



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0139995
(43) 공개일자 2020년12월15일

(51) 국제특허분류(Int. Cl.)
G06F 16/29 (2019.01) G06F 16/245 (2019.01)
G06F 16/28 (2019.01) G06N 3/02 (2019.01)
(52) CPC특허분류
G06F 16/29 (2019.01)
G06F 16/245 (2019.01)
(21) 출원번호 10-2019-0066607
(22) 출원일자 2019년06월05일
심사청구일자 2019년06월05일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
이경호
서울특별시 중구 다산로 32, 5동 1002호 (신당동, 남산타운)
김웅남
서울특별시 서대문구 거북골로20길 53-13, 401호 (북가좌동)
(74) 대리인
특허법인우인

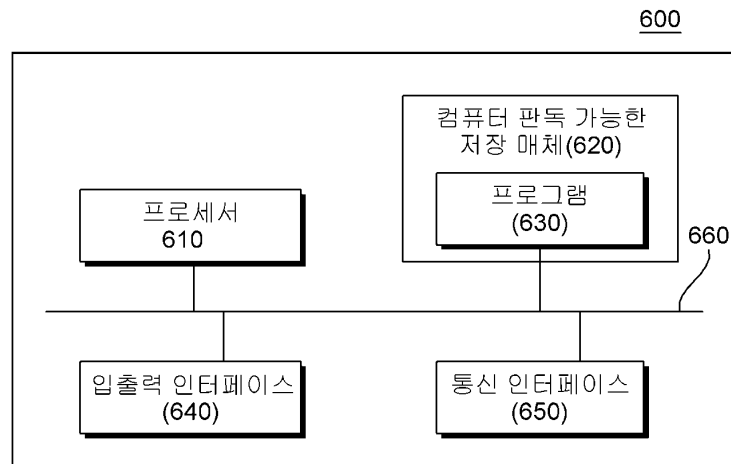
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치 및 방법

(57) 요약

본 발명의 실시예에 따른 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치는, 질의 지역 및 관심 지역들에 속하는 공간 엔티티들에 대하여 미리 정의된 계층적 유형 정보 및 상기 공간 엔티티들간 구조적인 관계에 따른 구조적 관계 정보를 기반으로 상기 공간 엔티티들을 벡터 공간으로 임베딩하는 공간 엔티티 임베딩부 및 상기 임베딩된 공간 엔티티 및 상기 임베딩된 공간 엔티티에 대한 지리적 정보를 이용하여 상기 관심 지역들에 대한 지역 유사성을 측정하는 지역 유사성 측정부를 포함할 수 있다.

대표도 - 도6



(52) CPC특허분류

G06F 16/283 (2019.01)

G06N 3/02 (2019.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	NRF-2016R1A2B4015873
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	한국연구재단 중견연구자지원사업
연구과제명	사물인터넷을 위한 복합 이벤트 처리 기반 개인맞춤 상황인지형 추천 시스템
기 여 율	1/1
과제수행기관명	연세대학교
연구기간	2016.06.01 ~ 2019.05.31

명세서

청구범위

청구항 1

질의 지역 및 관심 지역들에 속하는 공간 엔티티들에 대하여 미리 정의된 계층적 유형 정보 및 상기 공간 엔티티들간 구조적인 관계에 따른 구조적 관계 정보를 기반으로 상기 공간 엔티티들을 벡터 공간으로 임베딩하는 공간 엔티티 임베딩부; 및

상기 임베딩된 공간 엔티티 및 상기 임베딩된 공간 엔티티에 대한 지리적 정보를 이용하여 상기 관심 지역들에 대한 지역 유사성을 측정하는 지역 유사성 측정부;

를 포함하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 2

제1항에 있어서, 상기 공간 엔티티 임베딩부는,

상기 공간 엔티티들에 대한 계층적 유형의 관련성에 따른 계층적 유형 인식 거리 및 의미적 관련성을 기초로 상기 공간 엔티티들을 중 적어도 두 개의 공간 엔티티들을 상기 벡터 공간 상에 근접한 위치에 임베딩하거나 근접하지 않은 위치에 임베딩하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 3

제2항에 있어서,

상기 공간 엔티티 임베딩부는,

상기 공간 엔티티들 각각의 계층적 유형 정보에 따라 연결된 공간 엔티티들의 연결관계를 고려하여 상기 계층적 유형 인식 거리를 산출하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 4

제3항에 있어서,

상기 공간 엔티티 임베딩부는,

상기 계층적 유형 정보를 가지는 제1 공간 엔티티로 정의되는 제1 레벨의 제1 노드의 부모 노드(parent node)인 제2 레벨의 제1 노드와, 상기 제1 레벨의 제1 노드와는 다른 상기 계층적 유형 정보를 가지는 제2 공간 엔티티인 제1 레벨의 제2 노드의 부모 노드인 제2 레벨의 제2 노드가 동일한 노드라면, 상기 제2 레벨의 제1 노드 또는 상기 제2 레벨의 제2 노드의 정보량(information content)을 고려하여 상기 계층적 유형 인식 거리를 산출하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 5

제4항에 있어서,

상기 제2 레벨의 제1 노드와 상기 제2 레벨의 제2 노드가 상이한 노드이고, 상기 제2 레벨의 제1 노드의 부모 노드인 제3 레벨의 제1 노드와, 상기 제2 레벨의 제2 노드의 부모 노드인 제3 레벨의 제2 노드가 동일한 노드라면, 상기 제3 레벨의 제1 노드 또는 상기 제3 레벨의 제2 노드의 정보량을 고려하여 상기 계층적 유형 인식 거리를 산출하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 6

제5항에 있어서, 상기 공간 엔티티 임베딩부는,

상기 제1 레벨의 제1 노드와 상기 제1 레벨의 제2 노드 사이의 홉 수(hop count)를 고려하여 상기 의미적 관련

성을 산출하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 7

제1항에 있어서, 상기 지역 유사성 측정부는,

상기 관심 지역인 제1 지역에 속하고, 상기 공간 엔티티 임베딩부에 의해 임베딩된 제1 공간 엔티티를 나타내는 제1 텐서 및 상기 제1 지역과는 다른 관심 지역인 제2 지역에 속하고, 상기 공간 엔티티 임베딩부에 의해 임베딩된 제2 공간 엔티티를 나타내는 제2 텐서를 생성하는 텐서(tensor) 생성부; 및

상기 제1 텐서 및 제2 텐서를 인공 신경망에 입력하고, 상기 인공 신경망을 통해 제1 지역에 대한 제1 특징 정보 및 제2 지역에 대한 제2 특징 정보를 추출하는 특징 정보 추출부;

를 더 포함하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 8

제7항에 있어서,

상기 지역 유사성 측정부는,

상기 제1 특징 정보 및 제2 특징 정보를 이용하여 상기 제1 지역 및 제2 지역에 대한 지역 유사성을 측정하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 9

제7항에 있어서,

상기 인공 신경망은 컨볼루션 신경망(CNN)을 기반으로 하는 트리플렛 신경망(triplet network)이고,

상기 특징 정보 추출부는, 상기 질의 지역과 유사한 상기 제1 지역 및 상기 질의 지역 사이의 거리를 나타내는 제1 거리 정보가, 상기 제1 지역보다 상기 질의 지역과 덜 유사한 상기 제2 지역 및 상기 질의 지역 사이의 거리를 나타내는 제2 거리 정보보다 더 작은 값을 가지도록 비율 기반 손실함수를 이용하여 상기 트리플렛 신경망을 학습시키는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 10

제8항에 있어서,

상기 특징 정보 추출부는, 상기 제1 거리 정보, 상기 제2 거리 정보 및 상기 트리플렛 신경망에 대한 최소 제곱 편차의 정규화 항을 고려하는 상기 비율 기반 손실함수를 이용하여 상기 트리플렛 신경망을 학습시키는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치.

