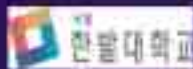


빅데이터를 활용한 정책분석 대상과제 선정방안 연구

한밭대학교 공공행정학과
권가석



2022.05.19

차례

1. 서론 : 배경 및 필요성
2. 정책분석 대상과제의 선정 방법론
3. 정책 빅데이터 관리 플랫폼 구축 방향
4. 향후 추진 과제

1. 서론 : 배경 및 필요성

1.1 정책과정에서의 빅데이터의 역할

○ 현행 정책 평가제도는 주로 각 부처가 연간 추진한 정책에 대해 사후적으로 평가하는 연단위 평가 기반으로 운영되고 있음. 그러나 행정 환경의 급변 등으로 적기 대응을 위한 수시과제 도출이 중요해 짐. 또한 경제사회적으로 정책적 영향(impact)이 큰 과제를 조기에 발견해 내는 것이 중요해 졌음. 이에 빅데이터나 인공지능을 활용한 정책 대상과제의 도출방식이 주목을 받고 있음.

1. 서론 : 배경 및 필요성

1.1 정책과정에서의 빅데이터의 역할

○ 해외 시스템에 대한 벤치마킹이나, 특정 분야의 빅데이터 분석을 통해 어떻게 개별 정책에 반영할 것인가에 대한 연구는 다수 있었음.

○ 빅데이터의 정책과정에서의 역할과 다양한 선행연구에도 불구하고, 정책과정 전반에서 빅데이터를 어떻게 구체적으로 적용해서 활용할 것인지와 시스템 구축에 대한 절차와 방안 제시는 충분히 이루어지고 있지 않음. 이에 본 연구는 빅데이터 분석 시스템을 활용하여 체계적이고 효율적인 정책분석 대상 과제 발굴 방안을 제시하고자 함.

〈그림 1〉 정책과정에서의 빅데이터의 역할
출처 : 한국정보화진흥원(2012), p.24.

1. 서론 : 배경 및 필요성

1.2 기존 빅데이터 분석 정책대응의 한계

가. 빅데이터와 빅데이터 분석

- 빅데이터는 방대한 양의 데이터로서 이를 분석하고 활용하는 기술을 포괄하여 지칭함 (이마숙 외, 2014). 이러한 빅데이터의 특성으로 큰 규모(Volume)와 다양성(Variety), 속도(Velocity)가 언급되며 (안창권·황승구, 2012), 최근의 페이스북, 트위터 등과 같은 소셜네트워크 서비스 상에서 즉각적으로 방대한 데이터를 수집하여 분석할 수 있는 시대가 도래하였음.



<그림 2> 빅데이터의 특성 - 3V (출처 : 안창권·황승구, 2012)

1. 서론 : 배경 및 필요성

1.2 기존 빅데이터 분석 정책대응의 한계

나. 일반적인 빅데이터 분석의 개요

- 본 연구에서는 기존 빅데이터 분석에서 일반적으로 이루어지는 방법론을 바탕으로 정책과제를 도출해 보았음. 보통 주어진 주제를 바탕으로 하여 신문기사, SNS, 웹사이트 등에서 텍스트 또는 말뭉치를 수집해 빈도 등을 수집하여 빈도분석과 네트워크 분석을 통해 정책이슈를 도출함. 다음 그림은 일반적인 텍스트 빅데이터 분석의 과정을 보여주고 있음.



1. 서론 : 배경 및 필요성

1.2 기존 빅데이터 분석 정책대응의 한계



1. 서론 : 배경 및 필요성

1.2 기존 빅데이터 분석 정책대응의 한계

다. 시범적 코로나 이슈 빅데이터 분석

- 본 시험적 분석에서는 코로나의 경우, 최근의 대표적인 정책현황 또는 정책문제로서 코로나19에 대한 분석을 진행해 보았음, 이를 통해 기존의 분석 방식이 가지는 의미와 한계에 대하여 논의하여 보고자 함.
- 다음 그림은 2020년 5월 코로나 관련 신문기사 7,011건의 텍스트를 분석해 본 결과임. 일반적으로 텍스트 분석을 통해 주제어의 빈도분석과 주제어 사이의 긴밀한 관계를 바탕으로 주제어별 군집을 분석하여 정책이슈를 도출하게 됨.

