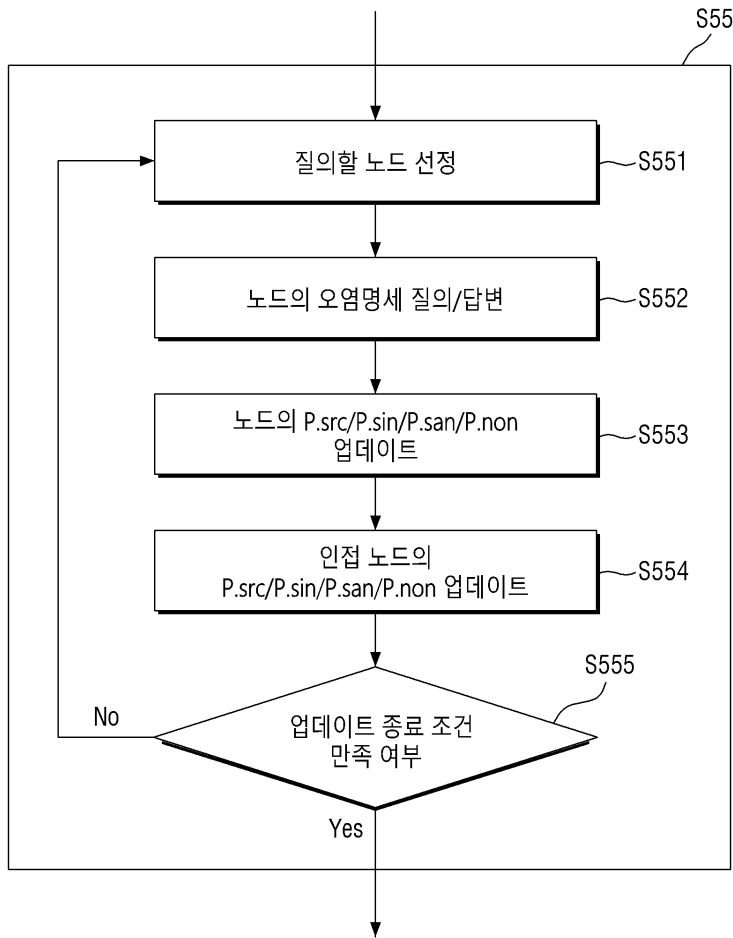
도면6



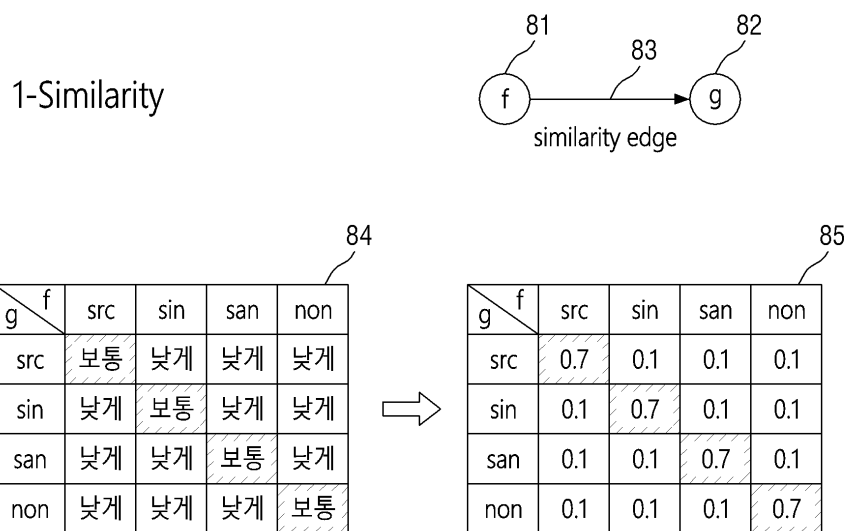
도면7



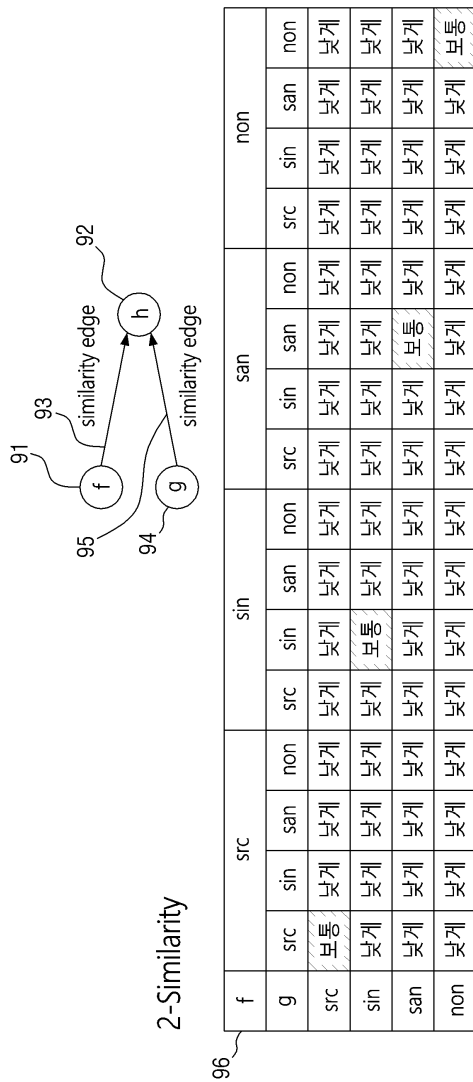
도면8



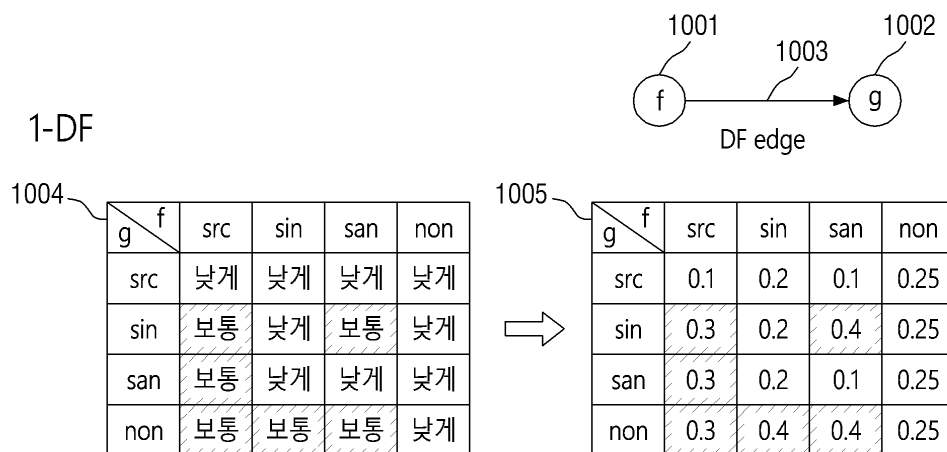
도면9



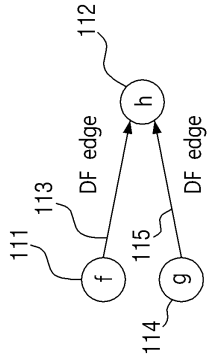
도면10



도면11



도면 12

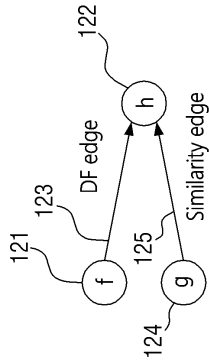


2-DF

[illegible]

f	src				sin				san				non			
	src	sin	san	non	src	sin	san	non	src	sin	san	non	src	sin	san	non
g																
src	0.1	0.2	0.1	0.1	0.25	0.25	0.25	0.25	0.1	0.25	0.1	0.1	0.1	0.25	0.1	0.1
sin	0.3	0.2	0.4	0.3	0.25	0.25	0.25	0.25	0.4	0.25	0.4	0.4	0.3	0.25	0.4	0.3
san	0.3	0.2	0.1	0.3	0.25	0.25	0.25	0.25	0.1	0.25	0.1	0.1	0.3	0.25	0.1	0.3
non	0.3	0.4	0.4	0.3	0.25	0.25	0.25	0.25	0.4	0.25	0.4	0.4	0.3	0.25	0.4	0.3

