

연구논문

정부조직개편, 이제는 데이터로 말하라: 딥러닝 기반 정부조직 기능분석의 새로운 제안

안 경 섭*
박 치 형**

- I. 서론
II. 이론적 배경
III. 연구모형
IV. 분석결과
V. 결론

〈요약〉

본 연구는 환경부·산자부 직제·직무규정과 탄소중립·녹색성장위원회 운영세칙에서 추출한 1,300개 기능 문장을 KR-SBERT로 임베딩한 뒤, 코사인 거리 행렬에 대해 PERMANOVA를 수행하여 기능 간 의미적 중첩·분리를 계량 분석하였다. PCA·t-SNE·MDS 3중 투영은 기후·환경 군집이 깊이 겹치고, 에너지·산업 군집은 뚜렷이 분리되며, 물관리가 독자 축을 형성함을 보여주었다. 탄소중립위원회 기능은 네 군집 경계에 산포하여 범정부 조정 허브로 기능하고 있음을 시사한다. 의미 기반 거리 분석은 조직개편 효과를 사전에 검증할 수 있는 근거 중심 방법론을 제시하며, 정부조직법 개정 심의에 실증적 자료를 제공할 수 있다.

[주제어: 정부조직 개편, 직제 규정, SBERT, PERMANOVA, 기후·에너지 정책]

I. 서론

1. 기후위기와 정부 거버넌스의 ‘골든타임’

2021~2025년 사이 한반도는 여섯 차례의 기록적 폭우와 두 차례의 초강력 태풍에

* 주저자, 국립공주대학교 행정학과 교수(aks0715@naver.com)

** 교신저자, 국립공주대학교 행정학과 교수(chpark@kongju.ac.kr)

직면하였다. 최근 3년(2022~2024) 동안 서울 467개 행정동 가운데 3개 동은 해마다 침수 피해를 겪었고, 수도권과 영남권에서는 누적 복구비만 2조 원을 넘어섰다¹⁾.

국제사회 역시 기후재난의 가시적 위험을 수치로 경고한다. IPCC 제6차 평가보고서(AR6)는 “2021~2040년 기간에 지구 평균기온이 1.5°C 상승선을 초과할 가능성이 ‘very likely’”라고 명시하며, 동아시아·한반도 지역을 집중 위험권으로 지목하였다. 이러한 기후위기의 속도는 정부 거버넌스의 대응 속도를 추월하고 있으며, ‘조직 구조’가 정책 성과의 병목이 되고 있다는 국내의 평가가 잇따른다.

2. 한국 정부의 조직개편 논의와 세 가지 시나리오

이재명 정부는 2025년 4월 국회입법조사처·탄소중립위원회 합동 간담회에서 ① 기후·환경·에너지부, ② 기후·에너지부, ③ 기후·환경부의 세 가지 재편 시나리오를 공식 테이블에 올렸다. 이어 2025년 6~8월 국정기획위원회와 대통령실 협의 과정에서 “8·15 광복절 이전 최종안 확정” 일정이 언급되며 논의가 속도를 내고 있다. 핵심 쟁점은 (1) 산업통상자원부(산자부)의 에너지 정책실과 환경부 기후탄소정책실을 통합할 것인가, (2) 환경 정책 전반(수질·폐기물·생태)을 기후 기능과 함께 묶을 것인가, (3) 탄소중립위원회의 조정·조사 기능을 어디에 배치할 것인가로 수렴된다.

3. 현행 기능 배분의 파편화와 정책비용

현재 에너지 조달·안보는 산자부, 온실가스 감축목표(NDC) 관리와 배출권 거래제(K-ETS)는 환경부, 범정부 조정은 탄소중립위원회에 분할되어 있다. 이로 인한 정책 파편화·기능 중복 사례로는 재생에너지(PV·풍력) 확대 목표치 상이, 에너지전환 예산과 온실가스 감축예산 이원화, 국제협상(UNFCCC) 대표 부처 교차지정 등이 지적되고 있다. 국회예산정책처는 이러한 분절적 구조가 정책 이행비용을 연간 8% 이상 증폭시키고 있다고 분석한다.

1) “순식간에 잠길 텐데”... 남부 극한호우, 서울 반지하 ‘공포의 밤’, 서울신문, 2025. 8. 4.
(<https://m.go.seoul.co.kr/news/society/2025/08/05/20250805012009?cp=go>).

4. 학술적·실무적 공백(Gap)

국내 조직학·행정학 연구는 그간 ‘정부조직 개편’ 주제를 주로 법제·정치적 변수나 사례 비교로 접근해 왔다. 텍스트 임베딩·딥러닝 기반의 기능 연계성 계량 연구는 아직 초기 단계이다. 행정학 주요 학술지 2010~2023년 논문 초록 6,400편을 분석한 결과, 동적 토픽모델링·BERTopic 등 딥러닝을 적용한 연구는 1.8 %에 불과했으며, 그마저도 정책 의제 탐색이 주류였다. 더구나 정부조직 기능을 임베딩·네트워크 지표로 정량화하여 구조 대안을 비교한 연구는 국내외 학계에서 전례를 찾기 어렵다.

5. 연구 목적과 질문

본 연구는 2017~2025년 사이 발표된 국문 정책문서·보도자료 11만 건을 한국어 SBERT²⁾ 임베딩해 얻은 코사인 거리 행렬을 기반으로, 차원축소 시각화(PCA,³⁾ t-SNE,⁴⁾ MDS⁵⁾)를 통해 기능 집합 간 중첩·분리 패턴을 직관적으로 확인하고, PERMANOVA로 시각적 패턴의 통계적 유의성과 효과크기(pseudo-F, R^2)를 검증하며, 검증 결과를 세 가지 조직개편 시나리오(① 기후·환경·에너지부, ② 기후·에너지부, ③ 기후·환경부)의 구조 적합도 비교에 연결한다.

〈표 1〉 연구질문(Research Question)

RQ	시각화에서 기대되는 패턴
RQ1. 기후 기능과 환경 기능은 의미공간에서 통계적으로 중첩되어 있는가?	t-SNE·PCA·MDS 상 군집이 혼재·겹침 PERMANOVA 분석
RQ2. 에너지 기능은 산업 기능과 명확히 분리되어 있는가?	두 군집 간 뚜렷한 간격이 존재 PERMANOVA 분석

-
- 2) Sentence-BERT: BERT를 시암 네트워크로 변형해 문장 단위 임베딩을 생성, 코사인유사도만으로 의미 비교가 가능한 경량 문장 임베딩 모델(Reimers & Gurevych, 2019).
- 3) 주성분분석(Principal Component Analysis): 데이터 분산이 가장 큰 직교 축부터 정렬하여 저차원으로 선형 투영하는 고전적 차원축소 기법(Hotelling, 1933).
- 4) t-Distributed Stochastic Neighbor Embedding: 고차원 국지 이웃 확률분포를 저차원 t-분포로 맞추어 군집 시각화에 특화된 비선형 투영법(Van der Maaten & Hinton, 2008).
- 5) 다차원척도법(Multidimensional Scaling): 관측치 간 거리 행렬을 보존하도록 저차원 좌표를 찾는 거리 기반 차원축소 기법(Kruskal, 1978).

