토픽모델링 기반의 학교폭력 사례 유형 연구

신승기
서울교육대학교 컴퓨터교육과

요약

본 연구는 학교현장에서 지속적으로 강조되고 있는 학교폭력 예방을 위한 방안을 도출하기 위한 노력의 일환으로 최근의 학교폭력 관련 이슈화되고 있는 주제를 데이터과학의 관점으로 살펴보고자 하였다. 특히, 온라인 SNS데이터를 활용하여 학교폭력 관련 게시물을 크롤링하고 토픽모델링 방법을 활용하여 유형별 특징을 살펴보고자 하였다. 토픽모델링 분석을 통해 도출된 주제별 키워드를 유형별로 정리한 결과를 통해 대체로 학교폭력의 피해 예방과 가해자 처벌 및 조치사안에 대한 내용으로 크게 3가지의 내용으로 구분할 수 있었다. 첫째, 학교폭력 예방활동에 대한 내용으로서 학교폭력예방을 위한 전문 기구들의 역할에 대한 내용이다. 둘째, 학교폭력에 대한 조치사항과 절차에 대한 내용으로 도출되었다. 셋째, 학교폭력의 최근 현안에 대한 내용에 대해서 살펴볼 수 있었다. 추후 연구에서는 데이터기반의 예측을 기반으로 당면하고 있는 사회적 문제해결에 활용하는 연구가 수행될 필요가 있다.

키워드 : 토픽모델링, LDA기법, 데이터기반연구, 데이터마이닝, 학교폭력

A Study on the Categorizes of School Bullying through Topic Modelling Method

Seungki Shin
Department of Computer Education, Seoul National University of Education

Abstract

As part of an effort to derive measures to prevent school violence, which is continuously emphasized in the school field, this study tried to examine the topic that has recently become an issue related to school violence from the perspective of data science. In particular, it was attempted to crawl posts related to school violence using online SNS data and examine the characteristics of each type by using the topic modeling method. As a result of arranging the keywords for each topic derived from the topic modeling analysis by type, it was possible to divide the contents into three main categories: prevention of school violence, punishment of perpetrators, and measures to be taken. First, as the contents of school violence prevention activities, it is the contents of the role of specialized organizations for the prevention of school violence. Second, it was derived from the contents of measures and procedures for school violence. Third, it was possible to examine the contents of recent issues of school violence. In future research, it is necessary to conduct research that is used to solve the social problems facing based on data-based prediction.

Keywords : Topic Modelling, LDA Method, Data-driven Research, Data Mininung, School Bullying

이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A5A8069141)
논문투고 : 2021-07-29
논문심사 : 2021-08-03
심사완료 : 2021-08-06

1. 서론

세계 많은 나라들은 우수한 인재를 양성하여 국가경쟁력을 높이기 위한 노력을 기울이고 있다. 우리나라의 경우에는 자원이 부족하고 빠른 경제성장을 이룩하기 위하여 교육에 대한 국가차원의 체계를 구축하고 우수한 교원양성을 기반으로 미래인재를 길러내는데 더욱 집중적인 투자를 하고 있다.

교육의 형태는 여러 가지로 구분될 수 있으나, 우리나라는 국가수준교육과정을 토대로 교과 및 내용요소에 대한 성취기준을 제시하고 있으며, 학교교육에서의 교실수업을 중심으로 교육이 실시되고 있다. 학교교육은 학급단위 학교생활을 통해 인지적 능력의 함양과 더불어 학생들이 함께 어울리며 정서적인 역량도 기르는데 초점을 두고 있다는 점에서 다양한 사회적 활동이 수행되고 학생들의 인지적 발달과 함께 사회적역량이 함께 길러지는 중요한 교육환경이라고 할 수 있다[7].

그러나 학교현장에서 많은 사회구성원들이 함께 생활하는 공동체라는 점에서 안타깝게도 학교폭력이 학교교육에서 발생하는 대표적으로 개선이 필요한 부분으로 제시되고 있다[4][6]. 학교폭력에 대해서는 엄정하게 대처하는 것을 원칙으로 하고 있으며, 학교폭력을 예방하고 피해학생을 보호하기 위한 “학교폭력예방 및 대책에 관한 법률”의 제정으로 2013년부터 매년 교육부에서는 17개 시도교육청을 통해 학교폭력 실태조사를 전수조사의 형태로 수행하고 있다[8][9][10].

교육부(2021)의 학교폭력 실태조사 결과 발표에 따르면 피해유형별 응답건수가 대체로 전년도에 비해 감소한 것으로 조사되었으나 사이버폭력의 경우에는 비중이 증가된 것으로 나타났다[9]. 학교폭력을 예방하고 대응을 강화하기 위해서는 최근의 학교폭력 사례에 대한 유형을 살펴보고 유형별 특징에 따른 대책 마련이 필요할 것이다.

