

Bigdata 연구방법론 활용방안

Proposal Seminar

2017. 4. 26

KEI

강성원

Random Thoughts

- 팀장이 뭐 하기는 해야 하는데 뭘하지?
 - 아는 것도 없고, 해 본것도 없고..
- 연구는 몇 차례나 시도도 못해보고 접고.... : 뭘놈의 문건이 이렇게 많아!
- 문건 만드면서 얻은 것을 써 먹으면서 할 수 있는것은??
 - 가만, 환경연구 + 빅데이터 ? 이거 많이 써 먹은 내용 같은데..
 - 이참에 정리?

1. 연구 목적

빅데이터분석기법 환경정책연구 적용가능성점검

- 문제제기: 빅데이터 연구에 대한 관심은 많지만 실제 연구 성과 희소
 - 기존 연구는 텍스트 마이닝을 설문조사 대용으로 사용하는 연구가 주종
 - (KEI) 점진적으로 분석기법을 직접 활용하는 연구가 출연(배현주, 이명진..)
- 환경정책연구에 빅데이터 분석기법을 어떻게 써야 하는지 고민 필요
 - 1) 빅데이터 분석기법의 장점: 패턴 파악, 예측
 - 2) 환경정책의 특징: 물리적 데이터 = 성과 지표, 지자체 중심, .etc
 - 1) + 2) 에서 시너지는 어떻게?
- 본 연구: 환경정책연구 특징+빅데이터 분석기법=> 시너지 가능성 파악
 - 환경정책연구: 환경정책 기능 도출 => 기능별 관련 연구 파악 => 연구특성 정리
 - 빅데이터 분석기법: 분석기법 정리 => 적용 가능 연구유형 파악

2. 연구내용 및 방법론

환경연구, 빅데이터 분석방법 특징 파악 => 적용방안 도출

1. 환경정책연구: 환경정책을 유형화 및 유형별 관련연구 특성 파악
 - 환경정책 유형화: 예산서 + 환경백서
 - 환경정책 연구 표본: 2016년 환경정책평가연구원 연구보고서
2. 빅데이터 분석방법: 주요 분석기법 정리 및 장점 파악
 - Supervised learning/Unsupervised learning 주요 기법 및 장점 소개
 - Deep Learning 별도 소개
3. 환경정책 유형별, 단계별 활용방안
 - 유형분류 특성에 기반한 적용방안 진단
 - 환경정책 수립-시행-평가 단계 별 적용방안 진단

1. 환경정책 유형별 관련연구 특징 추출

- 환경정책 유형화 => 유형별 연구 파악 => 연구 특징추출
 - 환경정책 도출과정 소개: '이슈 발굴 => 현황 진단 => 개선방안 '
 - 환경정책 유형화 1): 예산서 + 환경백서
 - 예산서의 구분을 따라서 환경백서에 소개된 환경정책을 재정리
 - 환경정책 유형화 2): 정책의 기능에 따른 분류를 시도
 - 환경오염물질 발생단계에 따라 정칙 기능 분류:
 - 억제/처리/예방/ 정책환경조성으로 분류(?)
 - 특징 파악: 정책 유형별 관련 연구를 수집하여 연구방법론 등 특징 파악
 - 2016년 한국환경정책평가연구원 보고서 자료 활용

환경정책 분류 기준(preliminary)

| | 억제 (규제, 조세) | 처리 (시설증축, 관리) | 예방 (예보, 환경보건..) | 환경조성.. (교육, 기술개발, 산업육성..) |
|------------------------------------|----------------|------------------|--------------------|------------------------------|
| 상하수도 | | | | |
| 수질 | | | | |
| 폐기물 | | | | |
| 대기 | | | | |
| 자연환경 | | | | |
| 환경일반 /보건/국제협력 환경교육.연구 /기타 | | | | |

| | 억제 (규제, 조세) | 처리 (시설증축, 관리) | 예방 (예보, 환경보전..) | 환경조성.. (교육, 기술개발, 산업 육성..) |
|------|---|---------------------------------|--------------------|----------------------------------|
| 상하수도 | 1. 상수원보호구역제도 2. 저수조 관리 3. 위생안전기준 인증 시행 4. 물수요관리 5. 공공하수처리시설통합운영관리 | 1. 상수도시설 확충 2. 하수도시설 확충 및 정비 | | 1. 물산업육성 2. 정수시설 운영관리 사제도 |