청구항 11

질의 지역 및 관심 지역들에 속하는 공간 엔티티들에 대하여 미리 정의된 계층적 유형 정보 및 상기 공간 엔티티들간 구조적인 관계에 따른 구조적 관계 정보를 기반으로 상기 공간 엔티티들을 벡터 공간으로 임베딩하는 단계; 및

상기 계층적 유형 정보 및 상기 구조적 관계 정보를 포함하는 상기 임베딩된 공간 엔티티 및 상기 임베딩된 공간 엔티티에 대한 지리적 정보를 이용하여 상기 관심 지역들에 대한 지역 유사성을 측정하는 단계;

를 포함하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 방법.

청구항 12

제11항에 있어서,

상기 공간 엔티티들을 벡터 공간으로 임베딩하는 단계는,

상기 공간 엔티티들에 대한 계층적 유형의 관련성에 따른 계층적 유형 인식 거리 및 의미적 관련성을 기초로 상기 공간 엔티티들을 중 적어도 두 개의 공간 엔티티들을 상기 벡터 공간 상에 근접한 위치에 임베딩하거나 근접하지 않은 위치에 임베딩하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을

측정하는 방법.

청구항 13

제12항에 있어서,

상기 공간 엔티티들을 벡터 공간으로 임베딩하는 단계는,

상기 공간 엔티티들 각각의 계층적 유형 정보에 따라 연결된 공간 엔티티들의 연결관계를 고려하여 상기 계층적 유형 인식 거리를 산출하는 것을 특징으로 하는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 방법.

청구항 14

제11항 내지 제13항에 따른 공간 엔티티의 지역 유사성을 측정하는 방법을 실행시키기 위하여 컴퓨터에서 판독 가능한 기록 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치 및 방법에 관한 것이다.

배경 기술

[0002] 지식 공유 커뮤니티의 증가 및 자동 정보 추출 기술의 발전에 따라 대규모 지식 베이스가 구축되고 있다. 예컨대 DBpedia, YAGO, Freebase 등의 지식 베이스는 방대한 양의 고품질의 의미적 데이터(Semantic Data)를 유지 관리하고 의미적 데이터는 WWW 표준인 RDF 데이터 모델을 준수한다.

[0003] 지식 베이스는 공간 데이터로 보강되고 있다. DBpedia의 11%은 공간 데이터를 포함하고 YAGO2의 59%는 공간 데이터를 포함한다. 공간 데이터의 증가는 지식 그래프를 대상으로 위치 기반 질의를 수행 가능하게 한다.

[0004] 공간 지식 그래프는 지리적 좌표, 계층적 유형, 구조적 관계 등 공간 엔티티(spatial entity)에 관련되는 풍부한 정보를 포함하고 있는데, 이는 도시 지역 지식을 탐구할 수 있는 기회를 제공할 수 있다. 하지만, 공간 지식 그래프의 기하급수적으로 증가하는 볼륨 및 그래프 구조적 복잡성 등으로 인해 비전문가인 일반 사용자가 공간 지식 그래프를 이해하는 것에는 다소 어려움이 있다.

[0005] 유사 지역 검색은 공간 데이터를 탐색할 수 있는 쉽지만 효율적인 방법이다. 이러한 유사 지역 검색은 하나의 질의 지역을 입력 받아, 목표 공간에서 질의 지역과 가장 유사한 top-k 개의 지역을 반환한다. 사람들은 보편적으로 도시지역의 작은 부분(예를 들어, 집 또는 직장 근처)에 대하여만 익숙하다. 유사 지역 검색은 서로 다른 지역에 대한 지식을 이전할 수 있는 간단한 방법을 제공할 수 있다.

[0006] 하지만, 종래의 유사 지역 검색 방법은 공간 엔티티의 계층적 유형을 정교하게 파악하지 못하고 공간 엔티티 사이의 관계도 포착하지 못하기 때문에 검색 결과의 정확도가 떨어진다는 단점이 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국 공개 특허 제10-2012-0104650호 (공개)

발명의 내용

해결하려는 과제

[0008] 상기 전술한 종래의 문제점을 해결하기 위한 본 발명의 실시예에 따른 공간 엔티티의 지역 유사성을 측정하는 장치 및 방법은 지역의 유사성 측정 결과의 정확도를 향상시킬 수 있도록 하는 것을 목적으로 한다.

과제의 해결 수단

- [0009] 상기 목적을 달성하기 위한 본 발명의 실시예에 따른 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치는, 질의 지역 및 관심 지역들에 속하는 공간 엔티티들에 대하여 미리 정의된 계층적 유형 정보 및 상기 공간 엔티티들간 구조적인 관계에 따른 구조적 관계 정보를 기반으로 상기 공간 엔티티들을 벡터 공간으로 임베딩하는 공간 엔티티 임베딩부 및 상기 임베딩된 공간 엔티티 및 상기 임베딩된 공간 엔티티에 대한 지리적 정보를 이용하여 상기 관심 지역들에 대한 지역 유사성을 측정하는 지역 유사성 측정부를 포함할 수 있다.
- [0010] 상술한 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 방법은 질의 지역 및 관심 지역들에 속하는 공간 엔티티들에 대하여 미리 정의된 계층적 유형 정보 및 상기 공간 엔티티들간 구조적인 관계에 따른 구조적 관계 정보를 기반으로 상기 공간 엔티티들을 벡터 공간으로 임베딩하는 단계 및 상기 계층적 유형 정보 및 상기 구조적 관계 정보를 포함하는 상기 임베딩된 공간 엔티티 및 상기 임베딩된 공간 엔티티에 대한 지리적 정보를 이용하여 상기 관심 지역들에 대한 지역 유사성을 측정하는 단계를 포함할 수 있다.
- [0011] 또한 본 발명은 상기한 방법에 따른 공간 엔티티의 지역 유사성을 측정하는 방법을 실행시키는 컴퓨터로 판독 가능한 기록매체에 저장된 컴퓨터 프로그램을 제안한다.

발명의 효과

- [0012] 본 발명의 실시예에 따른 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치는 지역의 유사성 측정 결과의 정확도를 향상시킬 수 있는 효과가 있다.

도면의 간단한 설명

- [0013] 도1은 일 실시예에 따라 계층적 유형과 관련한 동기화 시나리오에 대한 도면이다.
- 도2는 일 실시예에 따라 공간 엔티티들간의 구조적인 관계와 관련된 동기화 시나리오를 도시한 도면이다.
- 도3은 공간 엔티티 임베딩 과정에서의 표기법을 나타내는 표이다.
- 도4는 본 발명의 일 실시예에 따른 계층적 유형(C)을 나타내는 도면이다.
- 도5는 본 발명의 일 실시예에 따른 지역 유사성 학습과정에서 표기법을 나타내는 표이다.
- 도6은 본 발명의 일 실시예에 따른 공간 엔티티의 지역 유사성을 측정하는 장치의 구성을 개략적으로 도시한 블록도이다.
- 도7은 본 발명의 또 다른 실시예에 따른 공간 엔티티의 지역 유사성을 측정하는 장치의 구성을 개략적으로 도시한 블록도이다.
- 도8은 본 발명의 일 실시예에 따른 공간 엔티티의 지역 유사성을 측정하는 방법에 대하여 시간의 흐름에 따라 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.
- [0015] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계 없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.
- [0016] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함" 한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록"등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

