모바일 환경에서 딥러닝을 활용한 의미기반 이미지 어노테이션을 위한 이미지 태그 설계 및 구현

신윤미*, 안진현**, 임동혁***
*호서대학교 컴퓨터공학과
**제주대학교 경영정보학과
***호서대학교 컴퓨터정보공학부
e-mail : dhim@ hoseo.edu

Design and Implementation of Deep-Learning-Based Image Tag for Semantic Image Annotation in Mobile Environment

YoonMi Shin*, Jinhyun Ahn**, Dong-Hyuk Im***
*Department of Computer Engineering, Hoseo University
**Department of Management Information Systems, Jeju National University
***Division of Computer and Information Engineering, Hoseo University

요 약

모바일의 기술 발전과 소셜미디어 사용의 증가로 수없이 많은 멀티미디어 콘텐츠들이 생성되고 있다. 이러한 많은 양의 콘텐츠 중에서 사용자가 원하는 이미지를 효율적으로 찾기 위해 의미 기반 이미지 검색을 이용한다. 이 검색 기법은 이미지에 의미 있는 정보들을 이용하여 사용자가 찾고 자하는 이미지를 정확하게 찾을 수 있다. 본 연구에서는 모바일 환경에서 이미지가 가질 수 있는 의미적 정보를 어노테이션 하고 이와 더불어 모바일에 있는 이미지에 풍성한 어노테이션을 위해 딥러닝 기술을 이용하여 다양한 태그들을 자동 생성하도록 구현하였다. 이렇게 생성된 어노테이션 정보들은 의미적 기반 태그를 통해 RDF 트리플로 확장된다. SPARQL 질의어를 이용하여 의미 기반 이미지 검색을 할 수 있다.

1. 서론

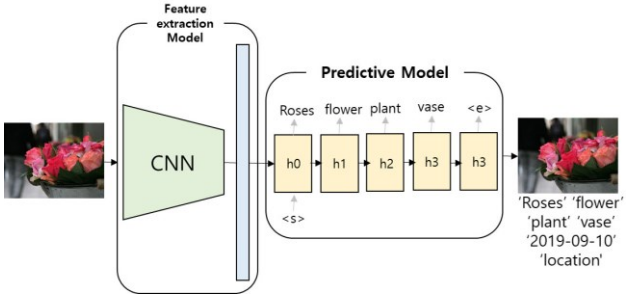
모바일 시장의 발달과 다양한 소셜 미디어의 활용 증가로 인하여 대량의 이미지 콘텐츠가 증가하고 있다. 따라서 이러한 많은 양의 이미지를 저장하고 관리하는 것이 중요하다[1,2]. 이를 위해 많은 양의 이미지 콘테츠 속에서 효율적으로 검색 할 수 있는 이미지 어노테이션 기법이 제안되었다[3,4]. 이미지 어노테이션 기법은 이미지가 가지고 있는 의미 정보들을 이미지와 함께 저장하여 이미지를 다양하게 표현하는 기법이다. 이 기법을 통해 많은 양의 이미지 데이터 속에서 사용자가 원하는 이미지를 정확히 찾을 수 있다.

이전 연구에서는 모바일 환경에서 얻은 어노테이션 정보를 이미지와 함께 저장한다[5]. 모바일 환경에서 이용할 시 이미지에 대한 상황 정보와 사용자가 입력한 태그 정보들을 이용하여 어노테이션을 한다. 모바일 환경에서 자동으로 얻어진 상황 정보는 온톨로지 언어인 RDF 트리플의 그래프 테이터로 변환하여 어노테이션 한다. 사용자가 직접 입력한 태그들은 DBPedia 를 이용하여 RDF 트리플로 모델링 한다.

본 연구에서는 모바일 기기에서의 어노테이션 및 검색 시스템 설계[5]를 기반으로 RNN(Recurrent Neural Network)[6]의 형태 중 One to Many 방식의 딥러닝 모델을 이용한다. 이 방식은 이미지 캡션 생성에서 많이 사용되는 모델로서 하나의 입력값으로 여러 개의 출력값을 얻을 수 있다. 이 모델을 추가하여 입력 이미지에 관련된 태그들을 자동으로 생성되는 형태로 확장하였다.

2. 제안한 시스템

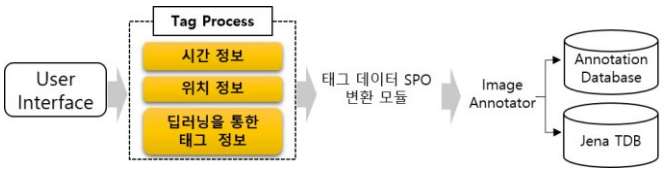
본 연구에서는 모바일 환경, 이미지 어노테이션 기술에 CNN(Convolutional Neural Network)과 Multimodal RNN(Multimodal Recurrent Neural Network)을 결합한 모델[7]을 추가로 사용한다.