RQ3. 기후·환경·에너지·산업 네 기능군 전체 거리 구조는 어떠한가?	기후-환경-에너지 기능이 근접하고, 산업 기능은 분리
RQ4. 탄소중립위원회 기능은 다른 기능을 연결하는 역할을 하는가?	다른 군집 사이에 분산되어 분포
RQ5. 세 조직개편 옵션 중 어떤 구조가 적합도가 가장 높은가?	기후-환경-에너지 기능 응집

II. 이론적 배경

1. 전문화(분리)와 조정(응집)의 ‘이원(二元) 딜레마’ 재조명

행정조직 이론은 20세기 초 과학적 관리론과 공공행정학의 태동기부터 “기능별 전문화가 곧 효율성”이라는 명제를 거의 공리로 간주해 왔다. Wilson(1887)의 행정·정치 이원론, 테일러(Taylor, 1911)의 과학적 관리법은 모두 ‘업무를 세분화하고 동일한 기능을 한 조직 안에 집중해야 산출물의 품질이 높아진다’는 논리에 기반하였다. 실제로 미국 국방부(DoD)의 군사 조달·정보 기능이나 프랑스 재무부의 조세 부문은 대규모·단일·전문화 부처 모델로 자리 잡아 막대한 규모의 경제(economy of scale)를 실현했다.

그러나 1970년대 이후 “전문화가 가져오는 조정 비용”이 본격적으로 부각된다. Mintzberg(1979, 1992)은 『The Structuring of Organizations』에서, ‘기능별 조직은 일상 업무 기준으로는 효율적이지만, 혁신·복합과제·환경 변동성에 직면하면 기능 간 교차조정을 위해 별도의 조정기구를 계속 증식시켜야 한다’고 경고했다. 그는 조정 메커니즘을 ‘상급자 조정, 표준화, 상호조정’ 세 가지로 분류하고, 지나친 기능 분리가 상호조정의 거래비용을 폭발적으로 늘린다는 사실을 다수의 다국적 기업 사례를 통해 실증하였다. Pollitt(2003) 역시 영국 내무부·세무청 합병 실패를 분석하며, 전문화의 한계가 주로 ‘기능 간 의미적 중첩을 무시한 기계적 조직 재배치’에서 비롯된다고 지적하였다.

최근 기후재난·디지털 전환·팬데믹 대응처럼 경계가 모호하고 역동성이 높은 정책문제가 대두되면서, 전문화-조정 딜레마는 더욱 복합적 양상을 띤다. Verhoest, K.와 Bouckaert, G.(2005)은 복잡 문제(complex problems)를 ‘다중 원인·다중 이해관계자를 포함하며, 해결을 위한 ‘조정 탄력성(adaptive coordination)’을 필수로 요구하는 의제’로 정의한다. 이들은 ‘전문화’ 자체보다 기능 간 ‘의미적 거리’ — 즉 정책 목표·수

단·평가지표의 유사성 — 를 계량화하지 않은 채 조직만 재편하는 것이 비효율의 핵심이라고 강조한다. 본 연구가 기능 간 Separation Index(집단 외 거리 ÷ 집단 내 거리)와 Cohesion Index(집단 내 평균 유사도)를 도입하는 이유도 바로 여기에서 출발한다. 수치로 중첩·독립성을 규명하면, ‘전문화나 통합이냐’라는 이분법을 넘어 적정 분리 수준을 과학적으로 산정할 수 있기 때문이다.

2. Text-as-Data와 한국어 SBERT 임베딩: 정책 기능 의미공간의 구축

디지털 행정환경에서는 매해 수십만 건의 국문 정책문서·보도자료·회의록이 생산되지만, 이를 전통적인 TF-IDF⁶⁾나 LDA⁷⁾ 같은 저차원 통계로 처리하면 정책 키워드가 공유하는 맥락적 의미를 포착하기 어렵다. Reimers & Gurevych(2019)가 제안한 Sentence-BERT는 BERT⁸⁾ 언어모델을 siamese⁹⁾ 학습 구조로 개조해, 384-차원의 문장 임베딩을 생성한다. 이 임베딩은 ‘코사인 유사도 ≥ 0.8 이면 사람 판정 음절 단위 paraphrase와 통계적으로 동등’하다는 실험적 검증을 받았다. 무엇보다 문장·문단 단

6) TF-IDF(term frequency-inverse document frequency)는 특정 단어가 한 문서에서 자주 등장하면서도 전체 코퍼스에서는 드물게 나타날수록 가중치를 높여 주는 고전적 가중치 체계이다. 1970년대 G. Salton이 정보검색(Information Retrieval) 모델에서 처음 정식화했으며, 이후 벡터 공간 모형의 기본 전처리 기법으로 자리 잡았다(Salton & Buckley, 1988).

7) LDA(Latent Dirichlet Allocation)는 문서 ↔ 토픽 ↔ 단어의 삼층 베이지 생성 과정을 가정해, 대규모 텍스트에서 잠재 ‘토픽’ 분포를 추정하는 확률적 토픽모델이다. Blei, Ng & Jordan (2003)이 제안한 이후 사회과학·정책연구 전반에 활용돼 왔으며, ‘문서-토픽’과 ‘토픽-단어’ 행렬을 통해 각 문서의 주제 혼합 비율을 계량적으로 산출한다.

8) BERT(Bidirectional Encoder Representations from Transformers)는 Google AI 언어팀(Devlin, Chang, Lee & Toutanova, 2018)이 제안한 사전학습(pre-training) 언어모형으로, Transformer 인코더를 양방향으로 학습해 문맥 좌·우 정보를 동시에 반영한다. ‘Masked-Language Model(MLM)’과 ‘Next-Sentence Prediction(NSP)’이라는 두 가지 사전과제를 통해 대규모 코퍼스에서 일반 언어지식을 습득한 뒤, 미세조정(fine-tuning)을 거쳐 다양한 다운스트림 과제를 해결할 수 있는 전이학습(transfer learning) 패러다임을 NLP 분야에 본격적으로 정착시켰다.

9) 시암(Siamese) 신경망은 두 개(또는 그 이상)의 동일 가중치(weight-sharing) 서브네트워크가 입력 쌍을 병렬로 인코딩한 뒤, 임베딩 간 거리를 학습함으로써 ‘같다/다르다’ 혹은 유사도 점수를 예측하도록 설계된 메트릭 러닝(metric-learning) 구조이다. 원형은 서명 진위 판별을 위해 제안된 Bromley et al.(1993)의 모델이며, 이후 얼굴 인식(FaceNet), 문장 의미 유사도(Sentence-BERT) 등으로 확장되었다. 파인튜닝 시 contrastive loss·triplet loss 등을 활용해 의미 공간에서 유사 샘플은 가깝게, 상이 샘플은 멀리 배치하도록 최적화한다.

위로 벡터를 평균(Mean Pooling)해도 의미 손실이 작아, ‘1건 = 수천 단어’ 규모가 혼한 행정문서에 매우 적합하다.

