

감정 일기를 통한 6가지 감정 분류 : LSTM모델 연구

이단비, 김가영, 윤예진, 이지은
부산대학교 IT응용공학과 학부생
{sweetrain, picapie, a8002165, jieun7056}@pusan.ac.kr

Classification of 6 Emotions with Emotion Diary : LSTM Model

Dan-Bi Lee, Ga-Yeong Kim, Ye-Jin Yoon, Ji-Eun Lee
School of Applied IT & Engineering, Pusan National University

요 약

이 연구는 사람의 감정 변화를 건강하게 파악하고 분석하기 위해 시작되었다. Natural Language Processing(NLP)는 컴퓨터가 인간의 언어를 이해하기 위해 개발된 자연어 처리 기술이다. 본 논문에서는 이 기술을 이용하여 Text Mining을 통해 사용자가 작성한 일기에 담긴 감정을 분석하고 LSTM 모델과 GRU 모델을 비교군으로 두어 두 모델 중 감정 분석에 더 적합한 모델을 찾는 과정을 거쳤다. 이 과정을 정확도가 더 높은 LSTM 모델을 사용하여 감정 분석 결과를 분류하였다.

1. 서론

사람의 감정은 상황이나 환경에 따라 다양하게 변한다. 이러한 감정을 잘 살피며 살아가는 것이 사람이 건강하게 살아가는 조건이다. 이러한 감정을 스스로 살피기 힘들 때 감정을 분석해주는 모델을 이용하여 감정을 살필 수 있도록 하여 사용자의 감정을 개선하는 데 도움을 준다.

최근 chatgpt의 등장으로 사람들의 AI에 대한 관심도가 증가하고 있다. 이러한 관심도의 증가로 인해 AI 기술을 사용하여 감정 분석을 하면 어떻게 생각하게 되었고 본문에 기술된 다양한 감정 분석 기술 중에서 RNN 계열 모델을 사용하게 되었다.

2. 감정 분석 적용 기술

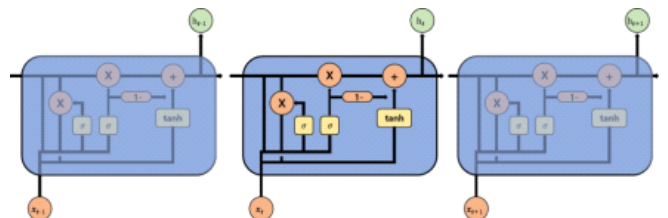
2.1 NLP(Natural Language Processing)

컴퓨터가 인간의 언어를 이해하도록 하기 위해 자연어처리(NLP) 기술[1]을 이용한다. 각 단계에서 이전 단계의 출력값을 다음 단계의 input으로 사용하여 순환적인 연산을 수행한다. 감정 분석 훈련 데이터는 문장과 그 문장이 지닌 감정으로 구성되며, 머신러닝 알고리즘은 이러한 데이터 세트를 읽고 문

장이 입력되면 그 문장의 감정을 반환하는 모델을 생성한다. 이렇게 문장이 입력되면 해당 입력값에 대한 레이블을 반환하는 종류의 모델로 자연어 처리 기술을 사용하였다.

2.2 Text Mining

'string punctuation'를 이용하여 물음표, 온점 등 불필요한 문장 부호를 사용하지 않도록 사용자가 입력한 데이터를 전처리한다. 데이터를 학습용과 테스트용으로 분류하는 과정에서 'stratify' 파라미터를 활용하여 층화추출을 통해 감정별 데이터가 골고루 포함될 수 있도록 한다. 모델 학습에는 RNN 계열 모델 중 LSTM 모델과 GRU 모델을 활용하여 정확도가 더 높은 모델을 선택하였다.



[그림 1] LSTM 모델의 구조

2.3 Sentiment Analysis

감정 분석은 주어진 텍스트의 극성을 분류하는 작업이다. 텍스트와 그에 수반되는 레이블이 주어지면 모델을 훈련하여 올바른 감정을 예측할 수 있다.

감정 분석 기술[2]은 기계 학습 접근 방식, 어휘 기반 접근 방식, 또는 하이브리드 방법으로 분류될 수 있다.

