

디지털 포렌식을 위한 Android 및 Windows 환경에서 카카오톡 아티팩트 분석 (II)

이나비*, 김광조*

*KAIST 정보보호대학원

Analysis of Artifacts in KakaoTalk under Android and Windows Environments for Digital Forensics (II)

Nabi Lee*, Kwangjo Kim*

*Graduate School of Information Security, KAIST.

요약

카카오톡은 국내에서 압도적인 점유율을 가진 채팅 애플리케이션으로 사용자의 스마트폰 또는 PC에 저장된 사용자의 송·수신 메시지는 사용자의 생활패턴, 심리상태 등을 확인할 수 있어 포렌식 분야에 적극 활용되고 있다. 하지만 카카오톡 메시지에 대한 심층학습 기반 감성분석은 극성 (긍정·부정)을 포함한 데이터 수집이 어렵다는 이유 등으로 지금까지 한 번도 시도된 적이 없었다. 이에 본 논문에서는 독자적으로 수집 및 전처리한 카카오톡 메시지 데이터셋을 Input, Hidden, Output 레이어로 구성된 기본적인 RNN (Recurrent Neural Network)으로 학습시켜, 카카오톡 메시지에 담긴 사용자의 주관 (최근 관심사, 심리적 상태, 채팅 참여자와의 밀접도 등)을 파악하고 이를 수사에 효과적으로 활용하는 방안을 살펴보았다.

I. 서론

카카오톡은 국내 스마트폰 사용자 중 97%가 애용하는 채팅앱으로 텍스트, 사진, 음성, 동영상, 화상통화 등 다양한 콘텐츠가 소통에 활용된다. 모든 송·수신 메시지는 사용자의 스마트폰과 PC에 저장되기 때문에 포렌식 연구에서 사용자의 스마트폰 또는 PC에 잔존하는 카카오톡 메시지 흔적 (아티팩트) 추출 및 분석은 대단히 중요하다.

선행연구 [1]에서 저자는 Android 6 (마시멜로) 환경의 모바일용 카카오톡과 Windows 10 환경의 PC용 카카오톡 아티팩트 유형과 속성 정보의 공통점·차이점을 분석하여 이를 효과적으로 포렌식에 활용하는 방안을 제시하였다. 또한 추출 가능한 데이터의 양이 방대하고 수사관의 재량에 따라 획득 가능한 정보의 수준에 격차가 발생하는 만큼 다양한 수사기법의 융합이 필요하다는데 공감대를 형성하였다.

본 논문에서는 연구 [1]에서 확인한 아티팩트 중 메시지 감성분석에 포커스를 맞추었다. 감성분석은 텍스트 마이닝 기법 중 하나로 opinion mining이라고도 불린다. 사용자가 생성

한 온라인 텍스트 속에 담긴 감성 (sentiment), 정서 (affect), 주관 (subjectivity), 또는 감정 (emotion)을 식별하기 위해 사용된다 [2]. 카카오톡 사용자는 자의든 타의든 일상생활의 많은 부분을 카카오톡을 통해 공유한다. 메시지를 송·수신하는 대상과 상황에 따라 사용자는 긍정·부정 등 감정의 흐름을 경험한다. 따라서 카카오톡 메시지 감성분석은 메시지에 담긴 사용자의 주관 - 최근 관심사, 심리적 상태, 채팅 참여자와의 밀접도 - 을 파악하고 이를 프로파일링 (profiling)에 활용할 수 있다.

II. 관련연구

지금까지 특정 대상에 대한 의견을 수집하고 분류하는 과정인 감성분석은 주로 마케팅 영역에서 활용되었다. 설문, 포럼, 블로그, 웹사이트 등에서 여론을 수집하고 그 흐름을 분석하여 시장의 요구에 기민하게 대응하기 위하여 감성분석을 활용하였다. 최근 한국어 대상 연구도 활발하다. 하지만 상업적인 목적 외에는 연구가 더딘 실정이며, 특히 카카오톡 등 SNS에 대한 감성분석은 연구가 희박하다. 데이터셋 구성에

어려움이 있기 때문이다.