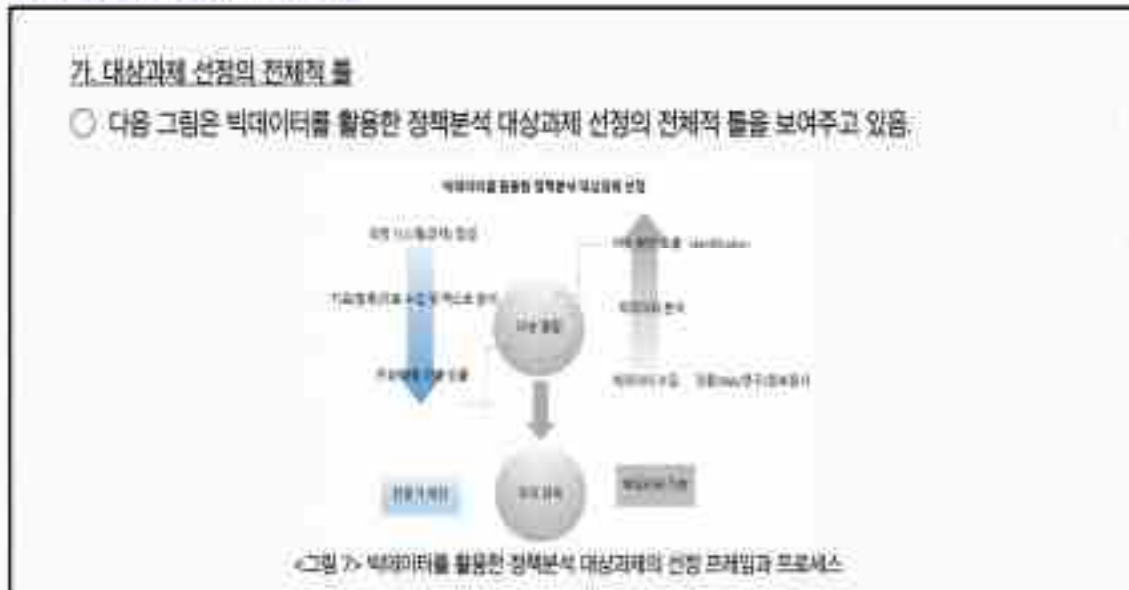
1. 서론 : 배경 및 필요성

1.3 본 연구의 특징과 주요내용



2. 정책분석 대상과제의 선정 방법론

2.1 정책분석 대상과제 선정 틀



2. 정책분석 대상과제의 선정 방법론

2.1 정책분석 대상과제 선정 틀

나. 전체적인 시스템 관점에서의 진단과 관리 (Top-down 관점) | 워크넷

- 국정관리 시스템 관점에서 국방, 복지, 과학기술 등의 분야별 또는 부처별로 세부 항목과 지표 체계를 구축하고, 선정된 관련 주제어를 기반으로 언론, SNS 등의 빅데이터를 분석하여 연결된 주제어를 도출하여 관련 대응 이슈를 도출하게 됨. 도출된 주제어나 네트워크에 대한 시사점 도출을 분야별 정책 전문가가 주도하게 됨.

다. 정보의 수집과 빅데이터 분석을 통한 이슈의 도출과 의제설정 (Bottom-up 관점)

- 언론, SNS, 정부문서, 학술데이터 등 다양한 데이터 범주별로 빈도수 등의 기준으로 하여 주요 주제어를 도출하게 됨. 이러한 주제어를 범주별로 비교하여 정책대응의 시차(time lag)나 대응 내용을 분석할 수 있게 됨.

2. 정책분석 대상과제의 선정 방법론

2.2 핵심어 선정 프로세스



2. 정책분석 대상과제의 선정 방법론

2.2 핵심어 선정 프로세스

- 아래 그림은 색인어 (index word)와 관련 핵심 이슈가 어떻게 연계되는지를 시각화하여 보여주고 있음.



