



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년06월02일

(11) 등록번호 10-2113969

(24) 등록일자 2020년05월15일

(51) 국제특허분류(Int. Cl.)
G06F 16/55 (2019.01) *G06F 16/583* (2019.01)
G06K 9/62 (2006.01) *G06N 3/08* (2006.01)
G06Q 50/14 (2012.01)

(52) CPC특허분류
G06F 16/55 (2019.01)
G06F 16/583 (2019.01)

(21) 출원번호 10-2020-0011747

(22) 출원일자 2020년01월31일

심사청구일자 2020년01월31일

(30) 우선권주장
 1020190134888 2019년10월28일 대한민국(KR)

(56) 선행기술조사문헌
 KR1020190107613 A
 (뒷면에 계속)

(73) 특허권자
 이화여자대학교 산학협력단
 서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)

(72) 발명자
 강영욱
 서울특별시 강남구 압구정로 151, 112동 304호 (신현대아파트)

조나혜
 서울특별시 서대문구 이화여대5길 23, 508호 (그레이크하우스)

(74) 대리인
 윤귀상

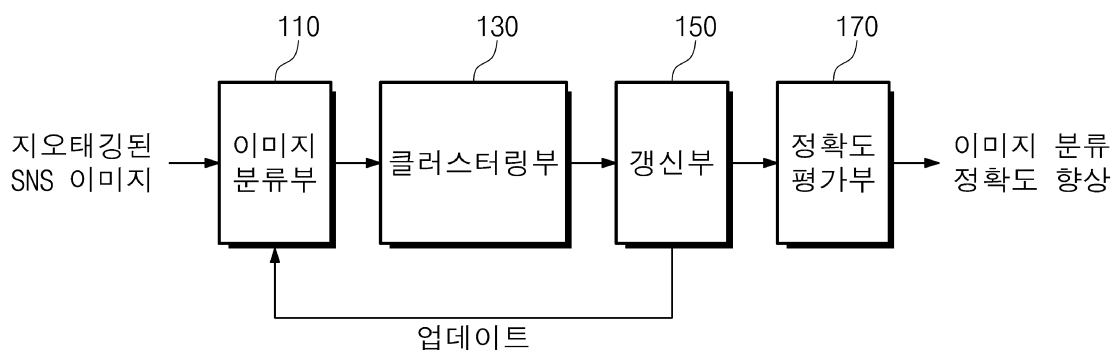
전체 청구항 수 : 총 9 항

심사관 : 박성수

(54) 발명의 명칭 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법, 이를 수행하기 위한 기록 매체 및 장치

(57) 요약

공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법은, 지오태깅되어 SNS에 게시된 이미지 데이터를 획득하여, 관광객의 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류하는 단계; 분류된 이미지 데이터에 라벨링된 카테고리 정보 (뒷면에 계속)

대 표 도 - 도1

를 이용하여, 밀도 기반 클러스터링(Density-based spatial clustering of applications with noise, DBSCAN)을 통해 각 카테고리 세부 항목별 클러스터를 추출하는 단계; 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하지 않는 경우, 상기 위치(Where) 정보를 기초로 상기 활동 정보(What)를 갱신하는 단계; 갱신된 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 재분류하는 단계; 및 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 단계;를 포함한다. 이에 따라, 카테고리가 라벨링된 사진 데이터의 DBSCAN을 통해 위치 정보를 확인하고, 오류가 있는 경우 위치 정보를 갱신하여 관광 목적 이미지의 분류 정확도를 향상시킬 수 있다.

(52) CPC특허분류

G06F 16/5866 (2019.01)
G06K 9/6267 (2013.01)
G06N 3/08 (2013.01)
G06Q 50/14 (2013.01)

(56) 선행기술조사문헌

JP2019164788 A
JP2011008635 A
KR1020170025454 A
KR101872733 B1
KR1020130110279 A
KR1020170107865 A
KR1020180130937 A
소셜 네트워크 데이터를 활용한 서울방문 관광객의 선호 관광지 시공간 특성 분석. 강영옥 외 4명. 서울도시연구 제20권 제1호. 2019.03. 논문 81-96쪽. 지능형 관광 서비스를 위한 관광 사진 분류체계 개발. 한국지도학회지 제19권 제3호 (p.87-101). 2019.12.
소셜 네트워크 데이터를 활용한 서울방문 관광객의 관광 특성 분석 -플리커 데이터를 중심으로-. 이화여자대학교. 김나연 외 3명. 2018.10.12.
Identifying Staying Places with Global Positioning System Movement Data Using 3D Density-based Spatial Clustering of Applications with Noise. 강영옥, 조나혜. 2019.10.31.
2019년 공간정보 국제컨퍼런스 인공지능과 디지털 트윈으로 여는 공간정보사회. 2019.08.08. 서울 코엑스. CNN딥러닝을 이용한 외국인 관광객의 서울 이미지 분석. 이화여자대학교 강영옥 교수.
CNN 딥러닝을 이용한 서울방문 관광객의 서울 이미지 분석. 강영옥 외 4명. 한국지형공간정보학회. 2018.11.
텍스트마이닝과 딥러닝 기술을 활용한 외국인 관광객의 국내 지역별 이미지 비교. 대한공간정보학회. 이화여자대학교. 강영옥 외 5명. 2019.11.
경험적 모델과 머신러닝 기법을 활용한 SNS 사용자 분류방법 비교: 플리커 데이터의 관광객 분류방법. 대한공간정보학회. 이화여자대학교. 강영옥 외 4명. 2019.07.
지오태깅된 플리커데이터의 사용자 분류 방법 연구: 관광객 추정방법. 대한공간정보학회. 이화여자대학교. 강영옥 외 4명. 2019.05.

이 발명을 지원한 국가연구개발사업

과제고유번호	2-2019-0601-000-1
부처명	국토교통부
연구관리전문기관	국토교통과학기술진흥원
연구사업명	국토교통기술촉진연구(R&D)
연구과제명	지능형 장소기반 관광 서비스 제공을 위한 합성곱 신경망을 활용한(Convolutional Neural Network, CNN) 이미지 분류모형 개발
기 여 율	1/1
주관기관	이화여자대학교 산학협력단
연구기간	2019.04.15 ~ 2019.12.31

명세서

청구범위

청구항 1

이미지 분류부에서 지오타깅(Geo tagging)되어 SNS에 게시된 이미지 데이터를 획득하여, 관광객의 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류하는 단계;

클러스터링부에서 분류된 이미지 데이터에 라벨링된 카테고리 정보를 이용하여, 밀도 기반 클러스터링(Density-based spatial clustering of applications with noise, DBSCAN)을 통해 각 카테고리 세부 항목별 클러스터를 추출하는 단계;

갱신부에서 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하지 않는 경우, 상기 위치(Where) 정보를 기초로 상기 활동 정보(What)를 갱신하는 단계; 및

정확도 평가부에서 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 단계;를 포함하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법.

청구항 2

삭제

청구항 3

제1항에 있어서, 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 단계는,

상기 정확도 평가부에서 건물, POI, 수치지도 데이터 중 적어도 하나의 정제된 값을 상기 실제 위치의 정답 셋으로 정의하는 단계를 포함하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법.

