Bigdata 연구방법론 활용방안

중간점검

2017. 5. 24

KEI

강성원

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

초록

본 연구는 빅데이터 분석 기법의 환경정책연구 적용 가능성을 점검한다. 빅데이터 분석에 대한 선풍적인 관심에도 불구하고 실제로 빅데이터 분석을 환경정책연구에 적용하는 사례는 충분하지 않으며, 기존 연구는 자연언어 분석을 언론매체에 적용하여 설문조사의 기능을 대체하는 연구에 집중되어 있는 상황이다. 그러나 빅데이터 분석기법은 대용량의 자료로부터 정교한 패턴 파악이 가능하다는 장점을 가지고 있다. 따라서 사실관계 파악 및 예측이 필요한 환경정책연구 분야에서는 빅데이터 분석기법을 사용하여 기존의 연구 성과를 보완 및 개선할 수 있는 가능성이 있다. 본 연구는 이러한 가능성을 개괄적으로 점검하는 연구이다. 구체적으로 본 연구에서는 2017년 환경부 예산서, 2016 환경백서 및 2016년 환경정책평가연구원 보고서를 이용하여 기존 환경정책연구의 특성을 정리하고, 빅데이터 연구기법을 활용한 연구가 환경정책의 수립 -집행-평가에 도움이 될 수 있는 영역을 파악하고자 한다.

목차

1. 서론
    1. 연구 목적 : 빅데이터 분석기법의 환경정책연구 적용 가능성 점검
    2. 기존 연구 : Hal Varian. 빅데이터 분석기법의 경제학 연구 적용 방법 essay
    3. 연구 내용 및 방법론: 문헌연구, 실험적 실증연구(?)
    4. 본문의 구성

2. 환경정책연구의 특성
    1. 환경정책 개관: 예산서/환경백서(May)
        i. 정책은 어떻게 개발되는가?
            1) 이슈 예측
            2) 진단
            3) 개선방안 => 정책
        ii. 영역(날줄) : 예산서 (배정액 순서대록..)
            1) 상하수도
            2) 수질
            3) 폐기물
            4) 대기
            5) 자연환경
            6) 환경일반: 보건/국제협력/환경연구. 교육/환경관
        iii. 정책의 특성(씨줄) ? 정책분류?
            1) 억제: 규제, 조세, (보조금)
            2) 처리: 폐기장, 하수처리장
            3) 예방: 예보, 환경보건
            4) 환경조성: 교육, 기술개발, 산업육성
    2. 환경정책연구 개관: 2016년 KEI 연구보고서
        i. 영역별 분포
        ii. 연구의 특성 : 자료의 특성

3. 빅데이터 분석 기법의 특성 (May)
    1. Machine Learning 개관
        i. General approach: predictive/test-train...
        ii. Category: Supervised/Unsupervised
        iii. Deep Learning: Why it is that important?
    2. Supervised Learning : 예측
        i. Regression
        ii. Logistic regression
        iii. Naive Bayes
        iv. Decision Tree
    3. Unsupervised learning : Pattern 파악
        i. K-nearest neighbor
        ii. Supporting Vector Mechanism
    4. 인과분석 적용방식: 변수소거법, Counter factual experiment
    5. Deep Learning: DNN, RNN, CNN
4. 빅데이터 분석기법의 환경정책연구 적용 가능성(June)
    1. 정책 수립단계 : 수요 파악, 예산 배정
    2. 정책 시행단계: MRV
    3. 정책 평가단계: 성과지표와 환경지표
    4. Experimental Study? (BOD 예측??)(July, August, September)

5. 결론 및 시사점

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/연구 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				