<그림 9> 색인어와 핵심이슈의 관계도 예

2. 정책분석 대상과제의 선정 방법론

2.2 핵심어 선정 프로세스

나. 데이터 범주별 비교분석을 통한 핵심이슈 도출

- 다음 그림은 사회적 이슈와 학술 연구, 정부정책 관련 빅데이터 대상별로 주제어와 군집을 대상으로 각각 핵심이슈를 도출하고 이들을 비교하여 이슈별로 시기별 시차(time lag)가 있는지, 정책공백이 있는지 분석해 낼 수 있을 것임.



<그림 10> 빅데이터 범주별 비교분석을 통한 핵심이슈 도출 (과학기술의 예)

2. 정책분석 대상과제의 선정 방법론

2.2 핵심어 선정 프로세스

- 다음 그림은 언론, SNS 등 비정형 데이터를 시공간 관점에서 분석하여 사회이슈를 분석하는 절차를 보여주고 있음. 그림은 사회이슈 분석모델의 예를 보여주고 있음. 절차에 따르면, SNS 등에서 비정형 빅데이터를 수집하게 되고, 노이즈 제거 등 전처리를 실시하게 됨.



<그림 11> 비정형 데이터 기반 사회이슈 분석 절차의 예
출처 : 홍영희(2014), p.40의 그림을 수정·보완함

※ 감성분석 : 텍스트에 나타난 특정 주제에 대한 사람들의 태도, 평가, 의견, 상황 등의 주관적 경향을 텍스트 분석을 통해 도출하는 것을 의미함

※ 토픽모델링 : 텍스트에서 출현한 단어의 사용 패턴을 기반으로 하여 해당 단어들을 대표하는 특정 미수, 주제 군집을 추출하는 분석 방법임. 토픽(topic)은 동시에 출현할 확률이 높고 단어들의 그룹이라고 할 수 있으며 유사한 의미를 가진다고 할 수 있음

2. 정책분석 대상과제의 선정 방법론

2.2 핵심어 선정 프로세스

- 다음 그림은 언론분석을 통한 사회 핵심이슈 도출의 과정을 보여주고 있음. 개별 키워드와 긴밀한 키워드의 그룹핑을 통한 분석이 진행됨.



<그림 12> 언론 텍스트를 통한 사회 핵심이슈 도출 과정
출처 : 장요한(2020), p.17의 그림을 수정·보완함

2. 정책분석 대상과제의 선정 방법론

2.2 핵심어 선정 프르세스

- 다음 그림은 텍스트 네트워크 분석의 절차를 보여주고 있음. 텍스트 네트워크는 앞에서 제시한 개별 키워드와 키워드 군집을 도출하는 핵심적인 방법론으로 알려져 있음. 데이터의 선정, 전처리, 데이터 분석의 절차로 이루어 짐.



<그림 13> 텍스트 네트워크 분석의 절차
출처 : 김강민(2020), p.5.

2. 정책분석 대상과제의 선정 방법론

2.3 핵심 이슈의 통합, 연계와 대응

가. 하향적 이슈와 상향적 이슈의 종합 : 범부처 협의체 구축 및 운영

- 하향적 방식과 상향적 방식을 통해 도출된 핵심이슈는 국무조정실을 중심으로 구축된 빅데이터 기반 핵심 이슈 관리의 범부처 협의체에 의하여 체계적으로 종합될 것임. 이를 통해 관련 부처, 협력 기관으로 연계하여 핵심이슈에 대한 분석 및 대응을 수행하게 됨.

나. 관련 부처의 연계와 대응

- 다음 그림은 빅데이터 기반 핵심이슈 관리의 부처별 연계도를 보여주고 있음. 범부처 협의체의 실행기구인 빅데이터 기반 전담기구는 협력부처와 기관, 담당자와 연계하여 관련 데이터의 수집 관리 뿐만 아니라 정책 대응과 관련되는 아이디어까지 제공할 수 있음.