| | 억제 (규제, 조세) | 처리 (시설증축, 관리) | 예방 (예보, 환경보전..) | 환경조성.. (교육, 기술개발, 산업육성..) |
|----|--|--|--|---|
| 수질 | <ol style="list-style-type: none"> 1. 먹는물 수질기준 2. 먹는 물 공동시설관리 3. 정수기 관리 4. 지하수수질관리 5. 가축분뇨관리 6. 수질오염물질 지정 및 배출기준 설정 7. 수질오염총량제 8. 유역통합관리 9. 수생태계 건강성 제고 10. 물이용부담금 11. 수변구역제도 12. 토지매수제도 13. 상수원관리지역 지원제도 14. 폐수배출시설 지정관리 15. 중점관리저수지 지정관리 16.배출부과금제도 17. 특정토양오염관리대상시설 지정 관리 18. 토양오염검사 19. 토양환경보전 자발적협약 20. 클린주유소 지정 21. 골프장 농약사용 제한 | <ol style="list-style-type: none"> 1.지하수오염방지시설 2.가축분뇨 공공처리시설 확충 3. 수질측정망 설치 운영 4. 환경기초시설, 자체처리시설 5. 폐수종말처리시설 6. 수질오염사고대응 7. 비점오염저감시설 8. 생태하천복원사업 | <ol style="list-style-type: none"> 1. 조류경보제도 2. 수질예보제도 | <ol style="list-style-type: none"> 1.수질개선부담금 2. 물환경관리기본계획 3. 수계관리기금 4. 수계관리위원회 5. 호소환경조사 6.녹조대응체계구축 7. 비점오염원관리종합대책 8. 새만금유역 수질개선종합대책 9. 토양측정망 운영 10. 토양오염실태조사 |

| | 억제 (규제, 조세) | 처리 (시설증축, 관리) | 예방 (예보, 환경 보건..) | 환경조성.. (교육, 기술개발, 산업육성..) |
|-----|--|--|------------------------|--|
| 폐기물 | <ol style="list-style-type: none"> 1. 수은배출관리 2. 유독물질지정관리 3. 화학물질 배출량 조사제도 4. 화학물질 배출 저감 자발적 협약 5. 하천하구쓰레기관리협약 6. 생산자책임재활용제도 7. 비용기보증금제도 8. 쓰레기종량제 9. 폐기물부담금 10. 1회용품, 제품포장 규제/자발적협약 11. 사업장폐기물감량화제도(모니터링) | <ol style="list-style-type: none"> 1.잔류성유기오염물질관리 2.재활용품 분리배출, 수거체계 효율화 3.음식물폐기물자원화 4.건설폐기물재활용대책(순환골재사용의무화) | | <ol style="list-style-type: none"> 1. 자원순환기본법 2. 나노물질 안전관리 3. 해양쓰레기대응센터 4. 해양쓰레기통합정보시스템 5. 재활용산업육성 6. 자원순환종합단지조성 7. 순환자원거래소 8. 폐자원에너지화시설확충 9.친환경에너지타운조성 10. 폐자원에너지화인력양성 11. 폐기물관리시스템 구축운영 |

| | 억제 (규제, 조세, 보조금) | 처리 (시설증축, 관리, 피해자 구제) | 예방 (예보, 환 경보건..) | 환경조성.. (교육, 기술개발, 산업육성..) |
|----|---|---|---------------------|---|
| 대기 | <ol style="list-style-type: none"> 1. 온실가스목표관리제 2. 온실가스배출권거래제 3. 자동차 온실가스/연비기준 4. 탄소포인트제도 5. 탄소성적표시제도 6. 그린카드제도 7. 석면해체사업장 감리인 지정제도 8. 건축물석면관리 9. 농어촌지역 슬레이트 관리 10. 대기오염물질, 배출허용기준 지정 11. 저황유 의무사용, 고체연료 사용금지, 청정연료 사용의무제도 12. 대기환경규제지역 지정관리 13. 대기오염물질 배출시설 설치 및 변경 허가-신고 제도 14. 대형배출사업장관리(굴뚝원격감시제도) 15. 비산먼지발생작업장관리 16. 휘발성유기화합물관리(특별대책지역, 대기환경 규제지역) 17. 악취배출원관리 18. 비산배출저감제도 (굴뚝이 아닌 배출) 19. 실내공기질관리(다중이용시설, 신축공동주택, 대중교통차량) | <ol style="list-style-type: none"> 1. 석면피해자구 제제도 | | <ol style="list-style-type: none"> 1. 온실가스관리체계구축 2. 무공해자동차보급/친환경자동차보 급 3. 기후변화특성화대학원지정운영 4. 지자체온실가스감축이행 및 평가 시스템 개발 보급 5. 기후변화 대응 종합계획 6. 수도권 대기환경관리 기본계획 |

