검색어 생성을 위한 딥 러닝 기반 문장 분석 연구

*나성원 **윤경로
건국대학교
*securityin4@naver.com

Deep Learning based Sentence Analysis for Query Generation

*Seong-Won, Na **Kyoungro, Yoon
Konkuk University

요약

최근 이미지의 Visual 정보를 추출하고 Multi label 분류를 통해 나온 결과의 상관관계를 modeling하여 문장으로 출력하는 CNN-RNN 아키텍처가 많은 발전을 이뤘다. 이 아키텍처의 출력은 이미지의 정보가 요약되어 문장으로 표현되기 때문에 Semantic정보가 풍부하여 유사 콘텐츠 검색에도 사용 가능하다. 하지만 결과 문장에 사람이 포함 되면 광범위한 검색 결과를 얻게 되고 부정확한 결과를 초래하게 된다.

이에 본 논문에서는 문장에서 사람을 인식하여 Identity를 부여함으로써 검색어를 좀 더 구체적으로 생성하고자 한다. 이 문제를 해결하기 위해 자연어 처리의 분야 중 하나인 개체명 인식(Named Entity Recognition) 문제로 다루며, 가장 많이 사용되고 있는 모델인 Bidirectional-LSTM-CRF와 CoNLL2003 dataset을 사용하여 수행 한다.

1. 서론

컴퓨터 비전 분야에서 이미지를 단순히 인식, 분류하는 문제는 ILSVRC15 dataset에 대해 Top5 error rate 3.08%를 달성하게 되면서 사람을 능가하는 성능에 도달하였다. 그에 따라 자연스럽게 단순 인식, 분류 문제를 다루는 것이 아닌 이미지내의 모든 object, properties, attributes, action 등을 처리하는 Multi-label classification, tagging, captioning과 같은 이미지 속성에 대해 보다 풍부한 설명을 생성하는 문제에 관심이 증가하였고, 많은 발전이 이루어 졌다[1]. 이 같은 문제를 다룰 때 가장 많이 사용되는 기술은 CNN-RNN구조로 CNN은 이미지의 feature를 추출해 RNN의 입력으로 사용하고, RNN은 concept prediction과 label-correlation modeling을 수행하여 정렬 된 문장을 생성하는 구조이다. 이 결과로 나온 문장은 이미지의 semantic한 정보를 담고 있기 때문에 유사 콘텐츠를 검색 시 검색어로 사용 가능하다. 예를 들어 “a man is playing tennis on a tennis court”와 같은 문장이 결과로 출력되면 특정한 선수가 아닌 남성이 테니스를 하고 있는 비교적 큰 category를 의미하는 검색어가 될 것이다. 위와 같은 문장을 보다 구체적인 검색어로 사용하기 위해 문장에서 사람을 인식하고, identity를 추가적으로 포함한 검색어를 생성하고자 한다. 여기에서 identity라고 하면 위 문장에서 ‘man’이라는 단어를 인식하여, ‘정현’이라는 고유 명사로 대체하는 것을 의미 한다.

자연어 처리의 task 중 하나인 개체명인식(Named Entity Recognition)은 문장에서 나타나는 고유한 의미를 가지는 명사를 인식하는 것으로 주로 4Class인 인명(Person), 지명(Location), 기관명(Organization), 기타(MISC)로 나눌 수 있다[2]. 이 NER 문제를 처리하는데 많이 사용되는 모델 중 하나인 Bidirectional-LSTM-CRF모델은 여러 분야에서 각광 받고 있는 Long Short-term Memory Network(LSTM)[3]기반 RNN이며, 기존의 문제였던 gradient vanishing문제를 해결하기 위해 제안 되었고, Sequence한 문제에서 강력한 성능을 보여 언어 모델, 음성 인식, 자연어 이해등과 같은 분야에서 많이 사용 되고 있다. 문장 분석은 위에서 설명한 모델을 사용하고, dataset으로는 CoNLL2003을 사용하여 학습 하였다.

2. 본론

2.1 Dataset

CoNLL2003의 NER dataset은 문장 단위로 구성 되어 있고, 문장을 구성하는 단어와 label이 쌍으로 이루어져 있으며, Network로 입력 시 하나의 쌍이 입력되는 구조이다. 하지만 우리가 원하는 사람을 지칭하는 Woman, Man, Player, Singer등과 같은 단어는 PER이라는 label이 달려 있지 않기 때문에 필요한 단어의 label을 직접 수정 하여 훈련을 진행 하였다. 우리 시스템의 목적은 사람을 인식하는 것이기 때문에 다른 label은 수정하지 않고 사용 되었다.

2.2 Bidirectional LSTM CRF 모델

RNN은 sequence data를 처리하는데 사용하는 Neural Network 이며, 이론적으로는 긴 의존성을 배울 수 있지만 실제로는 가장 최근 입력에 편향되는 경향이 있다. Long Short-Term Memory Network는 메모리 셀을 통하여 이 문제를 해결하기 위해 설계 되었으며 장거리 종속성을 포착가능 하다. RNN의 gradient vanishing 문제를 해결한 LSTM RNN은 다음과 같이 정의 된다.

$$i_t = \sigma(W_{xt}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i$$
$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f$$
$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o)$$
$$h_t = o_t \tanh(c_t)$$
$$y_t = g(W_{hy}h_t + b_y)$$

위 식에서 σ는 sigmoid 함수이고, I, f, o, c는 각각 input gate, forget gate, output gate, memory cell vector를 나타내며 각 벡터의 크기는 hidden layer 벡터 크기와 같다.

