

멀티 모달 기반의 뉴럴 네트워크를 이용한 SNS 콘텐츠 감정 분석

Sentiment analysis for SNS contents using multi-modal neural network

저자 (Authors)	김시진, 유준형, 손경아 Si-Jin Kim, Jun-Hyung Yu, Kyung-Ah Sohn
출처 (Source)	한국통신학회 학술대회논문집 , 2018.6, 886-887(2 pages) Proceedings of Symposium of the Korean Institute of communications and Information Sciences , 2018.6, 886-887(2 pages)
발행처 (Publisher)	한국통신학회 Korea Institute Of Communication Sciences
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07512781
APA Style	김시진, 유준형, 손경아 (2018). 멀티 모달 기반의 뉴럴 네트워크를 이용한 SNS 콘텐츠 감정 분석. 한국통신학회 학술대회논문집, 886-887
이용정보 (Accessed)	가천대학교 203.249.***.201 2019/09/29 19:00 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

멀티 모달 기반의 뉴럴 네트워크를 이용한 SNS 콘텐츠 감정 분석

김시진, 유준형, 손경아
아주대학교

{sijin0306, wnsquddb, kasohn}@ajou.ac.kr

Sentiment analysis for SNS contents using multi-modal neural network

Si-Jin Kim, Jun-Hyung Yu, Kyung-Ah Sohn
Ajou Univ.

요 약

본 논문에서는 소셜 네트워크 서비스(SNS) 콘텐츠에 나타나는 이미지와 텍스트 데이터를 복합적으로 사용하여 콘텐츠에 담겨있는 인간의 감정을 분석하는 멀티 모달 기반의 뉴럴 네트워크 모델을 구현하였다. 텍스트를 기반으로 하여 인간의 감정을 예측하였던 기존의 감정 분석 방법과 달리, 텍스트 타입과 이미지 타입의 데이터를 동시에 사용함으로써 보다 정확한 감정 분석을 돕는다. 연구에 사용한 SNS 콘텐츠는 Multilingual Visual Sentiment Ontology(MVSO)에 의해 제공되고 있다. Adjective Noun Pairs(ANP) sentiment score 이론을 적용하여 지도 학습 환경에서 모델 학습을 진행하였다. 텍스트 데이터와 이미지 데이터의 각 특성을 고려하여 텍스트 데이터에는 Long Short-Term Memory(LSTM)를, 이미지 데이터에는 Convolutional Neural Network(CNN)을 적용하였다. 텍스트와 이미지 데이터의 각 뉴럴 네트워크에서 1 차 feature vector 를 추출한 뒤, 이를 concatenation 하여 텍스트와 이미지 정보를 복합적으로 담고 있는 new feature vector 를 생성하였다. 이를 fully-connected layer 로 전달하여 멀티 모달 feature 데이터를 학습시켰고, 단일 타입의 데이터만을 사용하는 경우보다 멀티 타입의 데이터를 활용한 경우의 감정 분석 정확도가 상승하는 것을 확인할 수 있었다.

I. 서 론

인터넷과 스마트폰의 발전과 함께 SNS 를 사용하여 개인의 의사를 표명하고 대화를 나누는 것은 물론 타인과 감정을 공유하는 일이 더욱 빈번하게 발생하고 있다. SNS 를 사용하는 일이 많아지면서 이의 형태 역시 발전하였는데, 텍스트 형태의 단일 타입의 데이터를 장문으로 업로드하던 초기 형태에서 벗어나 이미지 혹은 동영상을 중심으로 하여 짧은 텍스트를 덧붙여 업로드하는 형태로 나아가고 있다. SNS 가 다양한 타입의 데이터를 담게 되면서 사용자의 감정을 더욱 풍부하게 드러낼 수 있게 되었고, 이는 감정 분석 분야에 있어서 중요한 의미를 갖는다. SNS 콘텐츠에 드러난 감정을 분석하여 마케팅에 사용하거나, 개인 맞춤형 시스템을 도입하는 등 다양한 분야에 그 결과를 적용할 수 있기 때문이다. 그런데 기존의 감정 분석은 텍스트 또는 이미지의 단일 데이터만을 기반으로 하여, 감정을 분석해내는 데 한계를 갖는다.

