

# 텍스트마이닝을 활용한 자살 관련 논문 토픽 모델링

조경원<sup>1\*</sup> · 김하영<sup>1</sup> · 김미리<sup>1</sup> · 우영운<sup>2</sup>

<sup>1</sup>고신대학교 의료경영학과

<sup>2</sup>동의대학교 응용소프트웨어공학과

## Topic Modeling of Suicide Papers using Text Mining

Kyoung Won Cho<sup>1\*</sup> · Ha-young Kim<sup>1</sup> · Mi-ri Kim<sup>1</sup> · Young Woon Woo<sup>2</sup>

<sup>1</sup>Kosin University, <sup>2</sup>Dong-eui University

E-mail : kwcho@kosin.ac.kr / ywoo@deu.ac.kr

### 요 약

본 연구에서는 지금까지 게재된 자살 관련 논문들을 대상으로 주제를 분류하고 주요 주제들의 비중과 20년 동안의 주제들의 변화 추이를 파악하는 것을 연구 목적으로 한다. 이를 위해 등재지 논문에 대한 정보 공유가 가장 활발하게 이루어지는 한국학술지인용색인(KCI)을 자료원으로 빅데이터 분석에 활용되는 텍스트마이닝 기법을 시행하였다. 이러한 시대적 변화에 따른 자살 관련 연구 주제 추이를 파악하는 연구는 향후 자살과 관련한 학문적 방향을 시대 흐름에 적응할 수 있는 전략을 수립하는 데 있어 기초자료가 될 것이다.

### ABSTRACT

The purpose of this study is to classify the topics related to the suicide papers published so far and to identify the proportions of the main topics and the trends of the topics over the past 20 years. For this purpose, a text mining technique used in big data analysis was used as a data base of the Korean Journal of Citation Index (KCI), where information sharing about the papers is most active. This study, which grasps the trends of suicide related research according to the changes of the times, will become a basic data for establishing a strategy to adapt the academic direction related to suicide in the future.

### 키워드

Papers on suicide, Text mining, Topic modeling, Topic Trend

### 1. 서 론

우리나라는 부진한 자살예방 사업으로 자살률이 OECD 국가 중 1위를 유지하다가 최근 우리나라보다 자살률이 높은 리투아니아가 가입함으로써 2위가 되었지만, 3위와는 큰 차이가 나는 것을 볼 수 있다[1]. 이에 사회 및 가정의 고통이 증가하고 있을 뿐 아니라 미래의 인적자원 감소로 인해 국가 경쟁력과 발전이 저해될 우려가 있다.

최근 우리나라 자살자는 거의 모든 계층에서 지속적으로 늘어나고 있다. 보건복지부에 따르면 2017년 기준으로 개인 특성별 자살 현황에서 연령 대별에서는 노년층에서 가장 자살률이 높았으며

특히 80대에서 두드러진 현상을 볼 수 있었다[2]. 노인자살을 유발하는 요인은 생물학적, 사회적, 심리적 등 복합적인 요인에 의해서 발생한다. 배우자와의 사별이나 자아존중감의 상실, 심신 건강의 악화, 경제적 어려움, 가족이나 사회적 지지의 저하, 젊은이들이 노인을 대하는 부정적인 태도 등이 노인의 자살률에 영향을 미친다[3].

또한, 우리나라 20세 이상 성인의 지난 1년간 자살 생각률은 14.8%이며, 우울감을 경험한 경우, 지각된 스트레스가 심한 경우, 여성, 65세 이상의 연령, 저학력, 미혼 등의 집단이 자살 생각률이 높은 것으로 나타났다[4].

우리나라의 자살 대응 정책의 현황을 보면 2004년 국가적 차원의 전략인 12월 제1차 국가 자살 예방 5개년 기본계획(2004년 - 2008년)을 발표하여

\* speaker

10대 과제를 제시하고 2005년 9월에는 12개의 세부사업 내용을 추가 발표하여 시행하였으나, 자살 예방을 위한 법률과 제도가 미비하여 재원 확보 및 중앙과 지역 및 공동체 간의 협력적 거버넌스가 구축되지 못하였다[5].

따라서, 본 연구에서는 이러한 문제를 해결하기 위한 기초 자료로 활용될 수 있도록 자살 관련 논문을 대상으로 빅데이터 분석을 활용하고자 한다. 지금까지 약 20년간 게재된 자살 관련 논문들의 주제와 그 비중에 대한 객관적인 검토를 한 연구는 찾아볼 수 없었다. 본 연구에서는 지금까지 게재된 자살 관련 논문들을 대상으로 주제를 분류하고 주요 주제들의 비중과 20년 동안의 주제들의 변화추이를 파악하는 것을 연구 목적으로 한다.

## II. 연구 방법

### 2.1 자료 수집

본 연구에서는 KCI에 등재된 논문들 중 “자살” 키워드로 검색된 논문 1,992개 중에서 제목이 없는 불분명한 논문과 중복된 논문을 제외한 2001년부터 2019년 현재까지의 논문 1,680개의 한글 초록 내용을 텍스트 마이닝 기법을 통해 분석하였다. 수집된 학문 분야별 학술지 개수 및 비율, 그리고 논문 개수 및 비율은 표 1과 같으며, 연도별 논문 개수는 표 2와 같다.

