

국회 활성화를 위한 국회의원 입법활동 평가모델

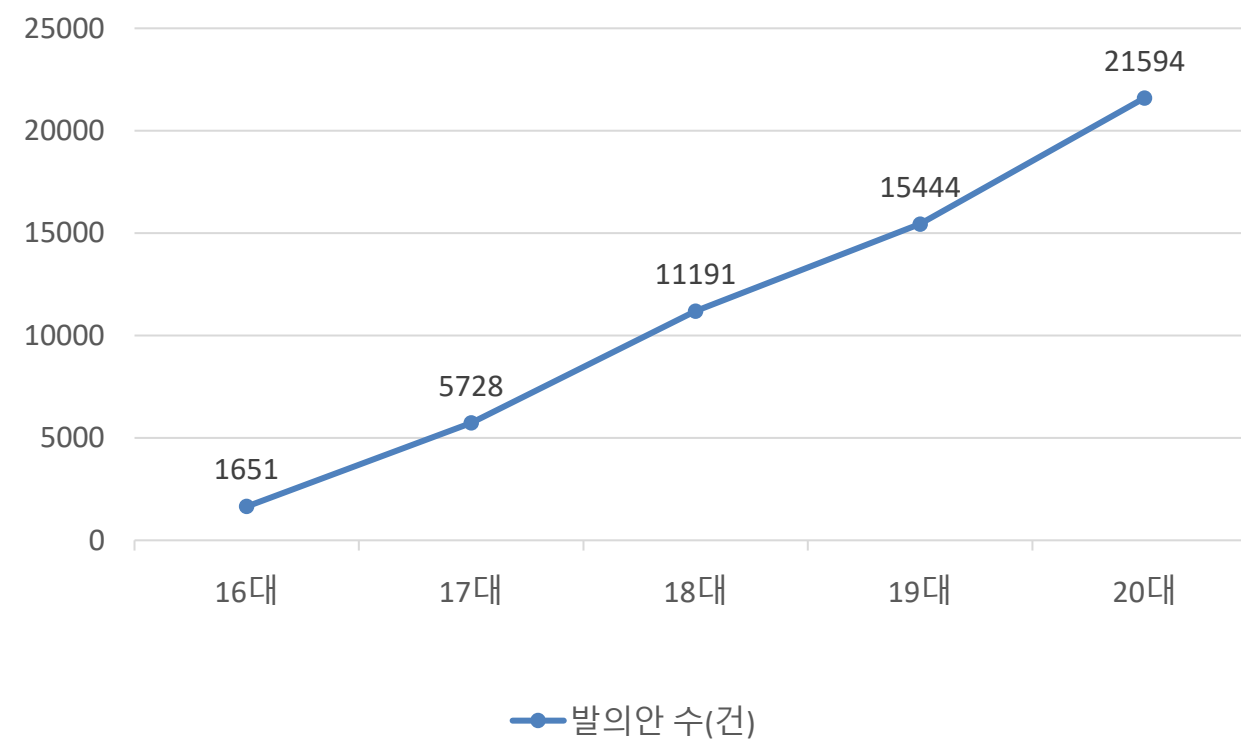
5팀 시계열자료분석팀



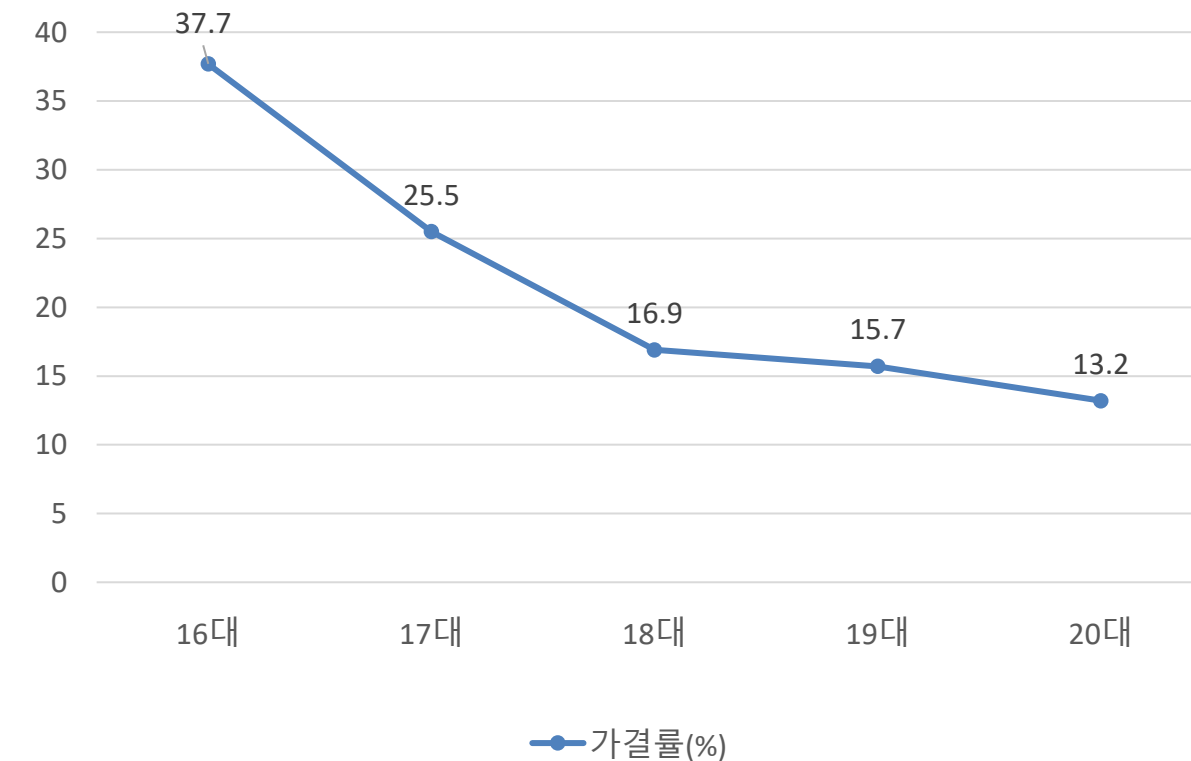
1. 주제 선정 배경

갈수록 증가하는 법률안 발의 건수와 대비되는 가결률의 감소

국회의원 회기별 법률안 발의 건수



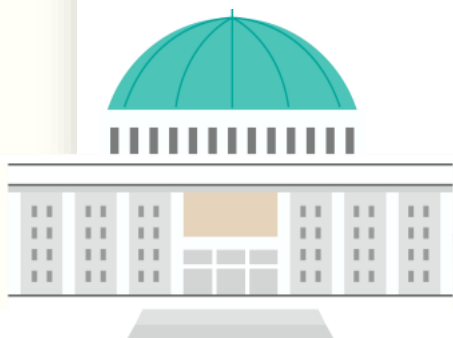
국회 회기별 법률안 가결률



가파르게 증가하는 법률안 발의 건수에 비해 **가결률**은 감소하는 추세를 보임!

발의안이 통과될 확률

자료 출처:의안정보시스템



1. 주제 선정 배경

갈수록 증가하는 법률안 발의 건수와 대비되는 가결률의 감소

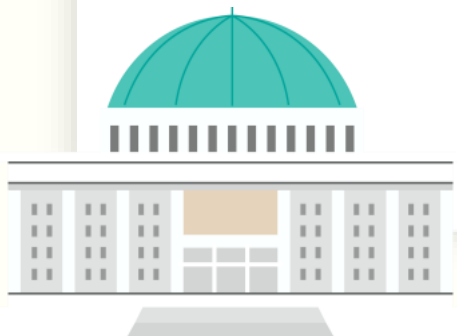
<20대 국회 소속정당별 의원안의 처리결과>

왜 이런 현상이 발생하는 걸까?