- [0017] 이하, 본 발명의 일 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략할 수 있다.
- [0018] 지식 공유 커뮤니티의 증가 및 자동 정보 추출 기술의 발전에 따라 대규모 지식 베이스가 구축되고 있고, 많은 공간 지식 그래프(spatial knowledge graphs, SKG)가 공간적으로 풍부한 DBpedia, YAGO, Freebase와 같은 지식 베이스를 이용할 수 있다.
- [0019] 공간 지식 그래프는 지리적 좌표, 계층적 유형, 구조적 관계 등 공간 엔티티(spatial entity)에 관련되는 풍부한 정보를 포함하고 있는데, 이는 도시 지역 지식 적으로 증가하는 볼륨 및 그래프 구조적 복잡성 등으로 인해 비전문가인 일반 사용자가 공간 지식 그래프를 이해하는 것에는 다소 어려움이 있다. 이와 같은 문제점을 해결하기 위해서는 유사 지역 검색(SRS, similar region search) 방법을 이용하여야 한다. 그러나 유사 지역 학습을 위한 기존의 접근 방식은 SKG에 포함된 풍부한 정보를 사용할 수 없으므로 부정확한 결과를 초래할 수 있다.
- [0020] 이에, 본 발명에서는 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치 및 방법을 제안한다. 본 발명의 실시예에 따른 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 장치 및 방법은 지역 유사성에 대한 공간 지식 표현 학습 방법(SKRL4RS, spatial knowledge representation learning method for region similarity)을 제공할 수 있다. 본 발명의 공간 엔티티의 지역 유사성을 측정하는 장치는 먼저, SKG의 공간 엔티티를 벡터 공간으로 인코딩하여 유용한 특징들을 용이하게 추출하고, 지리 정보와 함께 공간 엔티티의 임베딩을 통해 지역을 3차원 텐서로서 표현한다. 그리고, 3차원 텐서는 triplet network인 CNN 기반의 신경망에 입력되어 지역에 대한 특징 벡터를 학습시킨다. 이와 같은 방법으로, 본 발명의 장치는 관심 지역에 속하는 공간 엔티티들에 대한 계층적 유형, 의미 관련성 및 지역 내 공간 엔티티들간 상대적 위치를 추출할 수 있다.
- [0021] 여기서, 공간 지식 그래프(SKG)란 공간적으로 확장된 지식 그래프를 의미하고, 지식 그래프(KG, knowledge graph)란 다양한 소스로부터 축적한 시맨틱 검색 정보를 사용하여 검색결과를 향상시키는 것으로 정의되며, 공간 엔티티(ses, spatial entities)는 지리적 좌표와 연관되어 있는 개체를 의미한다.
- [0022] 공간 지식 그래프는 지리적 좌표, 계층적 유형, 구조적 관계 등 공간 엔티티(spatial entity)에 관련되는 풍부한 정보를 포함하고 있는데, 이와 같은 공간 지식 그래프는 특징 지역에 대한 지식을 탐구할 수 있는 기회를 제공할 수 있다. 하지만, 공간 지식 그래프의 기하급수적으로 증가하는 볼륨 및 그래프 구조적 복잡성 등으로 인해 비전문가인 일반 사용자가 공간 지식 그래프를 이해하는 것에는 다소 어려움이 있다.
- [0023] 본 발명의 공간 엔티티의 지역 유사성을 측정하는 장치에서 제공되는 유사 지역 검색(SRS)은 비전문가인 일반 사용자가 사용하기에도 용이할 뿐만 아니라 효과적인 공간 데이터를 탐색할 수 있는 방법이다. 본 발명의 장치는 일반적으로 직사각형으로 표현되는 질의 지역(query region)을 입력 받고, 관심 지역, 즉, 상기 질의 지역과의 유사성을 측정하기 위한 후보 지역들을 상기 질의 지역과 유사한 정도에 따라서 상기 관심 지역들을 분류하여 출력할 수 있다. 사람들은 대부분 자신의 거주지 또는 직장 근처의 지역과 같은 일부 특정 지역에 대해서만 친숙할 수 있다. 본 발명의 장치에서 제공되는 SRS는 지식을 상기 친숙한 지역 외에 다른 지역으로도 쉽게 전송하는 방법을 제공할 수 있다.
- [0024] 예를 들어, 부부가 최근에 새로운 도시로 이사 갔다고 가정하는 경우, 그들은 쇼핑, 레크리에이션(recreation), 음식, 편리한 교통 등 그들이 주로 데이트해왔던 장소와 유사한 지역을 찾길 원한다. 또 다른 예로 본 발명의 장치는 대기 오염 제어 및 사업장 선정과 같은 다양한 시나리오에 적용될 수 있다.
- [0025] 효과적인 유사 지역 검색을 실현하려면 첫째로 공간 지역 그래프에 포함된 공간 엔티티를 벡터 공간에 삽입하여 유용한 정보를 추출하고, 둘째로 각 지역 내의 공간 엔티티의 풍부한 정보를 보존하기 위해 지역 유사성을 모델링해야 한다.
- [0026] 최근 개발된 연구는 공간 엔티티의 유형(속성)을 one-hot 임베딩 기술을 사용하여 인코딩하고, 인코딩된 결과에 따라 공간 엔티티들의 지리적 좌표를 합하여 3차원 텐서로 표현한 후, 지역 유사성을 학습시키기 위하여 상기 3차원 텐서를 triplet network에 입력시킨다. 그러나, 이와 같은 기존 기술은 다음과 같은 한계점이 있다.
- [0027] 그 문제점으로는 첫째, 공간 엔티티의 유형을 정교하게 구별할 수 없다는 것이다. 도1은 질의 지역(R_q)과 유사한 지역을 검색하기 위한 유사 지역 검색 방법을 설명하기 위해 도시한 도면이다. 보다 자세하게는, 도1은 계층적 유형과 관련한 동기화 시나리오에 대한 도면이다.

