(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0094112
(43) 공개일자 2023년06월27일

(51) 국제특허분류(Int. Cl.)
G06F 21/55 (2013.01) *G06F 16/51* (2019.01)
G06T 7/11 (2017.01) *G06V 10/20* (2022.01)
G06V 10/40 (2022.01)
(52) CPC특허분류
G06F 21/554 (2013.01)
G06F 16/51 (2019.01)
(21) 출원번호 10-2022-0065156
(22) 출원일자 2022년05월27일
심사청구일자 2022년05월27일
(30) 우선권주장
1020210182605 2021년12월20일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
허준범
경기도 용인시 기흥구 보정로 30 (보정동, 동아솔레시티아파트) 119동 201호
김예림
경기도 김포시 사우중로3번길 34 (사우동) 대양캐슬 가동 303호
김민재
서울특별시 동대문구 왕산로 86 (용두동 동대문푸르지오시티) 102동 1312호
(74) 대리인
특허법인위솔

전체 청구항 수 : 총 14 항

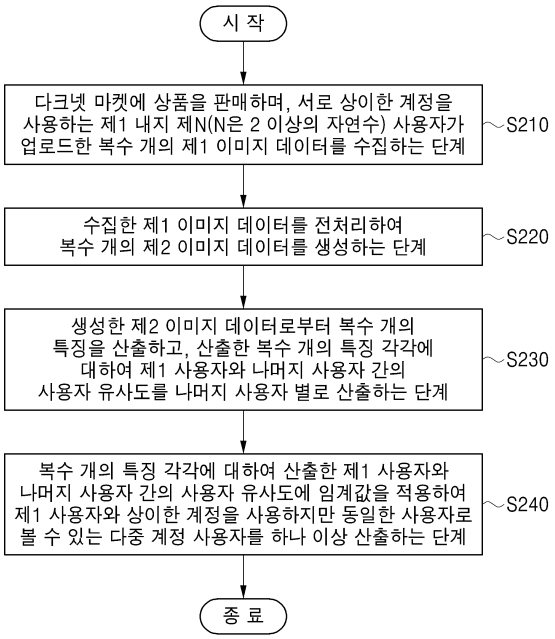
(54) 발명의 명칭 **다크넷 마켓 사용자의 다중 계정을 탐지하는 방법 및 이를 위한 장치**

(57) 요 약

본 발명의 일 실시 예에 따른 프로세서 및 메모리를 포함하는 장치를 통해 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법은 (a) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 단계, (b) 상기 수집한 제1 이미지 데이터를

(뒷면에 계속)

대 표 도 - 도2



전처리하여 복수 개의 제2 이미지 데이터를 생성하는 단계, (c) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 단계 및 (d) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 단계를 포함하며, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은, 상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상이다.

(52) CPC특허분류
G06T 7/11 (2017.01)
G06V 10/20 (2023.01)
G06V 10/40 (2023.01)
G06T 2210/12 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711126325
과제번호 2020-0-01819-002
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성(R&D)
연구과제명 ICT명품인재양성(고려대학교)
기 여 율 1/3
과제수행기관명 고려대학교산학협력단
연구기간 2021.01.01 ~ 2021.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1345342146
과제번호 2021R1A6A1A13044830
부처명 교육부
과제관리(전문)기관명 한국연구재단
연구사업명 이공학학술연구기반구축(R&D)
연구과제명 융합소프트웨어연구소
기 여 율 1/3
과제수행기관명 고려대학교
연구기간 2021.06.01 ~ 2022.02.28
이 발명을 지원한 국가연구개발사업
과제고유번호 1711139206
과제번호 2021-0-01810-001
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성(R&D)
연구과제명 양자세계를 연결하는 초신뢰 양자인터넷 요소 기술 개발
기 여 율 1/3
과제수행기관명 고려대학교산학협력단
연구기간 2021.07.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

프로세서 및 메모리를 포함하는 장치를 통해 다크넷 마켓(Darknet Market) 사용자의 다중 계정을 탐지하는 방법에 있어서,

(a) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 단계;

(b) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 단계;

(c) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 단계; 및

(d) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 사용자를 하나 이상 산출하는 단계;

를 포함하며,

상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은,

상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상인,

다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 2

제1항에 있어서,

상기 (a) 단계는,

(a-1) 상기 제1 내지 제N 사용자가 업로드한 복수 개의 제1 이미지 데이터를 상기 다크넷 마켓에서 직접 수집하는 단계; 및

(a-2) 상기 제1 내지 제N 사용자가 상기 다크넷 마켓에 업로드한 복수 개의 제1 이미지 데이터를 수집한 외부 서버로부터 수집하는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 3

제1항에 있어서,

상기 (b) 단계는,

(b-1) 상기 수집한 제1 이미지 데이터의 제1 전처리로 파일 형식을 JPG로 변환하여 상기 복수 개의 제2 이미지 데이터를 생성하는 단계;

(b-2) 상기 수집한 제1 이미지 데이터의 제2 전처리로 상기 제1 이미지 데이터 별로 사용자명(Username)에 대한 정보, 대분류 카테고리에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 및 사용자별 PGP Key에 대한 정보 중 어느 하나 이상을 추출하여 상기 복수 개의 제2 이미지 데이터를 생성하는 단계; 및

(b-3) 상기 수집한 제1 이미지 데이터의 제3 전처리로 상기 제1 이미지 데이터가 복수 개의 이미지 데이터가 연결된 이미지 데이터인 경우, 각각의 이미지 데이터가 포함하고 있는 상품 별로 분할(Cropping)하고, 해당 상품

명으로 분할한 이미지 데이터를 라벨링(Labeling)하여 상기 복수 개의 제2 이미지 데이터를 생성하는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 4

제1항에 있어서,

상기 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 상기 제2 이미지 데이터의 해시값에 대한 정보인 경우, 상기 (c) 단계는,

(c-1) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 복수 개의 해시 타입(Type)별 해시값을 산출하는 단계;

(c-2) 상기 산출한 복수 개의 해시 타입별 해시값을 이용하여 상기 복수 개의 해시 타입의 성능을 평가하는 단계; 및

(c-3) 상기 성능 평가 결과 가장 우수한 성능의 해시 타입의 해시값을 이용하여 상기 제1 사용자에 대한 제2 이미지 데이터의 해시값과 나머지 사용자에 대한 제2 이미지 데이터의 해시값에 따른 사용자 유사도인 해밍 디스턴스(Hamming distance)를 상기 나머지 사용자 별로 산출하는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 5

제4항에 있어서,

상기 (c-1) 단계의 해시값 산출은,

파이썬(Python)의 해시 함수(Hash function)을 이용하며,

상기 복수 개의 해시 타입은,

AHASH, PHASH, DHASH 및 WHASH중 어느 둘 이상인,

다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 6

제4항에 있어서,

상기 (c-2) 단계는,

(c-2-1) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 상기 복수 개의 해시 타입 별로 그룹핑하는 단계;

(c-2-2) 상기 복수 개의 해시 타입 별로 그룹핑한 각각의 그룹에 속하는 모든 제2 이미지 데이터에 대하여 페어(Pair)를 전부 생성하는 단계;

(c-2-3) 상기 전부 생성한 각각의 페어의 SSIM값을 산출하여 그룹별 평균 SSIM값을 산출하는 단계; 및

(c-2-4) 상기 산출한 그룹별 평균 SSIM 값 중, 가장 높은 평균 SSIM 값을 나타내는 그룹에 대한 해시 타입을 가장 우수한 성능의 해시 타입으로 평가하는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 7

제6항에 있어서,

상기 (c-2-3) 단계와 (c-2-4) 단계 사이에,

(c-2-3.5) 상기 산출한 그룹별 평균 SSIM 값에 대하여 각각의 그룹에 속하는 제2 이미지 데이터의 개수에 따라 가중치(Weight)를 부여하여 가중치가 부여된 그룹별 평균 SSIM 값(Weighted average SSIM)을 산출하는 단계;

를 더 포함하며,

이 경우 상기 (c-2-4) 단계의 평균 SSIM 값은 가중치가 부여된 평균 SSIM 값인,

다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 8

제1항에 있어서,

상기 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보 또는 상기 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 중 어느 하나인 경우, 상기 (c) 단계는,

(c-4) 상기 제1 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 제1 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 상기 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계;

(c-5) 상기 나머지 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 나머지 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 상기 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계; 및

(c-6) 상기 대분류 카테고리 또는 소분류 카테고리 별로 산출한 제1 사용자에 대한 비중과 나머지 사용자에 대한 비중을 이용하여 상기 제1 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중과 나머지 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중에 따른 사용자 유사도인 코사인 유사도(Cosine similarity)를 상기 나머지 사용자 별로 산출하는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 9

제8항에 있어서,

상기 (c-4) 단계 이전에,

(c-3.5) 상기 복수 개의 제2 이미지 데이터 중, 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터를 추출하는 단계; 및

(c-3.7) 상기 추출한 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터에 대하여 결여된 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보를 부여하는 단계;

중 어느 하나 이상을 더 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 10

제1항에 있어서,

상기 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보인 경우, 상기 (c) 단계는,

(c-7) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 포함하고 있는 텍스트 정보를 추출하는 단계; 및

(c-8) 상기 추출한 텍스트 정보를 이용하여 상기 제1 사용자에 대한 제2 이미지 데이터의 텍스트에 대한 정보와 나머지 사용자에 대한 제2 이미지 데이터의 텍스트에 대한 정보에 따른 사용자 유사도인 자카드 유사도(Jaccard similarity)를 상기 나머지 사용자 별로 산출하는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 11

제10항에 있어서,

상기 (c-7) 단계는,

(c-7-1) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 텍스트가 포함된 영역에 대한 바운딩 박스(Bounding

box)를 설정하는 단계; 및

(c-7-2) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 해상도를 향상시키는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 12

제1항에 있어서,

상기 (d) 단계는,

(d-1) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 사용자 유사도를 합산하여 나머지 사용자 별 유사도 스코어를 산출하는 단계;

(d-2) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 유사도 스코에 제1 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 가장 높은 사용자를 상기 다중 계정 사용자로 산출하는 단계; 및

(d-3) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 유사도 스코어에 제2임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 차순위로 높은 사용자를 다중 계정 사용자 후보로 산출하는 단계;

중 어느 하나 이상을 포함하는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법.

청구항 13

하나 이상의 프로세서;

네트워크 인터페이스;

상기 프로세서에 의해 수행되는 컴퓨터 프로그램을 로드(Load)하는 메모리; 및

대용량 네트워크 데이터 및 상기 컴퓨터 프로그램을 저장하는 스토리지를 포함하되,

상기 컴퓨터 프로그램은 상기 하나 이상의 프로세서에 의해,

(A) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 오퍼레이션;

(B) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 오퍼레이션;

(C) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 오퍼레이션; 및

(D) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 오퍼레이션;

을 포함하며,

상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은,

상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상인,

다크넷 마켓 사용자의 다중 계정을 탐지하는 장치.

청구항 14

컴퓨팅 장치와 결합하여,

(AA) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자

가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 단계;

(BB) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 단계;

(CC) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 단계; 및

(DD) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 단계;

를 포함하며,

상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은,

상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상인,

컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 발명은 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법 및 이를 위한 장치에 관한 것이다. 보다 자세하게는 다크넷 마켓에서 이미지 데이터를 수집하고 이로부터 다양한 특징을 추출/분석하여 동일 사용자에 대한 다중 계정을 탐지할 수 있는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 블록체인 기술에 기반한 암호화폐 거래가 늘어나면서 익명성과 추적이 어렵다는 특수성을 이용하여 불법적인 거래에 적극적으로 활용되고 있는바, 마약이나 불법 무기, 음란물, 위조 지폐 등의 거래가 이루어지는 다크넷 마켓이 그 대표적인 사례라고 볼 수 있다.

[0003] 이와 같은 다크넷 마켓에서 이루어지는 거래는 범죄 행위에 해당하며, 범죄 행위 색출을 위해 다크넷 마켓의 트래킹 및 분석이 필수적인바, 그에 따라 다크넷 마켓 사용자들에 대한 정보 수집이 요구된다.

[0004] 이러한 다크넷 마켓 사용자는 다크넷 마켓 상에서 상품을 판매하는 판매자 역할의 벤더(Vendor)와 이들로부터 상품을 구매하는 구매자 모두를 포함하는 광의의 개념이나, 불법적인 상품의 유통 과정과 경위를 추적할 수 있다는 측면에서 전자인 벤더에 대한 정보 수집이 보다 중요하다고 여겨지고 있다.

[0005] 그러나 실제 많은 수의 벤더들이 동일한 다크넷 마켓 내에서 다중 계정을 사용하거나 서로 상이한 마켓에서 다중 계정을 사용하고 있으므로 벤더들의 실질적인 분포와 그 숫자를 심도 있게 파악하기 어려운바, 이를 위해 종래에는 벤더들이 다크넷 마켓에 업로드한 이미지 데이터가 포함하고 있는 텍스트에 대한 정보, 예를 들어 상품이나 벤더에 대한 소개 등에 대한 정보를 산출하여 동일한 벤더의 다중 계정을 탐지하였으나, 이미지 데이터가 포함하는 텍스트에 대한 정보이기 때문에 내용이 지나치게 짧고 유사한 템플릿을 사용하기 때문에 텍스트 스타일을 파악하여 다중 계정을 탐지하기에는 어렵다는 문제점이 존재한다.

[0006] 아울러, 텍스트 기반의 종래의 탐지 방법은 영문 텍스트에 기반한 탐지 방법이기 때문에 언어에 따른 한계가 매우 컸으며, 다크넷 마켓의 성장으로 인해 영어 언어권의 국가뿐만 아니라 훨씬 다양한 언어권의 국가들에서도 다크넷 마켓이 신설되고 있기에 최신식 다크넷 마켓 생태계 분석에 적합하지 않다는 문제점 그리고 한 명의 벤더와 가장 유사한 벤더 한 명만 탐지할 수 있었기 때문에 해당 벤더가 3개 이상의 다중 계정을 사용하고 있는 경우에는 탐지할 수 없다는 문제점까지 존재한다.