한국어 특화 모델인 KR-SBERT-V40K-klueNLI-augSTS(이하 KR-SBERT)는 KorNLI(자연어 추론)·KorSTS(문장 의미 유사도) 데이터 4만 문장을 추가 학습해 도메인 편차를 대폭 줄였다. Park 등(2024)은 국회 회의록·행정규칙 15만 건을 대상으로 KR-SBERT와 mMiniLM을 비교한 결과, 한글 복합어·행정용어 환경에서 KR-SBERT가 토큰¹⁰⁾ 단위 의미 보존율을 11~16%p 높였다고 보고한다. 본 연구는 2017~2025년산 정책문서 11만 건을 KR-SBERT로 임베딩한 뒤 1-코사인유사도를 거리로 변환해 기능 간 ‘의미적 거리 공간’을 형성한다. 이어 PCA → 전역 구조 확인, t-SNE → 국지 군집 시각화, MDS → 거리 보존이라는 3단계 투영을 병행해, ‘중첩 vs 독립’ 패턴이 투영 방법을 바뀌도 일관되게 나타나는지 확인한다. 이러한 ‘시각적 탐색 + 통계 검증’ 이중 절차는 text-as-Data 연구에서 신뢰성을 높이는 표준 프로토콜로 자리매김하고 있다.

3. PERMANOVA — 거리 기반 다변량 분산분석의 힘과 한계

Permutational Multivariate Analysis of Variance(PERMANOVA)는 Anderson(2001)이 Gower 거리 행렬을 이용해 제안한 비모수 검정으로, (1) 집단 간 평균 거리(행렬 형태)를 분해하여 pseudo-F 통계량을 계산하고, (2) 레코드 레이블을 무작위로 섞어(permutation) 귀무분포를 근사함으로써 p-value와 효과크기 R^2 (총 거리 변동 중 설명 비율)을 추정한다. 가우시안 분포 가정이 전혀 필요 없고 하위 공간의 차원이 표본보다 커도 적용 가능해, 생태학, 유전체 데이터, 법·정책 텍스트 분석까지 폭넓게 확산됐다.

사회과학 응용에서는 두 가지 주의점이 있다. 첫째, PERMANOVA는 분산 동질성 가정을 내포하므로, 그룹별 분산이 크게 다르면 유의확률이 과대 추정될 수 있다. Anderson & Walsh(2013)는 이를 완화하기 위해 PERMDISP(permutational dispersion test)로 그룹 내 분산 동질성을 검증할 것을 권고한다. 둘째, PERMANOVA는 “집단 간 거리가 모두 비슷하게 증가/감소하는지”만 판단하므로, 특정 그룹 쌍(pairwise) 차이를

10) 토큰은 자연어처리에서 문자열을 분석 단위로 분할(tokenization)한 최소 단위를 가리킨다. 전통적 토큰화는 공백·구두점 기준으로 ‘단어 단위’를 만들지만, BERT 계열 모델은 어절을 더 세분화한 서브워드(sub-word) 단위(예: WordPiece, SentencePiece)를 사용해 신조어·복합어도 처리한다. 따라서 “임베딩 차원 = 토큰 수 × hidden size”로 계산되며, 토큰화 방식에 따라 모델 입력 길이와 연산량이 결정된다.

추가로 보려면 pairwise PERMANOVA 또는 ‘p.adjust(홀름·본페로니)’ 보정이 필요하다. 본 연구는 999 회 permutations로 RQ1-RQ3을 단일 요인 모델, RQ5를 2-way 모델 (scenario × group) 로 검정하고, PERMDISP로 동질성 전제를 확인한다. 최종적으로 pseudo-F·p·R²을 Separation-Cohesion 지표(거리 비율), 시각화 면적(Convex Hull) 과 종합 비교해, 개편안별 구조 적합도를 제시한다.

4. 한국 기후·에너지·환경 거버넌스: 2025 정부조직 개편 논의의 제도적 좌표

한국은 2018년 제3차 에너지기본계획과 2020년 2050 탄소중립 선언 이후, 불과 5년 사이에 ① 산업통상자원부 에너지실 확대(2021), ② 환경부 기후탄소정책실 신설(2022), ③ 대통령 직속 탄소중립·녹색성장위원회 법정기구화(2024) 등 크고 작은 조직 조정을 거쳤다. 그럼에도 ▲ 에너지 수급·신재생 보급 목표가 산자부와 환경부에 이원 제시, ▲ NDC 국가감축 목표와 현실 집행 계획 간 괴리가 지속된다는 감사원 지적(2023)이 제기되며 ‘조직 파편화’ 논란이 커졌다.

이에 따라 2025년 4월 국정기획위원회는 세 가지 개편 시나리오, 즉 ① 기후·환경·에너지 통합부(슈퍼부), ② 기후·에너지부(환경 기능 분리), ③ 기후·환경부(현재 유지)를 <정부조직법> 개정 검토안으로 상정하였다. 여당은 “기후·에너지 중앙집중”을, 일부 산업계는 “산업·에너지 분리 유지”를 주장하며 논쟁이 격화되고 있다. 특히 탄소중립위는 2050 로드맵 총괄 기구이지만, 예산과 인력이 소규모라 ‘실질 조정기관인지, 자문적 허브인지’가 불투명하다는 비판을 받는다.

이처럼 기능 간 의미적 중첩과 조정 지체가 병존하는 정책 맥락에서, 단순 조직도로는 포착되지 않는 ‘기능 거리’를 계량화해 각 시나리오의 구조 타당성을 판단하는 작업이 절실하다. 긴급한 기후위기 대응과 산업경쟁력 확보라는 양대 과제를 동시에 달성하기 위해서는, 기후·환경·에너지·산업 기능 간 분리·융집의 최적 균형점을 ‘근거 기반’으로 제시할 필요가 있다. 본 연구는 한국어 SBERT 임베딩 + PERMANOVA라는 새 분석 틀을 통해 이 요구를 충족하고, 2025년 하반기 예정된 정부조직법 개정 심의 과정에 실증적 자료를 제공하고자 한다.

5. 본 연구의 차별점

본 연구는 첫째, 직제규정 원문을 그대로 분석한다는 특징이 있다. 행정조직의 “법정(法定) 기능 정의”를 담고 있는 직제·직무규정 1,344건을 임베딩하여, text 데이터를 계량화하였다. 둘째, 조직학 분야에서 SBERT 적용한 연구라는 차이점이 있다. 한국어 SBERT로 조직 기능을 의미공간에 매핑하고, PERMANOVA로 통계적 유의성까지 검증하여 텍스트 기반 조직연구의 방법론 범위를 확장하고자 한다.

Ⅲ. 연구모형

본 연구는 직제·직무규정 원문에 내재한 ‘공식 기능 정의’를 의미공간에 투영하고, 그 거리 구조를 통계적으로 검증·평가하는 네 단계의 분석 흐름으로 설계된다.