도면13



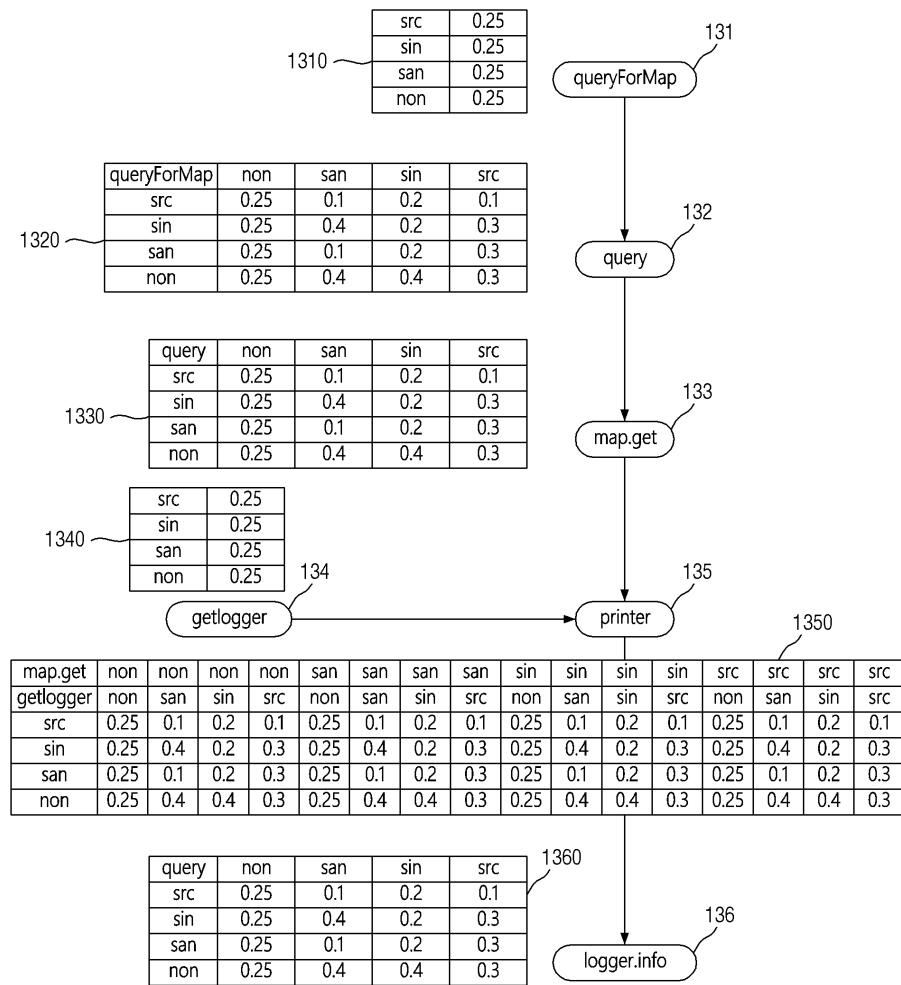
1-Similarity, 1-DF

f	src				sin				san				non			
	src	sin	san	non	src	sin	san	non	src	sin	san	non	src	sin	san	non
g	보통	낮게	낮게	낮게	보통	낮게	낮게	낮게	보통	낮게	낮게	낮게	보통	낮게	낮게	낮게
src	높음	보통	높음	높음	낮게	보통	낮게	낮게	높음	보통	높음	높음	보통	보통	높음	높음
sin	높음	보통	높음	높음	낮게	보통	낮게	낮게	높음	보통	높음	높음	보통	보통	높음	높음
san	높음	보통	높음	높음	낮게	보통	낮게	낮게	높음	보통	높음	높음	보통	보통	높음	높음
non	높음	보통	높음	높음	낮게	보통	낮게	낮게	높음	보통	높음	높음	보통	보통	높음	높음