[0007] 따라서 이미지 데이터로부터 산출한 텍스트에 대한 정보와 더불어 보다 다양한 특징(Feature)들에 대한 분석을 통해 언어에 따른 한계를 타파하고, 한 명의 벤더가 3개 이상의 다중 계정을 사용하고 있는 경우라 할지라도 동일한 벤더의 다중 계정 전부를 정확하게 탐지할 수 있는 새롭고 진보된 기술이 요구된다. 본 발명은 이에 관한

것이다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 대한민국 공개특허공보 제 10-2019-0054478호(2019.05.22)

발명의 내용

해결하려는 과제

[0009] 본 발명이 해결하고자 하는 기술적 과제는 이미지 데이터로부터 산출한 텍스트에 대한 정보와 더불어 보다 다양한 특징들에 대한 분석을 통해 종래 언어에 따른 한계를 타파할 수 있는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법 및 이를 위한 장치를 제공하는 것이다.

[0010] 본 발명이 해결하고자 하는 또 다른 기술적 과제는 사용자가 몇 개의 계정을 사용하는 지와 무관하게 동일한 사용자의 다중 계정 전부를 정확하게 탐지할 수 있는 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법 및 이를 위한 장치를 제공하는 것이다.

[0011] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0012] 상기 기술적 과제를 달성하기 위한 본 발명의 일 실시 예에 따른 프로세서 및 메모리를 포함하는 장치를 통해 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법은 (a) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 단계, (b) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 단계, (c) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 단계 및 (d) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 단계를 포함하며, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은, 상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상이다.

[0013] 일 실시 예에 따르면, 상기 (a) 단계는, (a-1) 상기 제1 내지 제N 사용자가 업로드한 복수 개의 제1 이미지 데이터를 상기 다크넷 마켓에서 직접 수집하는 단계 및 (a-2) 상기 제1 내지 제N 사용자가 상기 다크넷 마켓에 업로드한 복수 개의 제1 이미지 데이터를 수집한 외부 서버로부터 수집하는 단계 중 어느 하나 이상을 포함할 수 있다.

[0014] 일 실시 예에 따르면, 상기 (b) 단계는, (b-1) 상기 수집한 제1 이미지 데이터의 제1 전처리로 파일 형식을 JPG로 변환하여 상기 복수 개의 제2 이미지 데이터를 생성하는 단계, (b-2) 상기 수집한 제1 이미지 데이터의 제2 전처리로 상기 제1 이미지 데이터 별로 사용자명(Username)에 대한 정보, 대분류 카테고리에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 및 사용자별 PGP Key에 대한 정보 중 어느 하나 이상을 추출하여 상기 복수 개의 제2 이미지 데이터를 생성하는 단계 및 (b-3) 상기 수집한 제1 이미지 데이터의 제3 전처리로 상기 제1 이미지 데이터가 복수 개의 이미지 데이터가 연결된 이미지 데이터인 경우, 각각의 이미지 데이터가 포함하고 있는 상품 별로 분할(Cropping)하고, 해당 상품명으로 분할한 이미지 데이터를 라벨링(Labeling)하여 상기 복수 개의 제2 이미지 데이터를 생성하는 단계 중 어느 하나 이상을 포함할 수 있다.

[0015] 일 실시 예에 따르면, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 상기 제2 이미지 데이터의 해시값에 대한 정보인 경우, 상기 (c) 단계는, (c-1) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 복수 개의 해시 타입(Type)별 해시값을 산출하는 단계, (c-2) 상기 산출한 복수 개의 해시 타입별 해시값을 이용하여

상기 복수 개의 해시 타입의 성능을 평가하는 단계 및 (c-3) 상기 성능 평가 결과 가장 우수한 성능의 해시 타입의 해시값을 이용하여 상기 제1 사용자에 대한 제2 이미지 데이터의 해시값과 나머지 사용자에 대한 제2 이미지 데이터의 해시값에 따른 사용자 유사도인 해밍 디스턴스(Hamming distance)를 상기 나머지 사용자 별로 산출하는 단계 중 어느 하나 이상을 포함할 수 있다.

[0016] 일 실시 예에 따르면, 상기 (c-1) 단계의 해시값 산출은, 파이썬(Python)의 해시 함수(Hash function)을 이용하며, 상기 복수 개의 해시 타입은, AHASH, PHASH, DHASH 및 WHASH중 어느 둘 이상일 수 있다.

[0017] 일 실시 예에 따르면, 상기 (c-2) 단계는, (c-2-1) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 상기 복수 개의 해시 타입 별로 그룹핑하는 단계, (c-2-2) 상기 복수 개의 해시 타입 별로 그룹핑한 각각의 그룹에 속하는 모든 제2 이미지 데이터에 대하여 페어(Pair)를 전부 생성하는 단계, (c-2-3) 상기 전부 생성한 각각의 페어의 SSIM값을 산출하여 그룹별 평균 SSIM값을 산출하는 단계 및 (c-2-4) 상기 산출한 그룹별 평균 SSIM 값 중, 가장 높은 평균 SSIM 값을 나타내는 그룹에 대한 해시 타입을 가장 우수한 성능의 해시 타입으로 평가하는 단계 중 어느 하나 이상을 포함할 수 있다.

[0018] 일 실시 예에 따르면, 상기 (c-2-3) 단계와 (c-2-4) 단계 사이에, (c-2-3.5) 상기 산출한 그룹별 평균 SSIM 값에 대하여 각각의 그룹에 속하는 제2 이미지 데이터의 개수에 따라 가중치(Weight)를 부여하여 가중치가 부여된 그룹별 평균 SSIM 값(Weighted average SSIM)을 산출하는 단계를 더 포함하며, 이 경우 상기 (c-2-4) 단계의 평균 SSIM 값은 가중치가 부여된 평균 SSIM 값일 수 있다.

[0019] 일 실시 예에 따르면, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보 또는 상기 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 중 어느 하나인 경우, 상기 (c) 단계는, (c-4) 상기 제1 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 제1 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 상기 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계, (c-5) 상기 나머지 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 나머지 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 상기 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계 및 (c-6) 상기 대분류 카테고리 또는 소분류 카테고리 별로 산출한 제1 사용자에 대한 비중과 나머지 사용자에 대한 비중을 이용하여 상기 제1 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중과 나머지 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중에 따른 사용자 유사도인 코사인 유사도(Cosine similarity)를 상기 나머지 사용자 별로 산출하는 단계 중 어느 하나 이상을 포함할 수 있다.

[0020] 일 실시 예에 따르면, 상기 (c-4) 단계 이전에, (c-3.5) 상기 복수 개의 제2 이미지 데이터 중, 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터를 추출하는 단계 및 (c-3.7) 상기 추출한 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터에 대하여 결여된 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보를 부여하는 단계 중 어느 하나 이상을 더 포함할 수 있다.

[0021] 일 실시 예에 따르면, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보인 경우, 상기 (c) 단계는, (c-7) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 포함하고 있는 텍스트 정보를 추출하는 단계 및 (c-8) 상기 추출한 텍스트 정보를 이용하여 상기 제1 사용자에 대한 제2 이미지 데이터의 텍스트에 대한 정보와 나머지 사용자에 대한 제2 이미지 데이터의 텍스트에 대한 정보에 따른 사용자 유사도인 자카드 유사도(Jaccard similarity)를 상기 나머지 사용자 별로 산출하는 단계 중 어느 하나 이상을 포함할 수 있다.

[0022] 일 실시 예에 따르면, 상기 (c-7) 단계는, (c-7-1) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 텍스트가 포함된 영역에 대한 바운딩 박스(Bounding box)를 설정하는 단계 및 (c-7-2) 상기 복수 개의 제2 이미지 데이터 각각에 대하여 해상도를 향상시키는 단계 중 어느 하나 이상을 포함할 수 있다.

[0023] 일 실시 예에 따르면, 상기 (d) 단계는, (d-1) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 사용자 유사도를 합산하여 나머지 사용자 별 유사도 스코어를 산출하는 단계, (d-2) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 유사도 스코어에 제1 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 가장 높은 사용자를 상기 다중 계정 사용자로 산출

하는 단계 및 (d-3) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 유사도 스코어에 제2 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 차순위로 높은 사용자를 다중 계정 사용자 후보로 산출하는 단계 중 어느 하나 이상을 포함할 수 있다.

[0024] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치는 하나 이상의 프로세서, 네트워크 인터페이스, 상기 프로세서에 의해 수행되는 컴퓨터 프로그램을 로드(Load)하는 메모리 및 대용량 네트워크 데이터 및 상기 컴퓨터 프로그램을 저장하는 스토리지를 포함하되, 상기 컴퓨터 프로그램은 상기 하나 이상의 프로세서에 의해 (A) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 오퍼레이션, (B) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 오퍼레이션, (C) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 오퍼레이션 및 (D) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 오퍼레이션을 포함하며, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은, 상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상이다.

[0025] 상기 기술적 과제를 달성하기 위한 본 발명의 또 다른 실시 예에 따른 매체에 저장된 컴퓨터 프로그램은 컴퓨팅 장치와 결합하여, (AA) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 단계, (BB) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 단계, (CC) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 단계 및 (DD) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 단계를 포함하며, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은, 상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상이다.

발명의 효과

[0026] 상기와 같은 본 발명에 따르면, 이미지 데이터로부터 산출한 텍스트에 대한 정보와 더불어 해시값에 대한 정보, 상품이 속하는 대분류 카테고리에 대한 정보 및 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보까지 함께 분석하여 다중 계정 사용자를 탐지할 수 있는바, 종래 언어에 따른 한계를 타파할 수 있다는 효과가 있다.

[0027] 또한, 두 개의 임계값 적용으로 인해 가장 유사도가 높은 사용자뿐만 아니라 유사도가 높은 차순위 사용자들까지 탐지할 수 있는바, 사용자가 몇 개의 계정을 사용하는 지와 무관하게 동일한 사용자의 다중 계정 전부를 탐지할 수 있다는 효과가 있다.

[0028] 본 발명의 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해 될 수 있을 것이다.

도면의 간단한 설명

[0029] 도 1은 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치가 포함하는 전체 구성을 예시적으로 도시한 도면이다.

도 2는 본 발명의 제2 실시 예에 따른 암호화폐 지갑 주소를 클러스터링 하는 방법의 대표적인 단계를 나타낸 순서도이다.

도 3은 본 발명의 제2 실시 예에 따른 암호화폐 지갑 주소를 클러스터링 하는 방법에 있어서 S210 단계를 세분화한 순서도이다.

도 4는 주요 다크넷 마켓 별 이미지 데이터의 분포를 테이블도 도시한 도면이다.

도 5는 본 발명의 제2 실시 예에 따른 암호화폐 지갑 주소를 클러스터링 하는 방법에 있어서 S220 단계를 세분화한 순서도이다.

도 6은 Silkroad2에서 수집한 제1 이미지 데이터의 형식인 Base64형식을 Python base64 decoding tool을 이용하여 JPG 형식으로 제1 전처리하는 모습을 예시적으로 도시한 도면이다.

도 7은 복수 개의 이미지가 연결된 하나의 제1 이미지 데이터를 제3 전처리하여 분할하는 모습을 예시적으로 도시한 도면이다.

도 8은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 있어서, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터의 해시값에 대한 정보인 경우에 S230 단계를 구체화한 순서도이다.

도 9는 도 8에 도시된 S230-2 단계를 구체화한 순서도이다.

도 10은 제1 사용자와 제2 이미지 데이터의 해시값에 따른 사용자 유사도인 해밍 디스턴스, 보다 구체적으로 평균 해밍 디스턴스를 제2 사용자 내지 제N 사용자에 대하여 산출한 모습을 예시적으로 도시한 도면이다.

도 11은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 있어서, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보 또는 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보인 경우에 S230 단계를 구체화한 순서도이다.

도 12는 도 11에 도시된 순서도 앞단에 일부 단계가 추가된 순서도이다.

도 13은 다크넷 마켓의 상품 대분류 카테고리와 대분류 카테고리 중, 마약에 대한 소분류 카테고리를 예시적으로 도시한 도면이다.

도 14는 제1 사용자와 제2 이미지 데이터의 대분류 카테고리에 대한에 대한 정보에 따른 사용자 유사도인 코사인 유사도를 제2 사용자 내지 제N 사용자에 대하여 산출한 모습을 예시적으로 도시한 도면이다.

도 15는 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 있어서, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보인 경우에 S230 단계를 구체화한 순서도이다.

도 16은 제1 사용자와 제2 이미지 데이터의 텍스트에 대한 정보에 따른 사용자 유사도인 자카드 유사도를 제2 사용자 내지 제N 사용자에 대하여 산출한 모습을 예시적으로 도시한 도면이다.

도 17은 본 발명의 제2 실시 예에 따른 암호화폐 지갑 주소를 클러스터링 하는 방법에 있어서 S240 단계를 세분화한 순서도이다.

도 18 및 도 19는 에 Silkroad2에서 T_v가 20이며 중복되는 이미지 데이터를 그대로 둔 경우에 제1 임계값 및 제2 임계값 설정을 위한 ROC 커브 및 elbow point에 따른 임계값을 예시적으로 도시한 도면이다.

도 20은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법을 Silkroad2에, Agora에 그리고 이들 둘 모두에 적용하여 가장 유사한 다중 계정 사용자를 탐지한 결과를 도시한 테이블이다.

도 21은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법을 Silkroad2에, Agora에 그리고 이들 둘 모두에 적용하여 가장 유사한 사용자뿐만 아니라 차순위로 높은 사용자까지 탐지함에 관한 정확도 측정 결과를 도시한 테이블이다.

도 22는 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법을 Silkroad2에, Agora에 그리고 이들 둘 모두에 적용하여 2개 이상의 다중 계정을 사용하는 사용자를 탐지함에 관한 정확도 측정 결과를 도시한 테이블이다.

발명을 실시하기 위한 구체적인 내용

[0030] 본 발명의 목적과 기술적 구성 및 그에 따른 작용 효과에 관한 자세한 사항은 본 발명의 명세서에 첨부된 도면에 의거한 이하의 상세한 설명에 의해 보다 명확하게 이해될 것이다. 첨부된 도면을 참조하여 본 발명에 따른

실시 예를 상세하게 설명한다.

[0031] 본 명세서에서 개시되는 실시 예들은 본 발명의 범위를 한정하는 것으로 해석되거나 이용되지 않아야 할 것이다. 이 분야의 통상의 기술자에게 본 명세서의 실시 예를 포함한 설명은 다양한 응용을 갖는다는 것이 당연하다. 따라서, 본 발명의 상세한 설명에 기재된 임의의 실시 예들은 본 발명을 보다 잘 설명하기 위한 예시적인 것이며 본 발명의 범위가 실시 예들로 한정되는 것을 의도하지 않는다.