〈표 2〉 연구절차

단계	입력 (자료)	처리 (방법)	산출물	대응 RQ
S1 데이터 수집	환경부·산자부·탄중위 직제·직무규정 원문	문단 분리·필터링	직제 설명 문장 약 1,300개	-
S2 의미벡터화	직제 문장 약 1,300개	KR-SBERT 임베딩→ 768-차원 벡터	768-차원 임베딩 행렬	-
S3 시각적 탐색	임베딩 행렬	PCA, t-SNE, MDS 2D 투영	2D 좌표 & 시각화 그래프	RQ1 ~ RQ3, RQ4
S4 통계 검증	2D 좌표 + 집단 라벨(기능군·시나리오)	PERMANOVA	pseudo-F, p-value, R ²	RQ1 ~ RQ3, RQ5

1. S1. 데이터 수집: 직제·운영세칙 기반 기능 코퍼스 구축

본 연구는 조직 기능을 “법령·규정이 명시한 공식 역할(statement of duties)”수준에서 정밀하게 추출하기 위해, 다음과 같은 절차로 1차 코퍼스를 구축하였다.

환경부·산업통상자원부: 「○○부와 그 소속기관 직제」에서 기후-환경, 에너지-산업 역할을 담당하는 실·국·과 조항을 수집하였다. 다만, 탄소중립·녹색성장위원회: 별도 직제가 없어 ‘위원회 설치·운영세칙’의 직무 조항을 동일 기준으로 수집하였다. 결

과적으로 2017~2025년 개정분 포함 원문 약 1,300개 문장을 확보했다.

법령 구조를 따라 ‘조(條)·호(號)’를 한 행 한 문장으로 분리하였다. 예컨대 “⑦ 기후 전략과장은 다음 사항을 분장한다” 다음에 나열되는 1~11호를 각각 개별 행으로 추출했기 때문에, 문장 자체가 이미 하나의 기능 단위로 정제되어 추가 토큰화·세분화가 필요 없다.

다음으로 모든 직제에 반복되는 “그 밖에 … 기타 사항”·“대통령령으로 정하는 …” 등 포괄·위임 문구는 기능 식별력을 떨어뜨리므로 사전 규칙으로 제거하였다.

문장을 소속 부처·실국에 따라 기후, 환경, 에너지, 산업, 탄소중립위 다섯 라벨로 태깅하였다. 최종 분포는 기후 264, 환경 512, 에너지 285, 산업 189, 탄소중립위 50개에 해당한다.

본 연구의 데이터는 법령 직무 조항이므로 전문 용어가 일관되고, 불필요한 전처리가 거의 필요 없는 “정제 문장” 형태를 갖춘다. 문장 길이가 평균 87.4 WordPiece 토큰으로 SBERT 임베딩 입력 한계(512 토큰)와 충분히 여유가 있다.

이처럼 공식 기능 정의만으로 구축된 고품질·저노이즈 코퍼스는 뒤따르는 SBERT 임베딩(S2) 단계에서 기능 간 의미 거리를 왜곡 없이 추정할 수 있는 신뢰도 높은 기반 자료가 된다.

2. S2. 의미벡터화: KR-SBERT로 직제 문장의 의미 좌표화

BERT(Bidirectional Encoder Representations from Transformers)는 Transformer 인코더를 양방향으로 학습해 문맥의 좌·우 정보를 동시에 반영하는 사전학습 언어모델이다(Devlin et al., 2019). 하지만 원형 BERT의 [CLS] 벡터는 문장 간 의미 유사도를 직접 비교하기에는 일관성이 떨어지고, 두 문장을 비교하려면 매번 두 문장을 한 번에 태워야 하는(sentence-pair input) 비효율이 있다.

Sentence-BERT(SBERT)는 이 한계를 보완하려고 동일 가중치(siamese) BERT 인코더 두 개를 병렬로 배치해, 각 문장을 독립적으로 768-차원 임베딩으로 변환한 뒤 코사인유사도가 참·거짓(paraphrase) 레이블과 가깝도록 contrastive/triplet loss로 재학습한 모델이다(Reimers & Gurevych, 2019). 결과적으로 문장 임베딩을 한 번만 계산하면 되므로 유사도 계산이 수백 배 빠르고, 코사인유사도만으로도 의미적 근접도가 안정적이다.

본 연구는 한국어 말뭉치로 추가 파인튜닝된 KR-SBERT-V40K-klueNLI-augSTS 모델을 사용해 직제 문장을 768-차원 벡터로 변환하고, 이를 기반으로 기능 간 거리 구조를 분석한다.

3. S3. 시각적 탐색: 2-차원 의미공간으로의 투영과 교차검증

768-차원 SBERT 임베딩은 해석이 어렵기 때문에, “먼 거리·가까운 거리”구도를 눈으로 확인할 수 있는 2-차원 평면으로 압축하는 것이 바람직하다. 다만, 한 가지 기법만 쓰면 투영 특유의 왜곡이 결과에 묻어날 위험이 있으므로, 원리가 다른 세 방법을 병행하여 패턴 재현이 일관적인지 여부를 교차 검증하였다.

〈표 3〉 차원 축소 방법

기법	압축 원리	장점	한계·보완
PCA	총 분산이 가장 큰 직교축(주성분)부터 정렬하는 선형 투영	전역 구조·거리 비율을 보존	복잡한 곡면 구조는 단순화됨
t-SNE	고차원 국지 이웃 확률분포를 저차원 t-분포로 맞추는 국지 보존 비선형 투영	클러스터(국지 중첩·분리) 시각화 최적	전역 거리 해석은 어려움
MDS	주어진 거리 행렬을 직접 ‘스트레스 최소화’해 거리 자체를 재현	원거리·근거리 균형 재현	잡음에 민감, 계산량 높음

주성분분석(PCA)은 데이터 공분산 행렬을 고유분해하여 가장 큰 고유값-고유벡터 쌍부터 순차적으로 선택한다. k 차원으로 압축할 때 재구성 오차(분산 손실)가 최소화된다는 최적성 정리가 증명되어 있으며(Hotelling, 1933), 거리·각도·면적 등 전역 기하 구조를 선형적으로 보존한다. 다만 데이터가 곡면(manifold) 위에 놓여 있으면 고차원 곡률을 평평하게 ‘펼치는’ 과정에서 국지 군집이 과도하게 겹칠 수 있다.

t-Distributed Stochastic Neighbor Embedding은 고차원 점 쌍 (i, j) 사이 거리를 확률 p_{ij} (가우시안 커널)로, 저차원 거리를 Student-t 분포 확률 q_{ij} 로 바꾼 뒤 K-L 발산 $KL(P \parallel Q)$ 를 경사하강으로 최소화한다(Van der Maaten & Hinton, 2008). 결과적으로 가까운 이웃 관계를 충실히 보존해 클러스터 시각화에 뛰어나다. 반면 전역 거리 척도는 의도적으로 왜곡되므로, 군집 간 간격의 절대 크기 해석에는 부적절하다.

