

자연어 처리 인공지능 기술을 활용한 생활기록부 작성 효율성 제고 향상 연구

서정호*, 김 웅^o

*공주대학교 AI융합대학원,

^o고려대학교 일반대학원 산업경영공학부

e-mail: sjhmsmsmms@naver.com*, woong2241@korea.kr^o

A Study on the Improvement of the Efficiency of School Report Documentation Using Artificial Intelligence Technology in Natural Language Processing

Jung-Ho Seo*, Woong Kim^o

*Graduate School of AI Convergence, Gong-Ju National University,

^oIndustrial Management Engineering, Korea University

● 요약 ●

본 논문에서는 대입수시전형에서 생활기록부 입력은 대한민국 입시를 결정하는 매우 중요한 평가자료이다. 30명의 교사를 대상으로 실시한 설문조사를 통해서 교사들이 생활기록부를 작성하는데 가장 많이 참고하는 자료로는 수행평가 결과물, 발표내용, 보고서, 감상문 등의 학습 결과물(90%), 학생들이 작성한 자기평가서(73.3%), 관찰 평가지(50%)로 나타났으며, 80%(24명)의 교사들이 생활기록부를 작성하는데 고충을 겪고 있음을 확인할 수 있었다. 교사들이 느끼는 고충의 원인으로는 학생들의 개인별 특성 파악이 어려워 차별성있게 작성하는 것(76.7%)을 가장 힘들어 하였고, 작성해야 할 많은 수의 학생(60%), 문구를 만드는데 대해 부담(86.7%)을 느끼는 것으로 나타났다. 이 과정에서 교사의 전문성 뿐만 아니라 기계적이고 반복적인 작업도 많이 요구되고 있기 때문에, 생활기록부를 작성하는데에 도움을 줄 수 있는 프로그램 개발이 필요하다고 고안을 내었다. 교사들 역시 반복적이고 일률적인 생활기록부 작성에 도움을 줄 수 있는 프로그램이 있다면 유용하게 활용할 것이라는 응답이 90%였다. 따라서 본 연구에서 자연어 처리 인공지능 기술을 활용하여 교사들이 생활기록부를 작성하는데 있어 기계적이고 단순한 작업을 도와 주는 프로그램 개발에 대한 연구의 필요성을 제시하였다. 제안하는 프로그램은 학생들의 탐구보고서, 토론, 발표, 감상문 등의 생활기록부 작성 참고자료들을 텍스트로 변환하고 추상요약(Abstractive Summarization)을 통해 교사들이 효율적으로 작성하는데 활용될 수 있도록 설계하였다. 연구 결과 생활기록부 작성 참고자료를 텍스트로 변환하는 것과 추상요약을 할 수 있는 개방형 데이터셋까지는 확보하였다. 추상요약을 구현하는 방법에 대해서는 보다 심도 있는 추가연구가 필요하였다. 이를 통해 교사들이 교육 본질에 더욱 충실할 수 있는 환경을 마련하고, 내실 있는 생활기록부 작성이 공교육 신뢰 제고에 밑바탕이 되고자 한다.

키워드: 자연어처리(Natural Language Processing), 생활기록부(School Report Documentation), 추상요약(Abstractive Summarization)

I. Introduction

현재 대학입시전형은 크게 수시와 정시로 나뉜다. 정시는 수능 성적을 위주로 선발하며, 수시는 크게 내신 성적 위주로 평가하는 학생부 교과 전형과 내신 성적을 포함하여 다양한 교과, 비교과 활동 기록을 보는 학생부 종합 전형으로 나뉜다. 이는 내실 있는 생활기록부

를 통해서 대학입시전형에 큰 영향을 주는 것을 알 수 있다 [1]. 교사와 학생은 생활기록부 작성의 중요성을 잘 알고 있으나, 현실적으로는 학교 현장에서 교사가 모든 학생들을 의미있게 기록하는 것은 매우 어려운 일이다. 학생 별 생활기록부 작성란으로는 자율활동,

진로활동, 동아리활동, 교과세특, 행동발달 종합, 경우에 따라서는 개인별 세부특기사항까지 1,000바이트에서 1,500바이트까지 작성할 수 있다.

수업의 형태는 토론, 발표, 탐구활동이 현장에서 실천되고 있으나, 여전히 강의식 수업이 주를 이루고 있어 교사 한명이 학생들을 직접 다 관찰하고 작성하는 것은 매우 어려운 일이다. 설문조사 결과 대다수의 교사들은 생활기록부 작성에서 학생들이 작성한 자기평가서와 수행평가 결과물, 발표내용, 보고서, 감상문 등의 학습 결과물을 참고하는 것으로 나타났다.