[0032] 도면에 표시되고 아래에 설명되는 기능 블록들은 가능한 구현의 예들일 뿐이다. 다른 구현들에서는 상세한 설명의 사상 및 범위를 벗어나지 않는 범위에서 다른 기능 블록들이 사용될 수 있다. 또한, 본 발명의 하나 이상의 기능 블록이 개별 블록들로 표시되지만, 본 발명의 기능 블록들 중 하나 이상은 동일 기능을 실행하는 다양한 하드웨어 및 소프트웨어 구성들의 조합일 수 있다.

[0033] 또한, 어떤 구성요소들을 포함한다는 표현은 "개방형"의 표현으로서 해당 구성요소들이 존재하는 것을 단순히 지칭할 뿐이며, 추가적인 구성요소들을 배제하는 것으로 이해되어서는 안 된다.

[0034] 나아가 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급될 때에는, 그 다른 구성요소에 직접적으로 연결 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 한다.

[0035] 이하에서는 도면들을 참조하여 본 발명의 각 실시 예들에 대해 살펴보도록 한다.

[0036] 도1은 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100)가 포함하는 전체 구성을 예시적으로 도시한 도면이다.

[0037] 그러나 이는 본 발명의 목적을 달성하기 위한 바람직한 실시 예일 뿐이며, 필요에 따라 일부 구성이 추가되거나 삭제될 수 있고, 어느 한 구성이 수행하는 역할을 다른 구성이 함께 수행할 수도 있음은 물론이다.

[0038] 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100)는 프로세서(10), 네트워크 인터페이스(20), 메모리(30), 스토리지(40) 및 이들을 연결하는 데이터 버스(50)를 포함할 수 있으며, 기타 본 발명의 목적을 달성함에 있어 요구되는 부가적인 구성들을 더 포함할 수 있음은 물론이라 할 것이다.

[0039] 프로세서(10)는 각 구성의 전반적인 동작을 제어한다. 프로세서(10)는 CPU(Central Processing Unit), MPU(Micro Processer Unit), MCU(Micro Controller Unit) 또는 본 발명이 속하는 기술 분야에서 널리 알려져 있는 형태의 프로세서 중 어느 하나일 수 있다. 아울러, 프로세서(10)는 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법을 수행하기 위한 적어도 하나의 애플리케이션 또는 프로그램에 대한 연산을 수행할 수 있다.

[0040] 네트워크 인터페이스(20)는 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100)의 유무선 인터넷 통신을 지원하며, 그 밖의 공지의 통신 방식을 지원할 수도 있다. 따라서 네트워크 인터페이스(20)는 그에 따른 통신 모듈을 포함하여 구성될 수 있다.

[0041] 메모리(30)는 각종 정보, 명령 및/또는 정보를 저장하며, 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법을 수행하기 위해 스토리지(40)로부터 하나 이상의 컴퓨터 프로그램(41)을 로드할 수 있다. 도 1에서는 메모리(30)의 하나로 RAM을 도시하였으나 이와 더불어 다양한 저장 매체를 메모리(30)로 이용할 수 있음은 물론이다.

[0042] 스토리지(40)는 하나 이상의 컴퓨터 프로그램(41) 및 대용량 네트워크 정보(42)를 비임시적으로 저장할 수 있다. 이러한 스토리지(40)는 ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리 등과 같은 비휘발성 메모리, 하드 디스크, 착탈형 디스크, 또는 본 발명이 속하는 기술 분야에서 널리 알려져 있는 임의의 형태의 컴퓨터로 읽을 수 있는 기록 매체 중 어느 하나일 수 있다.

[0043] 컴퓨터 프로그램(41)은 메모리(30)에 로드되어, 하나 이상의 프로세서(10)에 의해 (A) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자 - 여기서 사용자는 다크넷 마켓에서 상품을 판매하는 판매자 역할의 벤더를 주로 의미하나, 특정 벤더가 다른 벤더로부터 구매자 입장에서 상품을 구매하는 경우도 있을 수 있으므로 판매자와 구매자 모두를 포함하는 광의의 개념이라 할 것이다 - 가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 오퍼레이션, (B) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 오퍼레이션, (C) 상기 생성한 제2 이미지 데이터로부터 복수 개의

특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 산출하는 오퍼레이션 및 (D) 상기 복수 개의 특징 각각에 대하여 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 오퍼레이션을 포함하며, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은, 상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상일 수 있다.

[0044] 이상 간단하게 언급한 컴퓨터 프로그램(41)이 수행하는 오퍼레이션은 컴퓨터 프로그램(41)의 일 기능으로 볼 수 있으며, 보다 자세한 설명은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 대한 설명에서 후술하도록 한다.

[0045] 데이터 버스(50)는 이상 설명한 프로세서(10), 네트워크 인터페이스(20), 메모리(30) 및 스토리지(40) 사이의 명령 및/또는 정보의 이동 경로가 된다.

[0046] 이상 설명한 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100)는 네트워크 기능을 보유한 서버일 수 있으며, 인하우스 시스템 및 공간 임대형 시스템 등과 같은 유형의 물리적인 서버와 무형의 클라우드(Cloud) 서버 등과 같이 실질적인 구현 형태는 무방하다 할 것이다.

[0047] 또한, 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100)는 서버뿐만 아니라 네트워크 기능을 보유한 사용자 단말(미도시)일 수도 있는바, 최근 사용자 단말의 성능이 비약적으로 향상됨에 따라 서버에 준하는 연산을 수행할 수 있기 때문이다. 사용자 단말은 예를 들어 스마트폰, PDA, PDP, 테블릿 PC, 스마트 워치, 스마트 글라스, 노트북 PC 등과 같이 휴대가 가능한 포터블(Portable) 단말일 수 있으며, 데스크톱 PC, 키오스크 등과 같은 설치형 단말일 수도 있다 할 것이다.

[0048] 이하, 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100)가 서버임을 전제로 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 대하여 도 2 내지 도 22를 참조하여 설명하도록 한다.

[0049] 도 2는 본 발명의 제2 실시 예에 따른 암호화폐 지갑 주소를 클러스터링 하는 방법의 대표적인 단계를 나타낸 순서도이다.

[0050] 그러나 이는 본 발명의 목적을 달성함에 있어서 바람직한 실시 예일 뿐이며, 필요에 따라 일부 단계가 추가 또는 삭제될 수 있음은 물론이고, 어느 한 단계가 다른 단계에 포함되어 수행될 수도 있다.

[0051] 한편, 각 단계는 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100, 이하 "장치"라 한다)를 통해 이루어지는 것을 전제로 한다.

[0052] 우선, 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집한다(S210).

[0053] 여기서 제1 내지 제N 사용자는 앞서 간단히 언급한 바와 같이 다크넷 마켓에서 상품을 판매하는 판매자 역할의 벤더를 주로 의미하나, 특정 벤더가 다른 벤더로부터 구매자 입장에서 상품을 구매하는 경우도 있을 수 있으므로 판매자와 구매자 모두를 포함하는 광의의 개념이라 할 것이며, N이 2 이상의 자연수인 것은 특정 사용자에 대한 다중 계정이라는 의미에 최소 두 개의 계정부터 포함되기에 설정한 범위라 볼 수 있다.

[0054] 수집한 복수 개의 제1 이미지 데이터는 이러한 제1 내지 제N 사용자가 다크넷 마켓에 업로드한 이미지 데이터일 수 있으며, 제1 이미지 데이터의 수집은 장치(100)가 직접 수집하거나 제1 이미지 데이터를 수집해둔 외부 서버(미도시)로부터 수집할 수 있다. 그에 따라 S210 단계는 도 3에 도시된 바와 같이 제1 내지 제N 사용자가 업로드한 복수 개의 제1 이미지 데이터를 다크넷 마켓에서 직접 수집하는 단계(S210-1) 및 제1 내지 제N 사용자가 다크넷 마켓에 업로드한 복수 개의 제1 이미지 데이터를 수집한 외부 서버, 예를 들어 다크넷 아카이브(Darknet Archive) 등으로부터 수집하는 단계(S210-2) 중 어느 하나 이상의 단계를 포함할 수 있으며, S210-1 단계에서의 수집은 크롤링(Crawling)의 의미를, S210-2 단계에서의 수집은 수신 또는 로딩(Loading)의 의미를 가진다 할 것이다.

[0055] 제1 이미지 데이터를 수집하는 장소인 다크넷 마켓은 특정 다크넷 마켓에 한정되지 않으며, 복수 개의 다크넷 마켓, 예를 들어 Abraxas, Agora, Alphabay, Blackbank, Cloudnine, Evolution, Silkroad2 및 Pandora 등과

같이 복수 개의 주요 다크넷 마켓으로부터 제1 이미지 데이터를 동시 다발적으로 수집할 수도 있고, 이 경우 장치(100)가 포함하는 프로세서(10)를 병렬 프로세싱이 가능한 것으로 구현한다면 제1 이미지 데이터의 수집 속도가 비약적으로 향상될 수 있을 것이다.

[0056] 더 나아가, 제1 이미지 데이터의 수집은 복수 개의 다크넷 마켓에 대하여 이루어지는 것이 바람직한 측면도 있는바, 주요 다크넷 마켓 별 이미지 데이터의 분포를 테이블로 도시한 도 4에 기재된 바와 같이 주요 다크넷 마켓 별로 분포되어 있는 이미지 데이터의 카테고리가 각각 상이하기에 최대한 다양한 다크넷 마켓에서 최대한 다양한 카테고리에 속하는 이미지 데이터를 충분하게 수집하여 동일 사용자에 대한 다중 계정을 최대한으로 탐지하는 것이 효율적일 수 있기 때문이다.

[0057] 한편, 수집하는 제1 이미지 데이터는 사용자 별로 구별하여 수집할 수 있는바, 예를 들어 제1 사용자가 업로드한 제1 이미지 데이터와 제N 사용자가 업로드한 제1 이미지 데이터를 수집하는 과정에서 처음부터 구별하여 수집할 수 있으며, 사용자와 무관하게 제1 이미지 데이터를 전체적으로 수집할 수도 있음은 물론이라 할 것이다.

[0058] 아울러, 수집하는 제1 이미지 데이터는 다크넷 마켓 별로 구별하여 수집할 수도 있는바, 예를 들어 Agora에 제1 사용자 내지 제N 사용자가 업로드한 제1 이미지 데이터와 Silkroad2에 제1 사용자 내지 제N 사용자 - 여기서 제1 사용자 내지 제N 사용자는 Agora에서의 제1 사용자 내지 제N 사용자와 상이한 사용자일 수 있으나, 설명의 편의를 위해 동일한 명칭을 부여하도록 한다 - 가 업로드한 제1 이미지 데이터를 수집하는 과정에서 처음부터 구별하여 수집할 수 있으며, 다크넷 마켓과 무관하게 제1 이미지 데이터를 전체적으로 수집할 수도 있음은 물론이라 할 것이다.

[0059] 더 나아가, 수집하는 제1 이미지 데이터는 상기 설명한 사용자 별 그리고 다크넷 마켓 별 수집하는 방법을 조합하여 특정 다크넷 마켓에 대하여 특정 사용자 별로 구별하여 수집할 수도 있음은 물론이라 할 것이다.

[0060] 복수 개의 제1 이미지 데이터를 수집했다면, 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성한다(S220).

[0061] 여기서 데이터의 전처리는 제1 전처리 내지 제3 전처리로 구체화될 수 있는바, S220 단계를 구체화한 순서도를 도시한 도 5를 참조하면, S220 단계는 수집한 제1 이미지 데이터의 제1 전처리로 파일 형식을 JPG로 변환하여 복수 개의 제2 이미지 데이터를 생성하는 단계(S220-1), 수집한 제1 이미지 데이터의 제2 전처리로 제1 이미지 데이터 별로 사용자명(Username)에 대한 정보, 대분류 카테고리에 대한 정보, 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 및 사용자별 PGP Key - 다크넷 마켓에서는 암호 소프트웨어인 PGP를 활용한 암호화된 이메일을 사용하게 되는데, 이러한 이메일을 전송할 때 사용하는 공개키를 PGP Key라 하며, 벤더는 구매자와의 커뮤니케이션을 위해 고유의 PGP Key를 지니게 되고, 암호화된 개인키로 구매자의 메시지를 확인할 수 있다 - 에 대한 정보 중 어느 하나 이상을 추출하여 복수 개의 제2 이미지 데이터를 생성하는 단계(S220-2) 및 수집한 제1 이미지 데이터의 제3 전처리로 제1 이미지 데이터가 복수 개의 이미지 데이터가 연결된 이미지 데이터인 경우, 각각의 이미지 데이터가 포함하고 있는 상품 별로 분할(Cropping)하고, 해당 상품명으로 분할한 이미지 데이터를 라벨링(Labeling)하여 상복수 개의 제2 이미지 데이터를 생성하는 단계(S220-3) 중 어느 하나 이상의 단계를 포함할 수 있다.

[0062] 우선, S220-1 단계에 대하여 설명하면 제1 전처리는 수집한 제1 이미지 데이터의 파일 형식을 하나로 통일화시키는 전처리로 볼 수 있는바, 사용자 별로 그리고 다크넷 마켓 별로 업로드한 제1 이미지 데이터의 파일 형식이 상이할 수 있으며, 이를 통일화 시키는 경우 데이터 처리에 용이함을 부여할 수 있기 때문이다.

[0063] 도 6은 Silkroad2에서 수집한 제1 이미지 데이터의 형식인 Base64형식을 Python base64 decoding tool을 이용하여 JPG 형식으로 제1 전처리하는 모습을 예시적으로 도시한 도면인바, 여기서 제1 전처리는 파일 형식의 변환으로 볼 수 있을 것이며, JPG는 파일 형식의 하나의 예일 뿐, 이에 반드시 한정하는 것은 아니라 할 것이다.