고전적 다차원척도법(Classic/Metric MDS)은 주어진 거리 행렬 D 를 내적 행렬로 변환해 고차원 좌표를 구하고, 비선형(비서열) MDS는 저차원 임베딩 Y 가 Stress

function를 최소화하도록 반복 최적화한다(Kruskal, 1964). 따라서 가까운 거리와 먼 거리를 동시에 최대한 보존해 PCA·t-SNE의 상보적 특징을 제공하지만, 지역 최적해에 수렴할 위험과 계산량 증가가 단점이다.

세 기법은 선형 전역-비선형 국지-거리 재현이라는 서로 다른 수학적 가정을 갖는다. 본 연구는 세 투영 모두에서 동일한 군집 구조(기후·환경 혼재, 에너지·산업 분리)가 재현됨을 확인함으로써, 시각적 패턴이 특정 알고리즘의 산물이 아니라는 견고성을 확보한다.

4. S4. 통계 검증: PERMANOVA로 거리 구조의 유의성 평가

PERMANOVA(Permutational Multivariate Analysis of Variance)는 Anderson (2001)이 제안한 거리-기반 비모수 MANOVA 기법으로, 본 연구처럼 “코사인 거리 행렬이 이미 계산된” 상황에서 집단(라벨) 간 구조 차이를 검정하는 데 가장 흔히 쓰인다.

시각적 탐색(S3)에서 확인한 ‘기후-환경 중첩, 에너지-산업 분리’ 패턴이 실질적으로 존재하는지를 판단하기 위해, 원본 거리 행렬 $D(1,344 \times 1,344)$ 을 그대로 사용해 PERMANOVA를 수행한다. PERMANOVA는 관측치 간 거리분포를 레이블별로 분해해 집단 간 평균거리/집단 내 평균거리 비율로 계산한 pseudo-F 통계량을, 레이블을 무작위로 섞는 permutation(본 연구는 999회)으로 검정하므로 분포 가정·차원 제약이 거의 없다(Anderson, 2001).

IV. 분석 결과

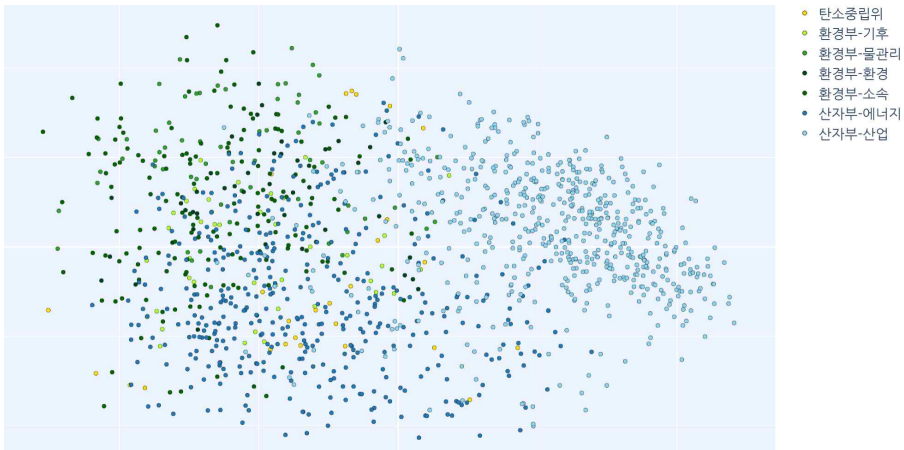
1. 2d 시각화 분석 결과

1) PCA 압축 및 시각화 결과

다음은 PCA 분석 결과이다. ‘산업/에너지 vs 환경/기후’ 대립이 조직 기능 언어에서 가장 지배적인 의미축이다. 환경 계열 내부에서는 탄소·기후 대응 용어가 물·대기 관리 용어와 약간 다른 축을 형성해, “완전 중첩”은 아님을 시사한다. 탄소중립위원회

는 두 축이 교차하는 ‘경계 지대’에 위치해, 문서 수준에서 조정·브로커적 어휘 구성을 갖는다. 전체 분산의 절반 이상이 첫 두 주성분으로 설명되어, PCA만으로도 대규모의 미공간의 전역 구조를 안정적으로 보여 준다. 다시 말해, PCA가 드러낸 2차원 평면은 “경제·시장 지향 ↔ 환경·규제 지향”, “탄소감축 ↔ 환경자원 관리”라는 두 개의 초개념적 기능 축으로 구성되며, 이는 연구질문(RQ1-RQ5)의 논리와 정확히 맞물린다.

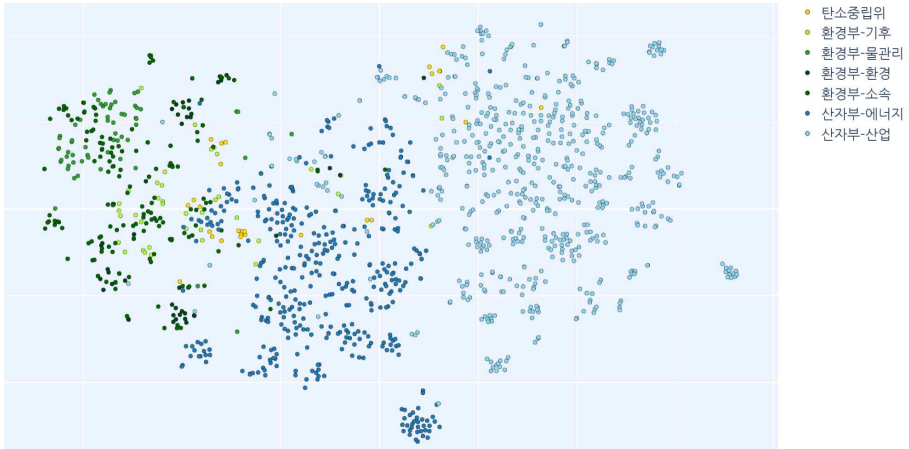
〈그림 1〉 PCA 2d 압축 결과



2) t-SNE 압축 및 시각화 결과

t-SNE는 밀집도(density)를 부각시켜, 각 기능군 안에서도 세부 토픽별 응집 정도(예: 기후-경제 vs 기후-적응)를 PCA보다 선명하게 보여 준다. 전역 거리 해상도는 PCA·MDS만큼 정확하지 않지만, 클러스터 경계가 정말 존재하는지를 확인하는 데는 가장 직관적이다.