본 연구에서는 온라인 SNS데이터를 활용하여 학교폭력의 피해사례를 수집하고 이를 토픽모델링 기반의 데이터 분석을 통해 학교폭력의 유형 및 유형별 특징을 살펴봄으로서 학교폭력의 예방 및 대응을 위한 데이터를 제공하고 학생들이 학교교육을 통해 정서적 안정과 유대감을 형성할 수 있는 공동체 의식 마련에 기여하고자 한다.

2. 이론적 배경

교육부에서는 “학교폭력예방 및 대책에 관한 법률”을 근거로 매년 학교폭력 실태조사를 전수조사의 형태로 실시하고 있다[9]. 아래의 <Table 1>은 2020년도의 학교폭력 전수조사의 결과를 피해유형별로 나타낸 자료로서, 8개 영역 중에서 사이버 폭력과 집단 따돌림의 피해비중이 증가됨을 살펴볼 수 있다[9].

<Table 1> Responses of School Bullying in 2020[1]

구분	2019년 (%)	2020년 (%)	전년대비 증감 (%)
언어폭력	35.6	33.6	-2.0
집단따돌림	23.2	26.0	+2.8
스토킹	8.7	6.7	-2.0
신체폭력	8.6	7.9	-0.7
사이버폭력	8.9	12.3	+3.4
금품갈취	6.3	5.4	-0.9
성폭력	3.9	3.7	-0.2
강요	4.9	4.4	-0.5

이와 같은 학교폭력의 세부적인 현황을 토대로 교육부에서는 관계부처와 함께 “학생 사이버폭력 예방 및 대응 실무협의체”를 구성하고 관련 부처 및 기관이 협력하여 사이버폭력 예방을 위해 노력하고 있다[10]. 현재의 청소년 세대들은 디지털 네이티브라고 불리우며 기성세대와는 생활환경이나 사고의 형태가 디지털을 기반으로 이루어진다는 점에서 최근의 메타버스와 같은 플랫폼이 인기가 있는 것도 관련이 있다[3][5][11].

온라인 환경은 MZ세대라고 불리우는 청소년 세대의 주요 생활공간이라는 점에서[2] 학교폭력을 예방하고 사이버폭력에 대한 피해사실을 조기에 인지하고 진단하기 위해서는 온라인 공간을 활용한 상담 등의 환경마련이 필요할 것이다[8]. 따라서 본 연구에서는 사이버 공간에서의 학교폭력과 관련된 게시물을 수집하고 최근의 학교폭력에 대한 특징을 살펴보기 위해 토픽모델링 기법을 활용하여 비지도학습 기반의 주요단어 및 주제를 살펴보고자 하였다. 특히, 단어의 출현빈도를 토대로 잠재적인 의미를 도출하는 기법인 LDA기반 토픽모델링을 활용하여[1] 주제별 의미를 도출하고자 하였다.

3. 연구방법

본 연구의 목적은 온라인 SNS의 공개데이터를 대상으로 데이터크롤링을 통해 학교폭력 관련 자료를 수집하고 토픽모델링 분석을 수행함으로서 학교폭력의 유형 및 유형별 의미를 살펴보는데 있다. 학교폭력 관련 온라인 SNS 데이터는 Facebook과 Instagram에서 수집하였으며, 최근 5년간의 데이터(2017년~2021년)를 기준으로 학교폭력에 대한 키워드 검색을 통한 게시물 수집 및 해쉬태그자료 크롤링이 수행되었다.

데이터크롤링을 통해 수집된 자료는 중복데이터를 확인하여 제거하는 과정과 KoNLPy를 활용한 형태소 분석을 통해 데이터전처리과정이 수행되었다. 토픽모델링을 수행하기 위하여 LDA(Latent Dirichlet Allocation)를 토대로 추출된 단어 중심의 데이터를 그룹화하고 유형화하여 비지도학습을 기반으로한 잠재적 의미도출이 실시되었다.

LDA기법을 적용하기 위해 α = 0.1, β = 0.01, K=10로 파라미터를 설정하였으며, 10개로 추출된 토픽에 대하여 정부와 학교현장에서 연관성 있는 8개의 주제를 선정하는 과정을 통해 학교폭력의 유형을 구분하고자 하였다.