청구항 4

제1항에 있어서,

상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하는 경우,

상기 정확도 평가부에서 상기 활동 정보(What)에 대한 정확한 이미지 데이터 분류로 판단하는 단계를 더 포함하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법.

청구항 5

제1항에 있어서, 상기 클러스터를 추출하는 단계는,

상기 클러스터링부에서 카테고리의 항목에 따라 밀도 기반 클러스터링의 파라미터인 Eps 및 minpts를 다르게 설정하는 단계를 포함하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법.

청구항 6

제1항, 제3항 내지 제5항 중 어느 하나의 항에 있어서,

공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법을 수행하기 위한 컴퓨터 프로그램이 기록된 컴퓨터로 판독 가능한 저장 매체.

청구항 7

지오태깅(Geo tagging)되어 SNS에 게시된 이미지 데이터를 획득하여, 관광객의 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류하는 이미지 분류부;

분류된 이미지 데이터에 라벨링된 카테고리 정보를 이용하여, 밀도 기반 클러스터링(Density-based spatial clustering of applications with noise, DBSCAN)을 통해 각 카테고리 세부 항목별 클러스터를 추출하는 클러스터링부;

상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하지 않는 경우, 상기 위치(Where) 정보를 기초로 상기 활동 정보(What)를 갱신하고, 갱신된 활동 정보(What)를 상기 이미지 분류부로 전달하는 갱신부; 및

상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 정확도 평가부;를 포함하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치.

청구항 8

삭제

청구항 9

제7항에 있어서, 상기 정확도 평가부는,

건물, POI, 수치지도 데이터 중 적어도 하나의 정제된 값을 상기 실제 위치의 정답 셋으로 정의하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치.

청구항 10

제7항에 있어서, 상기 갱신부는,

상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하는 경우, 상기 활동 정보(What)에 대한 정확한 이미지 데이터 분류로 판단하고, 상기 정확도 평가부로 전달하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치.

청구항 11

제7항에 있어서, 상기 클러스터링부는,

카테고리의 항목에 따라 밀도 기반 클러스터링의 파라미터인 Eps 및 minpts를 다르게 설정하는, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치.

발명의 설명

기술 분야

본 발명은 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법, 이를 수행

하기 위한 기록 매체 및 장치에 관한 것으로서, 더욱 상세하게는 대한민국을 방문한 관광객이 SNS에 게시한 사진을 합성곱 신경망을 통해 자동 분류한 경우에 관광 목적 이미지 분류의 정확도를 평가하고 이를 향상시키는 기술에 관한 것이다.

배경 기술

- [0003] CNN은 인공 신경망 중에서도 이미지 인식 문제를 해결하기 위해 특화된 딥 러닝 학습 모델 중 하나이며, 감독 학습(Supervised Learning)으로 모델을 훈련시킨다.
- [0004] CNN 모델 가운데에서도 구글에서 개발한 GoogleNet 구조의 Inception v3 모델은 ImageNet 웹사이트 (<http://image-net.org/>)에서 제공하는 총 14,197,122장의 사진을 27개의 상위 카테고리, 1,000개의 하위 카테고리로 분류해놓은 이미지 데이터를 훈련시켜 개발한 모델이다. Inception v3 모델은 소스코드가 널리 공개되어 있으며 높은 정확도 때문에 여러 연구에서 활용되고 있다.
- [0005] 그러나, Inception v3 모델을 사용하여 사진을 분류하면 ImageNet의 사진 분류카테고리를 그대로 활용하기 때문에 관광 목적에 맞게 분류되지 않으며, 일정 영역은 지나치게 상세하거나(예를 들어, 버섯의 종류, 파충류의 종류, 곰팡이 종류, 꽃 종류, 개 종류, 새 종류, 양서류 종류 등), 한국관광의 특성을 나타내지 못하는(예를 들어, 최근 관광객들이 한국에 와서 한복을 입고 사진을 찍는 모습을 볼 수 있는데 ImageNet 카테고리에는 한복이 없기 때문에 이런 사진들은 'kimono', 'hoopskirt', 'crinoline' 등의 카테고리로 분류됨) 문제점이 있다.
- [0006] 또한, 한국에서 나타나는 고유한 랜드마크(예를 들어, 동대문 디자인 플라자, 남산타워, 연등 등) 등을 분류해 내지 못하며, 훈련 데이터 셋이 해외에서 구축되어 저층 다세대 주택을 'prison'으로 분류하거나, 상점건물을 'theater'로 구분(Kim et al., 2019)하는 등의 오류가 발생한다.
- [0007] 따라서, 종래 기술을 활용하여 이미지를 분류하는 경우, 한국 관광의 특성을 제대로 반영하지 못하고, 잘못된 분류를 만들게 되므로, 국내의 관광 서비스를 위한 보다 정확한 이미지 분류 방법 및 분류 방법의 정확도 평가 개발이 필요한 실정이다. 또한, CNN을 통해 분류된 데이터가 정확하게 분류 되었는지를 확인하기 위해서 사용되는 수동 라벨링은 비효율적이며, 육안으로도 분류하기 어려운 한계가 있다.

선행기술문헌

특허문헌

- [0009] (특허문헌 0001) KR 10-2018-0101971 A
(특허문헌 0002) KR 10-2019-0047200 A
(특허문헌 0003) KR 10-2019-0072799 A

비특허문헌

- [0010] (비특허문헌 0001) Inception V3 모델을 활용한 관광사진 분류 및 정확도 평가, 조나혜 외, 2019년 대한공간정보학회 추계학술대회, 2019.11

발명의 내용

해결하려는 과제

- [0011] 이에, 본 발명의 기술적 과제는 이러한 점에서 착안된 것으로 본 발명의 목적은 사진 카테고리를 분류 및 유형화하여 지능형 장초기반 관광 서비스를 제공하기 위한 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법을 제공하는 것이다.
- [0012] 본 발명의 다른 목적은 상기 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법을 수행하기 위한 컴퓨터 프로그램이 기록된 기록 매체를 제공하는 것이다.
- [0013] 본 발명의 또 다른 목적은 상기 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도

향상 방법을 수행하기 위한 장치를 제공하는 것이다.