- [0028] 일 실시예인 도1을 참고하면, 질의 지역 R_q 는 화장품 쇼펍물, 한식당, 영화관, 이렇게 3가지의 공간 엔티티로 구성되어 있고, R_1 , R_2 , R_3 의 3개의 후보 지역이 있다고 가정할 때, 각 후보 지역은 질의 지역과 비교할 때 3개의 공간 엔티티를 포함하고, 공간 엔티티 중 하나는 다른 유형에 속한다.
- [0029] 이때, 기존의 측정 방법에 따르면, 지역 유사성 측정 결과는 $\text{sim}(R_q, R_1) = \text{sim}(R_q, R_2) = \text{sim}(R_q, R_3)$ 와 같이 두 지역 사이의 유사성을 도출하게 된다. 그러나, 도1에 나타난 바와 같이 R_3 후보 영역에는 공장(factory) 유형이 포함되어 있다. 공장(R_3)과 한식당(R_q) 간의 유사성은 유형 계층 구조에 따라 일식당(R_2)과 한식당(R_q)의 유사성보다 작게 설정되어야 한다. 즉, 공장 유형은 지역 선택 데이터의 유형에 있어 노이즈를 발생시킬 수 있는 부정적인 영향을 미칠 수 있기 때문에 지역 유사성 측정은, 예컨대 $\text{sim}(R_q, R_2) > \text{sim}(R_q, R_3)$ 와 같이 계층적 유형을 명확히 구별하여야 한다.
- [0030] 두번째로는, 기존의 측정 방법은 공간 엔티티들 간의 관계를 추출할 수 없다는 것이다. 도2는 일 실시예에 따라 공간 엔티티들 간의 관계에 대하여 설명하기 위해 도시한 도면으로, 도2는 도1과 같은 동기화 시나리오의 근본이 되는 SKG를 도시한 도면이다. 도2에 도시된 바와 같이, SKG는 dbp : affiliation의 관계와 같은 유형으로 공간 엔티티들간의 연결 관계를 정의할 수 있다.
- [0031] 기존의 측정 방법에 따르면, 지역 유사성 측정 결과는 $\text{sim}(R_q, R_1) = \text{sim}(R_q, R_2)$ 가 된다. 그러나, 도2에 도시된 바와 같이 R_q 와 R_1 은 내부의 모든 공간 엔티티들끼리 서로 연결되어 있지만, R_2 의 공간 엔티티들 사이에는 하나의 연결 관계만이 존재함을 확인할 수 있다. 즉, 지역 유사성 측정은 $\text{sim}(R_q, R_1) > \text{sim}(R_q, R_2) > \text{sim}(R_q, R_3) > \text{sim}(R_q, R_4)$ 의 결과를 도출하도록 설계되어야 한다.
- [0032] 전술한 바와 같은 기존 측정 방법의 문제점을 해결하기 위한 본 발명의 공간 엔티티의 지역 유사성을 측정하는 장치에 대하여 설명한다. 본 발명의 공간 엔티티의 지역 유사성을 측정하는 장치에 의해 제공되는 공간 지식 표현 학습 방법(SKRL4Rs)는 크게 공간 엔티티 임베딩하는 단계와, 지역 유사성을 학습시키는 단계, 이렇게 2개의 단계로 수행될 수 있다.
- [0033] 공간 엔티티 임베딩 과정을 통해, SKG의 엔티티는 엔티티들 각각의 계층적 유형에 대한 계층적 유형 정보 및 엔티티들간 구조적인 관계에 대한 구조적 관계 정보를 보존하면서 벡터 공간에 삽입될 수 있다. 이러한 임베딩 과정은 기존의 그래프 형태로 SKG를 처리하는 것보다 유용한 정보를 보다 용이하게 추출할 수 있다.
- [0034] 그 후에, 지역 유사성 학습 과정을 통해 학습된 임베딩(계층적 유형, 구조적인 관계) 및 지리적 좌표를 이용하여 측정 대상 지역(질의 지역, 관심 지역)을 3차원 텐서(tensor)로 나타낸다. 상기 3차원 텐서는 triplet network에 입력되어 triplet network에 의해 지역의 유사성을 학습할 수 있다.
- [0035] 여기서, 공간 지식 그래프(SKG)는 그래프 모델을 이용하여 정의될 수 있다. 본 발명에 따른 SKG는 방향 그래프 $G = (E, R)$ 로 정의할 수 있다. 여기서 E는 엔티티들의 집합(무리)을 의미하고, R은 엔티티들간 관계(edge)의 집합을 의미할 수 있다. 각 공간 엔티티는 유형 $c \in C$ 로 할당되며, 여기서 C는 계층적 유형을 의미한다. 공간 엔티티들의 집합($E^{??}$)은 E에 지리적 좌표가 추가된 것을 나타낸다. 지리적 좌표 se.x와 se.y는 공간 엔티티의 경도와 위도를 나타낸다.
- [0036] 질의 지역 또는 관심 지역과 같은 지역 R은 tb, bb, lb, rb(위, 아래, 왼쪽, 오른쪽)으로 바운딩되는 사각형 형상의 공간으로 이루어지는 것일 수 있다. 여기서, 지역 R은 상기 공간 엔티티의 집합($\mathcal{E}_R^{\otimes} \subset \mathcal{E}$)을 포함할 수 있다 ($\forall se \in \mathcal{E}_R^{\otimes}, lb \leq se.x \leq rb, bb \leq se.y \leq tb$).
- [0037] 본 발명에 따른 유사 지역 검색(SRS)은, 공간 지식 그래프 SKG, 목표 공간 영역 P, 질의 지역 R_q 가 주어지면, SKG로부터 유사성 측정 함수($\text{sim}(\cdot)$)를 학습하고, $\text{sim}(R_q, R_i) \geq \text{sim}(R_q, R_j), \forall R_i \in \mathcal{K}_R, \forall R_j \notin \mathcal{K}_R$ 가 되도록 목표 공간 영역 P의 위에 있는 top-k 지역 K_R 을 찾을 수 있다.
- [0038] 대규모 SKG는 전 세계적으로 배포되는 수많은 공간 엔티티들이 포함되어 있기 때문에 종래의 기술만으로는 SKG의 전체 공간안에서 유사 지역을 검색하는 데 어려움이 있다. 그리고, 사용자는 일반적으로 사용자가 거주하는 한정된 지역에 대해서 관심이 있다. 따라서, 본 발명의 장치는 목표 지리적 공간인 P를 검색 공간을 검색 공간

으로 제한하여 공간 개체의 지역 유사도를 측정할 수 있다. P는 직사각형의 공간으로 특정 도시의 최소 경계공간으로 설정될 수도 있다.

[0039] 본 발명에서는 관심 지역(R_1 , R_2)들간 유사성을 $\text{sim}(R_1, R_2)$ 으로 나타낼 수 있고, $\text{sim}(R_1, R_2)$ 은 각 지역 내의 공간 엔티티들에 대하여 계층적 유형 정보, 구조적 관계 정보, 그리고 공간 엔티티들간 상대적인 위치 정보를 반영한다.

[0040] 지역 유사성 측정은 정확한 지역 유사성 탐색(SRS)을 실현하기 위한 가장 중요한 요소이다. 본 발명에서의 장치는 SKG에 포함된 풍부한 정보들, 예를 들어 계층적 유형 정보, 구조적 관계 정보 및 공간 엔티티들의 지리적 위치를 이용한다.

[0041] SGK는 구조화된 데이터를 표현하는 데는 효과적이지만, 상징적 특성으로 인해 본래의 형태로 조작하는 것은 어렵다. 따라서, 본 발명의 장치는 먼저 공간 엔티티들을 벡터 공간에 임베딩하여 계층적 유형 정보 및 구조적 관계 정보를 유지하여 조작을 단순화한다. 그 후에, 공간 엔티티들의 벡터는 공간 크기로 풍부하게 표현되고, 이러한 공간 크기로 표현되는 각 지역의 공간 엔티티들의 벡터를 컨볼루션 신경망(CNN)에 입력하여, CNN을 통해 각 지역의 공간개체들에 대한 상대적인 위치에 대한 정보를 유지하고 있는 특징 벡터들을 추출할 수 있다.

[0042] 본 발명의 공간 지식 그래프를 대상으로 공간 엔티티의 지역 유사성을 측정하는 장치 및 방법의 보다 구체적인 동작 과정에 대하여 후술한다.

[0043] 먼저 수행되는 과정은 공간 엔티티를 임베딩하는 단계이다. 예컨대, 헤드 공간 엔티티(head spatial entity, h)와 테일 공간 엔티티(tail spatial entity, t)가 있다고 할 때, 헤드 공간 엔티티(h)와 테일 공간 엔티티(t)가 계층적 유형 또는 의미적으로 서로 관련된다면, 상기 헤드 공간 엔티티(h)와 테일 공간 엔티티(t)는 벡터 공간 상에 서로 가까운 위치에 임베딩되어야 한다.

[0044] 도3은 본 발명의 일 실시예에 따른 공간 엔티티 임베딩 과정에서의 표기법을 나타내는 표이다.