표 1. 학문 분야별 학술지와 논문 통계정보

| 학문 분야별    | 논문 개수        | 논문 비율       | 학술지 개수     | 학술지 비율      |
|-----------|--------------|-------------|------------|-------------|
| 인문학       | 201          | 10%         | 100        | 31%         |
| 사회과학      | 995          | 51%         | 100        | 31%         |
| 의약학       | 497          | 25%         | 61         | 18%         |
| 복합학       | 162          | 8%          | 26         | 8%          |
| 자연과학      | 68           | 3%          | 15         | 5%          |
| 공학        | 47           | 2%          | 10         | 3%          |
| 예술체육      | 20           | 1%          | 11         | 3%          |
| 농수해양      | 2            | 0%          | 2          | 1%          |
| <b>합계</b> | <b>1,992</b> | <b>100%</b> | <b>325</b> | <b>100%</b> |

표 2. 연도별 논문 개수

| 연도별       | 논문 개수 | 연도별          | 논문 개수 |
|-----------|-------|--------------|-------|
| 2019      | 40    | 2009         | 93    |
| 2018      | 179   | 2008         | 72    |
| 2017      | 183   | 2007         | 53    |
| 2016      | 225   | 2006         | 36    |
| 2015      | 203   | 2005         | 29    |
| 2014      | 216   | 2004         | 23    |
| 2013      | 178   | 2003         | 17    |
| 2012      | 161   | 2002         | 11    |
| 2011      | 147   | 2001         | 3     |
| 2010      | 121   | 1996         | 2     |
| <b>합계</b> |       | <b>1,992</b> |       |

### 2.2 데이터 처리 및 분석

수집된 논문들을 대상으로 LDA 분석[6]을 실시하여 토픽을 추출하고, 추출된 토픽들의 추이를 산출하였다. 이를 위해 먼저 토픽 수를 정하여야 한다. 본 연구에서는 perplexity를 이용하여 합리적인 토픽 수를 9개로 결정하였다[7][8]. 이에 따라 9개의 토픽을 나타내는 빈도수 기준 상위 단어 20개씩을 추출하였다. 토픽별 구성 단어의 연관성을 중심으로 해당 분야의 전문가 검토에 따라 토픽의 이름을 정하였으며, 9개 토픽의 연도에 따른 논문의 대표 토픽들의 변화를 파악하기 위해 연도별 토픽 추이를 그래프로 분석하였다.

## III. 연구 결과

“자살” 관련 연구에서 중심이 되는 토픽과 관련 핵심어를 추출하기 위하여 토픽 모델링을 실시한 결과는 표 3과 같다. 각 토픽은 LDA 기반 토픽 모델링 알고리즘을 사용하여 토픽들의 비중 순으로 제시하였다. 각 토픽을 나타내는 빈도수 기준 상위 20개 단어들을 각각 추출하였으며, 토픽별 상위 20개 단어들의 연관성을 중심으로 토픽의 이름을 정하였다.

토픽 모델링에 의한 토픽 분석 결과 알 수 있듯이, 우리나라가 당면하고 있는 자살에 관련된 주요 문제들이 명확하게 추출되고 있다는 점이다. 특히 ‘한국 사회의 자살 실태’, ‘노인 자살’, ‘청소년의 자살 생각’, ‘청소년기 정신적 스트레스’, ‘고위험군 정신질환자의 자살’의 4가지 주제가 10% 이상의 비중으로 상위 5가지 주제를 차지함을 알 수 있었다.

하지만 ‘대학생 부정적 스트레스 대처’, ‘자살 관련 생명보험’, ‘스트레스와 정서적 부적응’, ‘중·노년기 신체 상태, 사회적 위치 변화에 따른 부적응’의 4가지 주제도 비중이 높지는 않지만 명확하게 분리된 주제로 도출됨을 알 수 있어, 많은 분야의 학술 논문들이 자살과 관련된 주요 사회 이슈에 대한 연구들을 꾸준히 다루어 오고 있음을 알 수 있었다.

## IV. 결 론

본 논문에서는 “자살”의 제목으로 검색하여 추출된 KCI의 게재 논문들을 중심으로 데이터 분석을 실시하였다. “자살” 주제로 2001년부터 2019년 현재까지 게재된 논문들 중 국문 초록을 가지고 있는 1,680편을 분석에 활용하였다. 분석 결과, 수집된 논문들이 발행된 전체 기간에 걸쳐 우리나라 사회가 당면한 자살과 관련된 주요 이슈들이 모두 다루어지고 있음을 알 수 있었으며, 특히 9가지 주제들 중 5가지 주제가 10% 이상의 큰 비중으로 다루어지고 있으며, 나머지 4가지 주제는 상대적으로 낮은 비중으로 다루어지고 있음을 확인할 수 있었

다. 하지만 각 주제별로 상위 20개의 단어들을 면밀히 분석한 결과, 도출된 9가지 주제들은 비중에 관계없이 모두 명확하고 구분된 주제로 변별력을 충분히 가지고 있음을 파악할 수 있었다.