이를 바탕으로 생활기록부 내용으로 변환하여 기재하고 있는 것으로 나타났다. 이 과정에서 생활기록부 작성 참고자료를 정리하고, 문구로 변환하는 과정에서, 질적 측면의 전문성 외에도 양적 측면의 기계적이고 단순한 작업이 수반되어 작성에 매우 부담을 느끼고 있다. 따라서, 생활기록부 작성에 도움을 줄 수 있는 프로그램이 개발된다면, 교사들이 보다 교육 본질에 집중할 수 있을 것이다. 이에 본 연구는 자연어 처리 인공지능 기술을 활용하여 생활기록부 작성의 부담을 줄일 수 있는 방안을 연구하고자 한다.

II. Preliminaries

1. Related works

1.1 국내 동향

현재 우리나라 대학진학에서 수시로 입학하는 학생들은 학생부종합 전형에 필요한 학교생활기록부의 중요성이 매우 높기 때문에 데이터 마이닝 기반의 학교생활기록부 분석을 CBOW모델과 Skip-gram 모델을 활용한 키워드 추출을 한 연구가 있다 [2]. 대학입시에서 자기소개서와 학교생활기록부 평가를 위한 기계학습 모형을 개발하여, 평가자가 학교생활기록부를 짧은 시간에 읽고 그 속에 숨겨진 지원자의 잠재력과 능력을 판단할 수 있도록 하였다. 이 또한, 텍스트마이닝기법으로 글의 내용과 단어의 특성을 분석하여 제공하고 있다 [3]. 뿐만 아니라, 블록체인 기반으로 퍼블릭 및 허가형 블록체인을 이용하여 학교생활기록부 프레임워크를 제안하여, 4세대 나이스에 적합한 관리 방안도 제시되고 있다 [4]. 학생참여형 교과활동의 학교생활기록부 기재 유형 분석 연구를 통해서 단어의 빈도수 및 쓰임새를 확인 가능하다 [5].

III. The Proposed Scheme

1. Text Document Summary

텍스트 문서 요약이란, 복잡성과 문서의 길이를 줄이면서 원본의 필수적인 내용을 유지하는 것이다. 문서 요약은 크게 추상 요약 (abstractive summarization)과 추출 요약(extractive summarization)으로 나뉜다. 추상 요약은 원본 문서에서 나타나지 않는 단어와 구를 사용하여 요약문을 생성하는 것이며, 추출 요약은 원본 문서에서 추출된 문장으로 요약문을 생성하는 것이다.

1.1 추출 요약(extractive summarization)

추출 요약에서 가장 많이 활용되는 알고리즘은 그래프 기반의 텍스트랭크(TextRank)이다. 텍스트랭크는 기존의 페이지랭크(PageRank) 알고리즘에서 파생된 것으로 문장 간 같은 단어가 많이 언급될수록 가중치를 부여하여 문장 중요도 순위를 계산하는 알고리즘이다. 텍스트랭크 알고리즘은 이미 존재하는 문장이나 단어로만 구성하므로 모델의 언어 표현 능력이 제한된다는 한계가 있다.

1.2 추상 요약(abstractive summarization)

추상적 요약은 원문에 없던 문장이라도 핵심 문맥을 반영한 새로운 문장을 생성하기 때문에 마치 사람이 요약하는 것과 같은 방식으로 주로 인공 신경망을 사용하며 대표적인 모델로 seq2seq가 있다. 이와 같은 인공 신경망들은 기본적으로 지도 학습으로 추상 요약을 인공 신경망으로 훈련하기 위해서는 원문 뿐만 아니라 실제 요약문이라는 레이블 데이터가 있어야 한다.

2. Text Abstract Summarization Program-AI Hub

AI Hub는 한국지능정보사회진흥원(NIA)에서 운영하는 인공지능 기술 및 제품·서비스 개발에 필요한 AI 인프라(인공지능 학습용 데이터, AI SW API, 컴퓨팅 자원)를 지원함으로써 누구나 활용하고 참여하는 AI 통합 플랫폼이다. AI Hub 홈페이지 개방데이터에는 음성자연어 처리 매뉴가 있고 하위 매뉴에 논문자료요약, 도서자료요약, 한국어대화요약, 문서요약텍스트 등 추상추출을 위한 다양한 AI 데이터셋을 제공한다.