2. 정책분석 대상과제의 선정 방법론

2.3 핵심 이슈의 통합, 연계와 대응



3. 정책 빅데이터 관리 플랫폼 구축 방향

3.1 플랫폼의 개요와 구성요소

가. 정책 빅데이터 관리 플랫폼의 구성

○ 다음 그림은 (가칭) 정책 빅데이터 관리 플랫폼의 전체적인 개요와 구성을 보여주고 있음.



<그림 15> 정책 빅데이터 관리 플랫폼의 구성 출처 : 오미대(2019)의 p.35의 그림을 수정·보완함

3. 정책 빅데이터 관리 플랫폼 구축 방향

3.1 플랫폼의 개요와 구성요소

나. 정책 빅데이터의 수집, 관리, 활용 체계

- 다음 그림은 정책 빅데이터의 수집, 관리, 활용의 관점에서 그 체계를 보여주고 있음. 빅데이터의 수집에 있어서는 부처, 기관 등의 기관에서 수집을 하게 되고, 이는 전담기구가 관리하게 되며, 활용은 참여 부처와 연계하게 됨.



<그림 10> 빅데이터의 수집, 관리, 활용 체계

3. 정책 빅데이터 관리 플랫폼 구축 방향

3.2 플랫폼의 구축방안

- 다음 그림은 정책 빅데이터 관리 플랫폼의 운영체계를 체계, 과정, 운영시스템 세 수준에서 설명하고 있음.



<그림 11> 정책빅데이터 관리 플랫폼의 운영 개요

3. 정책 빅데이터 관리 플랫폼 구축 방향

3.3 플랫폼의 운영 방안과 거버넌스

- 다음 그림과 같이 정책빅데이터 플랫폼 운영 시스템 구축의 네 가지 과제를 제시할 수 있음.



<그림 1> 정책빅데이터 플랫폼 운영시스템 구축의 네 가지 과제
출처 : 오마에(2019), p.33.

4. 향후 추진 과제

4.1 시기별 추진과제

- 단기 과제 : 현재 운영 중인 '해안' 등 정책 관련 빅데이터 시스템의 한계를 보완하면서, 전문가 운영 위원회를 통해 시스템을 통한 분석결과와 활용을 극대화할 수 있는 방안을 마련하여야 할 것임.
- 장기 과제 : 정책 빅데이터 플랫폼 구축을 위한 비전과 전략, 세부 계획과 이를 뒷받침하는 추진주체의 결정, 협업체계, 예산 및 근거 규정 마련 등의 제도적 방안을 마련하여야 할 것임.

4. 향후 추진 과제

4.2 추진 시 고려사항

- 정책 키워드 온톨로지 구축 : 정부문서, 논문, 교과서 등의 텍스트마이닝을 통해 정책 키워드, 이슈, 범주를 포괄하는 온톨로지 구축이 병행되어야 함.
- 빅데이터 수집과 활용에 대한 우려 : 빅데이터의 수집에 있어 개인의 의사를 어디까지 인정해야 할 것인가 이슈가 있음. 또한 활용에 있어 편향된 (biased) 데이터가 수집되지 않도록 하는 장치 마련이 필요함.

※ 온톨로지(ontology) : 특정한 지식분야에 관련된 데이터 모델로서, 개념과 개념 사이의 관계와 이를 구성하는 어휘의 집합으로 구성됨. 어떤 주제나 개념이 전체 지식 체계에서 어떻게 자리매김 되는지 알려줌.