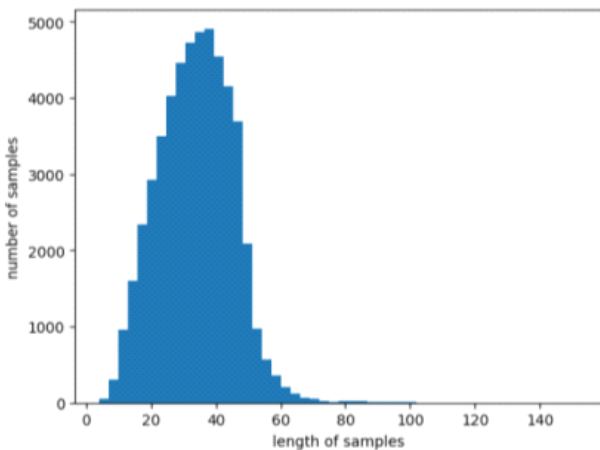
3. 실험 구성

3.1 데이터 소개

AI hub에서 제공하는 데이터 중 일반인 1,500명 대상의 인터뷰 및 클라우드 소싱을 수행한 ‘감성 대화 말뭉치’ 데이터[3]를 이용하였다. 총 51,630개의 데이터로 이루어져 있으며 ‘불안’, ‘분노’, ‘상처’, ‘슬픔’, ‘당황’, ‘기쁨’의 6개의 감정 대분류와 각각 대분류에 속해있는 여러 개의 감정 소분류로 라벨링 되어 있다.

3.2 데이터 전처리

타겟값인 ‘불안’, ‘분노’ 등 각 감정을 0부터 6까지의 숫자로 Label Encoding하고, 단어들을 숫자화하기 위해 형태소 분석기 Okt를 사용하여 토큰나이징 작업을 시행하였다. 학습을 위해 문장의 길이를 동일하게 맞추기 위해 문장의 최대 길이를 확인하고 패딩을 통해 빈 공간을 채운다.



[그림 2] 학습용 데이터의 문장 길이 시각화 결과

전처리한 데이터로 감정 분석을 위한 모델링을 진행한다. 본 연구에서는 한글 데이터를 정제하고 각 단어를 벡터화한 후 넘파이 파일로 저장하였다.

3.3 실험 결과

빠른 학습을 위해 LSTM과 GRU 모델을 간단하게 구축하여 텍스트 분류를 진행한 결과 정확도가 높지 않았다. 두 모델 중 테스트 데이터에 대한 정확도가 더 높았던 LSTM 모델을 기준으로 모델의 은닉층을 추가하고, 모델의 optimizer를 RMSprop에서 Adam(Adaptive Moment Estimation)으로 수정하고 dropout을 0.3으로 수정한 결과 기존 코드 대비 7% 이상의 정확도를 향상시킬 수 있었다.

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 152, 16)	172912
lstm (LSTM)	(None, 152, 128)	74240
lstm_1 (LSTM)	(None, 128)	131584
dense (Dense)	(None, 6)	774

Total params: 379510 (1.45 MB)
 Trainable params: 379510 (1.45 MB)
 Non-trainable params: 0 (0.00 Byte)

[그림 3] LSTM 모델 구축

4. 결론 및 향후 계획

본 논문에서는 자연어 처리 기술인 NLP와 RNN을 이용한 감정 분석 기술을 소개하였다. LSTM 모델과 GRU 모델을 비교 분석하였고 정확도가 더 높은 LSTM 모델을 사용하였다.

향후에는 Grid Search 방식을 이용하여 LSTM 모델의 최적의 하이퍼 파라미터를 탐색하고, 정확도를 90% 이상 향상시킨다. 정확도를 향상시킨 모델을 이용하여 사용자가 작성한 일기의 감정을 분석하고, 이를 기반으로 적절한 음악과 명언을 추천하여 감정 개선에 도움을 주는 어플리케이션을 개발한다.

※ 본 프로젝트는 과학기술정보통신부 정보통신 창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] Analysis of the Korean Tokenizing Library Module
 [2] Sentiments Analysis
<https://paperswithcode.com/task/sentiment-analysis>
 [3] Music Recommendation SNS(Social Network Service) Based on Emotion Analysis