| | 억제 (규제, 조세) | 처리 (시설증축, 관리) | 예방 (예보, 환경보전..) | 환경조성.. (교육, 기술개발, 산업육성..) |
|------|---|---|---|---|
| 자연환경 | <ol style="list-style-type: none"> 1. 4대 핵심생태축 보전 2. 생태계보전협력금 3. 생태계보호지역 지정 및 관리 4. 자연공원 지정 및 관리 5. 야생생물 보호. 관리 6. 해역이용협의 7. 환경관리해역 지정 및 관리 8. 해양환경기준 개선 및 설정 9. 연안오염총량관리제도 10. 연안. 해양보호구역 지정 | <ol style="list-style-type: none"> 1. 자연환경복원대책 2. 국립생물자연관 3. 국립생태원 4. 해양오염사고 영향조사 및 피해 보상 | <ol style="list-style-type: none"> 1. 전략환경영향평가 2. 환경영향평가 3. 소규모환경영향평가 4. 해역이용영향평가 5. 해안침수예상도 제작 6. 해양환경 오염사고 대응 역량 제고 | <ol style="list-style-type: none"> 1. 국토환경성평가지도 2. 환경영향평가정보지원시스템 3. 생태자연도 제작 및 갱신 4. 생물자원보전종합대책 5. 야생동식물보호기본계획 6. 국가 생물자원 관리체계 구축 7. 생태관광활성화 8. 생태탐방로조성 9. 해양환경측정망 운영 10. 해양생태계 및 종 조사 11. 국립해양생물자원관 12. 해양생명공학기술지원 |

| | 억제 (규제 조세) | 처리 (시설증축, 관리, 피해자구제) | 예방 (예보, 환경보건..) | 환경조성.. (교육, 기술개발, 산업육성..) |
|--|--|---|---------------------------------|---|
| 환경 일반 /보건/ 국제 협력 환경 교육. 연구/ 기타 | 1. 통합환경관리제도 2. 화학물질관리법 3. 화학물질등록및평가등에관한법률(화평법) 4. 라돈관리 5. 소음진동관리 6. 빛공해관리 | 1. 화학사고대응 정보시스템 2. 가습기살균제 피해자 지원 3. 환경오염피해구제제도 4. 환경책임보험 | 1. 환경성질환 예방 및 관리 2. 건강영향평가제도 | 1. 지속가능발전위원회운영 2. 지속가능발전기본계획 수립 3. 국가지속가능성 보고서작성 4. 환경기술개발사업 5. 환경산업체지원 6. 환경일자리창출 7. 국제환경협약대응 8. 지역별 국가간 국제환경협력강화 9. 국제기구 협력 10. 국제 환경무역협상대응 11. 환경보건센터 구축 및 추진 12. 국민환경보건건강영향조사추진 13. 환경보건기술개발 14. 석면관리기본계획 15. 환경분쟁조정제도 16. 환경교육 17. 환경협의회및 자문기구 운영 18. 저탄소친환경생활실천운동 19. 녹색제품구매제도 20. 환경라벨링제도 21. 친환경경영확산 22. 녹색금융 23. 신기술인증. 녹색기술인증 24. 공공데이터개방 25. 환경통계작성및 환경통계정보서비스강화 |