과제의 해결 수단

- [0015] 상기한 본 발명의 목적을 실현하기 위한 일 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법은, 지오타깅(Geo tagging)되어 SNS에 게시된 이미지 데이터를 획득하여, 관광객의 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류하는 단계; 분류된 이미지 데이터에 라벨링된 카테고리 정보를 이용하여, 밀도 기반 클러스터링(Density-based spatial clustering of applications with noise, DBSCAN)을 통해 각 카테고리 세부 항목별 클러스터를 추출하는 단계; 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하지 않는 경우, 상기 위치(Where) 정보를 기초로 상기 활동 정보(What)를 갱신하는 단계; 갱신된 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 재분류하는 단계; 및 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 단계;를 포함한다.
- [0016] 본 발명의 실시예에서, 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 단계는, 상기 활동 정보(What)의 혼동행렬 값과 상기 갱신된 활동 정보(What)의 혼동행렬 값을 비교 분석하는 단계를 포함할 수 있다.
- [0017] 본 발명의 실시예에서, 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 단계는, 건물, POI, 수치지도 데이터 중 적어도 하나의 정제된 값을 상기 실제 위치의 정답 셋으로 정의하는 단계를 포함할 수 있다.
- [0018] 본 발명의 실시예에서, 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하는 경우, 상기 활동 정보(What)에 대한 정확한 이미지 데이터 분류로 판단하는 단계를 더 포함할 수 있다.
- [0019] 본 발명의 실시예에서, 상기 클러스터를 추출하는 단계는, 카테고리의 항목에 따라 밀도 기반 클러스터링의 파라미터인 Eps 및 minpts를 다르게 설정하는 단계를 포함할 수 있다.
- [0020] 상기한 본 발명의 다른 목적을 실현하기 위한 일 실시예에 따른 컴퓨터로 판독 가능한 저장 매체에는, 상기 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법을 수행하기 위한 컴퓨터 프로그램이 기록되어 있다.
- [0021] 상기한 본 발명의 또 다른 목적을 실현하기 위한 일 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치는, 지오타깅(Geo tagging)되어 SNS에 게시된 이미지 데이터를 획득하여, 관광객의 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류하는 이미지 분류부; 분류된 이미지 데이터에 라벨링된 카테고리 정보를 이용하여, 밀도 기반 클러스터링(Density-based spatial clustering of applications with noise, DBSCAN)을 통해 각 카테고리 세부 항목별 클러스터를 추출하는 클러스터링부; 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하지 않는 경우, 상기 위치(Where) 정보를 기초로 상기 활동 정보(What)를 갱신하고, 갱신된 활동 정보(What)를 상기 이미지 분류부로 전달하는 갱신부; 및 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정하는 정확도 평가부;를 포함한다.
- [0022] 본 발명의 실시예에서, 상기 정확도 평가부는, 상기 활동 정보(What)의 혼동행렬 값과 상기 갱신된 활동 정보(What)의 혼동행렬 값을 비교 분석할 수 있다.
- [0023] 본 발명의 실시예에서, 상기 정확도 평가부는, 건물, POI, 수치지도 데이터 중 적어도 하나의 정제된 값을 상기 실제 위치의 정답 셋으로 정의할 수 있다.
- [0024] 본 발명의 실시예에서, 상기 갱신부는, 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하는 경우, 상기 활동 정보(What)에 대한 정확한 이미지 데이터 분류로 판단하고, 상기 정확도 평가부로 전달할 수 있다.
- [0025] 본 발명의 실시예에서, 상기 클러스터링부는, 카테고리의 항목에 따라 밀도 기반 클러스터링의 파라미터인 Eps 및 minpts를 다르게 설정할 수 있다.

발명의 효과

[0027] 이와 같은 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 분류 방법에 따르면, 카테고리가 라벨링된 사진 데이터의 DBSCAN을 통해 위치 정보를 확인하고, 오류가 있는 경우 위치 정보를 갱신하여 관광 목적 이미지의 분류 정확도를 향상시킬 수 있다.

도면의 간단한 설명

[0029] 도 1은 본 발명의 일 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치의 블록도이다.

도 2는 본 발명의 활동 정보의 분류 정확도 여부를 위치 정보를 통해 확인할 수 있는 일례를 설명하기 위한 도면이다.

도 3은 본 발명의 위치 정보를 활용한 이미지 분류 정확도 향상 및 라벨링(labeling) 업데이트의 일례를 보여주는 표이다.

도 4는 본 발명의 관광 목적 이미지 CNN 분류의 정확도 평가에 대한 일례를 보여주는 표이다.

도 5는 본 발명의 정확도 평가에서 사용하는 혼동행렬의 예시 표이다.

도 6은 본 발명의 정확도 평가에서 사용하는 What 정보에 대한 혼동행렬의 예시 표이다.

도 7은 본 발명의 정확도 평가에서 사용하는 Where 정보에 대한 혼동행렬의 예시 표이다.

도 8은 본 발명의 정확도 평가에서 사용하는 갱신된 What 정보에 대한 혼동행렬의 예시 표이다.

도 9는 본 발명의 일 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법의 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0030] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0031] 이하, 도면들을 참조하여 본 발명의 바람직한 실시예들을 보다 상세하게 설명하기로 한다.

[0032] 도 1은 본 발명의 일 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치의 블록도이다.

[0033] 본 발명에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치(10, 이하 장치)는 대한민국을 방문한 관광객이 SNS에 게시한 사진으로부터 수집한 이미지 데이터를 딥러닝 기술을 이용하여 분류한 경우에 분류의 정확도를 평가하고, 평가 결과 오류가 있는 경우 위치 정보를 갱신하여 관광 목적 이미지의 분류 정확도를 향상시키는 기술에 관한 것이다.

[0034] 도 1을 참조하면, 본 발명에 따른 장치(10)는 이미지 분류부(110), 클러스터링부(130), 갱신부(150) 및 정확도 평가부(170)를 포함한다.

[0035] 본 발명의 상기 장치(10)는 다른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상을 수행하기 위한 소프트웨어(애플리케이션)가 설치되어 실행될 수 있으며, 상기 이미지 분류부(110), 상기 클러스터링부(130), 상기 갱신부(150) 및 상기 정확도 평가부(170)의 구성은 상기 장치(10)에서 실행되는 상기 다른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상을 수행하기 위한 소프트웨어에 의해 제어될 수 있다.