[0064] S220-2 단계에 대하여 설명하면 제2 전처리는 제1 이미지 데이터로부터 후술할 사용자 유사도 산출 및 산출한 다중 계정 사용자를 검증하는데 활용될 수 있는 정보들을 사전에 추출하는 전처리로 볼 수 있는바, 사용자명(Username)에 대한 정보, 대분류 카테고리에 대한 정보, 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 및 사용자별 PGP Key에 대한 정보가 그것이며, 이 중에서 사용자별 PGP Key에 대한 정보는 앞서 도 3에 도시된 모든 다크넷 마켓의 이미지 데이터에서 공통적으로 존재함이 확인된 것이기에 PGP Key가 동일하면 동일한 사용자로 볼 수 있을 만큼의 중요한 정보이기 때문인바, 이는 후술할 각 특징 별 사용자 유사도 산출 결과에 따라 선정한 다중 계정 사용자의 검증에 활용될 수 있는 정보라 할 것이다.

[0065] S220-3 단계에 대하여 설명하면 제3 전처리는 제1 이미지 데이터가 한 개의 이미지 데이터라 할지라도 그 실제

는 복수 개의 이미지가 연결되거나 합쳐진 경우가 비일비재(非一非再)하므로 이를 상품 별로 분할하고 해당 상품명으로 분할한 이미지 데이터를 라벨링함으로써 후술할 복수 개의 특징 중, 상품이 속하는 대분류 카테고리에 대한 정보와 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 사용에 용이함을 부여하기 위한 전처리에 해당한다.

[0066] 도 7은 복수 개의 이미지가 연결된 하나의 제1 이미지 데이터를 제3 전처리하여 분할하는 모습을 예시적으로 도시한 도면인바, 각각의 이미지가 포함하고 있는 상품 별로 분할해야 하기 때문에 장치(100)의 프로세서(10)를 객체(Object) 인식이 가능한 인공지능 프로세서로 구현함이 바람직하다 할 것이다.

[0067] 다시 도 2에 대한 설명으로 돌아가도록 한다.

[0068] 제2 이미지 데이터를 생성했다면, 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 산출한 복수 개의 특징 각각에 대하여 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출한다(S230).

[0069] 여기서 제2 이미지 데이터로부터 산출한 복수 개의 특징은 제2 이미지 데이터의 해시값에 대한 정보, 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상일 수 있으며, 가급적 네 가지 특징 모두를 산출함이 가장 바람직하다고 할 것이다.

[0070] 제2 이미지 데이터로부터 산출하여 사용자 유사도를 산출하느데 이용되는 네 가지 특징은 종래 기술과 차별화되는 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법의 독자적인 기술적 특징인바, 이하 하나하나 설명하도록 한다.

[0071] 도 8은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 있어서, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터의 해시값에 대한 정보인 경우에 S230 단계를 구체화한 순서도이다.

[0072] 그러나 이는 본 발명의 목적을 달성함에 있어서 바람직한 실시 예일 뿐이며, 필요에 따라 일부 단계가 추가 또는 삭제될 수 있음은 물론이고, 어느 한 단계가 다른 단계에 포함되어 수행될 수도 있다.

[0073] 설명에 앞서 제2 이미지 데이터의 해시값에 대한 정보를 특징으로 산출하여 사용자 유사도를 산출하는 근거는 유사한 이미지 데이터의 해시값은 동일하기 때문에 이미지 데이터의 해시값을 이용하여 사용자들의 사진 유사도를 측정해 두 사용자가 유사한 사진을 많이 보유하고 있으면 사용자 간 유사도가 가장 높다고 하여 다중 계정 사용자로 볼 수 있다는 것이다.

[0074] 우선, 복수 개의 제2 이미지 데이터 각각에 대하여 복수 개의 해시 타입(Type)별 해시값을 산출한다(S230-1).

[0075] 여기서 해시값 산출은 공지된 툴인 파이썬의 해시 함수(Hash function) 등을 이용할 수 있으며, 복수 개의 해시 타입은 AHASH, PHASH, DHASH 및 WHASH중 어느 둘 이상이며, 해시값을 산출하는 해시 타입이 복수 개이기 때문에 둘 이상인 것이고 가급적 네 가지 해시 타입 별 해시값을 모두 산출하는 것이 가장 바람직하다 할 것이다.

[0076] 각각의 해시 타입에 대하여 설명하면, AHASH는 Average Hash로써 Color value과 Mean value를 비교하여 해시값을 산출하며, PHASH는 Perpectual Hash로써 AHASH와 비교했을 때 Robust한 알고리즘으로 감마 보정이 수행되거나 이미지에 컬러 히스토그램이 적용된 경우 허위 결측이 가능하고, DHASH는 Difference HASH로써 인접한 픽셀 간의 Gradient를 확인하여 구분해 해시값을 산출하며, 계산 속도가 빠르고 스케일이나 밝기, 대비, 감마 보정 등에도 영향을 받지 않고, WHASH는 Wavelet Hash로써 PHASH와 유사하지만 Discrete wallet transformation을 사용하여 이미지의 Frequency를 표현하는 해시 타입이다.

[0077] 한편, 각각의 해시 타입에 대한 해시 프로세스는 공지된 프로세스를 이용할 수 있는바, 예를 들어 AHASH의 경우 이미지 데이터의 디테일을 보여주는 High Frequency를 제거하고 이미지의 구조를 보여주는 Low Frequency만 남긴 후(이미지 크기 줄이기), 이미지 데이터를 그레이 스케일로 변경하며(색상 줄이기), 색상의 평균값을 산출하고(색상 평균값 구하기) Color Value가 평균값보다 큰지 아니면 작은지 산출하여 비트값을 산출하고(비트값 계산하기), 64bit를 64 bit integer로 변경하여 해시값을 산출할 수 있으며(해시값 구하기), 나머지 해시 타입들의 경우 역시 이와 유사한 공지된 프로세스를 이용할 수 있다.

[0078] 복수 개의 제2 이미지 데이터 각각에 대하여 복수 개의 해시 타입 별 해시값을 산출했다면, 산출한 복수 개의 해시 타입별 해시값을 이용하여 복수 개의 해시 타입의 성능을 평가한다(S230-2).

[0079] 해시 타입의 성능을 평가하는 것은 후술할 S230-3 단계에서 어떠한 해시 타입의 해시값을 이용하여 사용자 유사도를 산출할지 여부를 결정하기 위한 것인바, 동일한 제2 이미지 데이터를 동일한 해시값으로 변환해야 해시 함수의 성능이 우수하게 나올 수 있기 때문에 해시 타입 별로 해싱 성능에 대하여 평가하는 것이다.

[0080] 그에 따라 S230-2 단계는 도 9에 도시된 바와 같이 복수 개의 제2 이미지 데이터 각각에 대하여 복수 개의 해시 타입 별로 그룹핑하는 단계(S230-2-1), 복수 개의 해시 타입 별로 그룹핑한 각각의 그룹에 속하는 모든 제2 이미지 데이터에 대하여 페어(Pair)를 전부 생성하는 단계(S230-2-2), 전부 생성한 각각의 페어의 SSIM값을 산정하여 그룹별 평균 SSIM값을 산출하는 단계(S230-2-3) 및 산출한 그룹별 평균 SSIM 값 중, 가장 높은 평균 SSIM 값을 나타내는 그룹에 대한 해시 타입을 가장 우수한 성능의 해시 타입으로 평가하는 단계(S230-2-4) 중 어느 하나 이상의 단계를 포함할 수 있으며, 각각의 그룹에 속하는 제2 이미지 데이터의 개수가 상이할 수 있으므로 S230-2-3 단계와 S230-3-4 단계 사이에 산출한 그룹별 평균 SSIM 값에 대하여 각각의 그룹에 속하는 제2 이미지 데이터의 개수에 따라 가중치(Weight)를 부여하여 가중치가 부여된 그룹별 평균 SSIM 값(Weighted average SSIM)을 나머지 사용자 별로 산출하는 단계(S230-2-3.5)를 더 포함할 수 있고, 이 경우 S230-2-4 단계의 평균 SSIM 값은 가중치가 부여된 평균 SSIM 값일 수 있다.

[0081] 이와 같이 해시 타입을 평가함에 있어서 SSIM값을 이용하며, 그에 따라 이미지 데이터가 유사할수록 SSIM값이 1에 가까워지고 그렇지 않은 경우 0에 가까워지기 때문에 평균 SSIM값 또는 가중치가 부여된 평균 SSIM값이 1에 가장 가까운 해시 타입이 가장 우수한 성능을 발휘하는 해시 타입으로 평가될 수 있다.

[0082] 다시 도 8에 대한 설명으로 돌아가도록 한다.

[0083] 복수 개의 해시 타입의 성능을 평가했다면, 성능 평가 결과 가장 우수한 성능의 해시 타입의 해시값을 이용하여 제1 사용자에 대한 제2 이미지 데이터의 해시값과 나머지 사용자에 대한 제2 이미지 데이터의 해시값에 따른 사용자 유사도인 해밍 디스턴스(Hamming distance)를 나머지 사용자 별로 산출한다(S230-3).

[0084] 여기서 해밍 디스턴스는 같은 길이의 Binary string에 대하여 유사성을 비교하는 metric을 생성하는 기법으로서, string 간의 차이를 기준으로 하여 스코어를 산출하며, 스코어가 0일 때 두 string이 정확하게 매칭되는 것을 의미하고, 스코어가 16일 때 두 string이 전혀 매칭되지 않는 것을 의미하는 공지된 기법에 해당한다.

[0085] 이 경우, 제2 이미지 데이터로부터 산출한 특징이 제2 이미지 데이터의 해시값이기 때문에 해밍 디스턴스의 적용이 가능한 것이며, 보다 구체적으로 S230-2 단계에서의 해시 타입의 성능 평과 결과 가장 우수한 해시 타입의 해시값에 대하여 해밍 디스턴스를 적용해 해시값에 따른 사용자 유사도를 산출하는 것이다.

[0086] 해밍 디스턴스의 산출에 대하여 보다 구체적으로 설명하면, 제1 사용자에 대한 제2 이미지 데이터가 복수 개일 경우, 각각의 제2 이미지 데이터와 나머지 사용자 각각에 대한 제2 이미지 데이터 각각에 대하여 해밍 디스턴스를 개별적으로 산출하며, 제1 사용자에 대한 각각의 제2 이미지 데이터와의 관계에서 산출된 나머지 사용자 각각의 해밍 디스턴스 중 가장 낮은 해밍 디스턴스를 추출하고 이를 평균내어 나머지 사용자 별 평균 해밍 디스턴스를 산출한다.

[0087] 조금 더 쉽게 설명하면, 제1 사용자에 대하여 1 내지 10이라는 열 개의 제2 이미지 데이터에 대한 해시값을 산출한 상태이며, 제2 사용자에 대하여 a 내지 e라는 다섯 개의 제2 이미지 데이터에 대한 해시값을 산출한 상태라면, 제1 사용자의 제2 이미지 데이터 1과 제2 사용자의 제2 이미지 데이터 a의 해시값에 따른 해밍 디스턴스, 제2 사용자의 제2 이미지 데이터 1과 제2 사용자의 제2 이미지 데이터 b의 해시값에 따른 해밍 디스턴스 쪋 제1 사용자의 제2 이미지 데이터 1과 제2 사용자의 제2 이미지 데이터 e의 해시값에 따른 해밍 디스턴스를 산출하여 제1 사용자의 제2 이미지 데이터 1과의 관계에서 제2 사용자의 제2 이미지 데이터와 산출한 해밍 디스턴스 중 가장 낮은 해밍 디스턴스를 추출하고, 이러한 과정을 제1 사용자의 모든 제2 이미지 데이터에 대하여 반복하는 것이다. 이러한 예시에 의한다면 제1 사용자의 제2 이미지 데이터가 열 개이기에, 제2 사용자의 제2 이미지 데이터와의 관계에서 산출하여 추출한 가장 낮은 해밍 디스턴스 역시 열 개가 될 것이며, 추출한 열 개의 해밍 디스턴스에 대한 평균이 제2 사용자에 대한 평균 해밍 디스턴스가 되는 것이고, 이를 나머지 모든 사용자에 대하여 반복한다면 각각의 사용자에 대한 평균 해밍 디스턴스가 산출될 수 잇다.

[0088] 도 10에 제1 사용자와 제2 이미지 데이터의 해시값에 따른 사용자 유사도인 해밍 디스턴스, 보다 구체적으로 평균 해밍 디스턴스를 제2 사용자 내지 제N 사용자에 대하여 산출한 모습을 예시적으로 도시한바, 제2 사용자의 평균 해밍 디스턴스가 0.1로 다른 사용자들에 비하여 가장 낮기 때문에 제2 이미지 데이터로부터 산출한 복수 개의 특징 중, 제2 이미지 데이터의 해시값에 대한 정보를 이용하여 산출한 제1 사용자와 동일한 사용자로 볼

수 있는 다중 계정 사용자는 제2 사용자인 것으로 간주할 수 있다.

[0089] 그러나 이는 제2 이미지 데이터로부터 산출한 복수 개의 특징 중, 제2 이미지 데이터의 해시값에 대한 정보만을 이용하여 산출한 결과인바, 나머지 특징들도 이용할 수 있으므로 제2 사용자가 제1 사용자와 동일한 사용자라고 확정할 수는 없을 것이며, 이에 대해서는 S240 단계에서 후술하도록 한다.

[0090] 도 11은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 있어서, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보 또는 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보인 경우에 S230 단계를 구체화한 순서도이다.

[0091] 그러나 이는 본 발명의 목적을 달성함에 있어서 바람직한 실시 예일 뿐이며, 필요에 따라 일부 단계가 추가 또는 삭제될 수 있음은 물론이고, 어느 한 단계가 다른 단계에 포함되어 수행될 수도 있다.

[0092] 설명에 앞서 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보 또는 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보를 특징으로 산출하여 사용자 유사도를 산출하는 근거는 동일한 사용자는 동일한 카테고리에 속하는 상품을 판매하거나 상품의 카테고리 분포가 유사할 확률이 높다는 것이며, 이를 만족하는 경우 사용자 간 유사도가 가장 높다고 하여 다중 계정 사용자로 볼 수 있다는 것이다.

[0093] 우선, 제1 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 제1 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 대분류 카테고리 또는 소분류 카테고리 별로 산출한다(S230-4).