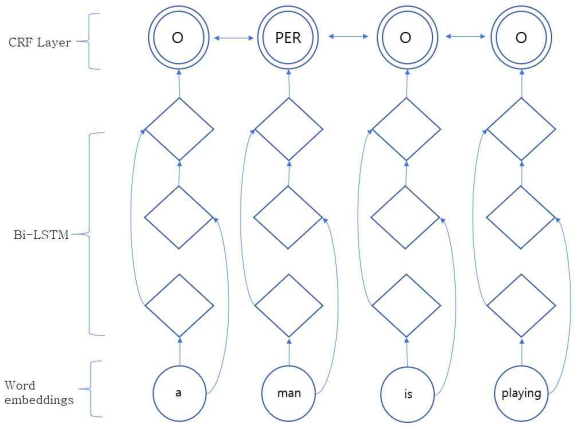

그림 1. Bidirectional LSTM CRF 구조[4]

그림 1은 Bidirectional LSTM CRF모델의 구조를 나타낸다. 기존 LSTM CRF와 달리 양방향으로 학습되기 때문에 현재 label 결정에 이전단어와 다음단어의 정보를 모두 볼 수 있다. Bidirectional-LSTM CRF 모델의 학습을 위해 Stochastic Gradient Descent(SGD)[5]알고리즘을 사용한다.

3. 실험

본 논문에서는 Bidirectional-LSTM-CRF모델에 사용한 영어 word embedding은 Collobert & Westone의 단어 표현을 사용하였고, feature embedding은 평균 0, 분산 0.01이 되도록 무작위로 초기화 시킨 값을 사용 하였다. 과적합 문제를 완화하기 위해 dropout rate는 0.5로 적용하였고, LSTM hidden layer의 dimension은 100으로 설정하여 진행 하였다. 문장에서 사람을 인식하는 것이 목적이기 때문에 dataset을 수정해서 training을 진행 하였고, 평가를 위해 CoLL2003 NER test dataset이 아닌 자체 dataset을 통해 평가 하였다. 이 dataset은 인터넷상에서 수집하였으며, 평가할 문장들은 사람이 포함 된 이미지를 CNN-RNN 아키텍처에 입력으로 사용하여, 처리 후 나온 출력 문장을 사용하였다.

표 1은 training dataset의 tag 변경 전, 후에 대해 학습한 model의 출력 결과를 보여준다. 사람을 표현하는 단어 전체가 아닌 특정단어에 대해 training set을 수정하였기 때문에 93% 이상의 높은 precision이 측정 되었다.

표 1. tag 변경 전과 후의 결과 비교

time	word	변경 전 tag	변경 후 tag
1	A	O	O
2	Man	O	B-PER
3	is	O	O
4	playing	O	O
5	Tennis	O	O
6	on	O	O
7	a	O	O
8	Tennis	B-LOC	B-LOC
9	Court	I-LOC	I-LOC

4. 결론

우리는 CNN-RNN 아키텍처의 출력인 Semantic한 문장을 좀 더 구체적인 query로 생성하기 위해 Bidirectional-LSTM-CRF 모델을 사용하여 문장내의 사람을 인식하는 실험을 진행 하였다. 특정 단어에 대해서만 label을 수정하였기 때문에 정확도는 높게 나왔지만 dataset을 더 크게 구성할 수 있다면 NER task에서 좀 더 넓은 범위의 person 인식을 할 수 있을 것으로 판단된다. 하지만 기존 Bidirectional LSTM CRF모델을 사용하여 실험하였기 때문에 기존에 person으로 잘못 인식한 단어들까지 identity를 갖게 되는 단점을 발견 하였다. 예를 들면 “a Woman is holding a Teddy Bear.”란 문장에서 ‘Woman’과 ‘Teddy Bear’ 두 단어가 PER로 인식 되면서 두 단어에 identity가 들어가게 되어 “IU is holding a IU”란 문장이 결과로 도출 되었다. 이러한 문제점을 해결하기 위해 추후 연구에서는 문장에서 하나의 술어 당 하나의 person만 인식하는 방법을 연구할 계획이다.

감사의 글

이 논문은 2018년 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (UHD 방송콘텐츠 기반 지능형 Dynamic Media 생성, 분배 및 소비 기술 개발)

참고문헌

[1] Liu, Feng, et al. "Semantic Regularisation for Recurrent Image Annotation." arXiv preprint arXiv:1611.05490 (2016).

[2] Huang, Zhiheng, Wei Xu, and Kai Yu. "Bidirectional LSTM-CRF models for sequence tagging." arXiv preprint arXiv:1508.01991 (2015).

[3] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[4] Lample, Guillaume, et al. "Neural architectures for named entity recognition." arXiv preprint arXiv:1603.01360 (2016).

[5] Bottou, Léon. "Large-scale machine learning with stochastic gradient descent." Proceedings of COMPSTAT'2010. Physica-Verlag HD, 2010. 177-186.