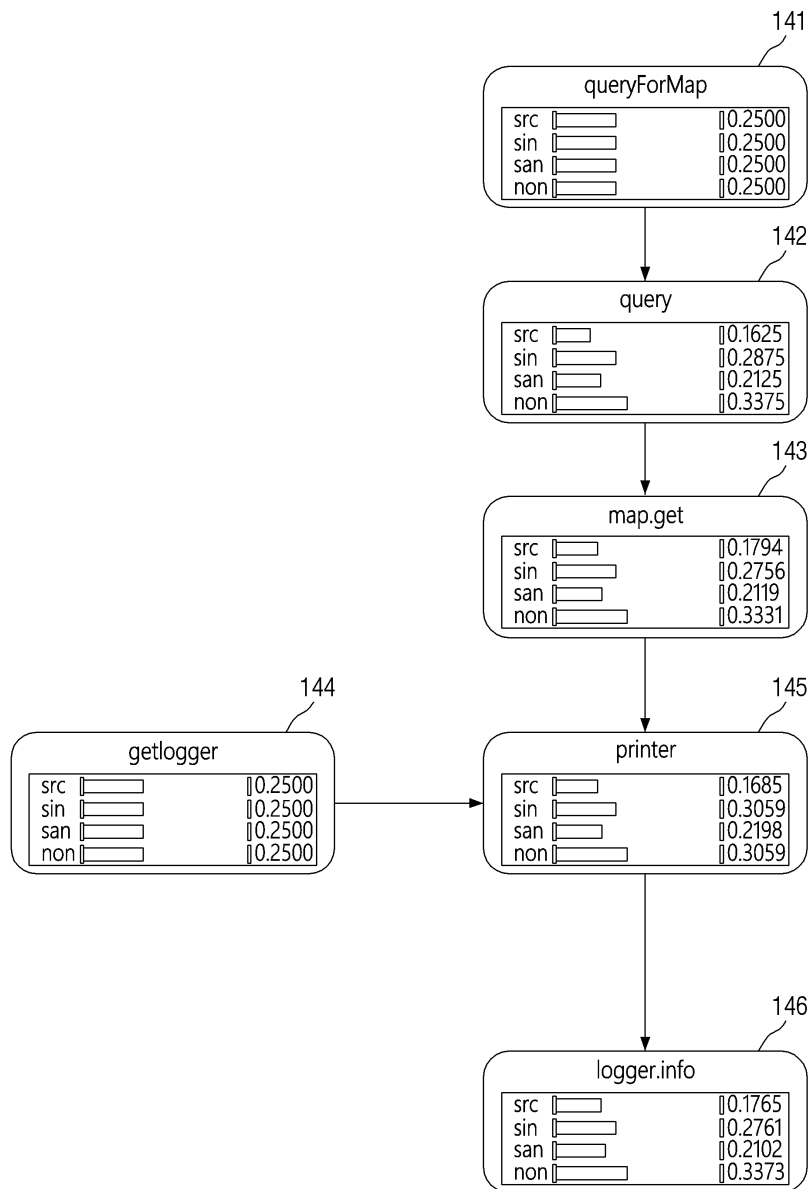


f	src				sin				san				non			
	src	sin	san	non	src	sin	san	non	src	sin	san	non	src	sin	san	non
g	0.1	0.2	0.1	0.1	0.4	0.2	0.2	0.2	0.2	0.2	0.1	0.05	0.1	0.2	0.1	0.1
src	0.3	0.4	0.35	0.2	0.2	0.4	0.2	0.2	0.35	0.4	0.35	0.3	0.3	0.4	0.35	0.2
sin	0.3	0.4	0.35	0.2	0.2	0.4	0.2	0.2	0.35	0.4	0.35	0.3	0.3	0.4	0.35	0.2
san	0.3	0.2	0.2	0.2	0.2	0.2	0.4	0.2	0.1	0.2	0.2	0.05	0.3	0.2	0.2	0.2
non	0.3	0.2	0.35	0.5	0.2	0.2	0.2	0.4	0.35	0.2	0.35	0.6	0.3	0.2	0.35	0.5

