

KoNLPy와 KoBERT를 활용한 키워드 및 감정분석 일기 서비스

이채원^o, 문미경^{*}

^o동서대학교 소프트웨어학과,

^{*}동서대학교 소프트웨어학과

e-mail: chawonlee@naver.com^o, mkmoon@dongseo.ac.kr^{*}

Keyword and Emotional Analysis Diary Service Using KoNLPy and KoBERT

ChaeWon Lee^o, Mikyeong Moon^{*}

^oDept of Software, Dongseo University,

^{*}Dept of Software, Dongseo University

● 요약 ●

최근 작성한 일기를 SNS에 올려 평범한 사람들이 음악, 음식, 사진 등 소소한 일상을 남기고 우울증 투병기를 공유하여 힘을 얻기도 하는 등 누가 시키지 않아도 일기를 작성하고 간직하는 사람들이 증가하고 있다. 이러한 변화로 일기는 하루의 일상을 기록하는 목적을 넘어 어떤 감정을 느꼈는지 알아차리고 자아를 성찰 및 탐구하는 단계로 발전하고 있다. 그러나 스스로 일기의 키워드를 분석하고 감정이 어떠한지 정확하게 아는 것은 어렵다. 이에 따라 본 논문에서는 제시한 문제를 해결하기 위한 방법으로 KoBERT와 KoNLPy를 활용한 키워드 및 감정분석 일기 서비스를 제안하였다. 본 연구의 키워드 및 감정분석 일기 서비스는 사용자가 무의식적으로 표현하는 텍스트 기반의 일기에서 자주 반복되는 키워드와 감정을 제공하여 자신의 감정 상태를 쉽게 인지하고 되돌아볼 수 있도록 제작하였다.

키워드: 일기(Diary), 키워드(Keyword), 감정(Emotion), KoBERT, KoNLPy

I. Introduction

자신의 일상과 감정을 정리하고 부족한 부분을 반성하는 등 다양한 이유로 일기를 작성하고 간직하는 사람들이 많아졌다. 특히 다양한 종류의 일기 작성 서비스가 증가함에 따라 경험한 사실을 작성하는 단순한 일상 일기에서 감정 일기의 형태로 발전하고 있다. 이를 통해 사람들이 자신의 감정 상태를 인지하는 것에 매우 관심이 있고 일기 작성을 통해 감정 분석을 스스로 시도하고 있다는 것을 알 수 있다. 하지만 스스로 일기의 키워드를 분석하고 감정이 어떠한지 정확하게 알아내는 것은 어렵다. 따라서 사용자가 작성한 일기에서 키워드와 감정을 쉽게 확인할 수 있는 시스템이 필요하다. 본 논문은 KoNLPy와 KoBERT를 통해 키워드와 감정을 분석하는 서비스에 관한 연구 내용을 기술한다. 이 서비스는 사용자가 작성한 텍스트 형식의 글에서 단어의 빈도수를 측정하여 키워드를 추출하고 미리 학습된 모델을 통해 감정을 분석한다.

II. Preliminaries

국내 카톡 감정 분석 '텍스트넷' 어플리케이션은 주고받은 카카오톡 대화를 기반으로 감정을 객관적으로 분석해주는 서비스이다. 사용자가 텍스트 내용을 직접 선택하여 원하는 부분을 감정분석 할 수도 있다. 네이버 요약 봇 서비스는 기사 본문의 키워드, 문장 중요도 등을 판단하여 최대 3문장까지 제공한다. 단, 문장이나 단어의 순서를 변경하거나 새롭게 추가하지 않는 등 원래 의미가 훼손되지 않도록 한다. 제시한 기능들을 통합하여 본 논문에서는 텍스트에서 키워드와 감정을 분석하여 제공하는 것을 주 기능으로 한다.

III. The Proposed Scheme

3-1. 키워드 추출

텍스트 형식의 글에서 키워드를 추출하기 위해 한국어 정보처리를 위한 Python 패키지인 KoNLPy를 사용한다. KoNLPy의 클래스 중 오픈소스 한국어 분석기이자 과거 트위터 형태소 분석기였던 Open Korean Text를 사용하였다. 텍스트 형식의 글에서 먼저 명사를

추출하였고 추출한 명사에서 내장 함수인 Count를 이용하여 빈도수를 계산하였다.

3-2. 감정 분석

텍스트 형식의 글에서 감정을 분석하기 위해 사전에 학습된 대용량 말뭉치 모델인 KoBERT를 사용한다. Table 1은 KoBERT에 학습시킨 감정 분류별 데이터 개수이다. 먼저 7가지 감정인 공포, 놀람, 분노, 슬픔, 중립, 행복, 혐오로 분류하였고 총 38,594개의 문장 데이터를 학습시켰다.

Table 1. Number of learning data by emotion classification

Classification	Number of data
Fear	5,468
Surprised	5,898
Anger	5,665
Sadness	5,267
Neutrality	4,830
Happiness	6,037
Disgust	5,429

3-3. 개발 결과

Fig 1.은 제안한 방법을 구현하여 실제로 텍스트 형식의 글을 입력해 키워드 추출과 감정분석을 진행한 결과이다. 사용자가 입력한 텍스트 형식의 글에서 키워드 (#생일, #친구, #가족)와 감정(Happiness)을 결과 값으로 제공한 것을 알 수 있다.

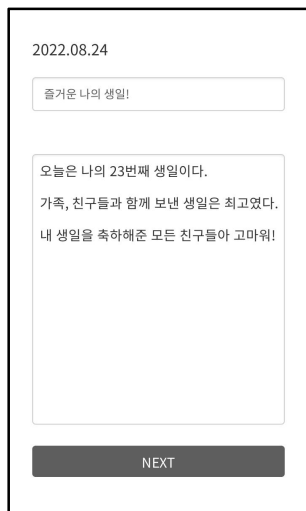


Fig. 1. Experimental Results

IV. Conclusions

본 논문에서는 KoNLPy와 KoBERT를 이용하여 사용자가 입력한 텍스트 형식의 글에서 키워드와 분석한 감정을 제공하는 서비스에 관해 기술하였다. 본 서비스를 통해 감정 상태를 쉽고 정확하게 인지와 되돌아 볼 수 있을 것으로 기대한다.

ACKNOWLEDGEMENT

본 연구는 2022년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2019-0-01817).

REFERENCES

- [1] Young-Jun Lee, Ho-Jin Choi. "Joint Learning-based KoBERT for Emotion Recognition in Korean," Proceedings of the Korean Information Science Society Conference, pp. 568-570, 2020.
- [2] Ji-Hoon Lee, Hyun-Sub Lee, Jin-Deog Kim. "A Study on the Use of Frequency of Morphology to Extract Similarities in Lecture Video," Proceedings of the Korean Institute of Information and Communication Sciences Conference Vol. 24, No. 1, pp. 9-11, 2020.