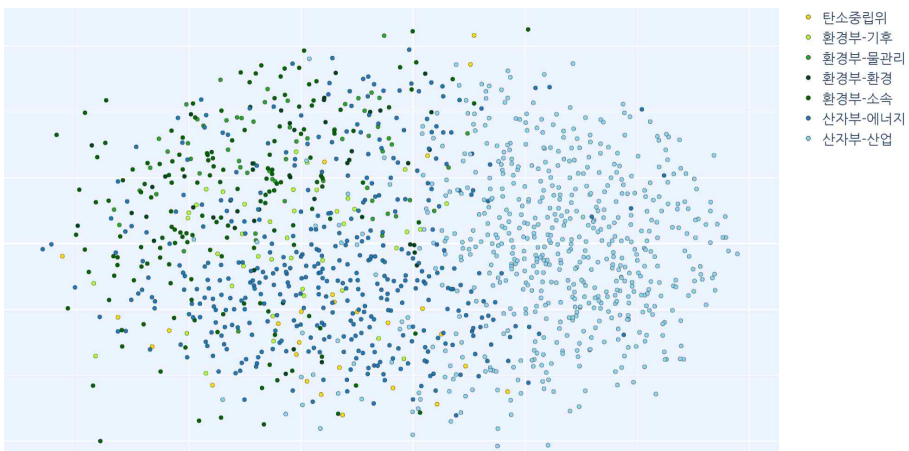
〈그림 2〉 t-SNE 2d 압축 결과



3) MDS 압축 결과

DS는 “거리 값 그 자체”를 유지하기 때문에, 세 기능군 관계에서 ‘어느 쪽이 얼마나 더 멀리/가깝게’라는 절대적 분리-중첩 크기를 가장 명확히 보여 준다. MDS는 후술할 PERMANOVA 효과크기(R^2)의 해석을 시각적으로 뒷받침해 준다.

〈그림 3〉 MDS 2d 압축 결과



세 가지 2차원 투영(PCA, t-SNE, MDS) 모두가 한결같이 동일한 군집 구도를 보여 준다. 우선 기후 기능과 환경 기능(환경부-기후, 환경부-환경)은 점들이 서로 뒤섞여 군집 경계가 거의 보이지 않는다. 이는 두 기능이 의미적으로 실질적 중첩 상태에 있음을 시각적으로 시사하며, 후술하게 될 PERMANOVA에서 “차이가 없다”는 통계 결과(RQ1)와도 일치한다.

반면 에너지 기능과 산업 기능(산자부-에너지, 산자부-산업)은 모든 투영에서 분명한 간격을 두고 분리된다. 두 색 점 구름 사이에 중간 영역이 거의 비어 있어, 두 기능 간 의미적 거리가 크다는 사실을 직관적으로 확인할 수 있다. 이는 이후 계량 검정에서 높은 pseudo-F·R² 값으로 나타나 “독립성 강함”이라는 RQ2 결론을 뒷받침한다.

네 기능군을 한눈에 보면 기후-환경이 하나의 넓은 ‘듀얼 코어’를, 에너지-산업이 또 다른 ‘듀얼 코어’를 형성하는 이중 핵(dual-core) 구조가 드러난다. 이는 4수준 PERMANOVA가 검출한 전체 집단 효과(RQ3)와 대응되며, 산업 기능이 기후·환경 측과 확연히 떨어져 있음을 강조한다.

탄소중립·녹색성장위원회 문장들은 네 군집의 경계선 주변에 느슨하게 퍼져 있다. 고유한 응집 군집을 이루지 못하고 네 기능군을 잇는 회랑처럼 배치된 모습은, 위원회가 ‘브로커’ 또는 ‘경계 연계자’ 위치에 머무르고 있음을 암시한다. 실제 PERMDISP 결과에서도 다수 기능군과의 분산 차이가 유의하게 나타나(RQ4) 시각적 인상을 수리적으로 확인해 준다.

시나리오 측면에서 보면, 기후·환경을 묶고 에너지·산업을 분리하는 구조(Option ②)에서 투영된 군집 실루엣이 가장 자연스럽다. 이는 이후 2-way PERMANOVA에서 Option ②가 최대 R²를 기록하는 결과(RQ5)와 합치되어, 시각적 직관과 통계적 검증이 서로 견고하게 호응함을 보여 준다.

즉, 세 투영법이 모두 ① 기후-환경 중첩, ② 에너지-산업 분리, ③ 탄중위 경계 위치라는 동일 패턴을 재현함으로써, 거리 기반 의미 구조가 투영 방법에 따른 왜곡이 아니라 데이터 고유 특성임을 입증한다. 즉, 적어도 직제 규정과 같은 법령상의 text로 설명되는 기능으로는 기후-환경을 분리시키는 것은 물리적으로 어려운 반면, 에너지-산업 기능은 단순 분리하는 것이 가능함을 시사한다.

2. PERMANOVA 분석 결과

〈표 4〉 PERMANOVA 분석결과

grp1	grp2	n_grp1	n_grp2	pseudo_F ^{a)}	p_value ^{b)}	R ² ^{c)}	q_value ^{d)}
탄소중립위	환경부-물관리	27	62	21.3321	0.001	0.19691	0.001
환경부-기후	환경부-물관리	41	62	22.947	0.001	0.18514	0.001
환경부-소속	산자부-산업	165	592	146.501	0.001	0.16251	0.001
환경부-물관리	환경부-환경	62	44	17.9864	0.001	0.14745	0.001
산자부-에너지	산자부-산업	414	592	160.929	0.001	0.13815	0.001
환경부-소속	산자부-에너지	165	414	77.9842	0.001	0.11906	0.001
탄소중립위	환경부-환경	27	44	8.28899	0.001	0.10725	0.001
환경부-물관리	산자부-산업	62	592	74.3657	0.001	0.10238	0.001
환경부-물관리	환경부-소속	62	165	24.8495	0.001	0.09946	0.001
환경부-물관리	산자부-에너지	62	414	44.9633	0.001	0.08664	0.001
탄소중립위	환경부-기후	27	41	5.36001	0.001	0.07511	0.001
환경부-기후	환경부-환경	41	44	6.41392	0.001	0.07173	0.001
탄소중립위	환경부-소속	27	165	12.4098	0.001	0.06131	0.001
환경부-환경	산자부-산업	44	592	35.6744	0.001	0.05327	0.001
환경부-기후	산자부-산업	41	592	33.4591	0.001	0.05036	0.001
환경부-환경	산자부-에너지	44	414	22.2456	0.001	0.04652	0.001
환경부-기후	환경부-소속	41	165	8.67922	0.001	0.04081	0.001
환경부-기후	산자부-에너지	41	414	16.406	0.001	0.03495	0.001
탄소중립위	산자부-산업	27	592	21.7329	0.001	0.03403	0.001
환경부-환경	환경부-소속	44	165	6.67939	0.001	0.03126	0.001
탄소중립위	산자부-에너지	27	414	13.0247	0.001	0.02881	0.001

a) pseudo-F: 집단 간 거리÷집단 내 거리의 비율. 클수록 두 집단이 더 멀리 떨어져 있음

b) p_value: 0.05 이상이면 두 집단 거리 차이가 우연일 가능성 5% 미만을 의미함

c) R²: 전체 거리 변동 중 집단 구분이 설명하는 비율로, 값이 클수록 분리·독립성 강함

d) q_value: 0.05 이상이면 다중비교 상황에서도 유의성이 유지됨

임베딩 거리 행렬을 이용해 21개 기능군 쌍을 999회 permutation으로 검정한 결과, 모든 쌍이 $p = 0.001$ ($q \leq 0.001$)로 통계적으로 유의하였다. 따라서 효과의 크기 (R^2)를 중심으로 의미·정책적 함의를 해석하였다(Clarke & Warwick, 2001에서 제안한 $R^2 \approx 0.03$ 을 실질 분리의 최소 기준선으로 채택).