[0094] 앞서 S220 단계, 보다 구체적으로 S220-2 단계에 대한 설명에서 제2 전처리로 대분류 카테고리에 대한 정보와 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보를 추출하여 제2 이미지 데이터를 생성했다고 했던바, S230-4 단계에서 이를 이용하는 것이며, 제1 이미지 데이터에 따라 대분류 카테고리에 대한 정보와 소분류 카테고리에 대한 정보가 존재하지 않을 수도 있으므로 도 12에 도시된 바와 같이 S230-4 단계 이전에 복수 개의 제2 이미지 데이터 중, 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터를 추출하는 단계(S230-3.6) 및 추출한 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터에 대하여 결여된 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보를 부여하는 단계(S230-3.7) 중 어느 하나 이상의 단계를 더 포함할 수 있으며, 이를 위해서는 장치(100)가 카테고리에 대한 정보 부여에 대한 학습이 완료되어 있어야 할 것이다. 예를 들어, 낮은 레이어는 이미지넷의 데이터를 입력하여 학습시키고 높은 레이어는 다크넷 사진을 입력하여 학습시킨 Pytorch module에 내장되어 있는 ResNet50 모델을 이용할 수 있을 것이며, 이 경우 이의 Final softmax 레이어를 다크넷 데이터 클래스를 다루는 Softmax 레이어로 교체하는 튜닝을 진행함으로써 본 발명의 일 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 커스터마이징할 수 있을 것이다.

[0095] 한편, 여기서의 비중은 제1 사용자가 판매하고 있는 전체 상품에서 해당 카테고리의 상품이 차지하는 비중을 의미하는바, 예를 들어, 제1 사용자가 판매하고 있는 전체 상품이 열 개이고, 이 중에서 마약 대분류 카테고리에 속하는 상품이 두 개라면, 비중은 0.2가 될 것이며, 이를 제1 사용자에 대한 제2 이미지 데이터 전부에 대하여 산출하는 것이다.

[0096] 도 13은 다크넷 마켓의 상품 대분류 카테고리와 대분류 카테고리 중, 마약에 대한 소분류 카테고리를 예시적으로 도시한 도면인바, 예시적인 것에 해당하므로 이에 반드시 한정되는 것은 아니라 할 것이나, 다크넷 마켓에서 거래되는 상품들 중, 마약을 제외한 나머지 대분류 카테고리에 속하는 상품들은 단일 소분류 카테고리를 가지고 있는 것이 대부분이기 때문에 대분류 카테고리와 소분류 카테고리가 동일하다고 할 것이다.

[0097] 이후, 나머지 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 나머지 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 상기 대분류 카테고리 또는 소분류 카테고리 별로 산출한다(S230-5).

[0098] S230-5 단계는 제1 사용자가 아니라 나머지 사용자에 대하여 비중을 산출하는 것을 제외하고 S230-4 단계에 대한 설명과 동일하기에 중복 서술을 방지하기 위해 자세한 설명은 생략하도록 한다.

[0099] 제1 사용자 그리고 나머지 사용자에 대한 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 모두 산출했다면, 대분류 카테고리 또는 소분류 카테고리 별로 산출한 제1 사용자에 대한 비중과 나머지 사용자에 대한 비중을 이용하여 제1 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중과 나머지 사용자에 대한 대분

류 카테고리 또는 소분류 카테고리 비중에 따른 사용자 유사도인 코사인 유사도(Cosine similarity)를 나머지 사용자 별로 산출한다(S230-6).

[0100] 앞서 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터의 해시값에 대한 정보인 경우, 사용자 유사도는 해밍 디스턴스였으나, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보 또는 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보 중 어느 하나인 경우, 사용자 유사도는 코사인 유사도라는 것이다.

[0101] 여기서 코사인 유사도는 두 벡터 간의 코사인 각도를 이용하여 구할 수 있는 두 벡터의 유사도를 의미하며, 두 벡터의 방향이 완전히 동일한 경우에는 1의 값을, 90° 의 각을 이루면 0의 값을, 180° 로 반대의 방향을 가지면 -1의 값을 가지게 되기에, 1에 가까울수록 유사도가 높다고 판단할 수 있는 공지된 기법에 해당한다.

[0102] 도 14에 제1 사용자와 제2 이미지 데이터의 대분류 카테고리에 대한에 대한 정보에 따른 사용자 유사도인 코사인 유사도를 제2 사용자 내지 제N 사용자에 대하여 산출한 모습을 예시적으로 도시한바, 제2 사용자에 대한 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리 별 비중과 제1 사용자의 대분류 카테고리별 비중이 일치하기에 코사인 유사도는 1임을 확인할 수 있으며, 나머지 사용자들의 경우 0에 가깝기에 제2 이미지 데이터로부터 산출한 복수 개의 특징 중, 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보를 이용하여 산출한 제1 사용자와 동일한 사용자로 볼 수 있는 다중 계정 사용자는 제2 사용자인 것으로 간주할 수 있다.

[0103] 그러나 이는 제2 이미지 데이터로부터 산출한 복수 개의 특징 중, 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보만을 이용하여 산출한 결과인바, 나머지 특징들도 이용할 수 있으므로 제2 사용자가 제1 사용자와 동일한 사용자라고 확정할 수는 없을 것이며, 이에 대해서는 S240 단계에서 후술하도록 한다.

[0104] 한편, 도 14에는 설명의 편의를 위해 대분류 카테고리에 대한 비중만을 도시하여 설명하였으나, 소분류 카테고리에 대한 비중 역시 개별적으로 산출하여 코사인 유사도를 산출함은 물론이라 할 것이며, 제1 사용자에 대한 제2 이미지 데이터의 대분류 카테고리의 비중은 나머지 사용자에 대한 제2 이미지 데이터의 대분류 카테고리와 코사인 유사도를 산출하고, 제1 사용자에 대한 제2 이미지 데이터의 소분류 카테고리의 비중은 나머지 사용자에 대한 제2 이미지 데이터의 소분류 카테고리와 코사인 유사도를 산출해야 하는 것 역시 물론이라 할 것이다.

[0105] 도 15는 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 있어서, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보인 경우에 S230 단계를 구체화한 순서도이다.

[0106] 그러나 이는 본 발명의 목적을 달성함에 있어서 바람직한 실시 예일 뿐이며, 필요에 따라 일부 단계가 추가 또는 삭제될 수 있음은 물론이고, 어느 한 단계가 다른 단계에 포함되어 수행될 수도 있다.

[0107] 설명에 앞서 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보를 특징으로 산출하여 사용자 유사도를 산출하는 근거는 동일한 사용자는 이미지 데이터를 업로드할 때 벤더 ID, 상품명 등과 같이 사용자를 식별할 수 있는 특징적인 텍스트를 기재하여 업로드하는 경우가 많이 때문이며, 이를 만족하는 경우 사용자 간 유사도가 가장 높다고 하여 다중 계정 사용자로 볼 수 있다는 것이다.

[0108] 우선, 복수 개의 제2 이미지 데이터 각각에 대하여 포함하고 있는 텍스트 정보를 추출한다(S230-7).

[0109] 이러한 S230-7 단계는 복수 개의 제2 이미지 데이터 각각에 대하여 텍스트가 포함된 영역에 대한 바운딩 박스(Bounding box)를 설정하는 단계(S230-7-1) 및 복수 개의 제2 이미지 데이터 각각에 대하여 해상도를 향상시키는 단계(S230-7-2) 중 어느 하나 이상의 단계를 포함할 수 있으며, 바운딩 박스를 설정함에 있어서 장치(100)는 TextuseNet model을, 해상도를 향상시킴에 있어서 장치(100)는 Super resolution tool 중, pre-trained SwinIR model을 이용할 수 있고, 텍스트 정보 추출에 있어서는 SATRN model을 이용할 수 있을 것이다.

[0110] 이후, 추출한 텍스트 정보를 이용하여 제1 사용자에 대한 제2 이미지 데이터의 텍스트에 대한 정보와 나머지 사용자에 대한 제2 이미지 데이터의 텍스트에 대한 정보에 따른 사용자 유사도인 자카드 유사도(Jaccard similarity)를 나머지 사용자 별로 산출한다(S230-8).

[0111] 앞서 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터의 해시값에 대한 정보인 경우, 사용자 유사도는 해밍 디스턴스였으며, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리에 대한 정보 또는 대분류 카테고리에 속하는 소분

류 카테고리에 대한 정보 중 어느 하나인 경우, 사용자 유사도는 코사인 유사도였고, 제2 이미지 데이터로부터 산출한 복수 개의 특징 중 하나가 제2 이미지 데이터의 텍스트에 대한 정보인 경우, 사용자 유사도는 자카드 유사도라는 것이다.

[0112] 여기서 자카드 유사도는 합집합에서 교집합의 비율을 구한다면 두 집합 A와 B의 유사도를 구할 수 있다는 것이 주된 아이디어의 시발점인 기법으로서, 자카드 유사도는 0과 1 사이의 값을 가지며 두 집합이 동일하다면 1의 값을, 두 집합의 공통 원소가 없다면 0의 값을 갖는 공지된 기법에 해당한다.

[0113] 도 16에 제1 사용자와 제2 이미지 데이터의 텍스트에 대한 정보에 따른 사용자 유사도인 자카드 유사도를 제2 사용자 내지 제N 사용자에 대하여 산출한 모습을 예시적으로 도시한바, 제2 사용자에 대한 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보가 제1 사용자에 대한 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보와 일치하기에 자카드 유사도는 1임을 확인할 수 있으며, 나머지 사용자들의 경우 0에 가깝기에 제2 이미지 데이터로부터 산출한 복수 개의 특징 중, 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보를 이용하여 산출한 제1 사용자와 동일한 사용자로 볼 수 있는 다중 계정 사용자는 제2 사용자인 것으로 간주할 수 있다.

[0114] 그러나 이는 제2 이미지 데이터로부터 산출한 복수 개의 특징 중, 제2 이미지 데이터가 텍스트에 대한 정보를 이용하여 산출한 결과인바, 나머지 특징들도 이용할 수 있으므로 제2 사용자가 제1 사용자와 동일한 사용자라고 확정할 수는 없을 것이며, 이에 대해서는 S240 단계에서 후술하도록 한다.

[0115] 다시 도 2에 대한 설명으로 돌아가도록 한다.

[0116] 제1 사용자와 나머지 사용자 간의 사용자 유사도를 산출했다면, 복수 개의 특징 각각에 대하여 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출한다(S240).

[0117] 앞서 각각의 특징들에 대한 설명 말미에, 각각의 특징과 나머지 사용자 별로 산출한 사용자 유사도에 따라 제1 사용자와 상이한 계정을 사용하되, 동일한 사용자로 간주할 수 있는 다중 계정 사용자를 산출할 수 있으나, 이에 대한 확정을 미루어 두었던 것이 S240 단계를 위함인바, S240 단계는 도 17에 도시된 바와 같이 복수 개의 특징 각각에 대하여 나머지 사용자 별로 산출한 사용자 유사도를 합산하여 나머지 사용자 별 유사도 스코어를 산출하는 단계(S240-1), 복수 개의 특징 각각에 대하여 나머지 사용자 별로 산출한 유사도 스코에 제1 임계값을 적용하여 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 가장 높은 사용자를 다중 계정 사용자로 산출하는 단계(S240-2) 및 복수 개의 특징 각각에 대하여 나머지 사용자 별로 산출한 유사도 스코어에 제2임계값을 적용하여 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 차순위로 높은 사용자를 다중 계정 사용자 후보로 산출하는 단계(S240-3) 중 어느 하나 이상의 단계를 포함할 수 있다.

[0118] 먼저 S240-1 단계에 대하여 설명하면, 앞서 설명한 바와 같이 코사인 유사도와 자카드 유사도는 그 값이 0 내지 1 사이인바 그대로 합산함에 있어서 문제되지 않으나 해밍 디스턴스의 경우 그 값이 0 내지 16 사이이므로 코사인 유사도 및 자카드 유사도와 스케일을 맞추기 위한 과정이 필요하며, 그 이후에 사용자 유사도를 합산해야 할 것이다.

[0119] 아울러, 사용자의 유사도를 합산하기 이전에 복수 개의 특징 별로 가중치를 부과한 상태에서 사용자 유사도를 합산할 수도 있을 것인바, 각각의 특징에 대한 가중치는 장치(100)의 관리자의 설정에 의해 조절 가능할 수 있을 것이나, 이미지 데이터를 이용하여 다중 계정을 탐지하는 기술이기 때문에 이미지 데이터 자체에 대한 특징으로 볼 수 있는 해시값에 대한 정보에 대한 가중치를 높게 설정하는 것이 바람직하다고 할 것이다.

[0120] 한편, S240-2 단계 및 S240-3 단계에서의 제1 임계값 및 제2 임계값에 대하여 설명하면, ROC 커브를 이용하여 False positive의 trade-off를 고려해 지정할 수 있으며, 보다 구체적으로 ROC 커브의 elbow point를 이용할 수 있고, 도 18 및 도 19에 Silkroad2에서 T_v가 20이며 중복되는 이미지 데이터를 그대로 둔 경우에 제1 임계값 및 제2 임계값 설정을 위한 ROC 커브 및 elbow point에 따른 임계값을 예시적으로 도시해 놓은바, 제2 임계값이 제1 임계값이 비하여 낮음을 확인할 수 있다 할 것이다.

[0121] 지금까지 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 대하여 설명하였다. 본 발명에 따르면, 이미지 데이터로부터 산출한 텍스트에 대한 정보와 더불어 해시값에 대한 정보, 상품이 속하는 대분류 카테고리에 대한 정보 및 대분류 카테고리에 속하는 소분류 카테고리에 대한 정보까지 함께 분석하여 다중 계정 사용자를 탐지할 수 있는바, 종래 언어에 따른 한계를 타파할 수 있다. 또한, 두 개의

임계값 적용으로 인해 가장 유사도가 높은 사용자뿐만 아니라 유사도가 높은 차순위 사용자들까지 탐지할 수 있는바, 사용자가 몇 개의 계정을 사용하는 지와 무관하게 동일한 사용자의 다중 계정 전부를 탐지할 수 있다.

[0122] 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 대한 성능 검증을 위해 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법을 통해 제1 사용자와 유사도가 가장 높다고 산출된 다른 사용자의 PGP Key를 비교하여 성능을 검증한 결과를 도 20 내지 도 22에 첨부하였다.