- [0036] 상기 장치(10)는 별도의 단말이거나 또는 단말의 일부 모듈일 수 있다. 또한, 상기 이미지 분류부(110), 상기 클러스터링부(130), 상기 갱신부(150) 및 상기 정확도 평가부(170)의 구성은 통합 모듈로 형성되거나, 하나 이상의 모듈로 이루어 질 수 있다. 그러나, 이와 반대로 각 구성은 별도의 모듈로 이루어질 수도 있다.
- [0037] 상기 장치(10)는 이동성을 갖거나 고정될 수 있다. 상기 장치(10)는, 서버(server) 또는 엔진(engine) 형태일 수 있으며, 디바이스(device), 기구(apparatus), 단말(terminal), UE(user equipment), MS(mobile station), 무선기기(wireless device), 휴대기기(handheld device) 등 다른 용어로 불릴 수 있다.
- [0038] 상기 장치(10)는 운영체제(Operation System; OS), 즉 시스템을 기반으로 다양한 소프트웨어를 실행하거나 제작할 수 있다. 상기 운영체제는 소프트웨어가 장치의 하드웨어를 사용할 수 있도록 하기 위한 시스템 프로그램으로서, 안드로이드 OS, iOS, 윈도우 모바일 OS, 바다 OS, 심비안 OS, 블랙베리 OS 등 모바일 컴퓨터 운영체제 및 윈도우 계열, 리눅스 계열, 유닉스 계열, MAC, AIX, HP-UX 등 컴퓨터 운영체제를 모두 포함할 수 있다.
- [0039] 상기 이미지 분류부(110)는 지오타깅(Geo tagging)되어 SNS에 게시된 이미지 데이터를 획득하여, 관광객의 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류한다.
- [0040] 지오타깅(Geo tagging)된 SNS 사진 데이터의 경우 위경도의 위치 정보와 사진이 촬영된 시간 정보가 기록되어 있어, 상기 시공간적 정보를 이용하여 사진 분류에 활용한다. SNS는 플리커(Flickr), 인스타그램, 페이스북, 포스퀘어(Foursquare), 카카오토리 등 사진을 공유하는 모든 어플리케이션을 포함할 수 있다.
- [0041] 합성곱 신경망(Convolutional Neural Network, 이하 CNN) 모델은 인공 신경망 중에서도 이미지 인식 문제를 해결하기 위해 특화된 딥 러닝 학습 모델 중 하나이며 감독 학습(Supervised Learning)으로 모델을 훈련시킨다.
- [0042] 본 발명에서는 관광 활동 카테고리(208개)를 생성하고, 구글의 Inception v3의 아키텍처를 활용하여 이를 재훈련시킨 모델을 활용한다. 이를 위해 관광객이 게시한 사진 약 16만여장을 활용하여 각 카테고리 항목별 다양한 각도의 이미지가 포함될 수 있도록 하였고, 새롭게 구축한 관광 이미지 훈련 데이터 셋을 활용하여 Inception v3 모델을 재훈련하였다.
- [0043] 특히, 마지막 Softmax layer를 fine-tuning하여 출력 개수를 변형하여 새롭게 수립한 관광 이미지 분류 체계의 카테고리 개수(208개)로 설정될 수 있도록 하였다.
- [0044] 본 발명의 일 실시예에서 제안하는 관광 목적의 분류 체계는 아래의 표 1과 같이 대분류 13개, 세분류 208개로 구성하였다. 그러나, 이는 일례에 불과하고 대분류와 세분류는 필요에 따라 삭제, 추가, 변경 등의 수정이 가능하다.

표 1

대분류	중분류	세분류
A.음식 /음료	A. Korean Food Best10	한국 대표음식(고기구이(AA01), 비빔밥(AA02), 치킨, (AA03) 삼계탕(AA04), 김밥(AA05), 떡볶이(AA06), 전 및 해물파전(AA07), 김치볶음밥(AA08), 김치찌개 및 부대찌개(AA09))
	B. Meal/Street Food	식사류(AB01), 길거리 음식(AB02)
	C. Restaurant	음식점 내부(AC01)
	D. Beverage	커피 및 음료(AD01), 캔 및 플라스틱 통(AD02)
	E. Cafe	카페 내부(AE01)
	F. Liquors	주류(AF01)
	G. Bakery/ Dessert	베이커리(AG01), 아이스크림(AG02), 사탕 및 초콜릿(AG03)
	H. Fruits/ Vegetables	과일(AH01), 채소(AH02)
	I. Food Truck	푸드트럭(AI01)
B.쇼핑	A. Traditional Marketplace	재래시장(BA01)
	B. Store/ Shop	쇼핑센터(BB01)
	C. Purchase	의류(BC01), 화장품(BC02), 가방/장갑/모자(BC03), 양말/신발(BC04), 액세서리(BC05), 문구 및 캐릭터(BC06)
	D. Packing/ Counter	장바구니 및 포장(BD01), 카운터/진열대(BD02)

C. 체험 / 활동	A. Leisure	수영장(CA01), 아이스링크장(CA02), 헬스(CA03), 썰매(CA04), 운동/스포츠(CA05), 오리배(CA06), 레일바이크(CA07)
	B. Traditional Experience	한복 체험(CB01), 바둑/장기/체스(CB02), 카드/화투(CB03)
	C. Relaxing	찜질방(CC01)
	D. Amusement Park	놀이공원(CD01), 아쿠아리움(CD02)
	E. Playground	놀이터(CE01)
	F. Entertainment	노래방(CF01), 오락실(CF02)
	G. Festival	각종 축제(CG01), 졸업식(CG02)
D. 문화 / 예술	A. Exhibition	무기(DA01), 전차/전투기(DA02), 밀랍인형 및 의복류(DA03), 지도/그림/병풍(DA04), 향아리 및 도자기(DA05), 관측기구(DA06), 장신구(DA07), 유물(DA08)
	B. Sculptures	조각상 및 각종 조형물(DB01), 박물관 및 전시관(DB02)
	C. Traditional Event	사물놀이(DC01), 근위대 및 취타대(DC02), 전통 공연(DC03)
	D. Performing Arts	공연장 입구(DD01), 공연장 내부(DD02)
	E. Sports	경기장(DE01), 경기 및 선수(DE02)
E. 도시 경관	A. Townscape	아파트 및 빌딩 경관(EA01), 주택 경관(EA02), 한옥(EA03)
	B. Wall Painting	벽화(EB01)
	C. Buddhist architecture	사찰(EC01), 불상(EC02), 불탑(EC03), 탑(EC04)
	D. Christian/ Catholic Architecture	교회/성당 내부(ED01), 서양식 건축물(ED02)
	E. Building Exterior	건물 외관(EE01)
	F. Building Interior	화단/계단(EF01), 건물 천장 및 바닥(EF02), 문/회전문(EF03)
	G. bridge/tunnel	교량/다리/가교(EG01), 터널(EG02)
	H. Road/ Street	도로/교차로/횡단보도(EH01), 거리/골목/길거리(EH02)
	I. Square/ Park	광장(EI01), 공원(EI02), 분수(EI03), 폭포(EI04)
	J. Stream	하천(EJ01)
	K. Tower	타워(EK01)
	L. Building Sculpture	거대 조형물(EL01)
	M. Nightscape	도시 야경(EM01), 네온사인/포장마차(EM02), 야간 고궁(EM03), 야간 불꽃놀이(EM04), 연등 및 등불축제(EM05)
	N. Cityscape Elements	공사장/운동장/주차장(EN01), 도시 시설물(EN02), 전봇대/가로등(EN03), 광고판/현수막(EN04), 홍보물/전단지(EN05)
F. 교통 수단	A. Car	차(FA01)
	B. Bus	버스(FB01)
	C. Metro	지하철(FC01)
	D. Train	기차(FD01)
	E. Plane	비행기(FE01)
	F. Bike/ Bicycle	오토바이/자전거(FF01)
	G. Ship	배(FG01)
	H. Vehicles' Equipment	차량 부속품 등(FH01)
G. 자연 환경	A. Weather	구름/하늘(GA01), 비/눈(GA02)
	B. River/ Lake	강/댐(GB01), 호수/연못(GB02)
	C. Mountain/ Valley	산/숲/숲길(GC01), 계곡/골짜기(GC02), 바위산(GC03), 동굴(GC04)
	D. Sea/ Seashore	바다/모래사장(GD01)
	E. Plain/ Farm	논/밭(GE01), 묘지/왕릉(GE02)
K. 동식물	A. Animal	강아지(KA01), 고양이(KA02), 동물(KA03), 물고기(KA04), 새(KA05)
	B. Vegetation	꽃(KB01), 나무(KB02), 풀/잎(KB03)
H. 인물	A. Person	사람/여러 사람(HA01), 집단/단체(HA02)
	B. Body	손/팔(HB01), 발/다리(HB02)
I. 궁 / 문화재 / 유적	A. Palace	궁/궁벽(IA01), 고궁 문/성 문(IA02), 문살/문창살(IA03), 지붕/서까래(IA04), 현판(IA05), 기념비(IA06)
	B. Relic	석상(IB01), 왕좌/옥좌(IB02), 전각(IB03), 종탑(IB04), 계단(IB05), 솟대(IB06), 성황당(IB07), 거대 석조물(IB08),
J. 사물	A. Things	전자기기 및 기계류(JA01), 가구/주방용품(JA02), 화장실 관련 물품(JA03), 사무용품(JA04), 악기류(JA05), 도구(JA06), 기타(JA07)