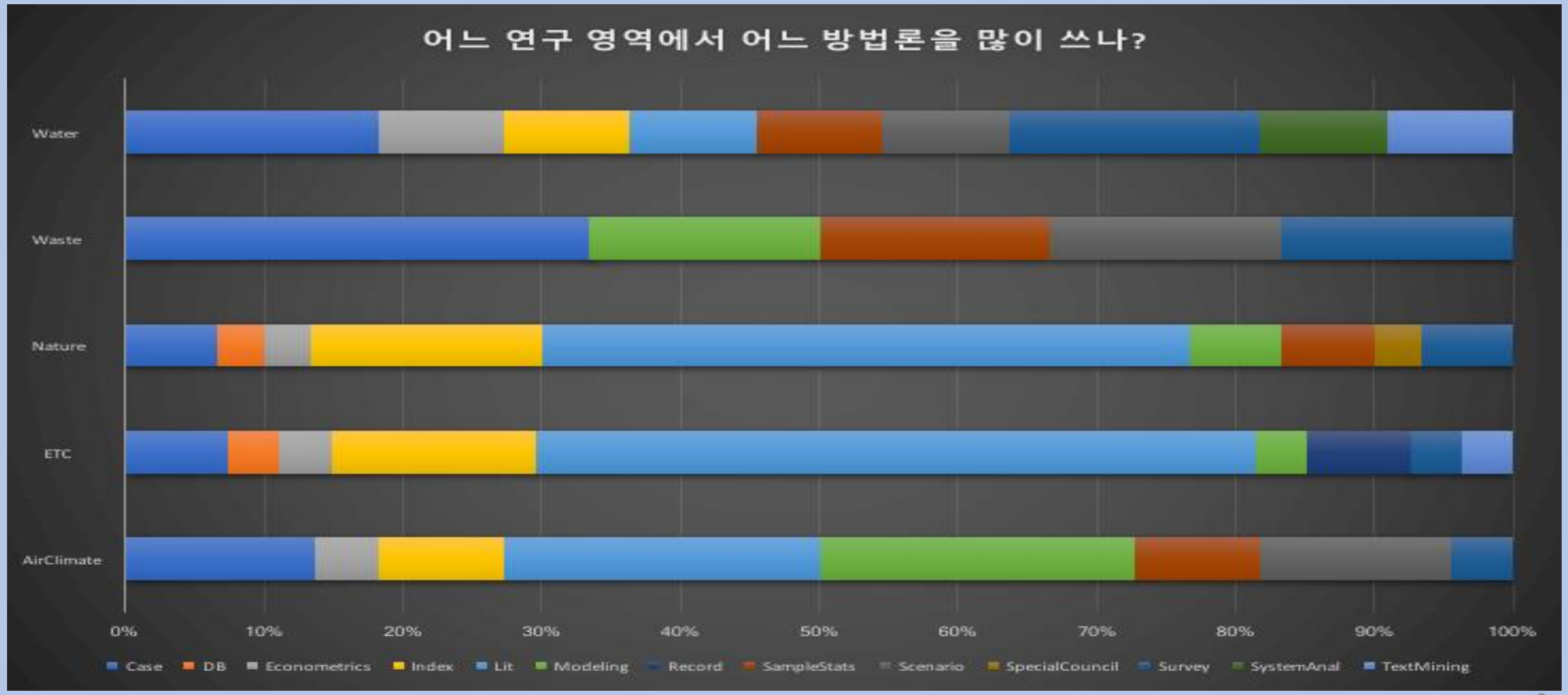
환경정책연구 방법론 분류기준

명칭	내용
Scenario	파라미터 값이 상이한 시나리오를 구축하고 관심 변수의 값을 시나리오에 따라 구하는 방법 (경제성., 수익성..)
Modeling	연역적 추론에 기반한 모형을 구축하고 이를 이용하여 관심 변수의 값을 구하는 방법 (일반균형, 산업연관분석)
Econometrics	계량경제학 모형을 사용하여 관심 변수의 값을 추정하는 방법
Lit	주제와 관련된 선행연구 및 사례에 관련된 문헌을 종합하여 정리하고 시사점을 도출하는 방법. 메타연구를 포함
SampleStats	특정한 방법론 없이 기초자료로부터 표본통계량을 조합하여 논거를 찾아내는 방식의 연구를 의미
Case	문헌조사 이외의 방법을 사용하여 사례를 조사하는 방법(인터뷰. 실측 등을 포함)
SystemAnal	사전적 모형 없이 직관적인 인과관계 네트워크 시스템 모형을 구축하여 분석하는 방법
Survey	설문조사
SpecialCouncil	전문가들로 구성된 패널의 의견을 종합하는 방법
Record	행사기록
DB	DB구축