상품평, 영화평점/리뷰는 사용자가 평가할 대상 (상품, 영화 등)이 존재하기 때문에 긍정·부정 등 사용자의 주관에 담긴 데이터를 수집하는데 어려움이 없다. 반면, 카카오톡, 트위터 등의 SNS는 다양한 주제로 개인의 의견을 표현할 뿐 그 의견의 극성 (긍정·부정 등)이 무엇인지, 무엇으로부터 기인했는지 우리는 주목하지 않는다. 하지만 포렌식 관점에서는 SNS상 문구 하나하나가 사용자의 상태를 유추할 수 있는 실마리가 된다.

연구 [3]에서는 트위터에서 한국어로 작성된 트윗의 긍정·부정 여부를 통계적 기반 기계학습을 통해 분류하였다. 감정분류 대상인 텍스트에서 단어를 추출하고 그 단어에 해당하는 사전에 정의된 감정값을 추출하여 문장의 감정 정도를 계산한 뒤 그 결과값으로 긍정·부정을 분류하였다. 실험결과 약 76%의 분류 정밀도 [Precision, 긍(부)정이라고 판정한 것 중 실제 긍(부)정 수]를 보이는 반면 재현율 [Recall, 전체 긍(부)정 수 중 판정된 긍(부)정 수]이 낮아 학습 방법의 업데이트가 요구되었다.

연구 [4]은 문맥 표현 중 하나인 ELMo (Embedding from Language Models) 기법을 소개하고 한국어 영화평 감성 분석을 위하여 NSMC (Naver Sentiment Movie Corpus)와 다음카카오 (DaumKakao) 데이터를 셀프 어텐션과 ELMo를 이용한 RNN (Recurrent Neural Network)으로 학습시키는 모델을 제안하였다. 기존 모델에 비해 성능이 좋음을 입증하였다.

본 연구에서는 카카오톡 메시지 감성 분석을 수행하기 위하여 독자적으로 수집 및 구축한 카카오톡 메시지 데이터를 말뭉치로 사용하였다. 긍정 또는 부정의 상태를 도출하는 분류문제이므로 many-to-one 문제로 정의하고 Input, Hidden, Output Layer로 구성된 기본적인 RNN 모델을 적용하였다. 학습모델은 새로운 말뭉치에 대한 감성 분석 가능성을 확인한 뒤 업데이트 예정이다.

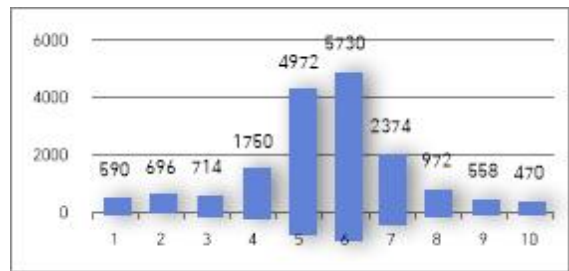
III. 연구설계

1.1 데이터 수집

카카오톡 메시지를 감성분석하려면 각각의 메시지는 긍정 또는 부정에 대한 정보를 포함하고 있어야 한다. 우선 우리는 2019. 7. 1. ~ 2019. 9. 30.까지의 카카오톡 메시지 18,826건을 수집한 뒤 메시지 작성자에게 메시지 작성 당시의 감정 상태를 수렴하여 1점~10점까지 (1

에 가까우면 부정적인 상태, 10에 가까우면 긍정적인 상태) 점수화하였다. 본 연구에서 필요한 것은 긍정 또는 부정에 대한 분류이므로 1점~10점까지 점수화된 메시지 중 4점~7점 구간은 긍정 또는 부정의 극성이 약하거나 정보 전달 목적의 메시지들로 극성을 전혀 띄고 있지 않다고 판단하여 제외하고 1점~3점 구간을 부정의 상태인 '0', 8점~10점 구간을 긍정의 상태인 '1'로 재분류하였다. 이렇게 분류된 메시지는 총 4,000건으로 긍정 2,000건 부정 2,000건 동일한 비율이다.