L.연수/교육/연구	A. Conference Hall	국제회의장(LA01), 세미나실/강의장(LA02),
	B. Classroom/ Reading room	교실(LB01), 독서실(LB02),
	C. Factory	공장(LC01)
M.숙박	A. Accommodation space	방(MA01), 화장실/욕실(MA02), 주방용품(MA03), 가구(MA04), 운동기구(MA05)
	B. Other space in the accommodation	프론트데스크(MB01)

- [0047] 상기 이미지 분류부(110)는 상기 CNN 모델의 학습 결과를 이용하여, 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류한다.
- [0048] 본 발명에서는 CNN 모델을 통하여 관광 이미지 분류를 학습하고, 이를 통해 관광 이미지를 자동으로 분류한다. 이미지 분류 단계에서 Softmax 층의 개수를 표 1과 같이 13개의 대분류 및 208개의 세분류로 구성된 관광 목적의 분류 항목으로 수정하여 데이터를 훈련하고, 이에 따라 분류된 이후 각 사진 데이터에 분류 label이 입력된다.
- [0049] 상기 클러스터링부(130)는 분류된 이미지 데이터에 라벨링된 카테고리 정보를 이용하여, 밀도 기반 클러스터링(Density-based spatial clustering of applications with noise, 이하 DBSCAN)을 통해 각 카테고리 세부 항목별 클러스터를 추출한다.
- [0050] 구체적으로, 본 발명에서는 각 카테고리 세부 항목별 DBSCAN의 머신러닝 기법을 활용하여 군집지역을 추출한다. 이는 '음식', '쇼핑', '체험' 등의 각각 카테고리별 주요 관광 클러스터를 발견하기 위함이다. 일 실시예로서, 카테고리의 항목별 특성에 따라 DBSCAN의 주요 파라미터인 eps와 minpts의 설정을 다르게 하여 다양한 스케일의 군집지역을 추출할 수 있다.
- [0051] 맥락(context) 정보들의 분류 유형에 따라 각각의 클러스터(cluster) 지역에 포함된 지점들과 군집에 속하지 않고 이례점(outlier)으로 분류되는 경우 맥락 정보의 값의 정확성 여부를 판단할 필요가 있다.
- [0052] 도 2는 본 발명의 활동 정보의 분류 정확도 여부를 위치 정보를 통해 확인할 수 있는 일례를 설명하기 위한 도면이다.
- [0053] 도 2를 참조하면, '창덕궁'이라는 POI(Point of Interest)의 면형 데이터와 '궁'으로 분류된 사진의 점형 데이터간 중첩분석을 수행하여, 이들의 관계가 면형 자료의 경계를 벗어난 점형 자료(disjoint)를 찾아내어 정확성 판단 후 자료 수정을 수행한다.
- [0054] '체험'의 '축제 및 이벤트', '도시 경관'의 '불꽃 축제', '야간 관광 활동' 등 각 카테고리별 DBSCAN을 통한 주요 클러스터(cluster)인 ROI(regions of interest)가 설정된 후, 클러스터 내에 포함되는 항목들은 해당 카테고리로 잘 분류된 사례로 판단한다.
- [0055] 반면, 이례점이 나타난 경우, 이례점에 라벨링된 활동 정보(What)의 분류 정확도 확인 및 수정 작업을 수행한다.
- [0056] 상기 갱신부(150)는 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하지 않는 경우, 상기 위치(Where) 정보를 기초로 상기 활동 정보(What)를 갱신한다. 또한, 상기 갱신부(150)는 갱신된 활동 정보(What)를 상기 이미지 분류부(110)로 전달한다.
- [0057] 사진의 위치 정보를 기반으로 DBSCAN에 의해 최종 추출된 위치 정보(Where)와 활동 정보(What)의 라벨링된 값을 비교 분석하여, 위치 정보와 활동 정보가 같은 유형일 경우 일치하는 것으로 처리한다.
- [0058] 그러나, 최종 추출된 위치 정보(Where)와 활동 정보(What)의 라벨링된 값의 차이가 있을 때 활동 정보가 이미지 분류시 오류가 있었던 것으로 판단하고 위치정보와 같은 값으로 수정한다.
- [0059] 도 3은 본 발명의 위치 정보를 활용한 이미지 분류 정확도 향상 및 라벨링(labeling) 업데이트의 일례를 보여주는 표이다. 도 4는 본 발명의 관광 목적 이미지 CNN 분류의 정확도 평가에 대한 일례를 보여주는 표이다.
- [0060] 예를 들어, 도 3과 같이 활동 정보(What)에는 "사찰 관람"으로 분류되어 입력되어 있지만, 위치 정보(Where)에는 "종로 창덕궁"임을 확인할 수 있다. 이러한 경우 이미지 분류 과정에서 문제가 있었음을 알 수 있기 때문에 CNN을 통한 이미지 분류를 재수행하여 정보를 갱신한다. 즉, 사찰 관람이 아닌 "궁의 서까래 관람"으로 내용이

업데이트 되었음을 확인할 수 있다.