2.1 논문자료요약

OA논문(18만 건)과 특허명세서(17만 건)를 대상으로 요약대상 원문과 요약문이 쌍으로 구성될 수 있도록 하였다. 학술논문은 한국학술지인용색인(KCI), 국립중앙도서관, KISTI 등의 관련 연구기관에서 제공하는 Open Access 한국어 학술 논문을 대상으로 수집하여 논문의 전체(초록)와 섹션(논문 원문 중 섹션 일부 Paragraph)에 해당하는 요약문을 작성, 활용서비스 데이터 제공하고 있다.

2.2 도서자료요약

추상요약 기술 개발에 활용할 수 있는 학습 데이터셋으로, 20만건의 한국어 도서자료 원문과 요약문으로 구성되어 있다.

2.3 한국어대화요약

대화 원문과 이에 대응하는 요약문 35만건을 활용하여 일상 대화, 토론, 회의록, 상담 내역 등 다양한 유형으로 구성되어 있다. 대화문 요약의 경우는 육하원칙 요소 중 2가지 요소 이상을 포함하며 누가에 해당하는 내용을 알 수 있으면 한 문장으로 요약하고, 대화문 내용에 등장하는 키워드 2개 이상을 포함하여 한 문장으로 요약한다.

2.4 문서요약텍스트

문서요약 텍스트 데이터셋은 원문데이터에 대한 추출요약과 추상요약으로 구성되며, 즉 하나의 데이터셋에 총 3건의 데이터(원문, 추출, 생성)가 포함된다. 원문데이터의 수량은 뉴스기사 30만 건, 기고문 6만 건, 잡지 1만 건, 법원 판결문 3만 건으로 총 40만 건이며, 이에 대한 추출요약과 생성요약도 각각 40만 건 씩 구축되었다. 뉴스기사는 요약 AI 알고리즘의 핵심 데이터로서 10개 언론사로부터 30만 건의 원문데이터를 확보하였으며, 이 중 종합면 30%, 정치 20%, 경제 20%, 사회 20%, 문화 및 스포츠 기타 10%의 비율로 구성되어 있다.

3. Analysis of Survey Results

충청남도 아산시 관내 일반고등학교 30명 교사를 대상으로 학생 생활기록부에 관한 설문조사를 실시하였다.

Table 1. Survey Results

설문 내용	선택지	결과
설문1. 현재 생활기록부에 작성할 수 있는 분량은 적정하다고 생각하십니까? 	매우 적다	0 (0%)
	적다	2 (6.7%)
	적절하다	10 (33.3%)
	많다	12 (40%)
	매우 많다	5 (16.7%)
	응답없음	1 (3.3%)
설문2. 현재 생활기록부 작성에 있어 느끼는 고충에 대해 선택해 주시기 바랍니다. 	매우 쉽다	0 (0%)
	쉽다	0 (0%)
	보통이다	6 (20%)
	힘들다	11 (36.7%)
	매우 힘들다	13 (43.3%)

설문3. 생활기록부 작성이 힘들다면 어떤 부분에서 가장 부담을 느끼는지 선택해 주시기 바랍니다.(중복선택) 	작성 소재	7 (23.3%)
	문구작성	4 (13.3%)
	학생수	18 (60%)
	특색있는 내용	23 (76.7%)
	기타	1 (3.3%)
설문4. 생활기록부를 쓰기 위한 문구를 변환하는데 느끼는 부담정도를 조사하는 설문입니다. 	매우 부담	15 (50%)
	부담	11 (36.7%)
	보통	3 (10%)
	부담없음	1 (3.3%)
	매우 부담없음	0 (3.3%)
설문5. 생활기록부 작성에 있어 참고하는 내용을 선택해 주시기 바랍니다.(중복선택) 	평가를 위해 개발된 학생 관찰 평가지	15 (50%)
	학생들의 자기 평가서	22 (73.3%)
	수행평가 결과물, 발표내용, 보고서, 감상문 등의 학습 결과물	27 (90%)
	교육과정의 상취기준과 성취수준	13 (43.3%)
	기타	0 (0%)
설문6. 학생들의 발표, 보고서, 감상문 등의 활동을 텍스트로 정리하고 그 내용을 요약해주는 프로그램이 개발된다면 생활기록부 작성 시 유용하게 활용 될 것이라고 생각하십니까? 	잘 활용되지 않을 것이다.	0 (0%)
	활용되지 않을 것이다.	2 (6.7%)
	보통이다.	1 (3.3%)
	대체로 활용될 것이다.	20 (66.7%)
	매우 유용하게 활용될 것이다.	7 (23.3%)

설문7. 설문6의 프로그램에서 추가되면 좋겠다고 생각하는 기능이 있다면 여러분의 의견을 자유롭게 작성해 주시기 바랍니다.