4. 향후 추진 과제

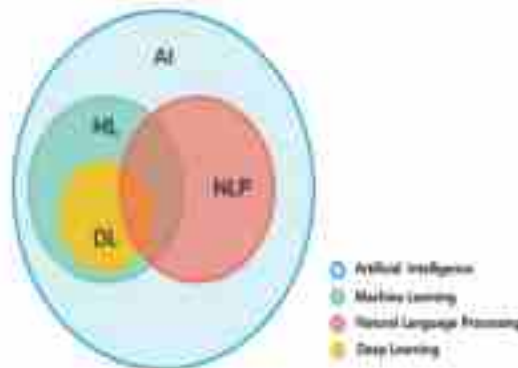
4.2 추진 시 고려사항

- 증거기반 정책 과정에 대한 인식 확산 : 빅데이터 기반의 정책의제 설정 및 정책결정 등 정책과정을 통해 정책의 과학성을 제고하여 정책과정 전반의 질을 제고할 수 있다는 인식을 확산하여야 함.
- 범정부 차원의 활용과 지방정부, 민간과의 연계 : 범정부적 차원에서 빅데이터를 활용하여 정책을 추진한다는 새로운 패러다임이 정착되고 지방정부, 민간의 데이터와 전문성을 연계하여 추진할 수 있는 플랫폼을 마련하여야 함.
- 법적 규제의 도출과 개선 : 빅데이터 기반 정책 시스템의 구축 및 운영과 관련된 규제를 도출하고 이를 개선하려는 노력을 경주해야 함. 예를 들어, 데이터의 인증과 표준화, 데이터 공개, 시스템의 연계 등의 문제가 해결되어야 함.

4. 향후 추진 과제

4.3 정책지능으로의 진화

- 인공지능(Artificial Intelligence)이라는 큰 범주에 기계학습(Machine Learning), 딥러닝(Dep Learning)이 있음. 다음 그림에 따르면 자연어를 처리(NLP)하는 데에 있어서, 앞에서 살펴본 토픽 분석, 감성분석 등이 있으며, 딥러닝 등 인공지능경향을 활용하는 방법도 적용이 될 수 있음.



<그림 19> 인공지능 활용의 기술적 영역 (출처: 공승민(2021), p.3에서 재인용)

4. 향후 추진 과제

4.3 정책지능으로의 진화

- 다음 그림은 인공지능경향을 활용하여 정책대응의 이슈리스트를 생산해 내는 과정을 보여주고 있음.



<그림 20> 정책이슈 도출을 위한 인공지능경향 적용 모델 예 (출처: 홍승현(2021), p.7 그림을 수정)

참고문헌

- 권기석 외(2014), 교육빅데이터를 활용한 마앤다 개발, 한국교육학술정보원
- 김강민 (2020), 엑스프로젝트워크 분석과검, 한국법제연구원 제1차 워크숍 (2020.3.20.)
- 성복준(2019), 빅데이터 분석을 적용한 정책 자문 연구, 국회입법조사처.
- 설지은 박기영(2014), 빅데이터를 활용한 정책 시뮬레이션과 시사점, 과학기술정책, 24(2), pp.94-106.
- 오아영(2019), 보건복지정책에서의 빅데이터 활용-전략과 과제, 미래의 초점, 보건복지포럼(2019.8)
- 유승희(2021), 인공지능과 자연어 처리기술 동향, 주간기술동향, 정보통신기획평가원
- 이미숙 외(2014), 빅데이터를 활용한 환경분야 정책수요 분석, 한국환경정책평가연구원
- 이현규(2015), 공공데이터와 빅데이터 활용사례 및 지방자치단체 추진전략, 지역정보화 동행분석, 11호, 한국지역정보개발원.
- 김요한(2020), 빅데이터를 이용한 국토 민생현황 모니터링 연구, 국토연구원
- 한국정보화진흥원(2012), 빅데이터 시대 / SNS의 진화와 공공정책, IT & Future Strategy, 제11호.
- 홍승현(2021), 인공지능법 기법의 최근태의 강박학습 알고리즘 개발 및 실증, 한양대 세미나자료, 2021. 12. 20.
- 홍영희(2014), 마앤 모니리움을 위한 맞춤형 빅데이터 시공간분석방법론 연구, 안양, 국토연구원
- Seo, I, Asnara, A. Y., and Kwon, K.-S. (2021), From knowledge arbitrager to policy entrepreneur?, Asian Journal of Innovation and Policy, 10(3).

Q&A