[0123] 여기서 도 20은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법을 Silkroad2에, Agora에 그리고 이들 둘 모두에 적용한 결과를 도시한 테이블인바, 정확도가 각각 0.985, 0.861, 0.908로 매우 높음을 확인할 수 있으며, 도 21은 가장 유사한 사용자뿐만 아니라 차순위로 높은 사용자까지 탐지함에 관한 정확도 측정 결과를 도시한 테이블인바, 정확도가 각각 0.953, 0.855, 0.897로 마찬가지로 매우 높음을 확인할 수 있고, 도 22는 2개 이상의 다중 계정을 사용하는 사용자를 탐지함에 관한 정확도 측정 결과를 도시한 테이블인바, 정확도가 각각 0.951, 0.77, 0.841로 마찬가지로 매우 높음을 확인할 수 있다.

[0124] 참고로 정확도 측정은 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 따라 동일한 사용자라고 산출된 사용자들에 대한 제2 이미지 데이터를 모두 합산하여 1/2씩 랜덤으로 나누고(Training set과 Test set), 본 발명을 적용하여 산출된 다중 계정 사용자 탐지 결과와 PGP Key에 따른 비교 결과를 비교하는 방식으로 진행하였다.

[0125] 한편, 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100) 및 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법은 모든 기술적 특징을 동일하게 포함하는 본 발명의 제3 실시 예에 따른 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램으로 구현할 수도 있는바, 이 경우 컴퓨팅 장치와 결합하여, (AA) 다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 단계, (BB) 상기 수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 단계, (CC) 상기 생성한 제2 이미지 데이터로부터 복수 개의 특징(Feature)을 산출하고, 상기 산출한 복수 개의 특징 각각에 대하여 상기 제1 사용자와 나머지 사용자 간의 사용자 유사도를 상기 나머지 사용자 별로 산출하는 단계 및 (DD) 상기 복수 개의 특징 각각에 대하여 상기 나머지 사용자 별로 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 상기 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 단계를 포함하며, 상기 제2 이미지 데이터로부터 산출한 복수 개의 특징은, 상기 제2 이미지 데이터의 해시값에 대한 정보, 상기 제2 이미지 데이터가 포함하고 있는 상품이 속하는 대분류 카테고리(Main Category)에 대한 정보, 상기 대분류 카테고리에 속하는 소분류 카테고리(Sub Category)에 대한 정보 및 상기 제2 이미지 데이터가 포함하고 있는 텍스트에 대한 정보 중 어느 하나 이상일 수 있을 것이며, 중복 서술을 위해 자세히 기재하지는 않았지만 본 발명의 제1 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치(100) 및 본 발명의 제2 실시 예에 따른 다크넷 마켓 사용자의 다중 계정을 탐지하는 방법에 적용된 모든 기술적 특징은 본 발명의 제3 실시 예에 따른 컴퓨터로 판독 가능한 매체에 저장된 컴퓨터 프로그램에 모두 동일하게 적용될 수 있음은 물론이라 할 것이다.

[0126] 이상 첨부된 도면을 참조하여 본 발명의 실시 예들을 설명하였지만, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시 예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

부호의 설명

[0127] 10: 프로세서

20: 네트워크 인터페이스

30: 메모리

40: 스토리지

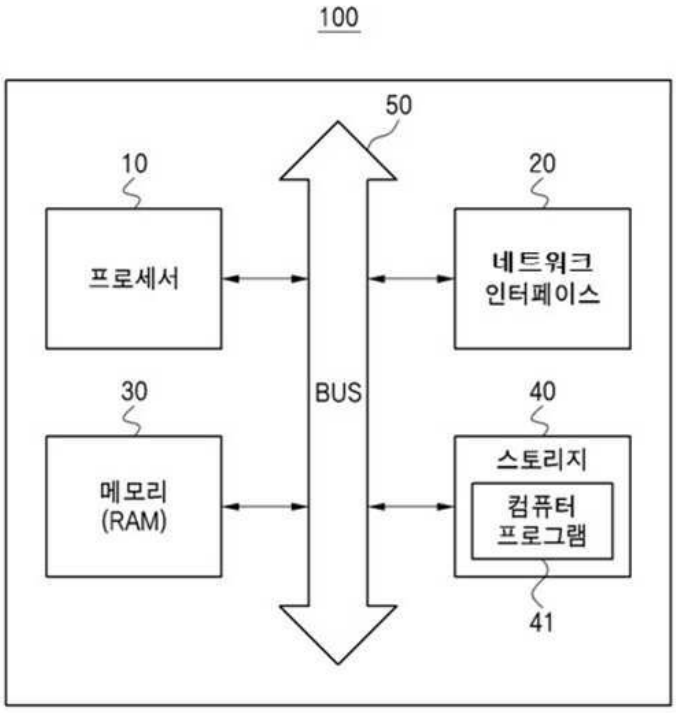
41: 컴퓨터 프로그램

50: 정보 버스

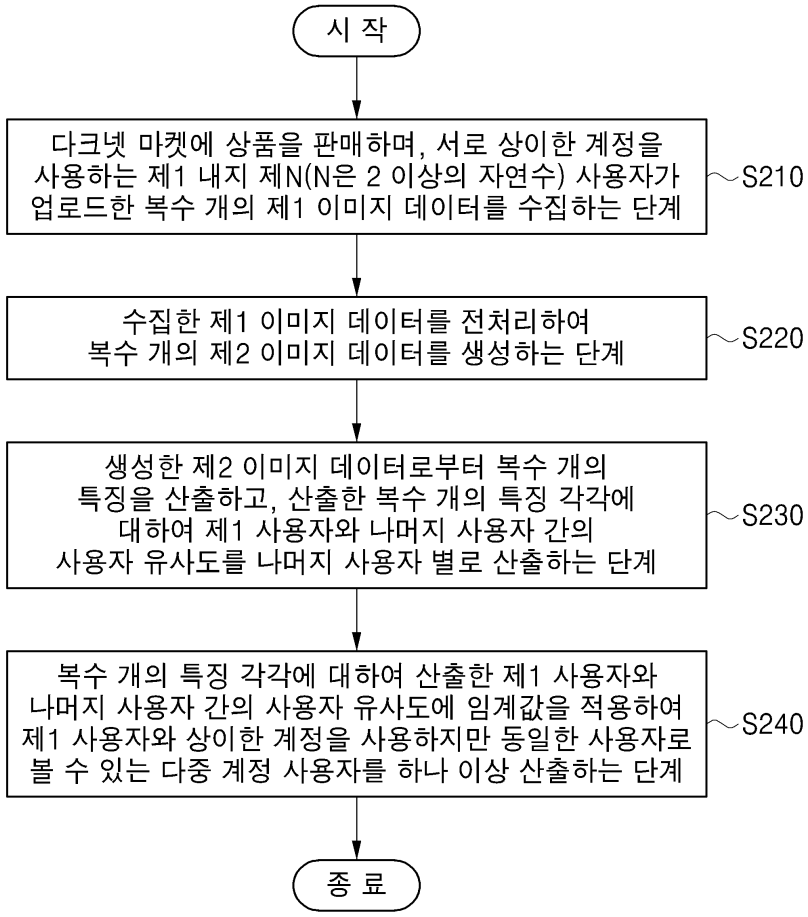
100: 다크넷 마켓 사용자의 다중 계정을 탐지하는 장치

도면

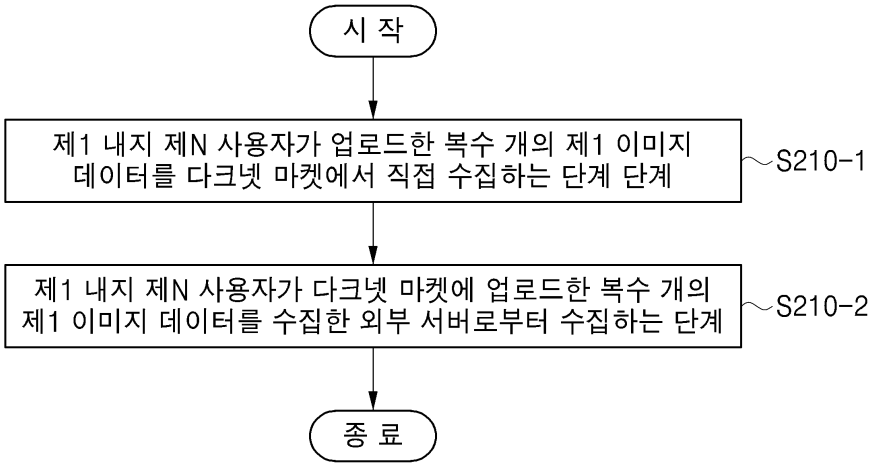
도면1



도면2


시 작
다크넷 마켓에 상품을 판매하며, 서로 상이한 계정을 사용하는 제1 내지 제N(N은 2 이상의 자연수) 사용자가 업로드한 복수 개의 제1 이미지 데이터를 수집하는 단계
S210
수집한 제1 이미지 데이터를 전처리하여 복수 개의 제2 이미지 데이터를 생성하는 단계
S220
생성한 제2 이미지 데이터로부터 복수 개의 특징을 산출하고, 산출한 복수 개의 특징 각각에 대하여 제1 사용자와 나머지 사용자 간의 사용자 유사도를 나머지 사용자 별로 산출하는 단계
S230
복수 개의 특징 각각에 대하여 산출한 제1 사용자와 나머지 사용자 간의 사용자 유사도에 임계값을 적용하여 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 다중 계정 사용자를 하나 이상 산출하는 단계
S240
종 료

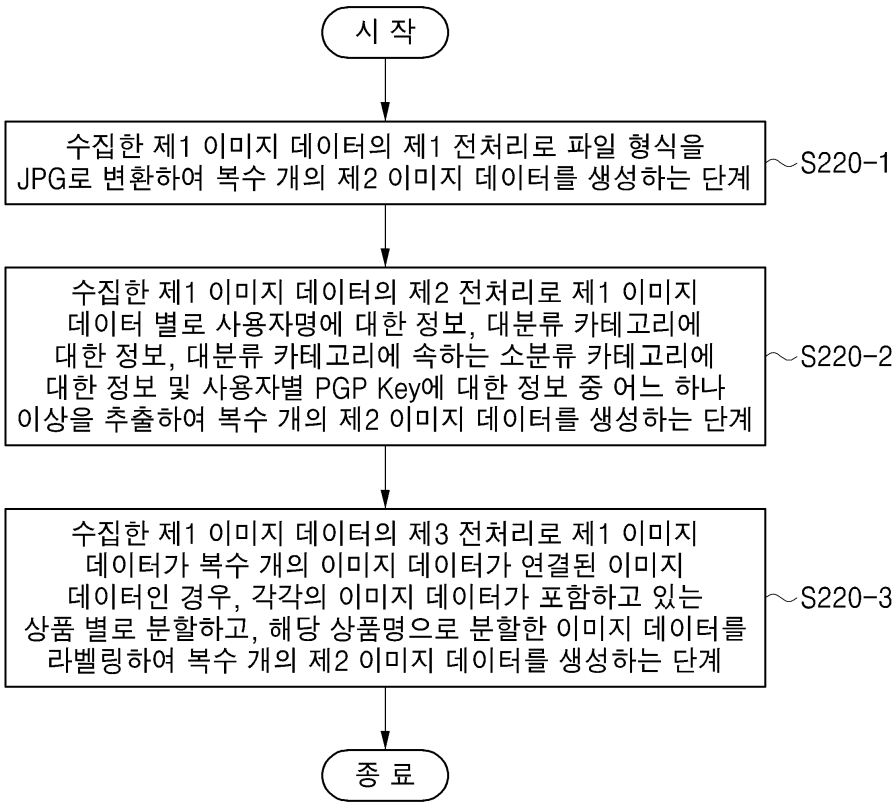
도면3


시 작
제1 내지 제N 사용자가 업로드한 복수 개의 제1 이미지 데이터를 다크넷 마켓에서 직접 수집하는 단계 단계
S210-1
제1 내지 제N 사용자가 다크넷 마켓에 업로드한 복수 개의 제1 이미지 데이터를 수집한 외부 서버로부터 수집하는 단계
S210-2
종 료

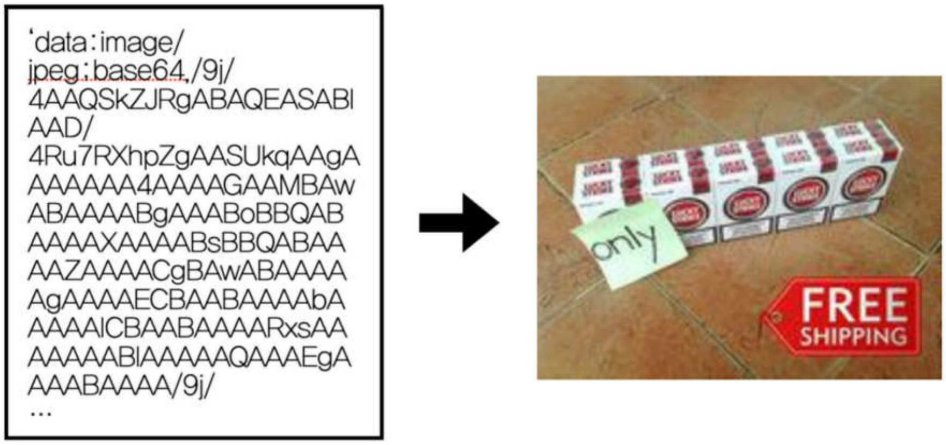
도면4

	Drug	Weapons	Digital goods	Electronics	Hacking	Others
Abraxas	O		O	O	O	O
Agora	O	O	O			O
Alphabay	O	O	O	O	O	O
Blackbank	O	O	O	O		O
Cloudnine	O		O			
Evolution	O	O	O	O	O	O
Silkroad2	O		O	O		O
Pandora	O		O			O