구분	전체	원안가결	왜 이런 현상이 발생하는 걸까?				수정반영폐기	철회	임기만료폐기
새누리당 / 자유한국당	6,241 (28.9%)	177 (2.8%)	267 (4.3%)	1 (0.0%)	36 (0.6%)	1,592 (25.5%)	8 (0.1%)	57 (0.9%)	4,103 (65.7%)
더불어민주당	11,067 (51.3%)	264 (2.4%)	427 (3.9%)	1 (0.0%)	54 (0.5%)	2,595 (23.5%)	20 (0.2%)	100 (0.9%)	7,606 (68.7%)
국민의당	3,508 (16.2%)	117 (3.3%)	156 (4.5%)	- (0.0%)	29 (0.8%)	776 (22.1%)	8 (0.2%)	51 (1.5%)	2,371 (67.6%)
정의당	405 (1.9%)	0 (0.0%)	- (0.0%)	- (0.0%)	1 (0.2%)	97 (23.7%)	5 (1.2%)	4 (1.0%)	301 (74.3%)
무소속	373 (1.7%)	11 (3.0%)	14 (3.8%)	- (0.0%)	3 (0.8%)	77 (20.6%)	- (0.0%)	3 (0.8%)	265 (71.1%)
의원안 합계	21,594 (100.0%)	569 (2.6%)	868 (4.0%)	2 (0.0%)	123 (0.6%)	5,130 (23.8%)	41 (0.2%)	215 (1.0%)	14,646 (67.8%)

기존 국회의원 평가세척이
발의안의 '질'이 아닌 '양'만을 고려하는
한계점을 지니기 때문

기존 국회의원 평가세척이
발의안의 '질'이 아닌 '양'만을 고려하는
한계점을 지니기 때문



1. 주제 선정 배경

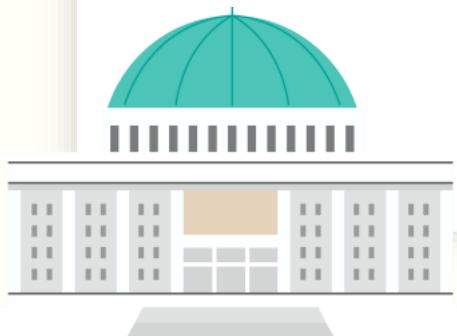
기대 효과

실적 채우기에 급급한 “**졸속법안**”을 발의 하는 것을 방지하고,
국민에게 필요한 “**실속 있는**” 법안을 발의하도록 유도

단순한 발의안 수가 아닌
발의안의 신설 조항 수를 고려하여 평가

뉴스기사 데이터를 활용하여
발의안의 사회동향 반영 정도 파악

국회의원 의정평가 모델로의 확장



2. 용어 정리 및 분석 흐름

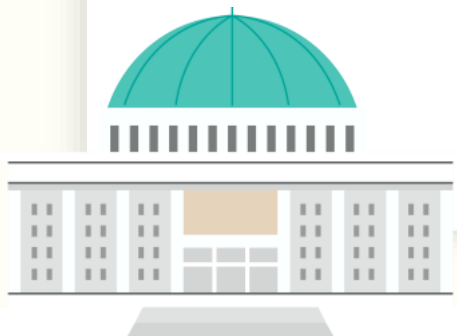
분석 흐름

법률발의안, 뉴스기사 데이터
수집 및 분석

각 발의안의 내용을 포함하는
발의안 데이터셋 구성

변수별 가중치를 산출하여
국회의원 별 최종 입법활동 점수 계산

발의안에서 유의미한 feature를 뽑아
각 국회의원에 대한 데이터셋으로 구성



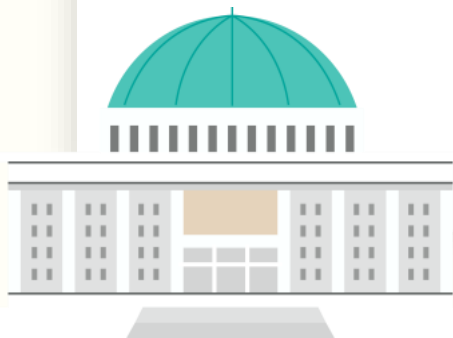
4. 데이터 활용

(1) 발의안 데이터 활용 | 1차 발의안 데이터셋

제안이유, 발의안 원문 pdf주소까지 함께 크롤링하여 수집

의안번호	링크	발의안명	대표발의자	공동발의자	소관위원회	발의날짜	본회의 심사결과	제anyi유	발의안 원문 pdf 주소
2024996	http://li...	집합건물 ...	김병관	고용진,김...	법제사법위원회	2020-05...	임기만료폐기	제anyi유 ...	javascrip...
2024995	http://li...	지방세법 ...	김병관	김철민,박...	행정안전위원회	2020-05...	임기만료폐기	제anyi유 ...	javascrip...
2024933	http://li...	상장회사 ...	채이배	김관영, 김...	정무위원회	2020-05...	임기만료폐기	제anyi유...	javascrip...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2020770	http://li...	대기환경...	윤상현	김성원,김...	환경노동위원회	2019-06...	대안반영폐기	제anyi유 ...	javascrip...