- [0061] 상기 정확도 평가부(170)는 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정한다.
- [0062] 본 발명에서는 카테고리 세부 항목별로 DBSCAN 분석을 통한 클러스터에 의해 갱신된 값과 건물/poi 또는 수치지도 데이터에서 확인할 수 있는 실제 위치 간 비교 분석(중첩 분석)을 통해 각각 라벨링된 값의 정확도를 분석할 수 있다.
- [0063] 특히, CNN을 통해 분류된 항목에 대한 혼동행렬 값과 데이터 갱신이 된 항목에 대한 혼동행렬 값을 비교 분석하여 정확도 향상을 측정한다. 즉, 도 3과 같이 "What" 컬럼과 "Where" 컬럼 간 일치하지 않는 경우는 정보가 갱신되도록 하며, 같은 경우는 그대로 데이터 입력이 되도록 한다.
- [0064] 일 실시예로서, 건물, POI, 수치지도 데이터 등으로부터 정제된 값을 정답 셋으로 정의하고, 딥러닝의 CNN 모델을 활용했을 때의 "What"컬럼의 정보와 DBSCAN 분석에 의해 도출된 "Where" 컬럼의 정보 간 일치 여부를 파악한다.
- [0065] 일치하지 않는 경우에 한해 "Where" 컬럼을 참조하여 "What" 컬럼의 정보를 갱신한다. 예를 들어, 도 4의 사진 번호 2, 3, 7, 10은 What과 Where 간 불일치한 4개의 항목임을 알 수 있다(실제 값과 비교했을 때 What 컬럼(이미지 분류 결과)의 정확도가 60%에 그침).
- [0066] 반면, DBSCAN을 통해 도출된 "Where"를 참고하여 정보를 갱신한 경우의 분류 정확도는 90%로 향상되어 높은 정확도를 확인할 수 있다.
- [0067] 여기서, 실제값(poi/건물/수치지도)과 What 컬럼 간 직접적인 비교를 하는 경우, 실제값의 데이터가 정제되지 않은 형태로 아주 자세한 값(예를 들어, 실제 해당 위치의 이름, 지명 등-인정전, 회정당, 돈화문, 조계사 템플스테이, 대설법전 등)을 가지기 때문에 Where 컬럼의 경우 DBSCAN을 통해 클러스터의 대표적인 명칭(예를 들어, 창덕궁, 조계사 등)으로 정제된 값을 가질 수 있다. 따라서, 본 발명과 같이 What 컬럼과 Where 컬럼간 비교분석이 더 용이하며, 속도가 빠르다.
- [0068] 건물, poi, 수치지도의 경우 오픈된 데이터(예를 들어, NSDI의 오픈마켓, DATA.go.kr, 국토정보플랫폼 NGII.go.kr 등)를 활용할 수 있다.
- [0069] 도 5는 본 발명의 정확도 평가에서 사용하는 혼동행렬의 예시 표이다. 도 6은 본 발명의 정확도 평가에서 사용하는 What 정보에 대한 혼동행렬의 예시 표이다. 도 7은 본 발명의 정확도 평가에서 사용하는 Where 정보에 대한 혼동행렬의 예시 표이다. 도 8은 본 발명의 정확도 평가에서 사용하는 갱신된 What 정보에 대한 혼동행렬의 예시 표이다.
- [0070] 본 발명에서 정확도 평가 수식은 도 5의 예시와 같은 혼동행렬(confusion matrix) 방법을 활용한다.
- [0071] 정확도는 전체 분류 결과에서 실제값과 예측값이 일치하게 분류할 확률로서 아래의 수학적 식 1과 같이 정의된다.
- [0072] [수학적 식 1]
- [0073]
$$(TP+TN)/(TP+FN+FP+TN)$$
- [0074] 민감도는 실제값 중 실제값으로 올바르게 분류할 확률로서 아래의 수학적 식 2와 같이 정의된다.
- [0075] [수학적 식 2]
- [0076]
$$TP/(TP+FN)$$
- [0077] 정밀도는 예측값 중 실제값과 일치하게 분류할 확률로서 아래의 수학적 식 3과 같이 정의된다.
- [0078] [수학적 식 3]
- [0079]
$$TP/(TP+FP)$$
- [0080] 각 컬럼별(What, Where, 갱신된 What) 정확도 평가 과정의 예시는 도 6 내지 도 8과 같다. 도 7과 도 8을 참조하면, Where, 갱신된 What 값은 동일함을 확인할 수 있다.
- [0081] 이미지 분류의 정확도를 판단하여, 정확한 분류라고 판단되면, 이미지 데이터를 입력되도록 한다. 정확도의 판단 값은 필요에 따라 조절 가능하다.

- [0082] 이에 따라, 카테고리가 라벨링된 사진 데이터의 DBSCAN을 통해 위치 정보를 확인하고, 오류가 있는 경우 위치 정보를 갱신하여 관광 목적 이미지의 분류 정확도를 향상시킬 수 있다.
- [0084] 도 9는 본 발명의 일 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법의 흐름도이다.
- [0085] 본 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법은, 도 1의 장치(10)와 실질적으로 동일한 구성에서 진행될 수 있다. 따라서, 도 1의 장치(10)와 동일한 구성요소는 동일한 도면부호를 부여하고, 반복되는 설명은 생략한다.
- [0086] 또한, 본 실시예에 따른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법은, 다른 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상을 수행하기 위한 소프트웨어(애플리케이션)에 의해 실행될 수 있다.
- [0087] 도 9를 참조하면, 본 실시예에 따른 합성곱 신경망을 활용한 SNS 데이터의 관광 목적 이미지 분류 방법은, 지오태깅(Geo tagging)되어 SNS에 게시된 이미지 데이터를 획득하여, 관광객의 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류한다(단계 S10).
- [0088] 지오태깅(Geo tagging)된 SNS 사진 데이터의 경우 위경도의 위치 정보와 사진이 촬영된 시간 정보가 기록되어 있어, 상기 시공간적 정보를 이용하여 사진 분류에 활용한다. SNS는 플리커(Flickr), 인스타그램, 페이스북, 포스퀘어(Foursquare), 카카오토리 등 사진을 공유하는 모든 어플리케이션을 포함할 수 있다.
- [0089] 합성곱 신경망(Convolutional Neural Network, 이하 CNN) 모델은 인공 신경망 중에서도 이미지 인식 문제를 해결하기 위해 특화된 딥 러닝 학습 모델 중 하나이며 감독 학습(Supervised Learning)으로 모델을 훈련시킨다.
- [0090] 본 발명에서는 관광 활동 카테고리(208개)를 생성하고, 구글의 Inception v3의 아키텍처를 활용하여 이를 재훈련시킨 모델을 활용한다. 이를 위해 관광객이 게시한 사진 약 16만여장을 활용하여 각 카테고리 항목별 다양한 각도의 이미지가 포함될 수 있도록 하였고, 새롭게 구축한 관광 이미지 훈련 데이터 셋을 활용하여 Inception v3 모델을 재훈련하였다.
- [0091] 특히, 마지막 Softmax layer를 fine-tuning하여 출력 개수를 변형하여 새롭게 수립한 관광 이미지 분류 체계의 카테고리 개수(208개)로 설정될 수 있도록 하였다.
- [0092] 본 발명의 일 실시예에서 제안하는 관광 목적의 분류 체계는 표 1과 같이 대분류 13개, 세분류 208개로 구성하였다. 그러나, 이는 일례에 불과하고 대분류와 세분류는 필요에 따라 삭제, 추가, 변경 등의 수정이 가능하다.
- [0093] 상기 CNN 모델의 학습 결과를 이용하여, 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 분류한다.
- [0094] 본 발명에서는 CNN 모델을 통하여 관광 이미지 분류를 학습하고, 이를 통해 관광 이미지를 자동으로 분류한다. 이미지 분류 단계에서 Softmax 층의 개수를 표 1과 같이 13개의 대분류 및 208개의 세분류로 구성된 관광 목적의 분류 항목으로 수정하여 데이터를 훈련하고, 이에 따라 분류된 이후 각 사진 데이터에 분류 label이 입력된다.
- [0095] 분류된 이미지 데이터에 라벨링된 카테고리 정보를 이용하여, 밀도 기반 클러스터링(Density-based spatial clustering of applications with noise, DBSCAN)을 통해 각 카테고리 세부 항목별 클러스터를 추출한다(단계 S20).
- [0096] 구체적으로, 본 발명에서는 각 카테고리 세부 항목별 DBSCAN의 머신러닝 기법을 활용하여 군집지역을 추출한다. 이는 '음식', '쇼핑', '체험' 등의 각각 카테고리별 주요 관광 클러스터를 발견하기 위함이다. 일 실시예로서, 카테고리의 항목별 특성에 따라 DBSCAN의 주요 파라미터인 eps와 minpts의 설정을 다르게 하여 다양한 스케일의 군집지역을 추출할 수 있다.
- [0097] 맥락(context) 정보들의 분류 유형에 따라 각각의 클러스터(cluster) 지역에 포함된 지점들과 군집에 속하지 않고 이례점(outlier)으로 분류되는 경우 맥락 정보의 값의 정확성 여부를 판단할 필요가 있다.
- [0098] '창덕궁'이라는 POI(Point of Interest)의 면형 데이터와 '궁'으로 분류된 사진의 점형 데이터간 중첩분석을 수행하여 이들의 관계가 면형 자료의 경계를 벗어난 점형 자료(disjoint)를 찾아내어(도 2 참조), 정확성 판단 후 자료 수정을 수행한다.
- [0099] '체험'의 '축제 및 이벤트', '도시 경관'의 '불꽃 축제', '야간 관광 활동' 등 각 카테고리별 DBSCAN을 통한 주