(그림 1)이미지 관련 태그 자동 생성 과정

(그림 1)은 입력 이미지와 관련된 자동 태그 생성 과정을 보여준다. 먼저 이미지가 입력값으로 들어가게 되면 CNN 을 이용하여 이미지의 특징을 추출하게 된다. 추출된 특징맵은 1 차원의 형태로 변형되는데 이 형태가 기존 CNN 모델의 완전 연결 계층이다. 완전 연결 계층은 RNN 의 히든 레이어의 초기값으로 들어가게 된다. RNN 의 첫번째 셀의 입력값은 문자토큰으로 <s>가 들어가게 된다. 이를 시작으로 이미지와 관련된 태그들을 예측하게 된다. <e>라는 문자토큰이 나오게 되면 예측은 종료된다. 이렇게 얻어진 태그들과 모바일 환경에서 시간, 장소 태그들까지 통합되어 이미지 태그 정보들을 갖는다.

위 과정을 추가한 전체 시스템 구조는 (그림 2)에서 보여준다.



(그림 2)전체 시스템 구조

사용자가 모바일 환경에서 원하는 이미지를 입력 한다. 입력 이미지에 대해 Tag Process 과정을 통해 시간, 위치 정보와 딥러닝 모델을 통해 태그 정보들을 자동으로 생성된다. 태그 데이터 SPO 변환 모듈에서 생성된 태그 정보들을 이용하여 어노테이션 RDF 트리플형태로 확장한다. 어노테이션 한 데이터는 MySQL 과 Jena TDB 에 저장된다. SPARQL 질의 시 저장 된 이미지 ID 를 찾기 위해 RDF 트리플을 사용하여 네임드 그래프 형태로 인덱싱 하여 Jena TDB 에 저장한다.

3. 시스템 구현

(그림 3)사진 촬영 (그림 4)이미지 태그 리스트

(그림 3)은 모바일 기기로 사진 찍기 또는 사진 등록을 하는 화면이다. 이 과정을 통해 사진을 입력하게 되면 (그림 4)처럼 딥러닝 과정을 통해 사진에 대해 자동으로 관련된 태크 정보 리스트를 보여준다. 추가로 사용자가 수동으로 이미지에 대한 태그를 입력할 수도 있다.

4. 결론

모바일 환경의 발전으로 이미지 콘텐츠가 증가하여 의미적인 이미지 검색이 중요해졌다. 본 연구에서는 모바일 환경에서 딥러닝 모델 중 CNN 과 Multimodal RNN 을 이용하여 이미지에 대한 태그 정보들을 자동으로 생성한다. 따라서 사용자가 직접 이미지에 태그를 입력하는 번거로운 단점을 보완한다. 또한 자동으로 생성 된 태그를 이용하여 풍부한 어노테이션을 구성 한다.

향후 과제로는 DBPedia 의 술어 정보를 사용하지 않고 태그 간에 술어 정보를 연결 할 수 있도록 딥러닝 모델을 이용하여 이미지에 대한 의미적 기반 태그를 자동으로 할 수 있도록 할 계획이다.

Acknowledgement

이 논문은 2017 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구 (No.NRF-2017R1C1B1003600)이며, 2018 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2018R1D1A1B07048380).
또한, 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2019-2018-0-01417).

참고문헌

[1]노승민, and 황인준. "멀티미디어 검색 시스템의 설계 및 구현." 정보과학회논문지: 데이타베이스 30.5 (2003): 494-506.
[2]이오준, et al. "소셜 빅데이터를 이용한 영화 흥행 요인 분석." 한국콘텐츠학회논문지 14.10 (2014): 527-538.
[3]Im, D. H., Park, G. D.: Linked tag: image annotation using semantic relationships between image tags. Multimed Tools Appl. April 2015, vol. 74, Issue 7, pp2273-2287 (2015)
[4]Im, Dong-Hyuk, and Geun-Duk Park. "STAG: semantic image annotation using relationships between tags." 2013 International Conference on Information Science and Applications (ICISA). IEEE, 2013.
[5]노현덕, 서광원, and 임동혁. "모바일 환경에서 의미 기반 이미지 어노테이션 및 검색." 멀티미디어학회논문지 19.8 (2016): 1498-1504.
[6]Mikolov, Tomáš , et al. "Recurrent neural network based language model." Eleventh annual conference of the international speech communication association. 2010.
[7]Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating

image descriptions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.