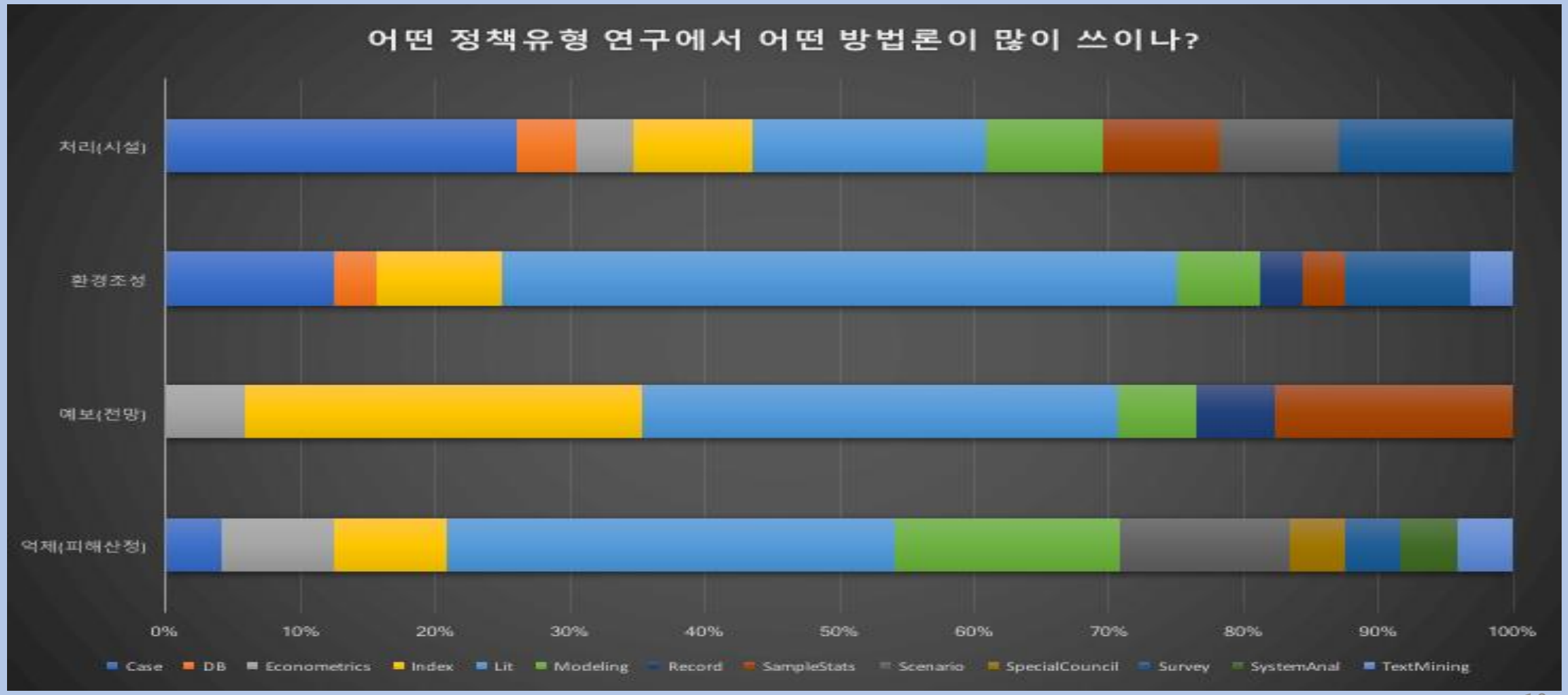
2016년 KEI 보고서 연구방법론 정리

	A	B	C	D
1	Section	Type	Name	Method
2	Water	Control	대도시지역의극한홍수로인한복합영향메커니즘분석(채여라)	TextMining
3	Water	Control	대도시지역의극한홍수로인한복합영향메커니즘분석(채여라)	SystemAnal
4	Water	Treatment	샛강관리및이용활성화방안연구강형식	Survey
5	Water	Treatment	물환경서비스와물인프라의지속가능성평가강형식	Case
6	Water	Treatment	물환경서비스와물인프라의지속가능성평가강형식	Survey
7	Water	Treatment	물환경서비스와물인프라의지속가능성평가강형식	Econometrics
8	Water	Treatment	물환경서비스와물인프라의지속가능성평가강형식	SampleStats
9	Water	Treatment	사회적투자수익률을고려한물환경인프라시설투자방향연구(류재나.강형식)	Scenario
10	Water	Treatment	가뭄단계에따른적응형가뭄관리정책연구:지역차원의비구조적가뭄대책을중심으로(김호정)	Case
11	Water	Treatment	사물인터넷을활용한스마트물환경관리방안및정책기반마련연구(한혜진)	Lit
12	Water	Preparation	물이용지속성의평가와미래전망(김익제)	Index
13	Waste	Treatment	공간정보를활용한재해폐기물성상별최적관리방안(조지혜)	Case
14	Waste	Treatment	공간정보를활용한재해폐기물성상별최적관리방안(조지혜)	SampleStats