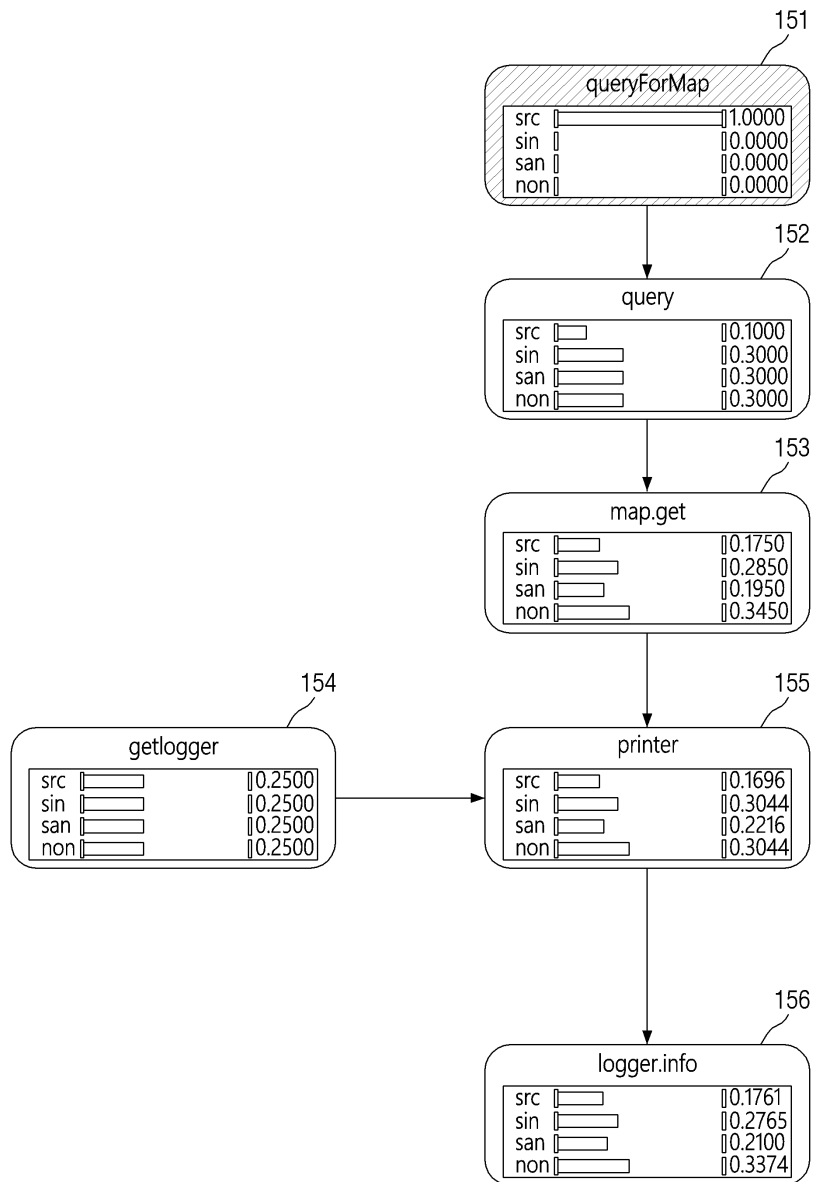
도면14



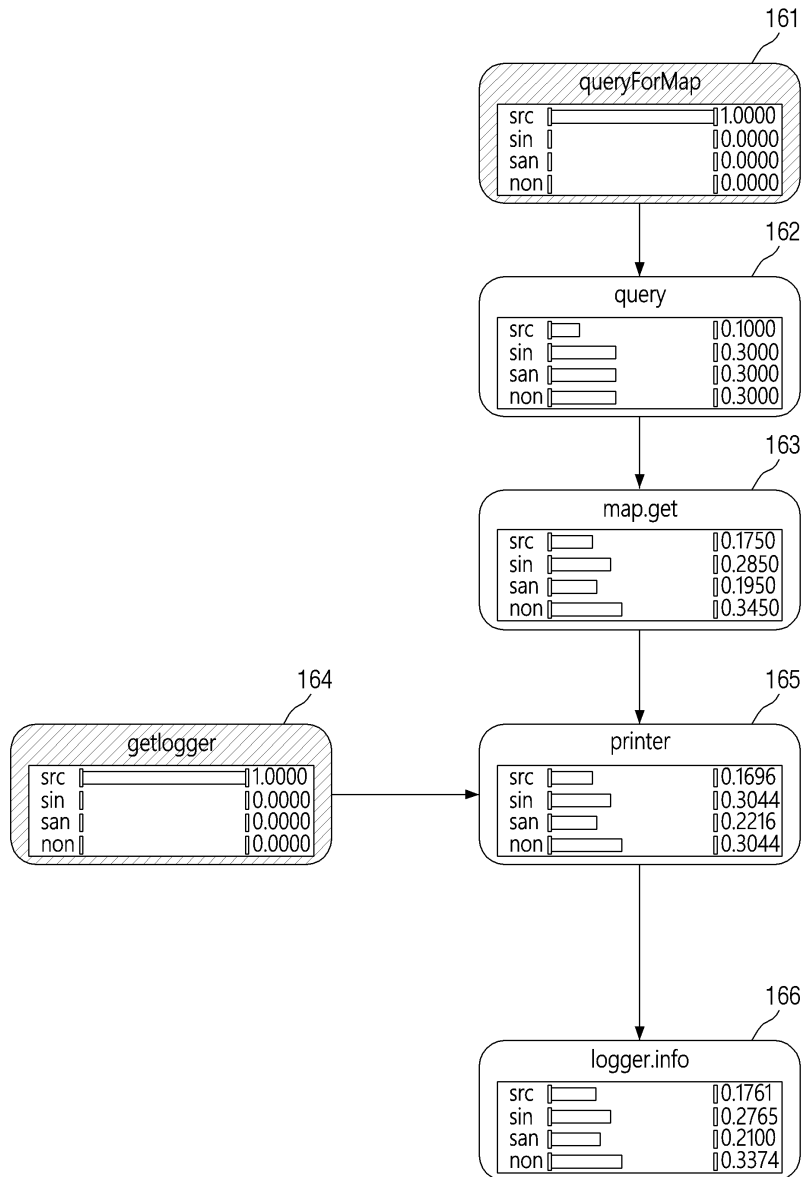