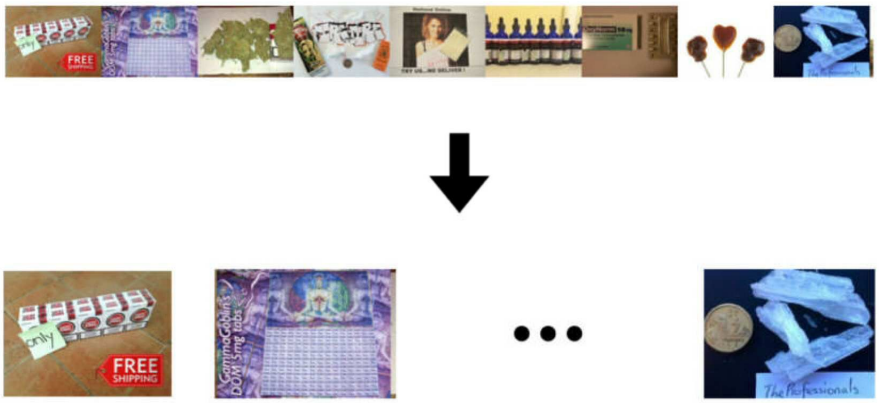
도면5



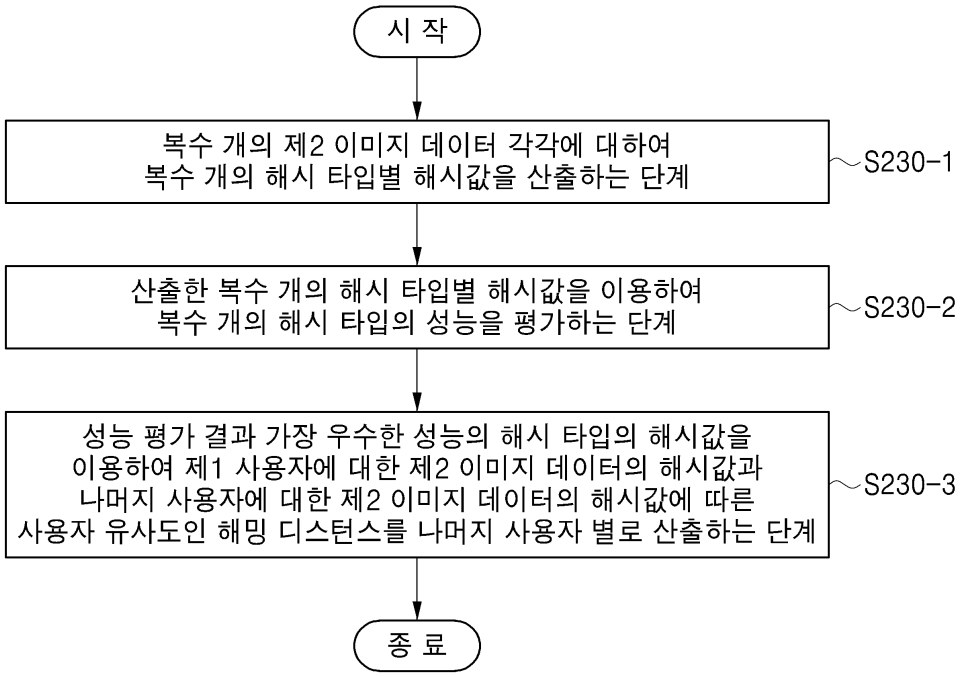
도면6


'
4AAQSkZJRgABAQEASABI
AAD/
4Ru7RXhpZgAASUkqAAgA
AAAAAA4AAAAGAAMBAw
ABAAAABgAAABoBBQAB
AAAAXAAAABsBBQABAA
AAZAAAACgBAwABAAAA
AgAAAAECBAABAAAAbA
AAAAICBAABAAAARxsAA
AAAAABIAAAAAQAAAEgA
AAABAAAA/9j/
...
only
FREE
SHIPPING

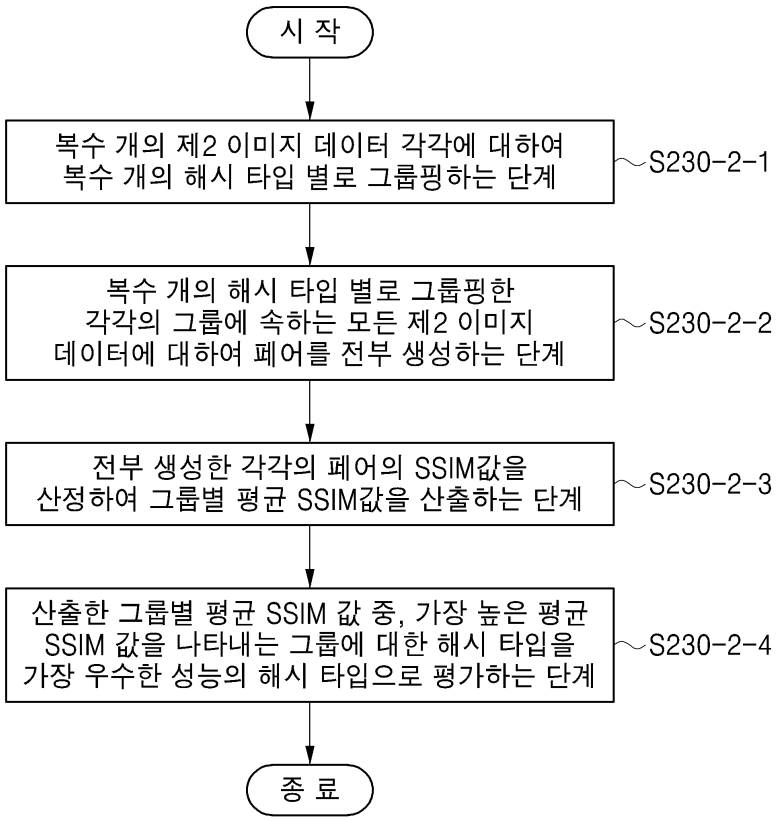
도면7


only
FREE
SHIPPING

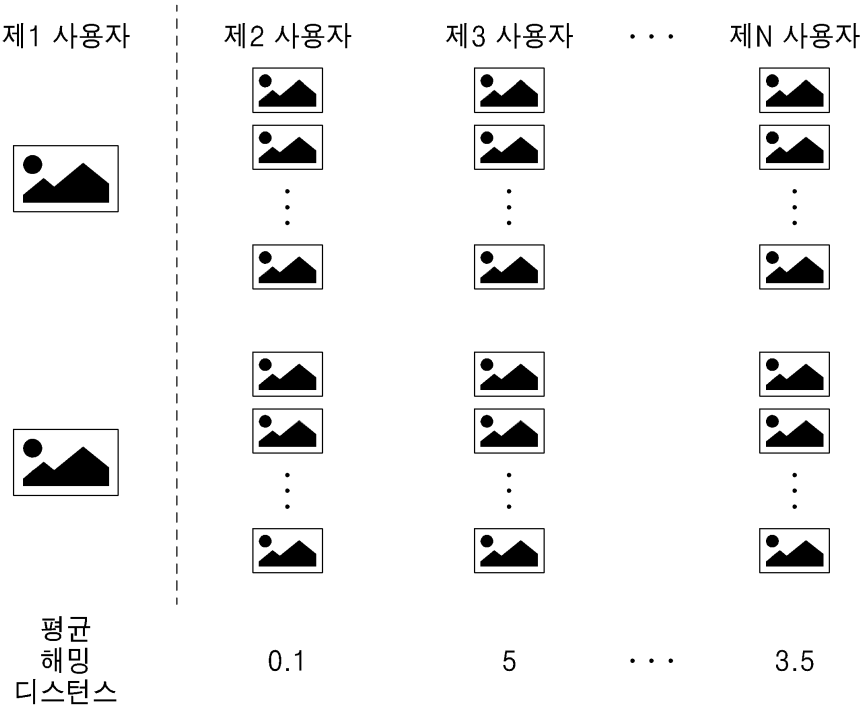
도면8


시 작
복수 개의 제2 이미지 데이터 각각에 대하여
복수 개의 해시 타입별 해시값을 산출하는 단계
S230-1
산출한 복수 개의 해시 타입별 해시값을 이용하여
복수 개의 해시 타입의 성능을 평가하는 단계
S230-2
성능 평가 결과 가장 우수한 성능의 해시 타입의 해시값을
이용하여 제1 사용자에 대한 제2 이미지 데이터의 해시값과
나머지 사용자에 대한 제2 이미지 데이터의 해시값에 따른
사용자 유사도인 해밍 디스턴스를 나머지 사용자 별로 산출하는 단계
S230-3
종 료

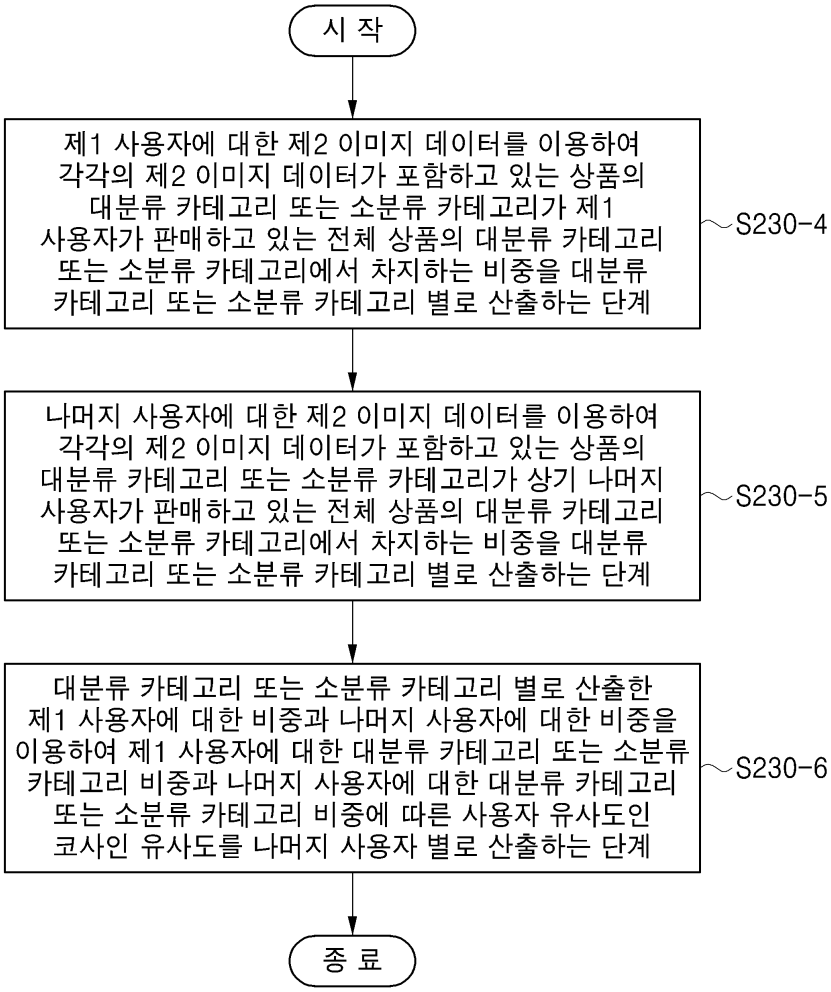
도면9


시 작
복수 개의 제2 이미지 데이터 각각에 대하여
복수 개의 해시 타입 별로 그룹핑하는 단계
S230-2-1
복수 개의 해시 타입 별로 그룹핑한
각각의 그룹에 속하는 모든 제2 이미지
데이터에 대하여 페어를 전부 생성하는 단계
S230-2-2
전부 생성한 각각의 페어의 SSIM값을
산정하여 그룹별 평균 SSIM값을 산출하는 단계
S230-2-3
산출한 그룹별 평균 SSIM 값 중, 가장 높은 평균
SSIM 값을 나타내는 그룹에 대한 해시 타입을
가장 우수한 성능의 해시 타입으로 평가하는 단계
S230-2-4
종 료

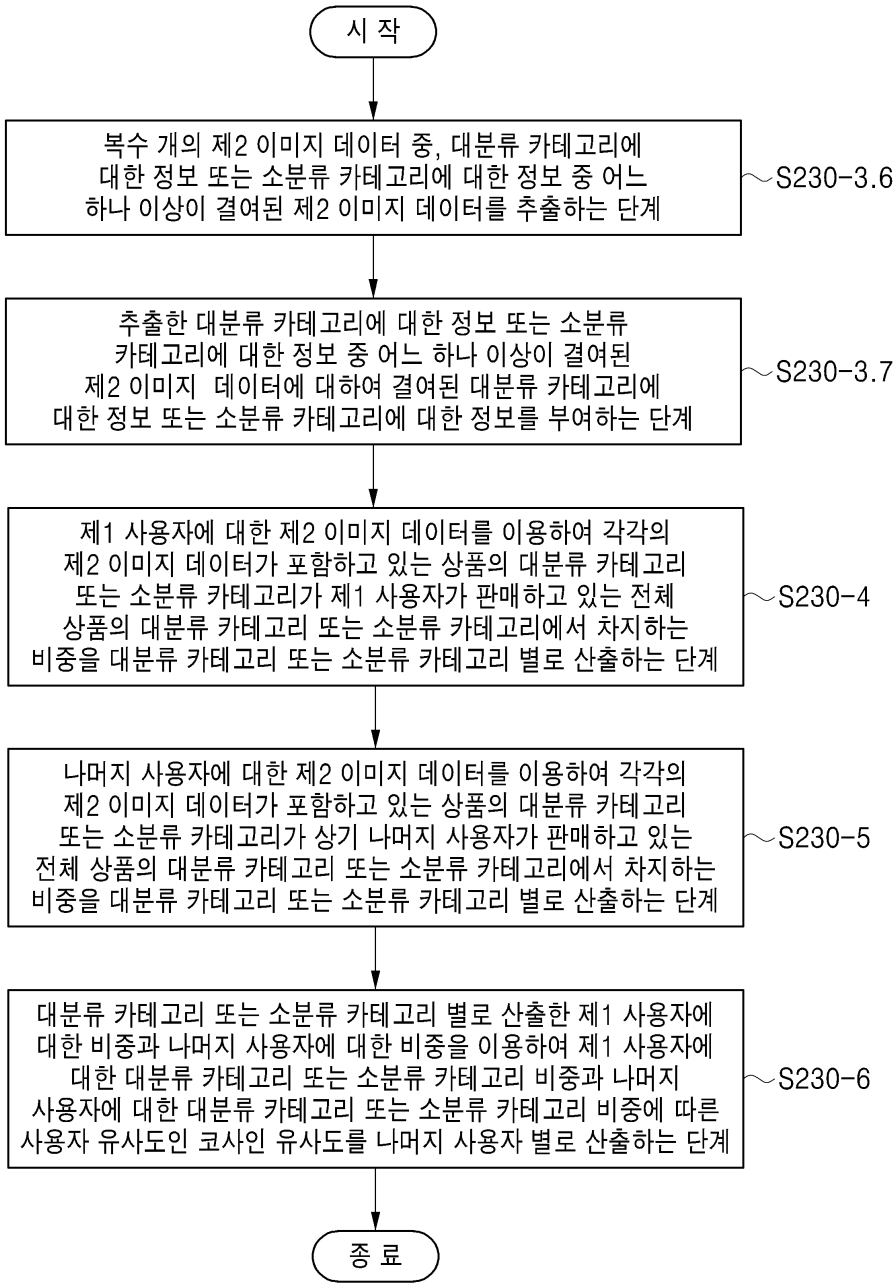
도면10



도면11


시 작
제1 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 제1 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계
S230-4
나머지 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 나머지 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계
S230-5
대분류 카테고리 또는 소분류 카테고리 별로 산출한 제1 사용자에 대한 비중과 나머지 사용자에 대한 비중을 이용하여 제1 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중과 나머지 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중에 따른 사용자 유사도인 코사인 유사도를 나머지 사용자 별로 산출하는 단계
S230-6
종 료