⋮



4. 데이터 활용

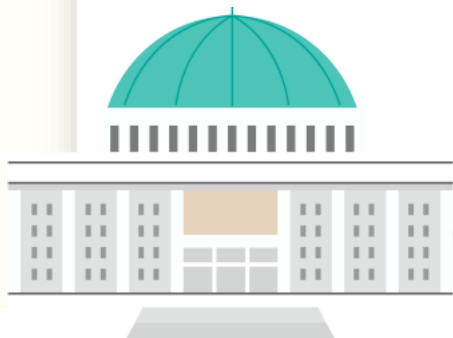
(1) 발의안 데이터 활용 | 1차 발의안 데이터셋

전처리

제목을 보고 신설 여부를 판단하여 이진 변수 추가
신설 법안 - 1 / 일부개정법률안 - 0



의안번호	링크	발의안명	대표발의자	공동발의자	소관위원회	발의날짜	본회의 심사결과	제안이유	발의안 원문 pdf 주소
2024996	http://li...	집합건물 ...	김병관	고용진,김...	법제사법위원회	2020-05...	임기만료폐기	제안이유 ...	javascrip...
2024995	http://li...	지방세법 ...	김병관	김철민,박...	행정안전위원회	2020-05...	임기만료폐기	제안이유 ...	javascrip...
2024933	http://li...	상장회사 ...	채이배	김관영, 김...	정무위원회	2020-05...	임기만료폐기	제안이유...	javascrip...
:	:	:	:	:	:	:	:	:	:
2020770	http://li...	대기환경...	윤상현	김성원,김...	환경노동위원회	2019-06...	대안반영폐기	제안이유 ...	javascrip...



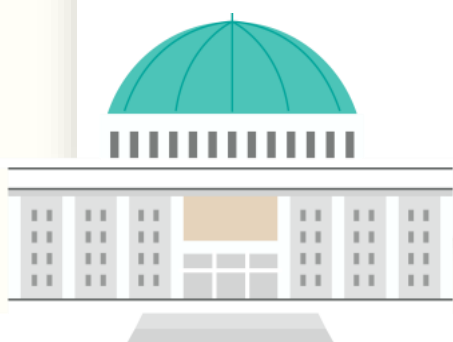
4. 데이터 활용

(1) 발의안 데이터 활용 | 1차 발의안 데이터셋

전처리

수집해둔 발의안 원문 pdf 주소를 이용하여
신·구조문대비표의 <신설> 및 <단서 신설>, <삭제> 조항의 개수 count
→ 각각 변수로 추가!

의안번호	링크	발의안명	대표발의자	공동발의자	소관위원회	발의날짜	본회의 심사결과	제안이유	발의안 원문 pdf 주소
2024996	http://li...	집합건물 ...	김병관	고용진,김...	법제사법위원회	2020-05...	임기만료폐기	제안이유 ...	javascrip...
2024995	http://li...	지방세법 ...	김병관	김철민,박...	행정안전위원회	2020-05...	임기만료폐기	제안이유 ...	javascrip...
2024933	http://li...	상장회사 ...	채이배	김관영, 김...	정무위원회	2020-05...	임기만료폐기	제안이유...	javascrip...
:	:	:	:	:	:	:	:	:	:
2020770	http://li...	대기환경...	윤상현	김성원,김...	환경노동위원회	2019-06...	대안반영폐기	제안이유 ...	javascrip...