4. 토픽모델링 분석 결과

학교폭력에 대한 온라인 데이터를 데이터크롤링을 통해 수집하고 토픽모델링을 통해 LDA기반의 비지도학습을 통한 유의미한 8개의 주제를 도출한 결과는 아래의 <Table 2>와 같이 나타났다. [주제 1]은 교육을 통한 학교폭력 예방의 전담기구로서 자치단체와 기관 및 경찰과 교육청의 역할에 대한 내용으로 살펴볼 수 있다. [주제 2]는 학생들의 피해 발생과 가해에 대한 위원회의 처벌이 증가됨에 따른 대책에 대한 현재의 전반적인 상황에 대한 내용이다.

<Table 2> Topic Modeling with LDA Method about School Bullying from Social Network Service Data

주제1	주제2	주제3	주제4	주제5	주제6	주제7	주제8
교육	학생	예방	학폭	상담	위원회	변호사	강사
0.042181	0.0549917	0.0654041	0.0280991	0.115597	0.0208456	0.0571389	0.0338509
예방	피해	활동	교사	청소년	교사	왕따	업체
0.0226361	0.0391913	0.0215811	0.0239378	0.108847	0.0137407	0.0179954	0.0280496
법	사건	캠페인	뉴스	센터	해결	만화	디지털
0.0159497	0.0246549	0.0209069	0.0197765	0.0286946	0.0118461	0.0179954	0.0203144
문화	대책	청소년	결과	복지	교육부	법률	예방
0.011835	0.0227589	0.0182101	0.0145749	0.0194138	0.0113725	0.0169375	0.0183806
자치단체	폭행	교육	징계	심리	가해자	예방	위탁
0.0113206	0.0189668	0.0155133	0.0145749	0.0177264	0.0108988	0.0127058	0.0174138
보호	발생	공연	요구	고민	처분	일상	강의
0.0108063	0.0132786	0.0141649	0.011454	0.0135078	0.0104252	0.0116478	0.0145131
기관	처벌	안전	대상	사이버	대책	피해자	가정
0.0102919	0.0126466	0.0134907	0.011454	0.0126641	0.00900419	0.0116478	0.0145131
경찰	증가	참여	언어	부모	사안	상담	인권
0.00874892	0.0126466	0.0101197	0.0104136	0.0118204	0.00900419	0.0105899	0.0116124
교육청	가해	경찰	논란	예방	조치	가해자	전문
0.00874892	0.0113826	0.00944554	0.00937331	0.0118204	0.00853053	0.0105899	0.00967861
전담	위원회	학생	생명	따돌림	절차	심의	컨설팅
0.00823458	0.00885454	0.00944554	0.00833299	0.0118204	0.00853053	0.00847404	0.00967861

[주제 3]은 학교폭력 예방을 위한 활동으로 청소년 교육과 공연 등을 통한 캠페인에 대한 내용으로 구성되어 있다. [주제 4]는 최근의 언어적 학교폭력에 대한 징계 요구와 관련된 내용이라고 할 수 있다. [주제 5]는 상담을 통해 피해학생을 조기에 발견하여 심리적인 지원을 실시하고 온라인에서의 학교폭력 피해 예방과 관련된 내용이다. [주제 6]은 학교폭력위원회에서의 운영 사례에 대한 내용으로 피해자 대책마련 및 가해자 처분 등의 해결 절차 및 조치 사안에 대한 내용이다. [주제 7]은 학교폭력 전문 변호사의 법률적 상담과 피해심의 및 예방에 대한 내용이다. [주제 8]은 학교폭력 예방을 위한 컨설팅과 온라인 위탁 강의 등 전문 업체에서의 강의진행과 관련된 내용으로 도출되었다.

5. 학교폭력의 유형별 주제 도출

토픽모델링 분석을 통해 도출된 주제별 키워드를 유형별로 정리한 결과는 아래의 <Table 3>과 같이 정리할 수 있다.

<Table 3> Results of Topic Modelling for School Bullying

Topic	Topics
1	학교폭력 예방 전담기구
2	학교폭력 피해 및 가해 증가 현황
3	학교폭력 예방을 위한 캠페인 활동
4	언어적 학교폭력의 최근 사건과 현안
5	상담을 통한 온라인 학교폭력 피해 예방
6	학교폭력위원회의 절차 및 조치 사안
7	학교폭력 전문 변호사의 법률적 절차
8	학교폭력 예방을 위한 전문외부기관

대체로 학교폭력의 피해 예방과 가해자 처벌 및 조치 사안에 대한 내용으로 도출된 것을 살펴볼 수 있으며, 크게 3가지의 내용으로 구분할 수 있다.

첫째, 학교폭력 예방활동이다. [주제 1]에서 살펴볼 수 있는 내용과 같이 학교폭력예방을 위한 전문기구로서 교육청과 경찰 및 자치단체가 함께 역할을 수행해야 하는 의미가 반영된 결과라고 할 수 있다. [주제 3]과 [주제 8]은 각각 교육공동체에서 수행되는 캠페인 활동과 전문강의로 진행되는 예방활동으로서 교육을 통해 학교폭력을 예방하고 피해사실이 있을 경우 조치사항 등에 대한 내용으로 수행되는 활동이다.