도면15



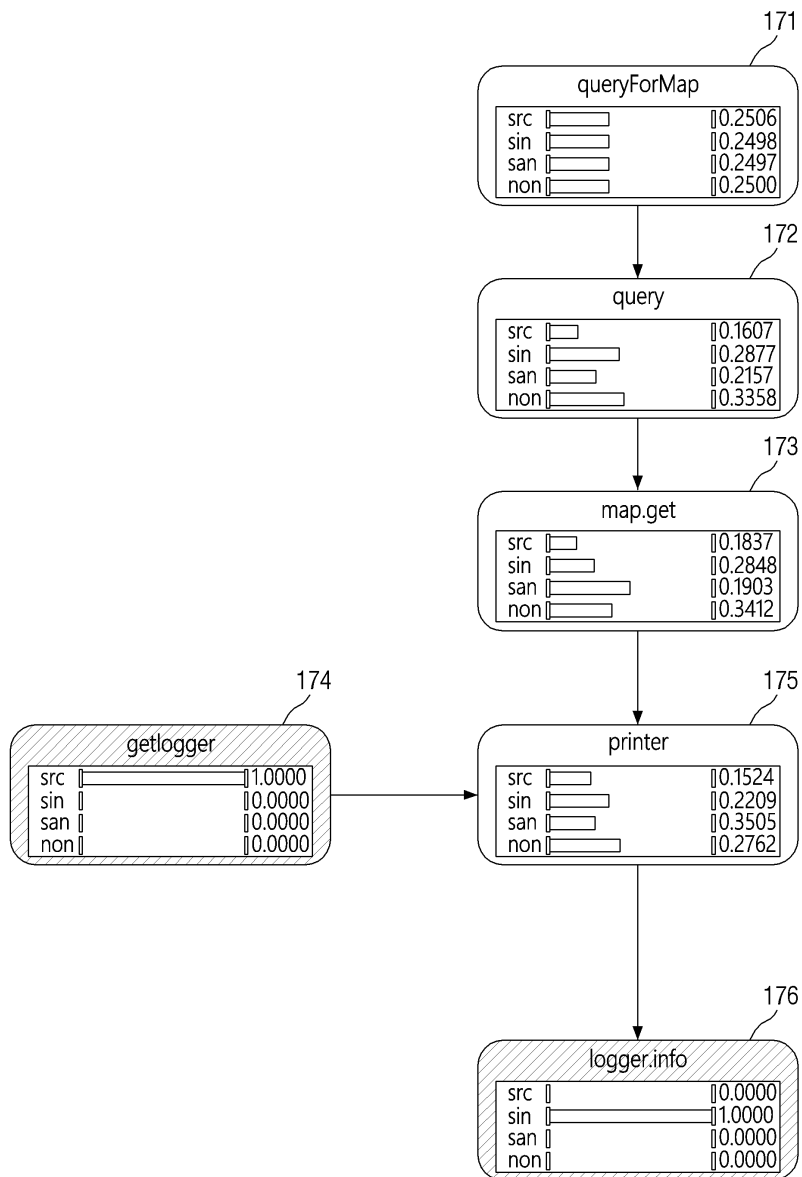
도면16



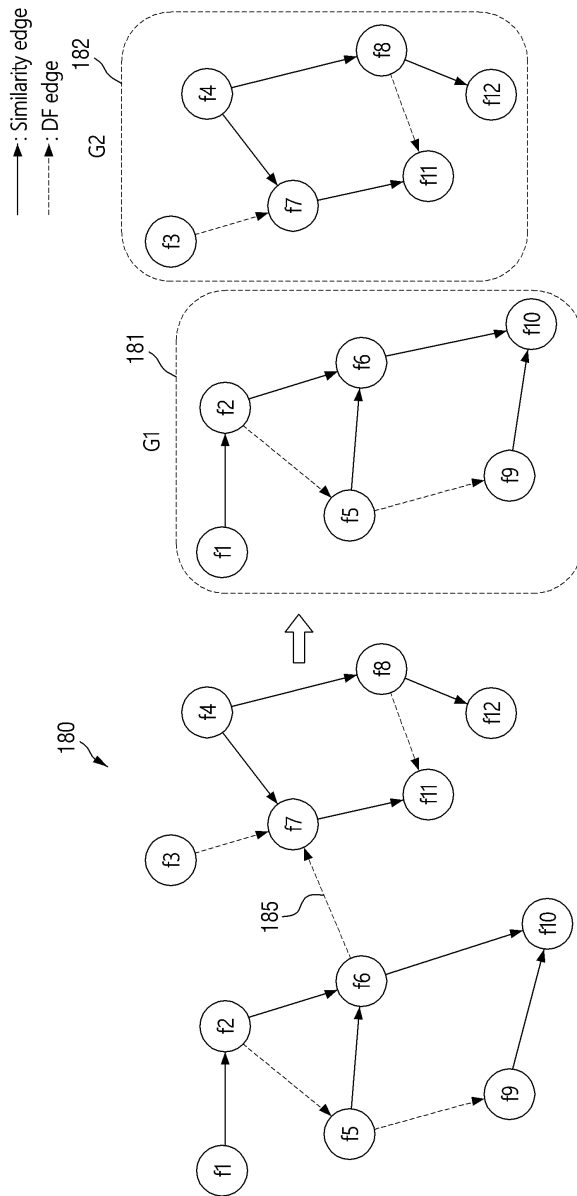