[0045] 본 발명에서 의미적 관련성(semantic relatedness)이라 함은 다음과 같이 정의될 수 있다. SKG 상에 두 개의 공간 엔티티(h, t)가 주어지면, 그에 따른 홉(hop, mh) 수의 최대치에 대한 임계치를 알 수 있게 되므로, 헤드 공간 엔티티(h)와 테일 공간 엔티티(t) 사이의 홉수를 기반으로 상기 두 공간 엔티티(h, t)의 의미적인 관련성에 대하여 판단할 수 있다.

[0046] 헤드 공간 엔티티(h)와 테일 공간 엔티티(t)의 벡터 표현인 h_e 와 t_e 사이의 거리를 결정하기 위한 계층적 유형 및 의미적 관련성 인식 거리 D_{ht} 는 아래 <수학식1>과 같이 정의될 수 있다.

수학식 1

$$D_{ht} = \frac{C_{ht}}{S_{ht}}$$

[0047]

여기서, C_{ht} 는 계층적 유형 인식 거리 이고, S_{ht} 는 h와 t의 의미적 관련성을 나타낸다.

[0049] 예를 들어, 계층적 유형에 있어 h와 t의 유형이 유사한 경우, h_e 와 t_e 는 벡터 공간 상에 서로 가까이 위치해야 한다. 이러한 이론적 근거에 따라 본 발명의 장치는 계층적 유형 인식 거리를 기반으로 하는 정보량(information content)를 이용하여 공간 개체들을 벡터 공간에 임베딩 할 수 있다.

[0050] 본 발명의 계층적 유형 인식 거리 C_{ht} 에는 아래 <수학식2>와 같이 표현될 수 있다.

수학식 2

$$C_{ht} = \gamma \times (1 - \frac{cSim(h, t)}{\max IC(\cdot)})$$

[0051]

[0052] 여기서, γ ($\gamma > 0$)는 거리 파라미터이고, $cSim(h, t)$ 는 h 와 t 사이의 계층적 유형의 유사성을 나타내며, $IC(\cdot)$ 는 각 유형에 대한 정보량을 의미한다. 정보량의 최대값을 나타내는 $IC(\cdot)$ 은 $cSim(h, t)$ 를 정규화하는데 사용되며, $cSim(h, t)$ 는 <수학식3>과 같이 정의될 수 있다.

수학식 3

$$cSim(h, t) = \max_{c \in Sub(h, t)} IC(c)$$

[0053]

$$IC(c) = -\log p(c)$$

[0054]

[0055] 여기서, $Sub(h, t)$ 는 계층적 유형과 관련한 h 와 t 에 따른 두 유형을 포함하는 유형 집합을 의미한다. 그리고, $p(c)$ 는 계층적 유형에서의 모든 유형을 방문(visiting)할 때 c 와 만나게 되는 확률을 나타낸다. 따라서, 유형 c 가 추상적일수록, 정보량은 더 줄어들게 된다.

[0056] 일반적으로 h 와 t 사이의 유사성은 가장 가까운 부모 노드(parent node)의 정보량에 따라 결정될 수 있다. 보다 구체적으로 설명하면, h 및 t 와 모두 관련되어 있는 부모 노드가 가까울수록 계층적 유형 유사성은 높아진다.

[0057] 그리고, 본 발명의 장치는 SKG의 계층적 유형은 비교적 정적이기 때문에, 공간 개체의 계층적 유형을 나타내는 각 노드의 정보량에 대하여 미리 계산하여 알 수 있다.

[0058] 도4는 본 발명의 일 실시예에 따른 계층적 유형(C)을 나타내는 도면이다. 도4를 참고하면, C는 총 17개의 유형 노드를 가지며, 각 노드는 각 유형의 정보량에 상응하는 값들을 포함한다. 근 노드(root node) c_0 ,와 관련하여, 모든 유형의 노드들은 상기 근 노드 c_0 와 연결되어야 하기 때문에, $p(c_0)=1$ 이므로 $IC(c_0) = -\log(1) = 0$ 이다. c_{111} 는 잎 노드(leaf node)이므로, 확률 $p(c_{111})$ 는 $(1/17)$ 이고, $IC(c_{111}) = -\log(1/17) = 1.230$ 이 된다. 또 다른 예로, 중간 노드인 c_{21} 에 있어서, $p(c_{21})$ 는 $(4/17)$ 이므로, $IC(c_{21}) = -\log(4/17) = 0.628$ 이 된다. 도4에 도시된 다른 잎 노드들의 정보량도 상기 설명한 원리와 같다.

[0059] 다음으로, 도4에 도시된 바와 같은 계층적 유형에 따라 C_{ht} 를 계산하는 방법에 대하여 설명한다. 예를 들어, 공간 엔티티 h 와 t 가 있다고 가정하는 경우, 헤드 공간 개체(h)에 따른 유형이 c_{112} 이고, 테일 공간 개체(t)에 따른 유형이 c_{121} 이며, 거리 파라미터(γ)가 1인 경우, c_1 이 c_{112} 와 c_{121} 를 연결하는 가장 가까운 부모 노드이기 때문에, $cSim(h, t) = IC(c_1) = 0.276$ 이 된다. 또한, $\max IC(\cdot)$ 는 계층적 유형에 따른 잎 노드의 정보량과 동일하다. 결과적으로, $C_{ht} = 1 \times \{1 - (0.276/1.230)\} = 0.775$ 가 된다.

[0060] D_{ht} 의 또 다른 중요한 요소는 h 와 t 간의 의미적 관련성을 나타내는 S_{ht} 이다. S_{ht} 는 D_{ht} 를 축소시키기 위한 축소 요소(narrowing factor)와 같다. 특히, h 와 t 가 의미적으로 관련성이 없다면, D_{ht} 는 C_{ht} 에 의해서만 결정될 수 있다. 이러한 경우가 아니라면, h 와 t 가 밀접할수록(h 와 t 사이의 홉이 작을수록) h_e 와 t_e 사이는 가 더 가깝게 설정된다. S_{ht} 는 아래 <수학식4>와 같이 정의될 수 있다

수학식 4

$$S_{ht} = (1 + \theta)^{mh+1-hop(h,t)}$$

[0061]

[0062] 여기서, mh 는 최대 홉수에 대한 임계치($\{1, 2, \dots, mh\}$)를 나타내고, $hop(h, t)$ 는 h 로부터 떨어져있는 t 의 홉수를 나타내며, 축소율 θ 은 $\theta \in (0, 1)$ 으로 설정될 수 있다. 또한, h 와 t 사이의 홉수가 상기 임계범위 내에 속하지 않는다면, h 와 t 는 관계가 없는 것으로 판단하고, 이에 따른 $hop(h, t)$ 는 $mh + 1$ 로 설정할 수 있다. 다음으로, D_{ht} 를 계산하는 방법에 대하여 설명한다.

[0063] 위 c_{112} 와 c_{121} 에 대한 실시예에 따라 두 개의 공간 엔티티 h 와 t 가 주어지고, m_h 는 3, θ 를 0.05로 설정하며, $\text{hop}(h, t)$ 가 2라고 가정할 때, $D_{ht} = 0.775/(1+0.05)^{(3+1+2)} = 0.703$ 이 된다.

[0064] 공간 엔티티에 대한 전술한 바와 같은 임베딩된 공간 개체들을 학습하기 위한, 손실 함수를 기반으로 하는 거리 (L_e)는 <수학식5>와 같이 정의될 수 있다.

수학식 5

$$\mathcal{L}_e = \sum_{(h,t) \in \mathcal{E}^{\otimes}} (\|h_e - t_e\| - D_{ht})^2$$

[0065]

[0066] 여기서, h_e 와 t_e 는 h 와 t 의 임베딩을 나타낸다.