4. 데이터 활용

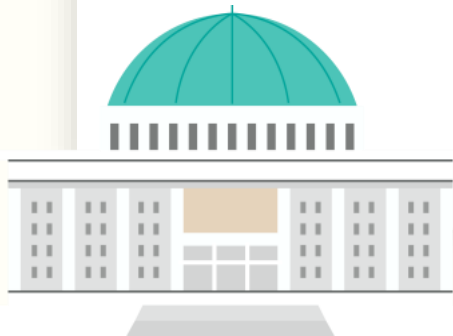
(1) 발의안 데이터 활용 | 1차 발의안 데이터셋

전처리



전처리 후 발의안 데이터셋

의안번호	발의안명	대표발의자	공동발의자	소관위원회	발의날짜	본회의 심사결과	제안이유	new	신설	삭제
2024996	집합건물 ...	김병관	고용진,김...	법제사법위원회	2020-05...	임기만료폐기	제안이유 ...	0	9	0
2024995	지방세법 ...	김병관	김철민,박...	행정안전위원회	2020-05...	임기만료폐기	제안이유 ...	0	0	0
2024933	상장회사 ...	채이배	김관영, 김...	정무위원회	2020-05...	임기만료폐기	제안이유...	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2020770	대기환경...	윤상현	김성원,김...	환경노동위원회	2019-06...	대안반영폐기	제안이유 ...	0	3	3



4. 데이터 활용

(1) 발의안 데이터 활용

발의안의 사회동향 반영 정도 파악

■ 제안이유

아파트 등 공동주택은 「공동주택관리법」에 의해 체계적으로 관리되고 있는 반면, 주상복합·상가·오피스텔 등 집합건물은 공적 관리기준이 미비하여 불투명하게 운영되고 있고 최근 주상복합·오피스텔 등 집합건물이 증가하고 규모가 대형화됨에 따라 거주민 간 혹은 당사자 간 다양한 형태의 분쟁과 갈등이 발생하고 있음.

특히 집합건물법은 「공동주택관리법」과 달리 관리비 정보공개 및 지자체의 감독 등 관리업무에 대한 세부사항에 대한 규정이 미흡하여 집합상가에서 관리비 관련 민원이 자주 발생하고 있음.

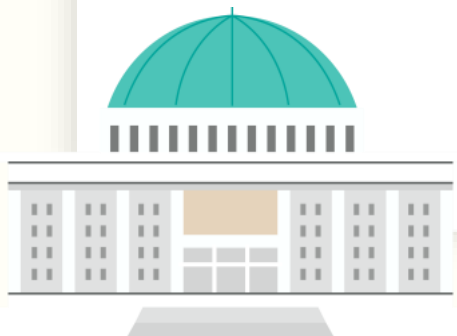
또한, 집합건물과 관련한 분쟁을 해결하기 위해 각 시·도에 '분쟁조정위원회'를 설치해 관리비의 사용내역과 관련된 분쟁을 조정하도록 하고 있지만, 조정의 강제력이 없어 실효성이 없는 실정임.

⋮



발의안 데이터 셋의 내용 열에서

"키워드 추출 시도"



4. 데이터 활용

(1) 발의안 데이터 활용

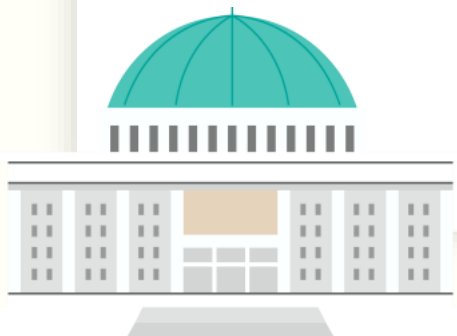
｜ 발의안의 사회동향 반영 정도 파악

[2] 텍스트 토큰화

텍스트를 토큰 단위로 쪼개는 것으로, 주로 형태소 기반 토큰 사용



Kiwi(Korean Intelligent Word Identifier) 엔진 기반의
파이썬 라이브러리 **KiwiPiePy** 사용하여 토큰화 진행 !