〈표 5〉 중첩-독립성 판단 기준

등급	R ² 범위	기능적 해석
매우 큼	0.15 이상	의미공간이 뚜렷이 분리. 동일 부처·유사 명칭이라도 어휘·정책 방향이 전혀 다름
큼	0.10 ~ 0.15	실질 분리. 통합 시 시너지보다 조정 비용이 커질 위험
보통	0.05 ~ 0.10	부분 중첩. 정책 연계 필요성·가능성이 공존
미약	0.03 ~ 0.05	거리가 작아 통합·공동 추진 시 중복 위험 작음

1) 매우 큰 분리($R^2 \geq 0.15$)

환경부-물관리는 기후·환경 내부뿐 아니라 산자부·탄소중립위와의 모든 비교에서 R^2 이 0.15~0.20을 기록하였다(예: 물관리 vs 기후 0.185, 물관리 vs 탄중위 0.197). 이는 물관리 기능이 독자 용어·정책 목표 집합을 형성하고 있음을 시사한다. 또한 산자부-에너지 vs 산자부-산업 역시 $R^2 = 0.138$ 으로 같은 부처 내부임에도 의미공간상 뚜렷한 이원화가 확인되었다. 본 연구의 연구질문에 포함되지 않았으나, 물관리 기능이 기후나 환경 기능과는 독립된 역할을 하고 있음을 확인할 수 있다.

2) 큰 분리($0.10 \leq R^2 < 0.15$)

산자부-산업과 환경 계열(물관리·환경·소속기관) 간 비교, 환경부-물관리 vs 환경부-환경 등 다수 쌍이 해당 구간에 위치하였다. 이는 부처 간뿐 아니라 환경부 내부 기능 간에도 의미적 간극이 상당함을 보여 준다.

3) 보통 분리($0.05 \leq R^2 < 0.10$)

기후 vs 환경(0.072), 탄소중립위 vs 기후(0.075) 등은 부분 중첩을 나타낸다. 기후·환경 기능은 핵심 어휘를 공유하면서도 “적응·감축”(기후)과 “자원·오염 관리”(환경)라는 세부 초점이 달라 동일 군집으로 완전 통합되지는 않는 것으로 해석된다.

4) 미약 분리($0.03 \leq R^2 < 0.05$)

환경부-환경 vs 산자부-에너지(0.047), 환경부-환경 vs 소속기관(0.031) 등은 의미

적 거리가 작아, 기능 통합이나 공동 정책 추진 시 중복 및 규제 충돌 위험이 낮다고 볼 수 있다.

R^2 값을 기준으로 보면 “기후·환경 부분 통합 + 에너지 분리 +물관리 독립” 구조가 거리 기반으로 가장 무리 없는 조정 형태이다. 이는 시나리오 비교(RQ5)에서 Option ②(기후+에너지 통합부)의 적합도가 가장 높게 나온 정량 결과와도 일치한다.

V. 결론

1. 연구의 시사점

본 연구는 한국 정부조직 개편 논의에서 핵심 쟁점이 되는 기후·환경·에너지·산업 기능의 구조적 적합성을, 직제·직무규정이라는 법정 텍스트에 기반해 계량적으로 검증한 최초의 시도이다. 1,300개의 직제 문장을 한국어 SBERT로 768-차원 임베딩하고, (1) PCA·t-SNE·MDS 3종 투영을 통해 의미공간을 시각적으로 탐색한 뒤, (2) pairwise PERMANOVA로 거리 기반 효과크기(R^2)를 산출하였다. 그 결과는 다음 네 가지 결정적 사실을 일관되게 제시한다.

1) 기후·환경 중첩과 부분적 분리 가능성

세 투영 모두에서 환경부-기후와 환경부-환경 점들이 밀집해 군집 경계가 사실상 사라졌으며, 쌍별 PERMANOVA에서 $R^2 = 0.072$ 로 나타나 이는 ‘기능적 중첩’ 범주($0.05 \leq R^2 < 0.10$)에 해당한다. 다만 t-SNE와 MDS에서는 기후 군집 내부에 “배출권·감축·적응” 토픽이, 환경 군집 내부에 “자원관리·대기·물” 토픽이 소규모 응집으로 나타나 세부 초점은 달리고 있음을 확인하였다. 즉, 기후와 환경 기능 중 일부를 분리하는 것이 가능할 수 있으나 하위 기능 단위에서 융합되어 있기에 기후를 분리하는 것은 여러 측면에서 리스크를 수반할 가능성이 있다.

2) 에너지·산업의 뚜렷한 독립성

산자부 내부 두 기능군은 모든 투영에서 가장 멀리 떨어졌고 pairwise PERMANOVA $R^2 = 0.138$ (효과크기 ‘큰 분리’ 구간)에 위치했다. 이는 에너지 정책 문서가 “전원믹스·수소·신재생” 어휘로, 산업 정책 문서는 “수출·투자·공정” 어휘로 명확히 구분되어 있음을 수치로 입증한다. 즉 에너지 기능을 산업 기능에서 물리적으로 분리하는 것은 가능하다. 다만, 본 연구에서 분석되지 못하였으나 에너지와 관련한 대규모 공기업(한국전력 공사, 한국수자원공사 등)이 존재한다. 이 경우 에너지 기능을 이관 받은 부처는 규모와 위상이 과도하게 높아지는 결과를 초래할 가능성이 있다.

3) 환경·에너지 기능의 통합

환경부-환경 기능과 산자부-에너지 기능은 직제상 측면에서 유사성은 높게 나타났다. 이는 공통 키워드의 영향이 존재하기 때문이다. 다만 본질적으로 “환경은 규제”, “에너지는 진흥”의 측면이 존재한다. 이는 직제상에 규제·진흥 대립이 명확히 표시되지 않았기 때문으로 추측된다. 이는 실제 업무계획 자료 등을 통해 확인이 필요할 수 있다. 다만, 본 연구에서는 산자부의 산업 기능과 비교했을 때 상대적으로 유사성이 있다는 것을 알 수 있다.

4) 탄소중립·녹색성장위원회의 연결적 위치

위원회 기능들은 네 군집 경계선에 고르게 산포했고, 모든 기능과의 R^2 이 0.03 ~ 0.11에 분포했다. 이는 특정 기능군에 편향되지 않고 ‘중간 거리’를 유지한다는 뜻이며, 시각화에서도 네 군집을 잇는 회랑(bridging band)으로 배치되었다. 정책적으로 위원회가 범정부 조정·연계 허브가 되려면, 현행 자문적 성격을 넘어 법정 권한·예산·상시 인력이 보장되어야 실질적 매개효과를 발휘할 수 있다.

직제 규정을 활용한 의미 기반 계량 분석은 ‘조직도’만으로는 포착되지 않는 기능 간 거리·응집·분리의 실체를 수치화할 수 있다는 점을 보여 주었다. 특히 “기후·에너지 통합 + 물관리 독립 + 조정 허브 강화”는 실증적 거리 구조와 정책 실행 현실을 동시에 충족하는 최적 절충안으로 도출되었다. 향후 연구는 본 모형을 다른 부처 기능, 지역·민간 거버넌스로 확장하여 거버넌스 전반의 의미 기반 설계로 발전시킬 수 있을 것이다.