[0067] 다음으로, 지역 유사성 학습과정에 대하여 설명한다.

[0068] 내부의 공간 엔티티들의 계층적 유형 및 의미적 관련성 외에도 상대적인 위치 정보는 각 지역을 나타내는 또 다른 중요한 요소이다. 예컨대, R_a 와 R_b 두 지역이 있고, 두 지역 모두 쇼핑 및 음식과 관련된 지역이며, R_a 와 R_b 는 제휴 관계로 서로 직접적으로 연결된 쇼핑물과 식당을 포함하고 동일한 유형이지만, R_a 의 쇼핑물은 식당과 가깝고, R_b 는 가까이 있지 않은 경우, 제한된 이동성을 가지는 고객들은 불편함 때문에 R_b 지역을 선택하지 않게 된다.

[0069] 본 발명의 장치는 triplet network를 이용하여 지역들의 특징 벡터를 학습함으로써 상기와 같은 문제점을 해결할 수 있다. 도5는 본 발명의 일 실시예에 따른 지역 유사성 학습과정에서 표기법을 나타내는 표이다. 본 발명의 장치는 콘볼루션 신경망을 기반으로 하되, 서로 다른 특징은 분류하면서, 구조가 유사하여 같은 공간 내에 있는 데이터는 떨어뜨리는 딥러닝 방법인 triplet network를 이용한다. Triplet network는 동일한 CNN 모델에서 기준이 되는 이미지와 이를 비교할 두 개의 이미지를 더 사용하는 것이다. Triplet network는 공유된 CNN의 세 개의 인스턴스(instance)들로 구성될 수 있다.

[0070] CNN은 기본적으로 지역 데이터를 학습시키는데 적합한 알고리즘이다. 특히, CNN을 통해 상기 가까운 영역에 속하는 내부 공간 엔티티들의 의미적 관련성 정보, 그리고 부분 영역에 대한 상대적인 위치들에 대한 정보를 포함(보존)한 채로, 일 지역의 부분 영역들에 대한 특징들을 추출할 수 있다. 그리고, CNN에서 사용된 특징 추출과정은 계층적인 방식으로 이루어지고, 이 방식은 공간 엔티티들을 지리적인 공간으로 결합하는 방식과 유사하다.

[0071] Triplet network를 이용하는 또 다른 중요한 이유는 지역 데이터에 대한 레이블(label)이 없다는 것이다. 여기서, 레이블이란 쇼핑 지역, 주거 지역, 관광 지역 등이 될 수 있다. 또한, 사람이 지역에 레이블을 할당하기 위한 기준을 설정하는 것은 어려운 일이다. 이와 같이 사람에게 의해 수동적으로 할당된 지역의 레이블은 매우 주관적일 수 있어 일관성에 대해 보장할 수 없는 문제가 있다. 그러나, 질의 지역 R_q 와 두 개의 후보 영역 R_1 과 R_2 가 주어지면 어느 지역이 더 R_q 와 유사한지 쉽게 알 수 있다. 본 발명에 따른 triplet network는 위에서 설명한 바와 같은 문제를 해결하기 위한 것이다.

[0072] 본 발명의 일 실시예에 따른 triplet network는 질의 x_q , 유사한 후보인 x_+ , 그리고 유사하지 않은 후보 x_- 를 입력 받을 수 있다. 여기서, x_q , x_+ , x_- 는 3D 텐서(tensor)로 구현될 수 있으며, 3D 텐서에서 두 개의 차원은 위도 정보와 경도 정보이고, 마지막 세번째 차원은 임베딩된 공간 엔티티 정보를 포함할 수 있다. 공유된 CNN($\text{Net}(\cdot)$)을 통해, 본 발명의 triplet network는 3개의 특징 맵, 질의 지역에 대한 특징 맵 $\text{Net}(x_q)$, 질의 지역과 유사한 후보 지역 R_+ 의 특징 맵 $\text{Net}(x_+)$, 및 질의 지역과 유사하지 않은 후보 지역 R_- 의 특징 맵 $\text{Net}(x_-)$ 를 생성할 수 있다. Triplet network는 제1 거리 $d_+(d_+ = \|\text{Net}(x_q) - \text{Net}(x_+)\|)$ 가 제2 거리 $d_-(d_- = \|\text{Net}(x_q) - \text{Net}(x_-)\|)$ 보다 작아지도록 학습시키는 것을 목적으로 한다. 아래 <수학식6>은 지역 유사성을 학습하기 위한 비율 기반 손실 함수(L_t)이다.

수학식 6

$$\mathcal{L}_t = \sum_{(R_q, R_+, R_-)} \max\{0, \frac{d_+}{d_+ - d_-} - \delta\} + \lambda \|Net(\cdot)\|_2$$

[0073]

[0074] 여기서, R_q 는 질의 지역을 나타내고, R_+ 는 질의 지역과 유사한 후보 지역을 나타내며, R_- 는 질의 지역과 유사하지 않은 후보 지역을 나타낸다. 또한, $\lambda \|Net(\cdot)\|_2$ 는 L_2 정규화 항(regularization term)을 나타낸다. 결과적으로 지역 유사성은 아래 <수학식7>과 같이 계산될 수 있다.

수학식 7

$$rSim(R_1, R_2) = \frac{1}{1 - \|v_1 - v_2\|_2}$$

[0075]

[0076] 여기서, v_1 은 $Net(\cdot)$ 에 의해 학습된 유사 후보 지역 R_1 의 특징 벡터를 의미하고, v_2 는 $Net(\cdot)$ 에 의해 학습된 R_q 와 유사하지 않은 후보 지역 R_2 의 특징 벡터를 의미한다.

[0077] Triplet network를 이용하여 지역의 유사성을 학습하기 위해서는 지역 데이터를 적절한 형태로 표현시켜야 한다. 본 발명의 장치에서는 지역 R 을 3차원 텐서 $\mathbb{R}^{w \times h \times s}$ 로 나타내고, 여기서 w 와 h 는 공간 크기(spatial dimension)을 나타내며, s 는 계층적 유형과 내부 공간 엔티티의 의미적 관련성에 대한 특징 크기를 나타낸다.

[0078] w 와 h 를 생성하기 위해, 본 발명의 장치는 먼저 지역 R 을 색인(index)하고, 지역 R 에 속하는 공간 엔티티들 각각 공간 엔티티의 지리적인 좌표와 관련하여 색인하기 위해 맵핑할 수 있다. 일 예로, 공간 엔티티의 색인 방법은 공간 타일링(space tiling) 방법을 이용할 수 있는데, 이러한 공간 타일링 방법은 이미 널리 공지된 기술에 해당하므로 이에 대한 자세한 설명은 생략한다.

[0079] 공간 타일링 방법을 통해 공간 엔티티들을 색인하는 방법에 대하여 설명한다. 먼저, 본 발명의 장치는 길이(w_{grid})와 너비(h_{grid})의 크기에 따라 동일한 길이로 2차원 공간을 격자형태로 나눈다. 그리고 위와 같은 격자들은 벡터($i \in (\mathbb{N}_w \times \mathbb{N}_h)$)들에 의해 인덱싱됨으로써 경도와 위도와 같은 주소 정보를 가지고 있는 버킷(bucket)으로 사용될 수 있다.

[0080] 따라서, 본 발명의 장치는 s 크기를 통해 R 지역의 각 격자 공간 내부에 임베딩된 공간 엔티티들을 합산 처리한다.