15	Waste	Treatment	폐자원흐름분석을통한전기전자제품의upcycling활성화방안(이희선)	Case
16	Waste	Treatment	나노폐기물의안전처리를위한관리전략수립연구(조지혜)	Scenario
17	Waste	Support	자원순환사회전환촉진을위한재활용산업활성화방안-재활용관리제도전환에따른영향분석(이소라)	Survey
18	Waste	Support	자원순환사회전환촉진을위한재활용산업활성화방안-재활용관리제도전환에따른영향분석(이소라)	Modeling
19	AirClimate	Control	빅데이터를이용한대기오염의건강영향평가및피해비용추정2(안소은)	Econometrics
20	AirClimate	Control	빅데이터를이용한대기오염의건강영향평가및피해비용추정2(안소은)	Survey
21	AirClimate	Control	온실가스감축정책평가를위한환경경제모형개발운용3(강성원박창석)	Modeling
22	AirClimate	Control	교통환경정책수립을위한대기환경개선효과추정방안연구:도로이동오염원을중심으로(한진석)	Scenario
23	AirClimate	Control	대기환경비용을고려한친환경차구매보조금실효성제고연구:차종별적정보조금수준분석을중심으로(한진석)	Modeling
24	AirClimate	Control	대기환경비용을고려한친환경차구매보조금실효성제고연구:차종별적정보조금수준분석을중심으로(한진석)	Scenario
25	AirClimate	Control	실도로에서경유차의대기오염물질초과배출에따른사회적비용연구(강광규)	Scenario

2016 KEI 보고서 연구 방법론 분포(영역)



2016년 KEI 보고서 연구방법론 분포(정책 유형)