둘째, 학교폭력에 대한 조치사항과 절차에 대한 내용이다. [주제 6]에서 살펴볼 수 있는 내용과 같이 학교 내에 설치되어 있는 학교폭력위원회에서 수행되는 절차와 조치 사안에 대한 내용과 함께, [주제 7]에서 제시된 법률적 절차를 통한 전문변호사의 피해심의 및 예방에 대한 내용으로 주제가 도출된 것을 알 수 있다. 상시 예방 및 조치사항으로 [주제 5]에서 제시된 상담을 통한 학교폭력 피해예방활동을 통해 관련 절차와 내용이 도출된 것을 살펴볼 수 있다.

셋째, 학교폭력의 최근 현안에 대한 내용으로서, [주제 2]와 [주제 4]에서 제시된 내용과 같이 최근에 발생하고 있는 학교폭력 피해 및 가해의 현황과 더불어 언어적 폭력 및 온라인에서 발생하는 학교폭력에 대한 최근 현안에 대한 최근의 사건에 대하여 도출된 결과를 살펴볼 수 있다.

6. 결론 및 시사점

본 연구는 학교현장에서 지속적으로 강조되고 있는 학교폭력 예방을 위한 방안을 도출하기 위한 노력의 일환으로 최근의 학교폭력 관련 이슈화되고 있는 주제를 데이터과학의 관점으로 살펴보고자 하였다. 특히, 현재 청소년들은 MZ세대로서 디지털 네이티브라는 특징을 갖고 있다는 점에서 온라인 SNS데이터를 활용하여 학교폭력 관련 게시물을 크롤링하고 토픽모델링 방법을 활용하여 유형별 특징을 살펴보고자 하였다.

데이터과학에서 수행되는 빅데이터 기반의 데이터기반 연구(Data-driven analysis)는 일상생활에서 겉으로 드러나지 않는 문자와 수치데이터를 토대로 경향성이나 유형별 특징을 살펴본다는 의미에서 교육현장에서 겉으로 쉽게 드러나지 않는 현상을 살펴볼 수 있는 의미있는 분석 방법이라고 할 수 있다. 토픽모델링을 통한 유형분석 뿐만 아니라 데이터기반의 예측을 기반으로 학교현장에서 사전에 예측하고 진단하는 과정을 통해 당면하고 있는 사회적 문제해결에 활용하는 연구가 수행될 필요가 있다.

참고문헌

[1] D. M. Blei, A. Y. Ng & M. I. Jordan. (2003). Latent Dirichlet allocation, *Journal of Machine Learning Research, 3*, 993-1022.

[2] Hernandez-de-Menendez, M., Díaz, C. A. E., & Morales-Menendez, R. (2020). Educational experiences with Generation Z. *International Journal on Interactive Design and Manufacturing*, 14(3), 847-859.

[3] Ng, W. (2012). Can we teach digital natives digital literacy?. *Computers & education, 59*(3), 1065-1078.

[4] Olweus, D. (1994). *Bullying at school. In Aggressive behavior* (pp. 97-130). Springer, Boston, MA.

[5] Prensky, M. (2001). *Digital natives, digital immigrants part 2: Do they really think differently?*. On the horizon.

[6] Sharp, S., & Smith, P. (2002). *School bullying: Insights and perspectives*. Routledge.

[7] Wren, D. J. (1999). School culture: Exploring the hidden curriculum. *Adolescence,* 34(135), 593-593.

[8] 교육부(2020). 제4차 학교폭력 예방 및 대책 5개년 기본계획. 교육부 정책자료(2020.12.21.)

[9] 교육부(2021). 2020년 학교폭력 실태조사 결과 발표. 교육부 보도자료(2021.01.21.).

[10] 교육부(2021). 학생 사이버폭력 예방과 대응 강화를 위해 7개 부처, 6개 기관이 힘을 모은다!. 교육부 보도자료(2021.05.26.)

[11] 이현정(2021). AI시대, 메타버스를 아우르는 새로운 공감개념 필요성에 대한 담론. 한국게임학회논문지. 21(3), 79-90.

저자소개

신 승 기

2017 University of Georgia (Ph.D.)

2016~2017 미국 칼빈슨 정부연구소 연구원

2019~2020 애리조나주립대학교 컴퓨터교육전공 교수

2020~현재 서울교육대학교 컴퓨터교육과 교수

관심분야: Computational Thinking, 인공지능교육, 보편적정보교육

e-mail: skshin@snue.ac.kr