4. 데이터 활용

(1) 발의안 데이터 활용

｜ 발의안의 사회동향 반영 정도 파악

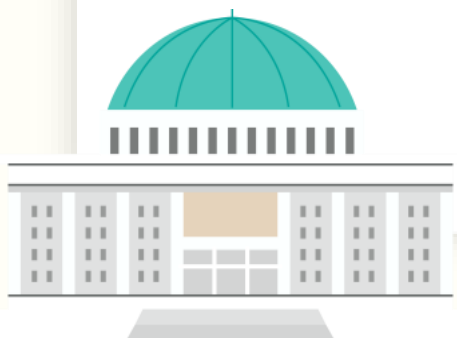
[2] 텍스트 토큰화

텍스트를 토큰 단위로 쪼개는 것으로, 주로 형태소 기반 토큰 사용



명사 추출 토큰화를 통해 전체 데이터에서 등장하는 명사의 빈도수 계산 후,
발의안 TEXT의 특성상 빈도수가 잦은 명사들 또한 제외

EX) 경우, 신설, 규정, 지원 위원회 ...



4. 데이터 활용

(1) 발의안 데이터 활용

발의안의 사회동향 반영 정도 파악

■ 제안이유

아파트 등 공동주택은 「공동주택관리법」에 의해 체계적으로 관리되고 있는 반면, 주상복합·상가·오피스텔 등 집합건물은 공적 관리기준이 미비하여 불투명하게 운영되고 있고 최근 주상복합·오피스텔 등 집합건물이 증가하고 규모가 대형화됨에 따라 거주민 간 혹은 당사자 간 다양한 형태의 분쟁과 갈등이 발생하고 있음.

특히 집합건물법은 「공동주택관리법」과 달리 관리비 정보공개 및 지자체의 감독 등 관리업무에 대한 세부사항에 대한 규정이 미흡하여 집합상가에서 관리비 관련 민원이 자주 발생하고 있음.

또한, 집합건물과 관련한 분쟁을 해결하기 위해 각 시·도에 '분쟁조정위원회'를 설치해 관리비의 사용내역과 관련된 분쟁을 조정하도록 하고 있지만, 조정의 강제력이 없어 실효성이 없는 실정임.

⋮



발의안명	topic
'집합건물의 소유 및 관리에 관한 법률 일부개정법률안'	['관리', '주택', '건물', '복합', '집합', '오피스텔', '주상', '아파트', '공동', '발생']
'지방세법 일부개정법률안'	['재산세', '주택', '가격', '부동산', '공정', '시장', '비율', '가액', '예측', '부과']
'상장회사법안'	['회사', '주주', '상장', '구조', '법률', '지배', '시장', '법령', '상법', '금융']

⋮



4. 데이터 활용

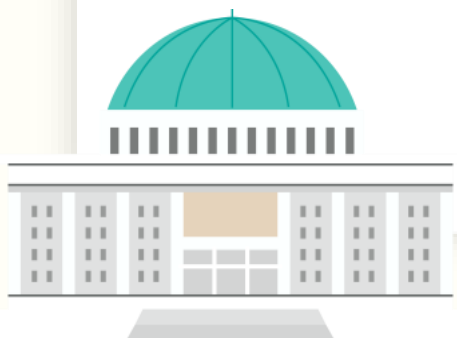
(2) 뉴스 데이터 활용 | 입법의 목표란?

입법의 목표

“입법의 목표는 국민의 입법역사와 동시대를 살아가는 사람들의 사회적 가치를 적절히 조화하여 민주적 정당성이 확보된 실효성·효율성이 있는 사회적 공익실현이 가능한 질적 입법의 달성에 있는 것”



고인석, “입법과정의 개선을 통한 ‘최적화 법률’의 달성방안 연구, 법이론실무연구 제 10권 제4호, 2022.11.