도면17



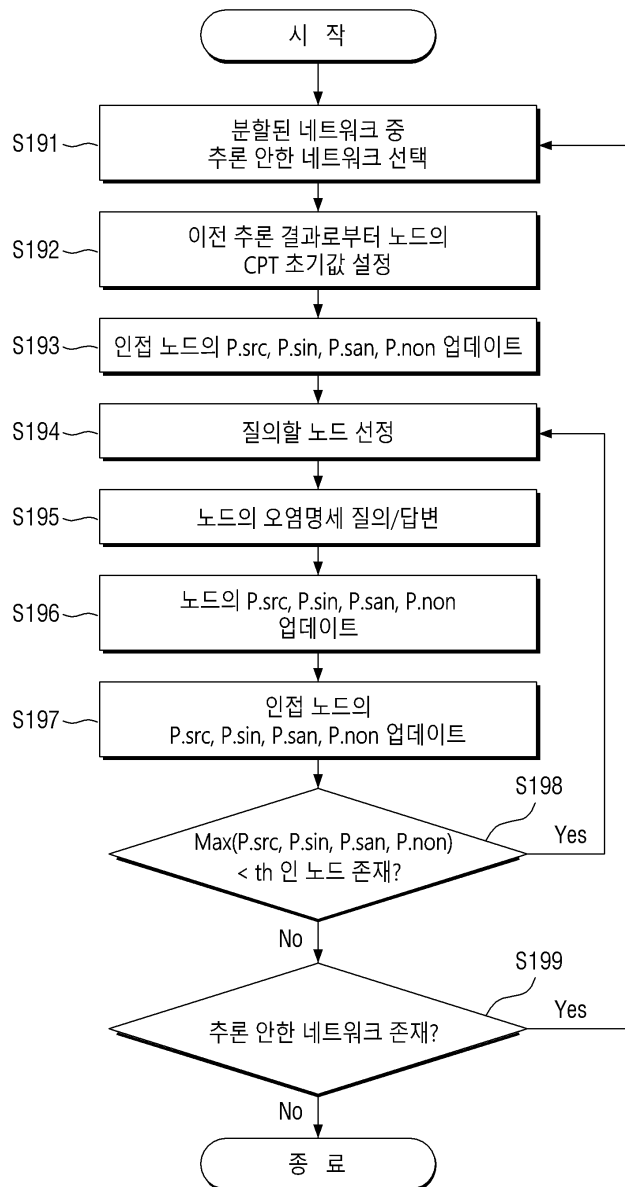
도면18



도면19



도면20



도면21