① 자기평가보고서, 교사의 교과평가와 관련된 성취수준내용이 함유된 자료
 ② 학생의 활동 상황 기록에 대해 교사와 피드백을 주고받을 수 있는 기능
 ③ 학생의 포트폴리오를 분석하여 진로 성향을 분석하여 안내하면 좋겠다.
 ④ 대략적인 문구 틀이 있으면 좋겠습니다.
 ⑤ 같은 주제나 내용으로 자료를 만들더라도 의미가 같지만 다른 뜻을 가진 단어들로 한 반의 요약 문장이 구성되었으면 좋겠음
 ⑥ 교사가 발표, 보고서, 감상문 정도는 요약해서 정리할 수 있어야 하는 것 아닌가요? 혹시 학생부 개별화에 역행하지는 않을지 걱정이네요.
 지금도 학생부 성의 없이, 쓰는 방법도 모르고 쓰는 경우가 많은데, 물론 저도 반성! 학생부 작성 방법을 자주 연수해 주는 것이 나올 듯요

IV. Conclusions

1. Conclusions

본 연구를 통해 다음과 같은 결론을 도출하였다. 첫째, 교사들은 생활기록부 작성의 중요성을 매우 잘 인지하고 있지만, 작성하는 과정에서 매우 큰 부담을 느끼고 고충을 겪고 있다. 둘째, 교사들이 생활기록부 작성참고자료를 바탕으로 문구로 변환하는 과정에는 기계적이고 단순한 작업이 있고 이를 보완해주는 프로그램이 개발된다면 매우 효용성이 클 것이다.

셋째, 교사들의 생활기록부 작성에서 느끼는 고충의 원인과 실현가능 자연어 처리 인공지능 기술을 고려하여 추상요약 인공지능 기술을 활용한 프로그램이 효과적일 것으로 추론할 수 있다. 넷째, 추상요약 인공지능 기술 활용에 있어 개방데이터로 데이터셋은 확보할 수 있으나 매뉴얼을 확보하기가 어려웠다.

2. Suggestion

대학 입시와 맞물려 학생생활기록부는 학생들의 대학교에 진학하기 위해 반드시 필요하다. 교사들도 이러한 상황을 매우 잘 인지하고 있지만, 교사들이 작성하기에는 역부족이다. 기록 과정에서 기계적이고 반복적인 작업도 많이 요구되고 있다. 따라서 자연어 처리 인공지능 기술을 활용하여 기계적이고 단순한 작업을 도와 주는 프로그램이 개발된다면 교사가 보다 교육 본질에 집중할 수 있을 것이다. 이는 곧 더 많은 교사가 효율적으로 내실있는 생활기록부 작성이 될 것이며, 나아가 학생과 학부모들의 신뢰를 회복하고 교사의 권위를 바로 세우는 토대가 될 수 있을 것이다. 마지막으로 학생의 입장에서 자연어 처리 인공지능 기술은 학생들의 포트폴리오를 효율적으로 관리할 수 있도록 도움을 주며, 분류 인공지능 알고리즘 등을 통해 진로에 대해 추천, 안내, 분석 등의 서비스를 제공할 수 있을 것이다.

REFERENCES

- [1] Oh Seong-bae. "Analysis of academic achievement and school life by admissions process of college students," Korean Educational Issues Study, 34.3: pp. 157-175, 2016.
- [2] Yang Jin-woo, et al. "A data mining-based school life record analysis system for college entrance exams," Journal of the Korean Contents Association, 21.2: pp. 49-58, 2021.
- [3] Lim Jun-muk, et al. "Machine learning model for evaluation of self-introduction and school life record book in college entrance examination," Proceedings of the Fall Conference of the Korean Society of Industrial Engineers, pp. 3029-3033, 2020.
- [4] Jeong Seung-Wook, Kim Doo-yeon . "Blockchain-Based Student's School Record Book System Management," Journal of the Korean Telecommunications Society, 46.10: pp. 1762-1770, 2021.
- [5] Lim Jin-teak et al. "An Analysis of the Record Types of Student-Participatory Curriculum Activities in School Transcripts," 6: pp. 65-98, 2018.