Figure 1. 카카오톡 메시지 감정분포



1.2 데이터 점검 및 탐색

우리가 수집한 카카오톡 메시지 18,826건 중 약 80% (15,060건)가 연구 [1]에서 Axion Process를 통해 수집한 csv 파일이고, 나머지는 카카오톡 메시지 내보내기 기능을 통해 수집한 txt 파일이다. 이를 Tab으로 구분한 txt 파일 형태로 통일하였다. 또한 데이터는 user, date, contents, label 네 개의 열로 이루어져 있는데 user는 메시지 작성자, date는 메시지 작성 일시, contents는 메시지 내용, label은 긍정(1), 부정(0)인지를 나타낸다.

1.3 데이터 전처리 및 정제

카카오톡은 일상생활 대화부터 업무, 동호회, 관심사별 오픈채팅 등 사용자에게 따라 그 내용과 유형이 다양하고 외부에 유출될 경우 사생활이 침해받을 수 있는 내용도 다수 있기 때문에 특정 개인·집단을 직·간접적으로 지칭 및 유추 가능할 수 있는 내용에 대한 정제가 필요하다. facebook 메신저, 인스타그램 등 여느 SNS와 같이 이모티콘, 띄어쓰기, 'ㅋㅋ' 등 분리된 자음·모음 등에 대한 선별적인 정제도 요구된다.

1.3.1 형태소 분석 및 품사 태깅

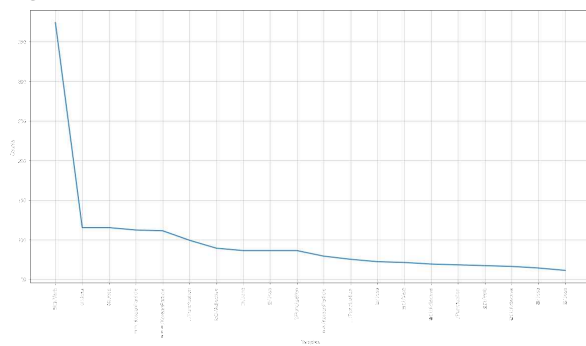
데이터는 KoNLPy 라이브러리의 okt 클래스를 이용하여 형태소 분석 및 품사 태깅을 하였다 [5]. 카카오톡 메시지는 1.3.1에서 설명한대로 특성상 맞춤법이나 띄어쓰기 등이 제대로 되어

있지 않은 경우가 많기 때문에 정확한 분류를 위하여 띄어쓰기, 맞춤법 등을 정제해주는 KoNLPy를 이용하였다. KoNLPy 라이브러리를 통해 분석된 형태소 하나하나를 데이터의 토큰(token)이라 한다.

1.3.2 중복토큰 분석

분석한 데이터의 토큰 중 일부는 중복된 경우가 있을 수 있으므로 nltk 라이브러리를 통해서 전처리를 하였다. 가령 카카오톡 메시지 4,000건을 okt로 형태소를 분석해보면 전체 토큰의 개수는 24,496개이지만 중복을 제외한 토큰의 개수는 4,910건이다. 출현 빈도가 높은 상위 토큰 20개를 보면 주로 ‘이’, ‘에’ 같은 조사 또는 ‘ㄱ’ 였다.

Figure 2. 출현빈도 높은 상위 토큰 20개



1.3.3 데이터 벡터화

분석된 토큰으로 데이터를 벡터화한다. 원 핫 인코딩 대신 CountVectorization을 사용하였다. CountVectorization은 Scikit-Learn 라이브러리의 문서 전처리 클래스 중 하나이다. 문서 집합에서 단어 토큰을 생성하고 각 단어의 수를 세어 각 문서를 BOW (Bag of Words) 인코딩 벡터로 변환한다.