이에 본 논문에서는 최근 들어 다양한 분야에서 강력한 퍼포먼스가 입증되고 있는 딥러닝 모델을 사용하여 멀티 모달 네트워크 기반의 감정 분석을 진행하였다.

II. 본론

2.1 데이터 탐색

본 논문에서는 Columbia Univ.에서 제공하는 MVSO[4] 데이터를 기반으로 연구를 진행하였다. MVSO 데이터셋은 Visual Affect Around the World[1]에서 소개된 SNS 콘텐츠 기반의 데이터셋으로, 15,600 개의 감점 컨셉을 기반으로 이미지가 가지고 있는 감정을 분석하는 detector 의 학습에 사용되었다. 태그 등의 텍스트 데이터를 바탕으로 감정 컨셉(ANP)을 이미지에 할당하고 이를 기반으로 해당 콘텐츠에 sentiment score 를 부여하는 프레임워크를 갖고 있다. <Figure 1>은 Flickr(www.flickr.com)의 게시물로 텍스트와 이미지가 쌍을 이루고 있다. xml 형태로 제공되는 정보 중 'raw_text', 'token', 'image_url' 등 분석에 필요한 정보를 뽑아 페어링되는 이미지를 key 값으로 하여 딕셔너리 형태로 가공하였다. 추가로 ANP sentiscore 데이터와 결합시켜 sentiscore 가 0 보다 작은 경우 부정, sentiscore 가 0 과 같거나 큰 경우 긍정으로 labeling 하였다. (ANP value 범위: $-2 < \text{ANP} < 2$) 10,000 개의 데이터 중 긍정 콘텐츠는 5,114 개, 부정 콘텐츠는 4886 개이다.



Figure 1 기본 데이터셋 예시

2.2 데이터 전처리

비정형의 SNS 데이터를 학습하기 위해 전처리를 진행하였다. 게시물의 제목과 내용이 중복되지 않도록 텍스트를 가져온 후, 연구 결과를 명확히 하기 위해, 텍스트를 구성하는 단어의 개수가 5 개 보다 많고 150 개 보다 작은 샘플의 경우만을 대상으로 하였다. Numpy[5], Pandas[6] 그리고 nltk[5] 패키지를 사용하여 영문 텍스트 데이터를 소문자로 통일하고 알파벳을 제외한 기호들과 불용어를 모두 제거하였다. 이미지 데이터는 크롤링을 통해 수집하였으며, (64,64,3) 형태로 resizing 을 진행하였다. 이미지 데이터의 sentiment labeling 은 쌍을 이루는 텍스트의 sentiment 값을 사용하였다.

2.3 LSTM network 와 CNN 모델 학습

모델을 구성하는 데 파이썬 keras[7] 패키지를 사용하였다. keras[7]의 Tokenizer 를 사용하여 텍스트를 토큰화하고 빈도수가 높은 20000 개의 단어에 대해서만 indexing 하여 학습을 진행하였다. training set 의 크기는 6,700, validation set 의 크기 1,500, test_set 의 크기 1,800 으로 나눠 각각 모델 학습, 평가, 테스트에 사용하였다. 텍스트와 페어링되는 이미지에 대하여 LSTM 모델을 구성할 때와 동일한 패키지를 사용하여 CNN 모델을 구성하였다. CNN 의 성능을 향상시키기 위해 pretrained model VGG16 을 적용하였으나, 정확도에 큰 영향을 미치지 않아 본 연구의 트레이닝 데이터를 사용하여 모델을 구성하였다. 데이터셋의 크기 구성은 텍스트 데이터와 동일하다.