[0081] 도6은 도1 내지 도5를 참고하여 상술한 바와 같은 동작 과정을 수행하는 공간 엔티티의 지역 유사성을 측정하는 장치를 예시한 블록도이다. 공간 엔티티의 지역 유사성을 측정하는 장치(600)는 적어도 하나의 프로세서(610), 컴퓨터 판독 가능한 저장매체(620) 및 통신 버스(660)를 포함한다.

[0082] 프로세서(610)는 공간 엔티티의 지역 유사성을 측정하는 장치(600)로 동작하도록 제어할 수 있다. 예컨대, 프로세서(610)는 컴퓨터 판독 가능한 저장 매체(620)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 컴퓨터 실행 가능 명령어는 프로세서(610)에 의해 실행되는 경우 공간 엔티티의 지역 유사성을 측정하는 장치(600)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.

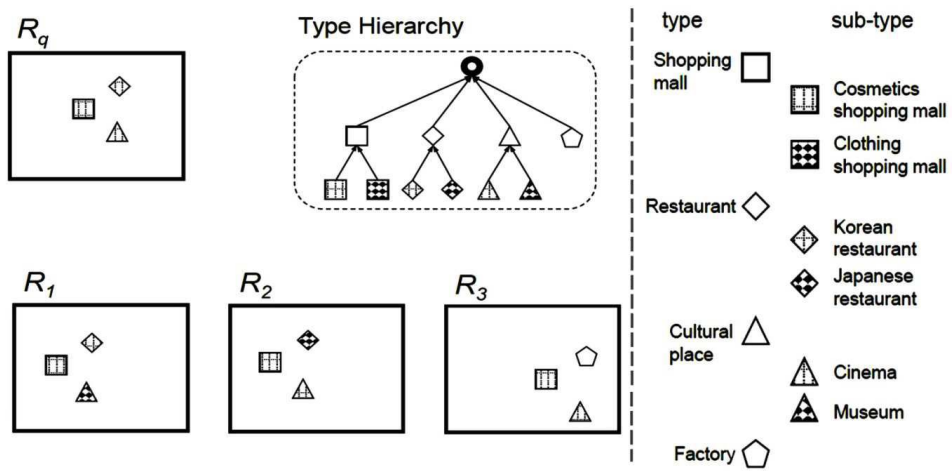
[0083] 컴퓨터 판독 가능한 저장 매체(620)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능한 저장 매체(620)에 저장된 프로그램(630)은 프로세서(610)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독한 가능 저장 매체(620)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 공

간 엔티티의 지역 유사성을 측정하는 장치(600)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

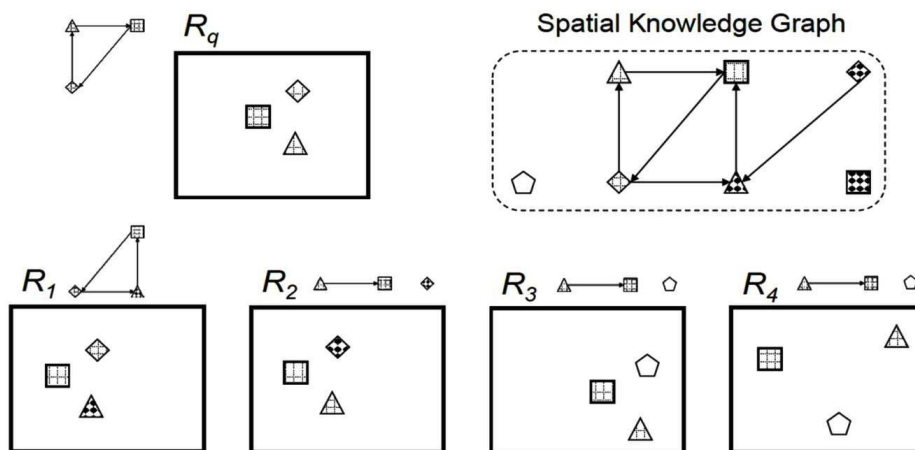
- [0084] 통신 버스(660)는 프로세서(610), 컴퓨터 판독 가능한 저장 매체(620)를 포함하여 공간 엔티티의 지역 유사성을 측정하는 장치(600)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0085] 공간 엔티티의 지역 유사성을 측정하는 장치(600)는 또한 하나 이상의 입출력 장치를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(640) 및 하나 이상의 통신 인터페이스(650)를 포함할 수 있다. 입출력 인터페이스(640) 및 통신 인터페이스(650)는 통신 버스(660)에 연결된다. 입출력 장치(미도시)는 입출력 인터페이스(640)를 통해 공간 엔티티의 지역 유사성을 측정하는 장치(600)의 다른 컴포넌트들에 연결될 수 있다.
- [0086] 도7은 공간 엔티티의 지역 유사성을 측정하는 장치(600, 700)의 프로세서(610)의 구성을 개략적으로 도시한 블록도이다. 도7에 도시된 바와 같이 공간 엔티티의 지역 유사성을 측정하는 장치(700)의 프로세서(620)는 공간 엔티티 임베딩부(710) 및 지역 유사성 측정부(720)를 포함하여 구성될 수 있다.
- [0087] 본 발명의 실시예에 따른 공간 엔티티 임베딩부(710)는 질의 지역 및 관심 지역들에 속하는 공간 엔티티들에 대하여 미리 정의된 계층적 유형 정보 및 상기 공간 엔티티들간 구조적인 관계에 따른 구조적 관계 정보를 기반으로 상기 공간 엔티티들을 벡터 공간으로 임베딩할 수 있다. 공간 엔티티 임베딩부(710)의 동작 구성은 도3 내지 도4를 참조하여 공간 엔티티를 임베딩하는 단계에 대하여 상술한 내용에 대응된다.
- [0088] 그리고, 본 발명의 지역 유사성 측정부(720)는 상기 임베딩된 공간 엔티티 및 상기 임베딩된 공간 엔티티에 대한 지리적 정보를 이용하여 상기 관심 지역들에 대한 지역 유사성을 측정할 수 있다. 지역 유사성 측정부(720)의 동작 구성은 도4 내지 도5를 참조하여 지역 유사성 학습과정에 대하여 상술한 내용에 대응된다.
- [0089] 도8은 본 발명의 일 실시예에 따른 공간 지식 그래프를 이용하여 공간 엔티티의 지역 유사성을 측정하는 방법의 시간의 흐름에 따라 도시한 흐름도이다.
- [0090] 도8을 참조하면, 먼저 S810 단계에서 공간 엔티티 임베딩부는 질의 지역 및 관심 지역들에 속하는 공간 개체들에 대하여 미리 정의된 계층적 유형 정보 및 상기 공간 개체들간 구조적인 관계에 따른 구조적 관계 정보를 기반으로 상기 공간 개체들을 벡터 공간으로 임베딩한다. S810 단계는 도3 내지 도4의 참조설명에서 상세하게 설명한 바 자세한 설명은 생략한다.
- [0091] 그리고, S820 단계에서 지역 유사성 측정부가 상기 계층적 유형 정보 및 상기 구조적 관계 정보를 포함하는 상기 임베딩된 공간 개체 및 상기 임베딩된 공간 개체에 대한 지리적 정보를 이용하여 상기 관심 지역들에 대한 지역 유사성을 측정한다. S820 단계는 도4 내지 도5의 참조설명에서 상세하게 설명한 바 자세한 설명은 생략한다.
- [0092] 이상에서 설명한 본 발명의 실시예를 구성하는 모든 구성요소들이 하나로 결합하거나 결합하여 동작하는 것으로 기재되어 있다고 해서, 본 발명이 반드시 이러한 실시예에 한정되는 것은 아니다. 즉, 본 발명의 목적 범위 안에서라면, 그 모든 구성요소들이 하나 이상으로 선택적으로 결합하여 동작할 수도 있다. 또한, 그 모든 구성요소들이 각각 하나의 독립적인 하드웨어로 구현될 수 있지만, 각 구성요소들의 그 일부 또는 전부가 선택적으로 조합되어 하나 또는 복수개의 하드웨어에서 조합된 일부 또는 전부의 기능을 수행하는 프로그램 모듈을 갖는 컴퓨터 프로그램으로서 구현될 수도 있다. 또한, 이와 같은 컴퓨터 프로그램은 USB 메모리, CD 디스크, 플래쉬 메모리 등과 같은 컴퓨터가 읽을 수 있는 기록매체(Computer Readable Media)에 저장되어 컴퓨터에 의하여 읽혀지고 실행됨으로써, 본 발명의 실시예를 구현할 수 있다. 컴퓨터 프로그램의 기록매체로서는 자기 기록매체, 광 기록매체 등이 포함될 수 있다.
- [0093] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위 내에서 다양한 수정, 변경 및 치환이 가능할 것이다. 따라서, 본 발명에 개시된 실시예 및 첨부된 도면들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예 및 첨부된 도면에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구 범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리 범위에 포함되는 것으로 해석되어야 할 것이다.