도면12


시 작
복수 개의 제2 이미지 데이터 중, 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터를 추출하는 단계
S230-3.6
추출한 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보 중 어느 하나 이상이 결여된 제2 이미지 데이터에 대하여 결여된 대분류 카테고리에 대한 정보 또는 소분류 카테고리에 대한 정보를 부여하는 단계
S230-3.7
제1 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 제1 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계
S230-4
나머지 사용자에 대한 제2 이미지 데이터를 이용하여 각각의 제2 이미지 데이터가 포함하고 있는 상품의 대분류 카테고리 또는 소분류 카테고리가 상기 나머지 사용자가 판매하고 있는 전체 상품의 대분류 카테고리 또는 소분류 카테고리에서 차지하는 비중을 대분류 카테고리 또는 소분류 카테고리 별로 산출하는 단계
S230-5
대분류 카테고리 또는 소분류 카테고리 별로 산출한 제1 사용자에 대한 비중과 나머지 사용자에 대한 비중을 이용하여 제1 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중과 나머지 사용자에 대한 대분류 카테고리 또는 소분류 카테고리 비중에 따른 사용자 유사도인 코사인 유사도를 나머지 사용자 별로 산출하는 단계
S230-6
종 료

도면13

대분류 카테고리	
1	apparel
2	art
3	books
4	chemicals
5	computer
6	custom
7	data
8	digital
9	drug
10	electronics
11	erotica
12	forgeries
13	hardware
14	jewelry
15	lab
16	listings
17	lotteries
8	money
19	other
20	packaging
21	services
22	writing

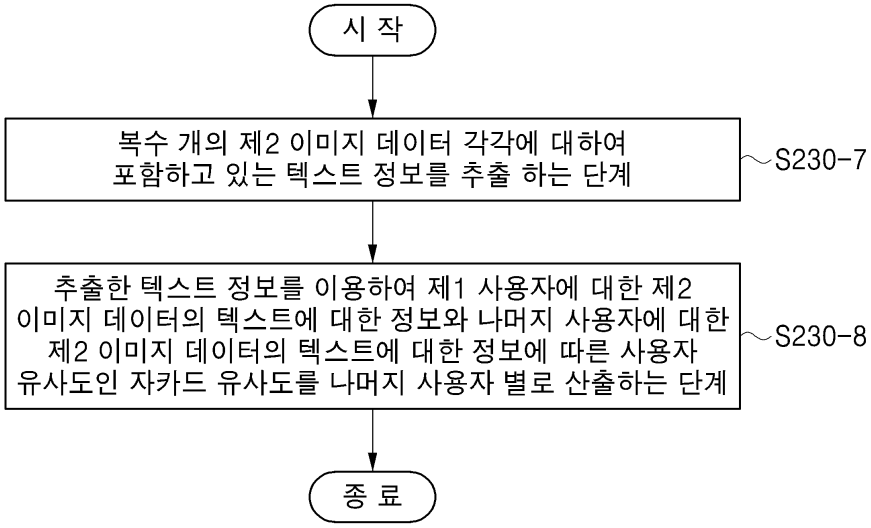
소분류 카테고리	
1	barbiturates
2	benzos
3	cannabis
4	dissociatives
5	ecstasy
6	grinders
7	opioids
8	other
9	paper
10	paraphernalia
11	psychedelics
12	precursors
13	prescription
14	RCs
15	scales
16	smoked
17	stashes
8	steriodpeds
19	sitmulants
20	weight loss

도면14

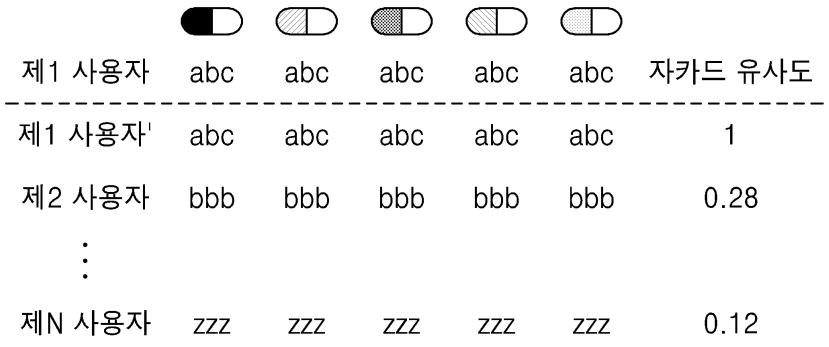


제1 사용자	0.7	0.1	0.05	0	0.15	코사인 유사도
제1 사용자'	0.7	0.1	0.05	0	0.15	1
제2 사용자	0.2	0.1	0.7	0	0	0.56
⋮						
제N 사용자	0.1	0.7	0	0.2	0	0.31

도면15

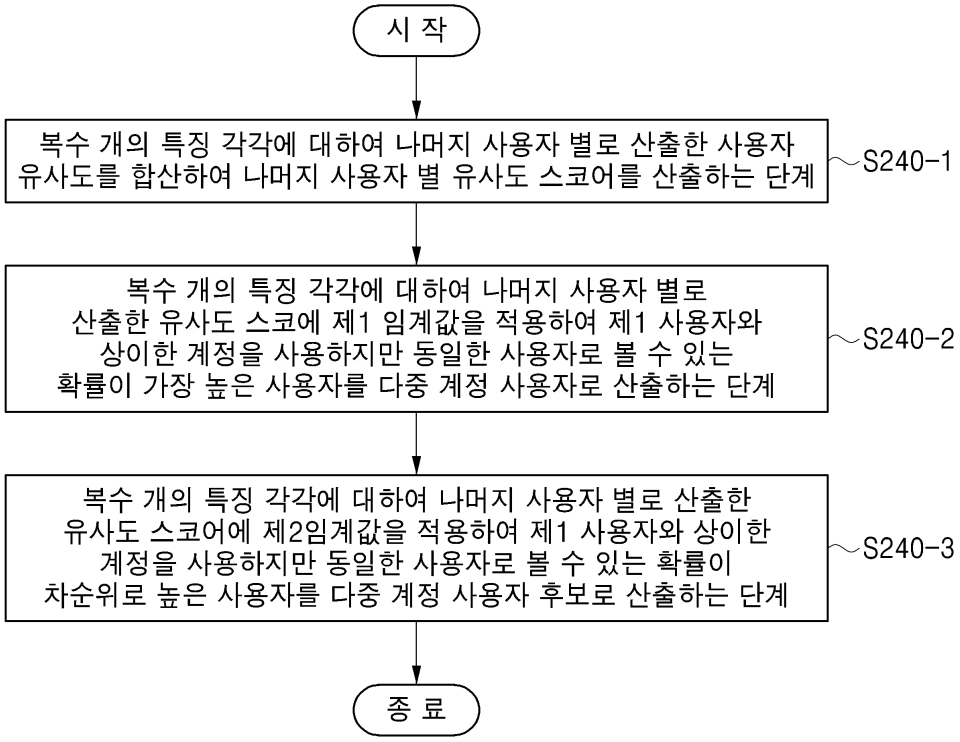


도면16

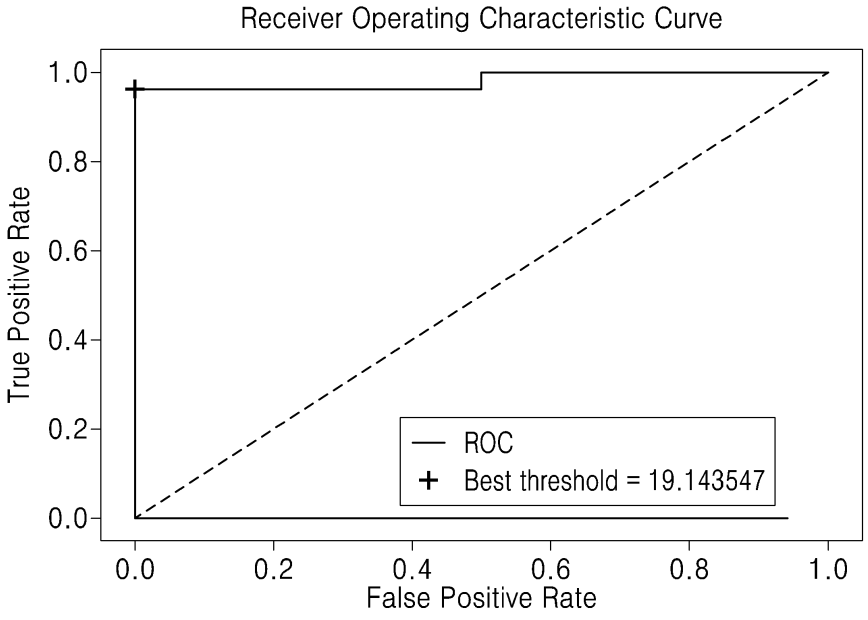


제1 사용자	abc	abc	abc	abc	abc	자카드 유사도
제1 사용자'	abc	abc	abc	abc	abc	1
제2 사용자	bbb	bbb	bbb	bbb	bbb	0.28
⋮						
제N 사용자	zzz	zzz	zzz	zzz	zzz	0.12

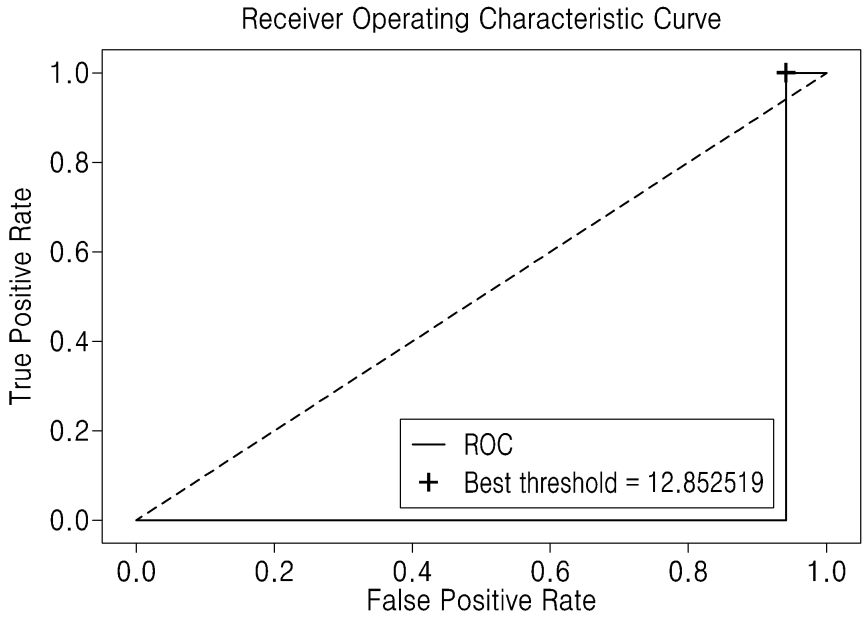
도면17


시 작
복수 개의 특징 각각에 대하여 나머지 사용자 별로 산출한 사용자 유사도를 합산하여 나머지 사용자 별 유사도 스코어를 산출하는 단계
S240-1
복수 개의 특징 각각에 대하여 나머지 사용자 별로 산출한 유사도 스코에 제1 임계값을 적용하여 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 가장 높은 사용자를 다중 계정 사용자로 산출하는 단계
S240-2
복수 개의 특징 각각에 대하여 나머지 사용자 별로 산출한 유사도 스코어에 제2임계값을 적용하여 제1 사용자와 상이한 계정을 사용하지만 동일한 사용자로 볼 수 있는 확률이 차순위로 높은 사용자를 다중 계정 사용자 후보로 산출하는 단계
S240-3
종 료

도면18


Receiver Operating Characteristic Curve
True Positive Rate
False Positive Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
ROC
Best threshold = 19.143547

도면19



도면20

모델	Silkroad2		Agora		Silkroad2&Agora	
	Accuracy	F-score	Accuracy	F-score	Accuracy	F-score
대분류 카테고리 정보	0.17	0.140	0.177	0.142	0.212	0.183
소분류 카테고리 정보	0.39	0.291	0.386	0.283	0.427	0.321
대/소분류 카테고리 정보	0.395	.293	0.38	0.279	0.433	0.329
텍스트 정보	0.82	0.772	0.57	0.469	0.623	0.536
해시값 정보	0.96	0.949	0.684	0.657	0.768	0.724
대/소분류 카테고리 정보 +해시값 정보	0.975	0.965	0.835	0.796	0.88	0.84
본 발명의 제2 실시 예	**0.985**	0.978	**0.861**	0.823	**0.908**	0.872

도면21

마켓	Candidate vendor	Precision	Recall	F-score	Accuracy
Silkroad2	–	0.976	0.969	0.973	0.948
	O	0.953	1.0	0.976	**0.953**
Agora	–	0.847	0.956	0.898	0.82
	O	0.953	1.0	0.921	**0.855**
Silkroad2 &Agora	–	0.872	0.972	0.92	0.853
	O	0.897	1.0	0.946	**0.897**

도면22

마켓	Pairs	종래 기술		본 발명	
		Match	Precision	Match	Precision
Silkroad2	277	275 (99%)	0.993	253 (91%)	**0.951**
Agora	242	214 (88%)	0.911	134 (55%)	**0.77**
Silkroad2 &Agora	519	470 (91%)	0.94	307 (59%)	**0.841**