표 3. 토픽 분석 결과

| 토픽명                           | 비중   | 토픽별 상위 20개 단어  |
|-------------------------------|------|--|
| 청소년기 정신적 스트레스                 | 15.6 | 청소년, 자살생각, 학교폭력, 중학생, 경험, 피해경험, 부모, 학대, 여아, 고등학생, 학업스트레스, 학대경험, 우울, 스트레스, 자아존중감, 우울감, 고등부, 자살충동, 폭력, 성별          |
| 스트레스와 정서적 부적응                 | 2.5  | 병사, 우울, 사회, 간호대학생, 자아존중감, 자살관련행동, 프로그램, 극통, 완벽주의, 자살생각, 복무, 중년 여성, 문제해결능력, 심리적, 부적응, 소외감, 자살사고, 생활, 생활스트레스, 자살의도 |
| 대학생 부정적 스트레스 대처               | 4.2  | 대학생, 취업스트레스, 좌절, 자살생각, 소속감, 외상, 스트레스, 우울, 삶, 자살사고, 정서조절, 대인관계, 생활스트레스, 지각, 외상경험, 대학, 대처, 성장, 우울성향, 의미            |
| 자살 관련 생명보험                    | 3.1  | 보험자, 판결, 피보험자, 대법원, 보험금, 약관, 보험사고, 재해, 재해사망보험금, 조력자살, 고의, 보험약관, 규정, 선고, 의사, 정신질환, 사망, 생명보험, 판례, 환자               |
| 중노년기 신체 상태, 사회적 위치 변화에 따른 부적응 | 2    | 수치심, 회복탄력성, 내면화, 중년기, 자살생각, 상실감, 위기감, 자기비난, 장애수용, 자살사고, 고립감, 실행력, 자살실행력, 잠재계층, 사회참여, 긍정성, 도장애인, 독거노인, 대학생, 우울    |
| 고위험군 정신질환자의 자살                | 10.4 | 환자, 우울증, 증상, 자살보도, 자살 시도, 인터넷, 보도, 자살행동, 청소년, 치료, 응급실, 자살시도자, 수면, 충동성, 임상, 불안, 알코올, 장애, 자살자, 행동                  |
| 한국 사회의 자살 실태                  | 24   | 자살률, 죽음, 사회, 삶, 자살자, 자살예방, 경험, 청소년, 고통, 개인, 한국, 태도, 생명, 인식, 여성, 우리나라, 가족, 문화, 노인, 한국사회                           |
| 노인 자살                         | 19.6 | 노인, 자살생각, 우울, 독거노인, 사회, 스트레스, 노인자살, 건강상태, 우울증, 우울감, 경험, 삶, 가족, 연령, 성별, 자살위험, 성인, 절망감, 생각, 고독감                    |
| 청소년의 자살 생각                    | 18.5 | 청소년, 자살생각, 우울, 스트레스, 고등학생, 대학생, 중학생, 프로그램, 중독, 무망감, 부모, 여아, 학생, 자살충동, 자아탄력성, 자아존중감, 스마트폰, 고등부, 사회, 삶             |

References

[1] Korea Joongang Daily. 13 years 'OECD suicide rate first place' Korea suddenly came down to 2nd place [Internet]. Available: <https://news.joins.com/article/22824428>.

[2] Korea Suicide Prevention Center. Suicide status by individual characteristics [Internet]. Available: <http://spckorea-stat.or.kr/korea02.do>.

[3] C. Y. Koo, J. S. Kim and J. Yu, "PA Study on Factors Influencing Elders' Suicidal Ideation: Focused on Comparison of Gender Differences," *Journal of Korean Academic Community Health Nurse*, vol. 25, no. 1, pp. 24-32, Mar. 2014.

[4] E. Park and S. J. Choi, "Prevalence of Suicidal Ideation and related Risk Factors among Korean Adults," *Journal of Korean Academic Psychiatric Mental Health Nurse*, vol. 22, no. 2, pp. 88-96, June 2013.

[5] S. Y. Won, "Problems of Suicide Prevention and Improvement Tasks," National Assembly Research Service: Seoul, Report 117-20110315, Mar. 2011.

[6] K. Y. Cho, S. K. Bae and Y. W. Woo, "Analysis on Topic Trends and Topic Modeling of KSHSM Journal Papers using Text Mining," *The Korean Journal of Health Service Management*, vol. 11, no. 4, pp. 213 - 224, Dec. 2017.

[7] C. H. Nam, "An Illustrative Application of Topic Modeling Method to a Farmer's Diary," *Comparative Culture Study*, vol. 22, no. 1, pp. 89-135, Jan. 2016.

[8] J.-Y. Cho and K. Y. Cho, "Topic Modeling on the Adolescent Problem Using Text Mining," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 12, pp. 1589-1595, Dec. 2018.