도면

도면1



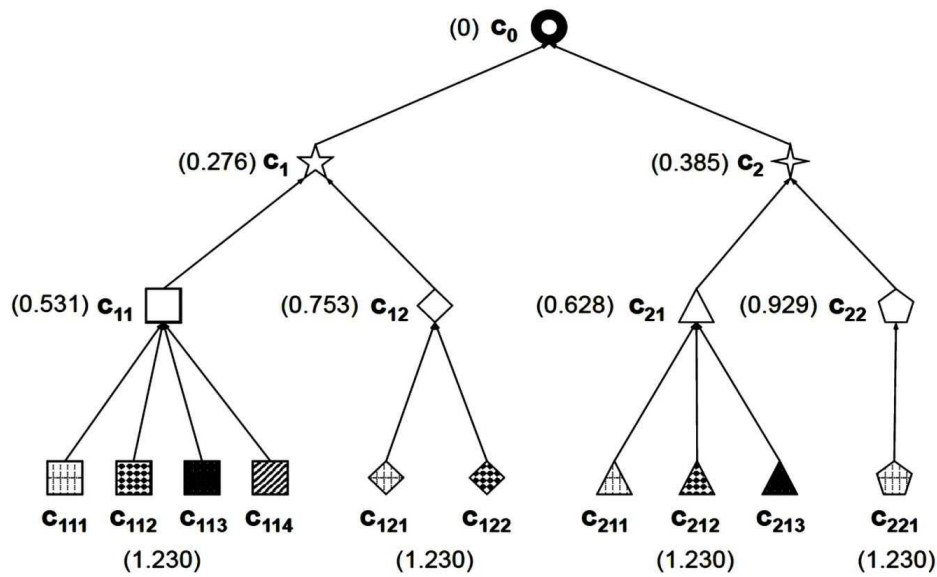
도면2



도면3

Symbol	Description
mh	a threshold for the max number of hops
D_{ht}	a hierarchical type and semantic relatedness aware distance between h and t
C_{ht}	a hierarchical type aware distance between h and t
S_{ht}	the semantic relatedness between h and t
h_e, t_e	the vector representations of h and t
γ	a distancing parameter for C_{ht}
$cSim(h, t)$	the hierarchical type similarity between h and t
$IC(\cdot)$	the information content of a type
$Sub(h, t)$	a set of types that subsumes both the types of h and t
$p(c)$	probability of encountering c w.r.t. a type hierarchy
$hop(h, t)$	the number of hops that t is away from h
θ	a narrowing rate for S_{ht}
\mathcal{L}_e	a distance based loss function for spatial entity embedding

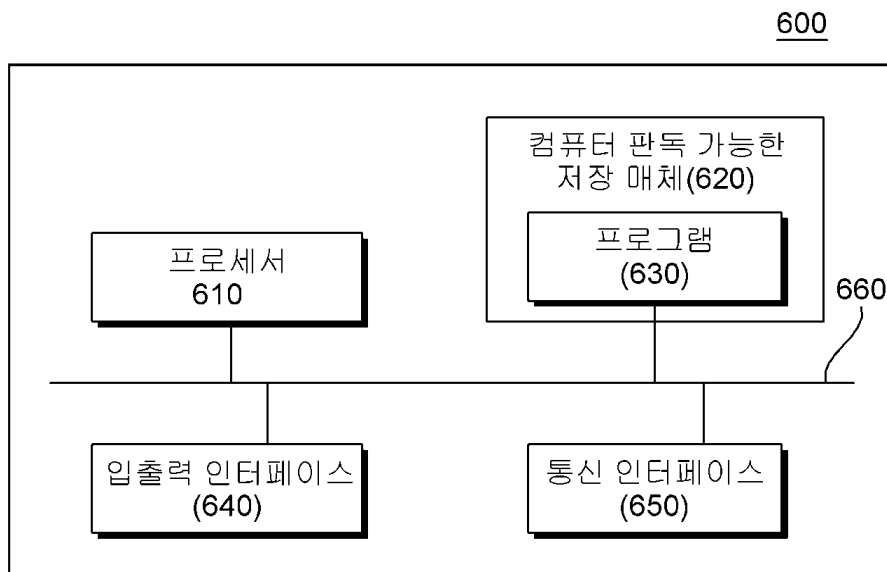
도면4



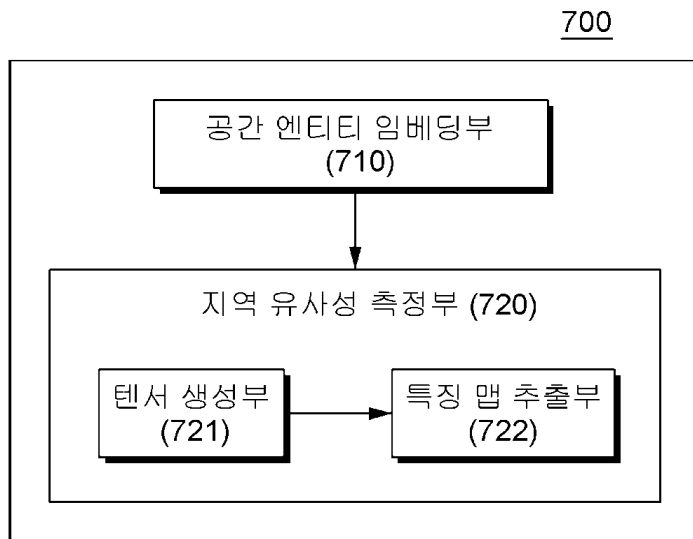
도면5

Symbol	Description
$Net(\cdot)$	a shared convolutional neural network
R_+/R_-	a region that is similar/dissimilar to R_q
d_+/d_-	the Euclidean distance between the vector representations of R_q and R_+/R_-
\mathcal{L}_t	a ratio-based loss function for learning region similarity
δ	a gap parameter for \mathcal{L}_t
λ	a regularization parameter for \mathcal{L}_t
wid_{grid}	the width of a grid
len_{grid}	the length of a grid
Δ	the granularity factors for space tiling
$rSim(\cdot)$	the region similarity measure
v_i	the feature vector of region R_i learned by $Net(\cdot)$

도면6



도면7



도면8