2.4 Multi-modal neural network 구성

LSTM 과 CNN 에서 각각 텍스트 feature 와 이미지 feature 를 추출하여 128 차원으로 flattening 하는 과정을 거친 후 feature concatenation 을 진행하였고, 생성된 256 차원의 feature 를 최종 feature 로 사용하였다. 이를 3 층의 fully-connected 네트워크 모델에 input 으로 하여 최종 feature 를 학습시킨 후 테스트를 진행하였다. <figure2>는 본 실험에서 구성한 multi-modal neural network 의 구조를 보여준다.

III. 실험 결과

실험 결과 텍스트 혹은 이미지 데이터의 단일 타입 데이터만을 사용하여 ‘긍정’ 혹은 ‘부정’의 감정을 분류한 경우보다 두 데이터의 정보를 복합적으로 활용한 경우의 감정 분석 정확도가 높다는 것을 확인할 수 있었다. <table 1>에 따르면 text 혹은 image 의 단일 정보 feature 만을 사용한 경우 약 78.6%와 52.3%의 감정 분류 성능을 보인 반면, lstm 과 cnn 을 통해 추출한 text, image 결합 feature 를 사용한 경우 87.3%의 향상된 성능을 보였다.

CNN 의 경우 object detection, classification 등의 분야에서 MNIST 등의 데이터셋에 관하여 성능이 입증된 pretrained VGG16 모델을 적용해 보았으나, 감정을 기준으로 분류하는 데이터에 있어서는 50%의 정확도를 보였다.

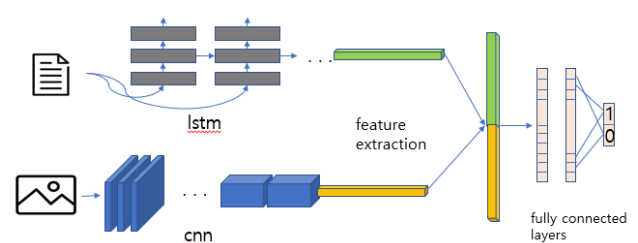


Figure 2 multi-modal neural network

Methods		Acc.
Baseline model	Text only(CNN)	0.695
	Image only(CNN)	0.631
	Text + Image (CNN feature + CNN feature)	0.847
Multimodal model	Text only(LSTM)	0.7863
	Image only(CNN)	0.5238
	Text + Image (LSTM feature + CNN feature)	0.8727

Table 1 모델 별 감정 분류 정확도

IV. 결론

자연어 처리와 감정 분석 분야에서도 딥러닝 기술이 주목되고 있고 성능을 입증하고 있으나, SNS 같은 비정형 데이터의 다양한 타입의 데이터 구성을 고려한 감정 분석 프레임이 필요하다. 본 연구에서는 위의 문제를 고려하여 진행되었으며, SNS 콘텐츠의 두가지 타입의 데이터가 감정을 분류하는 데 상호 보완적인 역할을 하고, 또한 텍스트와 이미지 각각의 네트워크 모델에서 추출한 feature 를 단순 concatenation 하는 것만으로도 감정 분류의 정확도를 약 9%(text only 비교) 가량 향상시킬 수 있음을 제시한다.

ACKNOWLEDGMENT

“본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW 중심대학지원사업의 연구결과로 수행되었음” (2015-0-00908)

참 고 문 헌

- [1] Jou, Brendan, et al. "Visual affect around the world: A large scale multilingual visual sentiment ontology." ACM, 2015.
- [2] Gers, Felix A., Jurgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM". pp.850-855. 1999.
- [3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems. 2012.
- [4] Dalmia, Vaidehi, Hongyi Liu, and Shih-Fu Chang. "Columbia MVSO Image Sentiment Dataset." arXiv preprint arXiv:1611.04455 2016.
- [5] E. Jones, T. Oliphant, P. Peterson, and others, "{SciPy}: Open source scientific tools for {Python}.".
- [6] "pandas: Python Data Analysis Library." 2012.
- [7] Chollet, Francois. "Keras". 2015