요 클러스터(cluster)인 ROI(regions of interest)가 설정된 후, 클러스터 내에 포함되는 항목들은 해당 카테고리 잘 분류된 사례로 판단한다.

- [0100] 반면, 이례점이 나타난 경우, 이례점에 라벨링된 활동 정보(What)의 분류 정확도 확인 및 수정 작업을 수행한다.
- [0101] 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여(단계 S30), 일치하지 않는 경우 상기 위치(Where) 정보를 기초로 상기 활동 정보(What)를 갱신한다(단계 S40).
- [0102] 반면, 상기 클러스터로부터 추출된 위치(Where) 정보와 상기 활동 정보(What)의 라벨링된 값을 비교하여 일치하는 경우, 상기 활동 정보(What)에 대한 정확한 이미지 데이터 분류로 판단하여 해당 이미지에 대한 데이터를 입력하도록 한다.
- [0103] 사진의 위치 정보를 기반으로 DBSCAN에 의해 최종 추출된 위치 정보(Where)와 활동 정보(What)의 라벨링된 값을 비교 분석하여, 위치 정보와 활동 정보가 같은 유형일 경우 일치하는 것으로 처리한다.
- [0104] 그러나, 최종 추출된 위치 정보(Where)와 활동 정보(What)의 라벨링된 값의 차이가 있을 때 활동 정보가 이미지 분류시 오류가 있었던 것으로 판단하고 위치정보와 같은 값으로 수정한다.
- [0105] 예를 들어, 활동 정보(What)에는 "사찰 관람"으로 분류되어 입력되어 있지만, 위치 정보(Where)에는 "종로 창덕궁"임을 확인할 수 있다. 이러한 경우 이미지 분류 과정에서 문제가 있었음을 알 수 있기 때문에 CNN을 통한 이미지 분류를 재수행하여 정보를 갱신한다. 즉, 사찰 관람이 아닌 "궁의 서까래 관람"으로 내용이 업데이트 되었음을 확인할 수 있다(도 3 참조).
- [0106] 갱신된 활동 정보(What)를 기초로 합성곱 신경망(CNN)을 통해 관광 목적의 이미지 전용 분류 체계에 따라 이미지 데이터를 재분류한다(단계 S10).
- [0107] 상기 갱신된 활동 정보(What) 및 실제 위치의 중첩 분석을 통해 이미지 데이터 분류의 정확도를 측정한다(단계 S50).
- [0108] 본 발명에서는 카테고리 세부 항목별로 DBSCAN 분석을 통한 클러스터에 의해 갱신된 값과 건물/poi 또는 수치지도 데이터에서 확인할 수 있는 실제 위치 간 비교 분석(중첩 분석)을 통해 각각 라벨링된 값의 정확도를 분석할 수 있다.
- [0109] 특히, CNN을 통해 분류된 항목에 대한 혼동행렬 값과 데이터 갱신이 된 항목에 대한 혼동행렬 값을 비교 분석하여 정확도 향상을 측정한다. 즉, 도 3과 같이 "What" 컬럼과 "Where" 컬럼 간 일치하지 않는 경우는 정보가 갱신되도록 하며, 같은 경우는 그대로 데이터 입력이 되도록 한다.
- [0110] 일 실시예로서, 건물, POI, 수치지도 데이터 등으로부터 정제된 값을 정답 셋으로 정의하고, 딥러닝의 CNN 모델을 활용했을 때의 "What"컬럼의 정보와 DBSCAN 분석에 의해 도출된 "Where" 컬럼의 정보 간 일치 여부를 파악한다.
- [0111] 일치하지 않는 경우에 한해 "Where" 컬럼을 참조하여 "What" 컬럼의 정보를 갱신한다. 예를 들어, 도 4의 사진 번호 2, 3, 7, 10은 What과 Where 간 불일치한 4개의 항목임을 알 수 있다(실제 값과 비교했을 때 What 컬럼(이미지 분류 결과)의 정확도가 60%에 그침).
- [0112] 반면, DBSCAN을 통해 도출된 "Where"를 참고하여 정보를 갱신한 경우의 분류 정확도는 90%로 향상되어 높은 정확도를 확인할 수 있다.
- [0113] 여기서, 실제값(poi/건물/수치지도)과 What 컬럼 간 직접적인 비교를 하는 경우, 실제값의 데이터가 정제되지 않은 형태로 아주 자세한 값(예를 들어, 실제 해당 위치의 이름, 지명 등-인정전, 회정당, 돈화문, 조계사 템플스테이, 대설법전 등)을 가지기 때문에 Where 컬럼의 경우 DBSCAN을 통해 클러스터의 대표적인 명칭(예를 들어, 창덕궁, 조계사 등)으로 정제된 값을 가질 수 있다. 따라서, 본 발명과 같이 What 컬럼과 Where 컬럼간 비교분석이 더 용이하며, 속도가 빠르다.
- [0114] 건물, poi, 수치지도의 경우 오픈된 데이터(예를 들어, NSDI의 오픈마켓, DATA.go.kr, 국토정보플랫폼 NGII.go.kr 등)를 활용할 수 있다.
- [0115] 본 발명에서 정확도 평가 수식은 도 5의 예시와 같은 혼동행렬(confusion matrix) 방법을 활용한다.
- [0116] 정확도를 판단하여, 정확한 정보라고 판단되면, 이미지 데이터를 입력되도록 한다. 정확도의 판단 값은 필요에

따라 조절 가능하다.