영역/유형별로 많이 쓰는 방법론

영역

Section	maxm
기후대기	Modeling
환경일반	Lit
자연환경	Lit
폐기물	Case
상하수도수질	Case

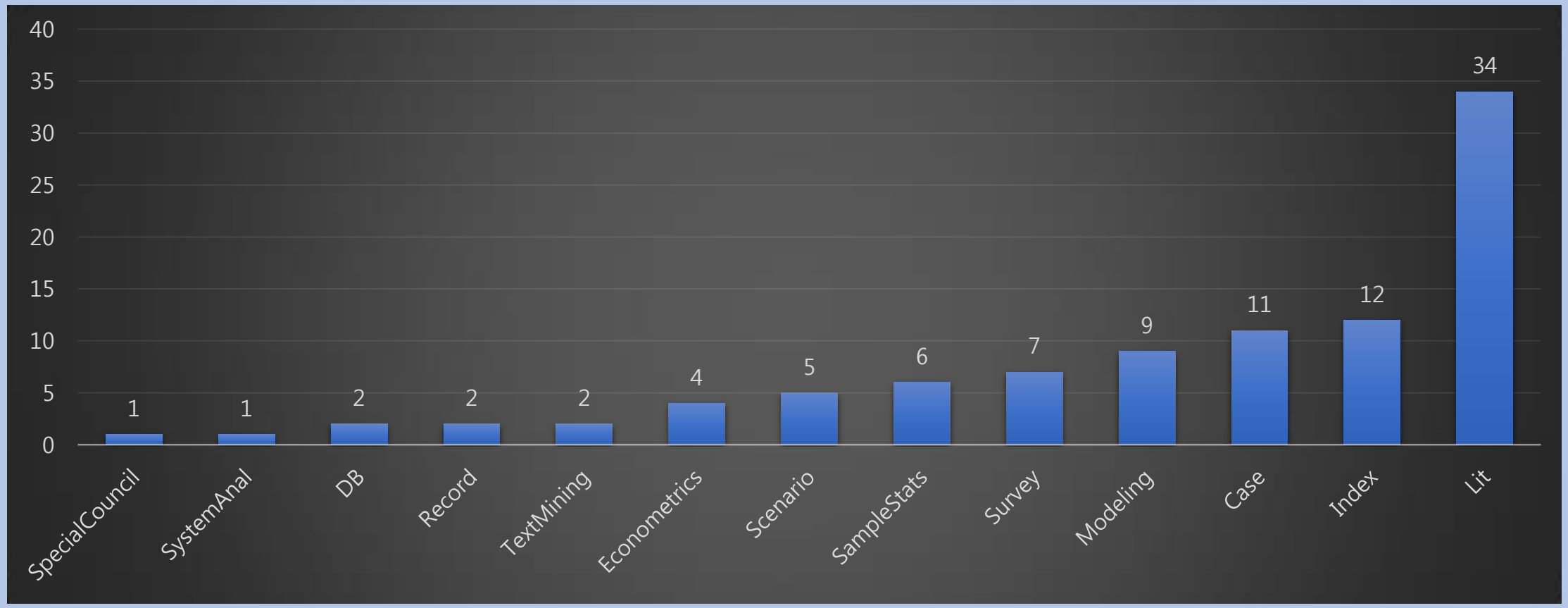
유형

Type	maxm
억제(피해산정)	Lit
예보(전망)	Lit
환경조성	Lit
처리(시설)	Index

문제점

- 환경 연구 분야 세분 필요: 자연환경/환경일반 너무 많은 내용
  - 자연환경: 생태계/토양-지하수/화학물질/환경영향평가
  - 환경일반: 산업/경제/외교/'전략'/행정
- 연구의 대표성 문제: 문헌연구 >> 정량적 연구
  - 기초연구의 대부분이 문헌연구: 기초연구를 제외하고 접근?
  - 정량적 연구 방법론 적용이 가능한 연구 vs. 현재 정량적 연구 방법론을 적용하고 있는 연구를 구분해서 접근할 필요
  - 학술논문? (가장 인용이 많이 되는 100개?)

방법론 분포


40
35
30
25
20
15
10
5
0
1
1
2
2
2
4
5
6
7
9
11
12
34
SpecialCouncil
SystemAnal
DB
Record
TextMining
Econometrics
Scenario
SampleStats
Survey
Modeling
Case
Index
Lit

절	절	5월	6월	7월	8월	9월	10월	11월	12월	현황(5/24)
1. 서론	1) 필요성 및 연구 목적									집필 필요
	2) 선행연구									
	3) 연구내용 및 방법론									집필 필요
	4) 본문 내용									집필 필요
2. 환경정책연구의 특성	1) 환경정책 개관									집필 필요
	2) 환경정책연구 개관									집필 필요
3. 빅데이터 분석기법의 특징	1) Machine Learning 개관							후속 조치		미진행
	2) Supervised Learning									미진행
	3) Unsupervised Learning									미진행
	4) Deep Learning									미진행
	5) 인과분석									미진행
4. 빅데이터 분석기법의 환경정책연구 적용 가능성	1) 정책 수립단계									미진행
	2) 정책 시행단계									미진행
	3) 정책 평가단계									미진행
	4) 실험적 분석 예?									미진행
5. 결론 및 시사점										미진행

Questions and Comments