2. 빅데이터 분석기법(Machine Learning) 소개

- Machine Learning 을 분류하고 유형별 장점을 소개
 - Machine Learning 개관 : Predictive analysis의 특징을 소개하고 분류를 시도
- Supervised Learning : 예측 기능
 - Regression
 - Logistic Regression/Naïve Bayes/Decision Tree
- Unsupervised Learning : 패턴 파악 기능
 - k-mean cluster/Supporting Vector Mechanism
- Deep Learning : DNN, CNN,RNN/최신조류(증강학습, 진화적 방법론)
- 인과분석: 변수소거법/Counter Factual experiment

3. 빅데이터 분석기법 환경연구 적용가능성

- 정책 유형별, 단계별 적용 방식 도출

- 정책 수립 단계 : 예측 기능 활용

- 예 1) 환경위험 예측을 통한 정책수요 파악 및 예산 배정

- 예 2) 환경관련 문헌 분석을 통한 민간 환경정책 수요 파악

- 정책 시행 단계 : 패턴 분석 기능 활용

- 예 1) [억제]오염원 주변환경 패턴 분석을 활용한 환경규제 모니터링

- 예 2) [환경조성]보조금 지급과 기업성과 패턴 분석을 활용한 환경산업지원 성과 모니터링

- 정책 평가 단계: 패턴 분석 기능 활용

- 예) [처리,시설]정책 성과 지표와 환경 지표간 패턴 파악 및 인과분석

4. 방법론: 문헌연구 + 실험적 실증연구(?)

- 환경연구특성 분석 + Machine Learning 특성 분석: 문헌연구
 - 환경연구 특성 분석 : 2016년 KEI 연구보고서, 2017년 환경부 예산서, 2016년 환경부 환경백서
 - Machine Learning 특성 분석: 교과서 + 최신연구동향
 - 교과서
 - [\[ISLR\]Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani\(2009\) Introduction to Statistical Learning. Springer](#)
 - [\[APM\] Kuhn, Max, and Kjell Johnson. Applied predictive modeling. Vol. 26. New York: Springer, 2013.](#)
 - [\[ESL\] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. Springer, Berlin: Springer series in statistics, 2001.](#)
 - [Iang Goodfellow, Yoshuan Bengio and Aaron Courville \(2017\) Deep Learning . MIT Press](#)
 - 최신연구동향 : [ARVIX.org](#)
- 실험적 실증연구(?) : BOD 추정 및 수질오염국고보조사업 지출과 패턴 분석
 - 본문에 소개한 기법을 사용하는 실제 분석 예를 삽입(???)
 - Wang, Heyi, and Yi Gao. "Elman's Recurrent neural network Applied to Forecasting the quality of water Diversion in the Water Source Of Lake Taihu." *Energy Procedia* 11 (2011): 2139-2147.

3. 연구관리

| 절 | 절 | 5월 | 6월 | 7월 | 8월 | 9월 | 10월 | 11월 | 12월 |
|-----------------------------|--------------------------|----|----|----|----|----|-----|-------|-----|
| 1. 서론 | 1) 필요성 및 연구 목적 | | | | | | | | |
| | 2) 선행연구 | | | | | | | | |
| | 3) 연구내용 및 방법론 | | | | | | | | |
| | 4) 본문 내용 | | | | | | | | |
| 2. 환경정책연구의 특성 | 1) 환경정책 개관 | | | | | | | | |
| | 2) 환경정책연구 개관 | | | | | | | | |
| 3. 빅데이터 분석기법의 특징 | 1) Machine Learning 개관 | | | | | | | 후속 조치 | |
| | 2) Supervised Learning | | | | | | | | |
| | 3) Unsupervised Learning | | | | | | | | |
| | 4) Deep Learning | | | | | | | | |
| | 5) 인과분석 | | | | | | | | |
| 4. 빅데이터 분석기법의 환경정책연구 적용 가능성 | 1) 정책 수립단계 | | | | | | | | |
| | 2) 정책 시행단계 | | | | | | | | |
| | 3) 정책 평가단계 | | | | | | | | |
| | 4) 실험적 분석 예? | | | | | | | | |
| 5. 결론 및 시사점 | | | | | | | | | |

- Work in Progress report : [Big team 출판블로그](https://keibigdata.github.io/)
(<https://keibigdata.github.io/>)
 - [초록 및 목차 게시 예정](#)
- Comments and Suggestions

