

텍스트 마이닝을 이용한 시대별 청소년 문제 토픽 분석

조경원¹ · 조주연^{2*}

¹고신대학교 · ²신라대학교

Topic Analysis on the Adolescent Problem Using Text Mining

Kyoung Won Cho¹ · Cho, Ju-Yeon^{2*}

¹Kosin University · ²Silla University

E-mail : twinsy2k@hanmail.net / kwcho@kosin.ac.kr

ABSTRACT

This research was conducted to identify adolescent problems in internet articles. This research defines adolescent problems as diverse issues related to adolescents and examine how it was dealt in the media to find out how different categories and the aspect of adolescent problems are changing by time. The result of the research was that in 1990's, education policy and family were mainly dealt with when it came to adolescent problems. As the era is changing, adolescent problems were far diversified compared to the past, and each problems are dealt with similar importance. This research is significant in that it does not only examine the social trend adolescent problems but also expand the range of adolescent counselling and utilizes quantitative analysis in considering diversity to provide new information.

Key words

Adolescent problems, Data mining, Topic modeling, Big data, Internet new article

I. 서 론

본 연구는 데이터마이닝을 활용하여 청소년의 문제가 미디어 상에서 어떻게 변화되어져 왔는지 알아보고, 청소년문제 및 문제행동 관련한 요인들을 사회적 사건을 다루는 미디어에서 도출하고자 한다. 최근 청소년 문제의 특징은 연령이 낮아지고 개인적 가정적 문제에서 전통적 가치관을 위협하는 강력범죄까지 다양화되어 하나의 사회적 문제로 조망 받고 있다(맹창영, 2013). 청소년문제와 청소년 문제행동의 구분은 개념상 상호 배타적이지 않고 구분하기도 어렵고, 모호하다는 점이 다수의 공통적인 지적사항이기도 하다(김정열, 2017; 신정업, 2017; 천정웅, 2017). 청소년문제 및 문제행동 관련한 요인들을 파악하기 위해 정보의 원천인 온라인 상 데이터를 수집하여 현재 다양하게 활용되고 있는 빅데이터 분석법 중 하나인 데이터마이닝을 활용하여 이슈를 도출하고 이슈 내 핵심 키워드의 빈도와 문서수를 통해 주요 토픽을 도출하여 청소년 문제의 패턴을 찾아보고자 한다.

II. 연구방법

분석을 위해 수집한 데이터는 조선일보와 동아일보는 1993년부터 2018년까지, 중앙일보는 1966년부터 2018년까지의 기사이다. 조선일보는 10,585건, 동아일보는 5,478건, 중앙일보는 901건으로 총 16,964건의 기사가 검색되었고, 웹크롤링을 통해 기사작성연도, 기사제목, 기사본문을 수집하였다. 데이터 전처리를 위하여 웹크롤링을 활용하여 수집된 데이터 중 기사본문의 형태소 분석을 실행하여 명사를 추출하여 기사본문을 대체하였다. LDA 기법을 활용하여 토픽 모델링을 실시하였고 각 토픽을 나타내는 주요구성단어 20개씩을 추출하였으며, 주제별 구성단어의 연관성을 중심으로 연구자가 토픽의 이름을 정하였다.

III. 연구결과

1993-1997년(1구간)의 특징은 '교육정책'과 '가정'이라는 두 토픽이 다른 토픽에 비해 청소년문제로 대부분의 비중을 차지한다는 것이다. 1998-2002년(2구간)은 '청소년상담'과 '성범죄'가 연도별로 동일한 패턴을 보이고 '무료지원프로그램' 함께 토픽화 되었다. 2003-2007년(3구간)의 토픽에서는 '사회문제'가 최상위토픽을 차지하였다. 2008-2012년(4