1.4 데이터 모델링 및 학습

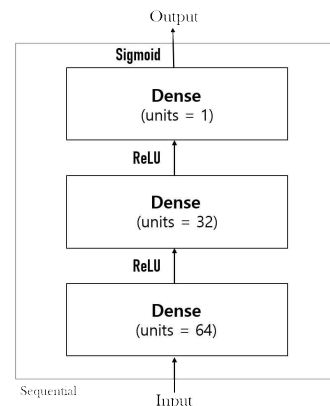
인공신경망 (Artificial Neural Network)을 구현하는 여러 가지 Neural Network 생성용 library 중 본 연구에서는 Tensorflow와 Keras를 사용하였다. 네트워크 내부 레이어 (layer) 구성은 Keras 내부 모듈 중 Sequential과 Dense를 사용하였다. 데이터는 카카오톡 메시지 4,000건 중 training dataset과 validation dataset을 8:2 비율로 각각 3200건, 800건으로 나누어 활용하였다.

1.4.1 모델 정의

Fig. 3과 같이 Sequential 모듈을 통해 신경망 Template를 만들고 training dataset을 Input 레

이어에 넣은 뒤 각각의 뉴런을 활성화시킨다. activation 함수로는 'ReLU (Rectifier Function)'를 사용하였다. 이를 통해 얻은 우선 예측값을 실제 결과값과 비교하여 그 차이를 계산한다 (forward propagation). 당연히 오차가 발생하기 때문에 우리는 이 차이를 최소화하기 위하여 예측값을 다시 구하는 과정을 반복하게 되는데 이때 MSProp 옵티마이저를 통해서 경사하강법으로 오차를 최소화하는 과정을 거친다. Output 레이어에서는 이진 분류에 적합한 sigmoid 함수와 validation dataset을 사용하여 학습한 모델의 성능을 검증한다. 손실함수로는 binary_crossentropy를 사용하였다.

Figure 3. 카카오톡 메시지 감성분석 모델



1.4.2 모델 구현 (학습) 결과

카카오톡 메시지 4,000건에 대해 Fig. 3 모델을 기반으로 학습할 경우 (batch size: 256, epoch: 10) 정확도는 최초 60%에서 점차 상승하여 약 94%에 도달하였으며 손실률은 최초 70%에서 7%대까지 감소하는 등 카카오톡 메시지 감성분석에 해당 모델이 유용하게 활용될 수 있음을 보였다.

Figure 4. 학습모델의 정확도

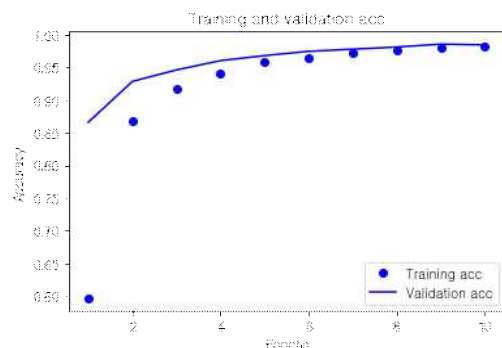
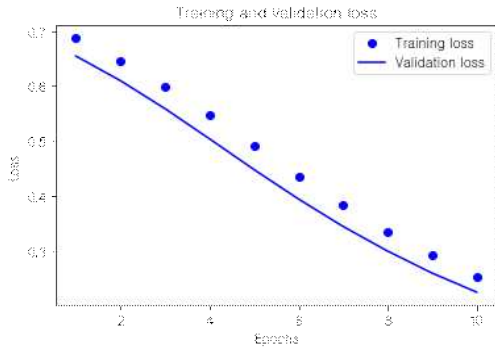


Figure 5. 학습모델의 손실률



IV. 평가

1.4.1의 모델의 성능평가를 위하여 우리는 train dataset 및 validation dataset과 중복되지 않는 카카오톡 메시지 200건 (긍정, 부정 각각 100건)을 추가로 준비하여 평가 (정확도, 정밀도, 민감도, 특이도)를 진행하였다 [6].