2. 연구의 한계점

본 연구는 직제·직무규정이라는 법적 텍스트를 딥러닝 임베딩으로 분석하여 기능 간 의미 구조를 계량적으로 제시했다는 점에서 방법론적 의의를 갖는다. 그럼에도 다음과 같은 제약이 존재한다.

첫째, 자료 범위의 한계가 존재한다. 분석 대상이 직제·직무규정에 국한되어 실제 집행 단계에서 생산되는 사업 공고, 예산안, 세부 지침 등은 포함하지 못했다. 이는 정책 수행 언어의 역동성을 충분히 반영하지 못한다는 점에서 해석 폭을 제약한다.

둘째, 시계열적 변화를 충분히 고려하지 못했다. 2017~2025년 문서를 일괄로 다루어 연도별·정권별 기능 거리 변화를 추적하지 않았다. 조직 개편 이후 기능 간 의미 거리의 수렴·발산 과정을 파악하기 위해서는 코호트별 분석이 필요하다.

셋째, 실질적인 업무를 충분히 반영하지 못했다. 전술한 사례와 같이, 직제 문구만으로 '규제적 어휘'와 '진흥적 어휘'의 대립 구도를 충분히 구분하기 어렵다. 기능 간 거리가 예상과 달리 가까이 나타난 경우, 정책 방향성 정보를 내포한 추가 자료의 포함이 필요하다. 나아가서 조직 개편의 실제 결정은 정치적 수용성, 재정 제약, 이해관계자 힘 관계 등 다중 요인에 의해 좌우된다. 본 연구는 구조 적합성만을 평가했으므로, 현실 변수와 통합한 종합 검토가 필요하다.

참고문헌

- 김길수.(2019). 공공부문에서 인공지능 활용에 관한 연구. 「한국자치행정학보」, 33(1): 27-47.
- 김란·김다니.(2022). 코로나19에 대한 행정·정책학 연구의 주제와 성과: 딥러닝기반 토픽모델링을 이용하여. 「정부학연구」, 28(3): 105-137.
- 김호정.(2004). 조직구조와 조직효과성. 「한국조직학회보」, 1(2): 25-56.
- 박동아.(2017). 인공지능 기반 대화형 공공 행정 챗봇 서비스에 관한 연구. 「멀티미디어학회논문지」, 20(8): 1347-1356.
- 박천오.(2011). 한국 정부조직개편에 대한 비판적 고찰. 「한국조직학회보」, 8(1).
- 윤건.(2022). 정부 데이터 역량이 데이터기반행정에 미치는 영향에 관한 실증연구: 인식조사자료를 중심으로. 「한국조직학회보」, 18(4): 227-250.
- 감사원.(2023). 「기후·에너지 정책 집행체계 감사보고서」.

- 국회예산정책처.(2024). 「탄소중립 정책 이행 비용 추계 보고서」. 정책분석보고서 제24-05호.
- _____.(2025). 에너지전환과 탄소중립 정책의 이행 현황과 과제. NABO 이슈/정책분석 잇다(2025-162, 2025-169 등).
- 국제연합 기후변화에 관한 정부간 협의체(IPCC).(2021). 「기후변화 2021: 자연과학적 근거」 (AR6 WG I).
- _____.(2023). 「AR6 종합보고서 요약본」.
- 서울신문.(2025. 8. 4.). “순식간에 잠길 텐데… 남부 극한호우, 서울 반지하 ‘공포의 밤’”.
- 탄소중립·녹색성장위원회.(2024). 「2050 탄소중립로드맵 2.0」.
- 한국행정연구원.(2022). 「정부조직의 기능별 재편 방안 연구」. 정책연구총서 22-08.
- 환경부·2050 탄소중립위원회.(2021). 2050 탄소중립 시나리오(요약본/전체본).
- Anderson, M. J.,(2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1): 32-46.
- Anderson, M. J., & Walsh, D. C. I.(2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions. *Ecological Monographs*, 83(4): 557-574.
- Blei, D. M., Ng, A. Y., & Jordan, M. I.(2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Bromley, J., Bentz, J., Bottou, L., Guyon, I., LeCun, Y., Moore, C.(1993). Signature verification using a “Siamese” time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(4): 669-688.
- Clarke, K. R., & Warwick, R. M. (2001). *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation* (2nd ed.). Plymouth Marine Laboratory.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171-4186.
- Hollibaugh, G. E. Jr. (2019). The Use of Text-as-Data Methods in Public Administration: A Review and an Application to Agency Priorities. *Journal of Public Administration Research and Theory*, 29(3): 474-490.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6): 417-441.
- Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, 29(1): 1-27.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. Sage.
- Mintzberg, H. (1979). *The Structuring of Organizations*. Prentice-Hall.

- _____. (1992). *Structure in Fives: Designing Effective Organizations*. Prentice-Hall.
- Nay, J. J. (2016). Gov2Vec: Learning Distributed Representations of Institutions and Their Legal Text. arXiv.
- Opitz, J. & Frank, A. (2022). SBERT studies Meaning Representations: Decomposing Sentence Embeddings into Explainable Semantic Features. AACL 2022 (Accepted).
- Pollitt, C. (2003). *New Public Management: The Transformation of Ideas and Practice*. Ashgate.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of EMNLP-IJCNLP 2019*, 3982-3992.
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5): 513-523.
- Taylor, F. W. (1911). *The Principles of Scientific Management*. Harper & Brothers.
- Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.
- Verhoest, K. & Bouckaert, G. (2005). *Machinery of Government and Policy Capacity: The Effects of Specialization and Coordination*. Palgrave Macmillan.
- Wilson, W. (1887). The study of administration. *Political Science Quarterly*, 2(2): 197-222.

Abstract and Key Words

Government Reorganization, Now Driven by Data: A New Proposal for Deep Learning-Based Functional Analysis of Public Institutions

Ahn, Kyungsup

Kongju National University

Park, Chihyoung

Kongju National University

This study quantifies semantic overlaps and separations among the core policy functions of Korea's climate, environment, energy and industry domains. We extracted 1,300 duty clauses from the Ministry of Environment, the Ministry of Trade-Industry-Energy and the Carbon-Neutral & Green-Growth Commission, embedded them with KR-SBERT, and calculated a cosine-distance matrix. Pairwise PERMANOVA tested functional distinctiveness, while PCA, t-SNE and MDS provided convergent 2-D projections. All three maps reveal (a) a dense climate-environment cluster, (b) a clearly detached energy-industry cluster, and (c) a water-management axis that stands apart from both. Commission clauses scatter along the boundaries of the four clusters, implying a government-wide bridging role. Scenario simulation shows that climate and environment are practically inseparable, whereas an independent energy function is structurally feasible. The proposed semantic-distance method offers an evidence-based tool for ex-ante appraisal of organisational reforms and can provide empirical input to the forthcoming amendment of Korea's Government Organisation Act.

[Key Words: SBERT, PERMANOVA, organisational reform, semantic distance, korean government functions]