* Corresponding Author

표 1. 시대별 토픽분석

토픽 (1993-1997)	비중	토픽 (1998-2002)	비중	토픽 (2003-2007)	비중	토픽 (2008-2012)	비중	토픽 (2013-2018)	비중
교육정책	0.510	교육	0.358	사회문제	0.318	문화컨텐츠(문학)	0.145	성폭력	0.141
가정	0.230	정치	0.230	가정문제	0.274	교육정책	0.126	가정	0.124
학교폭력	0.052	청소년상담	0.103	정부지원	0.138	문화컨텐츠 (대중문화)	0.124	사회공헌활동	0.119
문화컨텐츠(방송)	0.049	무료자원프로그램	0.064	축구스포츠	0.048	진로프로그램	0.122	대한민국역사의 식	0.111
문화컨텐츠(문학)	0.032	성범죄	0.057	건강	0.033	교육지원사업	0.118	정부지원정책	0.096
청소년건강유해	0.021	문화컨텐츠(음악)	0.036	입시	0.032	청소년기 건강	0.068	엔터테인먼트(영상)	0.075
문화컨텐츠(공연)	0.016	인터넷서비스	0.032	문화컨텐츠(문학)	0.031	스포츠	0.051	문화컨텐츠	0.072
청소년상담	0.016	월드컵	0.018	문화컨텐츠(종합)	0.029	음란물	0.040	건강	0.050
문화컨텐츠(영상)	0.015	학교	0.016	정보통신	0.024	학교폭력	0.040	스포츠	0.061
약자보호 정책	0.012	청소년건강유해 물	0.016	학교폭력	0.017	성폭력	0.038	아동보호정책	0.033
인터넷활용	0.011	문화컨텐츠(영상)	0.014	논술	0.015	가정	0.035	입시	0.030
음란물	0.008	음주흡연문제	0.012	게임중독	0.009	청소년상담	0.024	북한청소년인권 문제	0.021
정신건강	0.007	한일관계	0.011	정치	0.009	학업	0.023	청소년건강유해 행태	0.021
청소년 유해물	0.007	입시	0.011	청소년역사의식	0.009	게임중독	0.016	학교폭력	0.013
청소년문제	0.007	문화컨텐츠(문학)	0.008	성매매	0.008	청소년중독	0.011	음란물	0.012
스포츠	0.005	게임중독	0.007	청소년흡연	0.004	청소년건강유해	0.010	게임중독	0.010
대북문제	0.002	돈	0.006	존속상해	0.002	북한문제	0.008	아동청소년 성범죄	0.009

구간)의 토픽 특징은 2003년 ‘문화컨텐츠(문학)’이 토픽의 높은 비중을 차지하였다가 2009년부터 전체 토픽이 골고루 기사화된 양상을 보이며, 이 현상은 4구간이 끝나는 2012년까지 지속되었다. 2013-2018년(5구간)의 토픽 특징은 전체 토픽이 골고루 분포되어 있고 사회적 이슈와 청소년문제가 함께 움직인다는 것이다.

청소년문제를 다룬 기사들을 통해 1993년에서 2018년까지 시대적 변화에 따른 청소년 문제의 변화 양상을 파악하고자 하였다. 이 연구를 통해 청소년문제의 해결을 위한 새로운 방향과 전략을 설정하는데 기초자료로 활용될 수 있을 것이라 기대한다.

IV. 결 론

1993-1997년(1구간)에서 2013-2018년(5구간)까지 전 구간에 걸쳐 청소년문제의 대상이나 장소, 관련 주요문제는 공통으로 나타났다. 즉, 전체 기간 토픽을 비교했을 때 상위순위의 단어들은 말, 학생, 학교, 부모, 교육, 사회, 정부, 프로그램 등으로 학생이나 부모, 정부는 청소년문제의 주체와 주요 관련자로, 학교, 사회는 청소년문제의 주요 장소인 것으로 파악된다. 구간마다 정부, 국가, 대통령, 선거, 정책, 후보가 나타나는데, 이는 4년 및 5년마다 총선과 대선이 있었고, 공약에 청소년문제가 언급되었기 때문으로 해석된다. 1993-1997년(1구간)에서 5구간(2013-2018년)으로 갈수록 이전 구간에 비해 토픽이 다양해지고 5구간(2013-2018년)에서는 토픽 간 유사한 비중을 보이고 있다. 이는 청소년문제의 토픽이 다양해짐을 의미한다.

결론적으로 본 연구는 인터넷 뉴스 사이트에서

References

- [1] C. -Y. Maeng, “Study on the Causes and Prevention of Juvenile Crime,” M.S. Thesis. Department of National Defence & Police administration, The Graduate School Konyang University, 2013.
- [2] J. Y. Kim, *Youth Problems*. Seoul: Knowledge Community, 2017.
- [3] D. U. Shin, “The Adolescents' Problem Behavior Pattern and Educational Guidance Method at School,” *The Journal of Welfare and Counselling Education*, vol. 6, no.2, pp.157-185, 2017.
- [4] J. W. Chun, *Youth work practice : positive perspective*, Seoul: Shinjeong, 2017.