Table 1. 카카오톡 메시지 감성분석모델 성능평가

구 분		실제 정답	
		긍정 (Positive)	부정 (Negative)
실험 결과	긍정 (Positive)	69	17
	부정 (Negative)	31	83

1.1 정확도 (Accuracy)

카카오톡 메시지 200건 중 152건의 메시지 극성을 제대로 판별하여 정확도는 76%이다.

1.2 정밀도 (Precision)

긍정으로 분류된 메시지 86건 중 실제 긍정 메시지인 경우는 69건으로 긍정에 대한 분류 정밀도는 약 80.2%이며, 부정으로 분류된 메시지 114건 중 실제 부정 메시지인 경우는 83건으로 부정에 대한 분류 정밀도는 약 72.8%인 것으로 확인되었다.

1.3 민감도 (Sensitivity)

재현율 (Recall)로도 불린다. 실제 긍정 메시지 100건 중에서 긍정으로 제대로 분류된 메시지는 69건으로 69%의 민감도를 보였다.

1.4 특이도 (Specificity, True Negative Rate)

실제 부정 메시지 100건 중에서 부정으로 제대로 분류된 메시지는 83건으로 83%의 특이도를 보였다.

카카오톡 메시지 감성분석모델에 대한 평가 결과, 학습단계에서의 정확도 대비 약 18% 감소한 76%를 기록하였다. 분류 정밀도, 민감도, 특이도 역시 70%~80% 수준을 보였으며, 긍정 보다는 부정의 감정이 담긴 메시지 분류에 상대적으로 높은 성능을 보임을 확인하였다.

V. 결론

본 연구에서는 지금까지 카카오톡 메시지 감성분석을 위하여 18,826건의 카카오톡 메시지의 극성을 분류하고 극성이 뚜렷한 메시지 4,000건을 선별하여 이를 Input으로 하는 RNN 모델을 설계 및 구현하였다. 또한, 학습이 완료된 모델에 대해 카카오톡 메시지 200건을 추가 수집하여 성능평가를 진행한 결과 76%의 정확도를 보였다. 학습단계에서의 정확도 94% 대비 정확도가 감소한 이유는 감성분석모델의 한계 때문인 것으로 판단된다. 기본적인 RNN 구조를 기반으로 모델링하였기 때문에 메시지 길이가 길거나 판별이 어려운 것이 나오면 생략해버리는 경우가 발생할 수 있다. 이에 향후에는 LSTM (Long Short Term Memory)을 활용한 카카오톡 메시지 감성분석모델을 설계 및 구현 후, 본 연구의 모델과 성능을 비교할 필요가 있다.

[참고문헌]

- [1] 이나비, and 김광조. “디지털 포렌식을 위한 Android 및 Windows 환경에서 카카오톡 메시지의 아티팩트 분석(I).”
- [2] 김유영, and 송민. “영화 리뷰 감성분석을 위한 텍스트 마이닝 기반 감성 분류기 구축.” 지능정보연구 22.3 (2016): 71-89.
- [3] 임좌상, and 김진만. “한국어 트위터의 감정 분류를 위한 기계학습의 실증적 비교.” 멀티미디어학회논문지 17.2 (2014): 232-239.
- [4] 박천음, et al. “문맥 표현과 셀프 어텐션을 이용한 한국어 영화평 감성 분석.” 정보과학회논문지 46.9 (2019): 901-908.
- [5] 박은정, 조성준, “KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지”, 제 26회 한글 및 한국어 정보처리 학술대회 논문집, 2014.
- [6] Olson, David L.; and Delen, Dursun (2008); Advanced Data Mining Techniques, Springer, 1st edition (February 1, 2008), page 138, ISBN 3-540-76916-1