- [0117] 이에 따라, 카테고리가 라벨링된 사진 데이터의 DBSCAN을 통해 위치 정보를 확인하고, 오류가 있는 경우 위치 정보를 갱신하여 관광 목적 이미지의 분류 정확도를 향상시킬 수 있다.
- [0118] 이와 같은, 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법은 애플리케이션으로 구현되거나 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다.
- [0119] 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다.
- [0120] 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.
- [0121] 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0122] 이상에서는 실시예들을 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

산업상 이용가능성

- [0124] 본 발명은 관광 분야에 적합한 분류를 통해 보다 정확도 높은 이미지 분류가 가능하고, 관광 이미지 카테고리의 분류 및 유형화한다. 이를 통해, 카테고리가 라벨링된 사진 데이터의 DBSCAN을 통해 위치 정보를 확인하고, 오류가 있는 경우 위치 정보를 갱신하여 관광 목적 이미지의 분류 정확도를 향상시킬 수 있다.
- [0125] 또한, 본 발명에서 새로 개발한 이미지 분류 체계 모델의 재훈련 및 이와 관련된 소스코드를 통해 다양한 영역의 AI 기술에 응용하고 활용할 수 있을 것으로 예상된다.

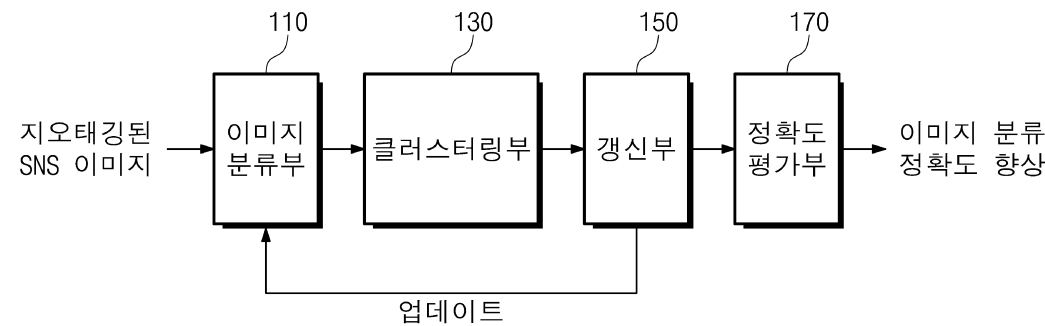
부호의 설명

- [0127] 10: 공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 장치
- 110: 이미지 분류부
- 130: 클러스터링부
- 150: 갱신부
- 170: 정확도 평가부

도면

도면1

10



도면2



도면3

NO	What?	Who?	When?	Where?
1	아쿠아리움 체험	ID01	오후12~2시	강남 코엑스
2	~~사찰 관람~~	~~ID01~~	<del>오후3~5시</del>	~~종로 창덕궁~~
3	궁의 서까래 관람	ID01	오후3~5시	종로 창덕궁

업데이트

도면4

NO	POI/건물/수치지도 (정답/ 실제값)	What (a단계:CNN)	Where (c단계:DBSCAN)	What (c단계에서 업데이트)
1	궁	궁	창덕궁	궁
2	사찰	기와지붕	조계사	사찰
3	사찰	기와지붕	조계사	사찰
4	한옥	기와지붕	북촌 한옥마을	기와지붕
5	궁	궁	경복궁	궁
6	사찰	사찰	월명사	사찰
7	궁 (실제값이 틀린 경우)	정자	북촌 한옥마을	한옥
8	궁	궁	경복궁	궁
9	궁	궁	덕수궁	궁
10	사찰	기와지붕	- (poi 정보 획득 실패)	기와지붕
	정답(90%)	7/10=70%	9/10=90%	9/10=90%

도면5

		예측값		Total
		Positive	Negative	
실제값	True	TP (True Positive)	FN (False Negative)	P
	False	FP (False positive)	TN (True Negative)	N
Total		P'	N'	P+N

도면6

		예측값		Total
		Positive	Negative	
실제값	True	6	3	9
	False	0	1	1
Total		6	4	10

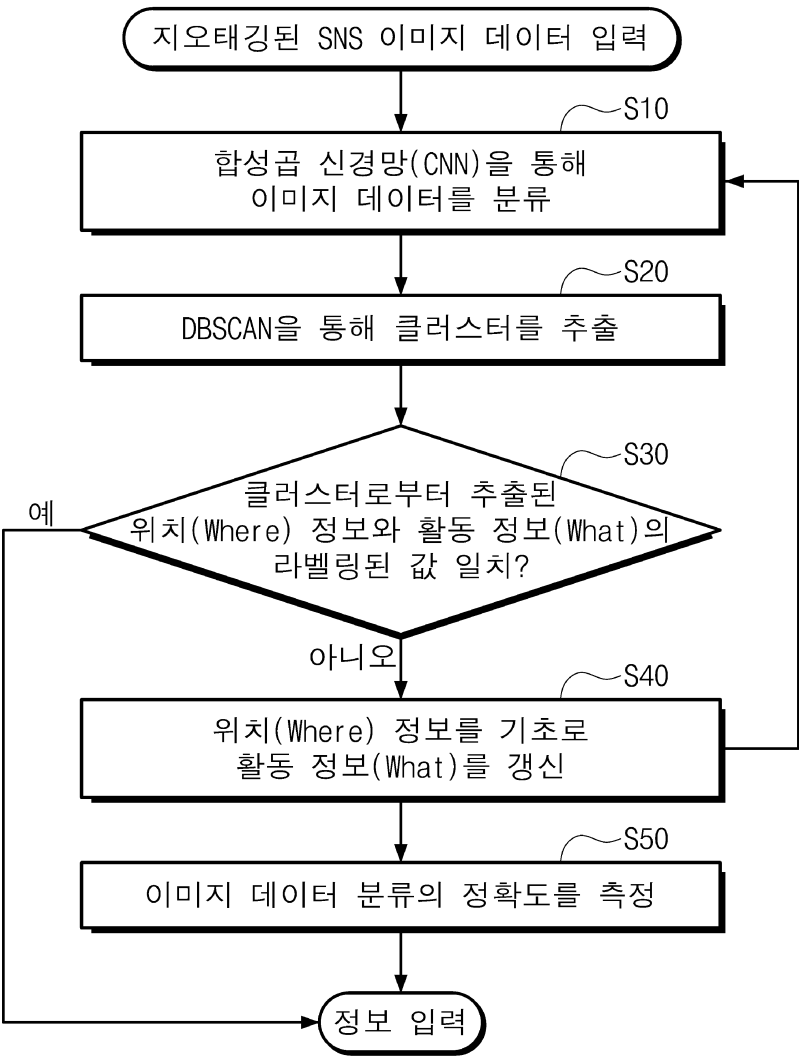
도면7

		예측값		Total
		Positive	Negative	
실제값	True	8	1	9
	False	0	1	1
Total		8	2	10

도면8

		예측값		Total
		Positive	Negative	
실제값	True	8	1	9
	False	0	1	1
Total		8	2	10

도면9



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 제6항

【변경전】

상기 합성곱 신경망을 활용한 SNS 데이터의 관광 목적 이미지 분류 방법

【변경후】

공간 정보 딥러닝 기술을 활용한 관광 목적 SNS 이미지 데이터 분류의 정확도 향상 방법