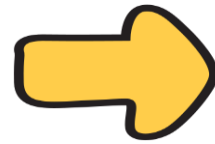


4. 데이터 활용

(2) 뉴스 데이터 활용 | 뉴스데이터 활용의 목적

국민의
입법 의사

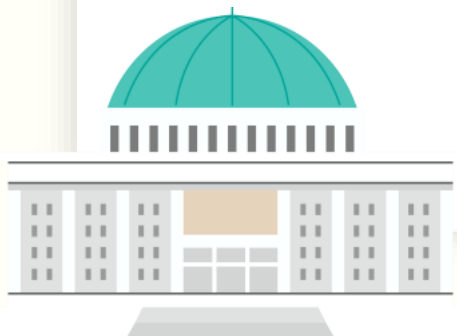
사회적 가치



뉴스데이터를 통해서
입법 목표의 구성요소인
국민의 입법 의사와 사회적 가치를 파악



뉴스데이터와 발의안과의 비교를 통해
발의안이 **민주적 정당성**을
가지는지 확인



4. 데이터 활용

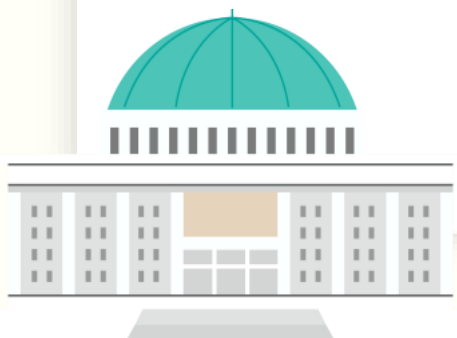
(2) 뉴스 데이터 활용 | 데이터 수집

■ 빅카인즈 확인

기간, 언론사, 뉴스 대주제, 대주제 별 소주제로 데이터를 구분하는 것을 확인

⋮

빅카인즈 내에서 기사별 키워드(**특성추출**) 제공하는 것을 이용하기로 결정

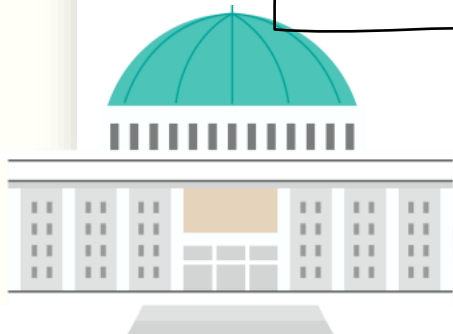
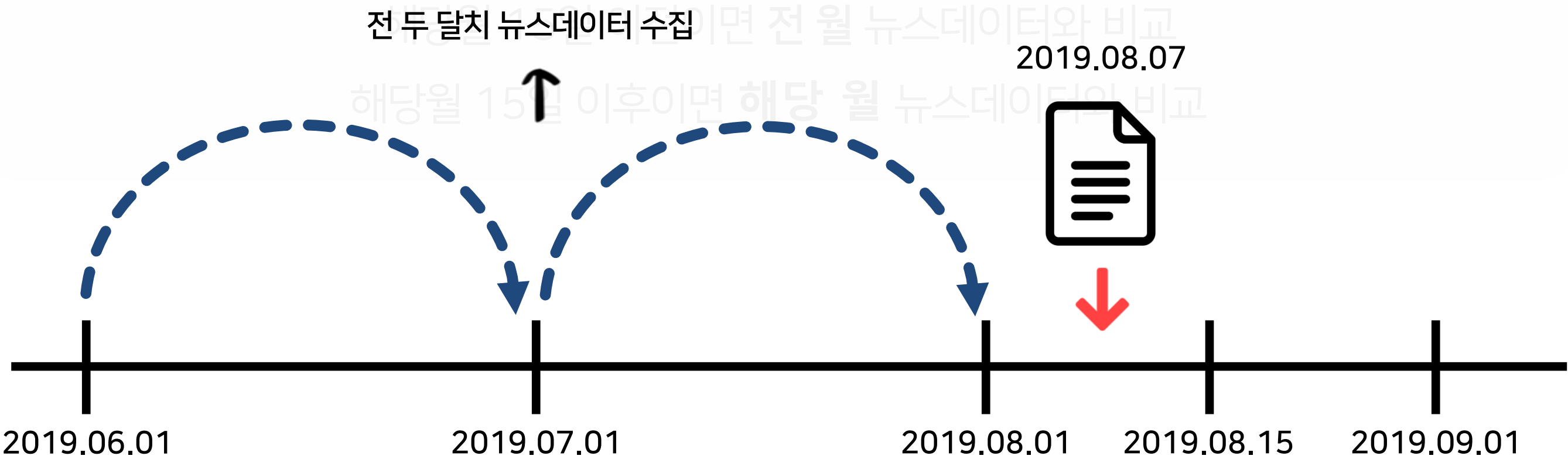


4. 데이터 활용

(2) 뉴스 데이터 활용 | 데이터 수집

| 수집범위 설정 | 기간

① 발의안이 해당 월 15일 이전이면 전 두 달치 뉴스데이터와 비교

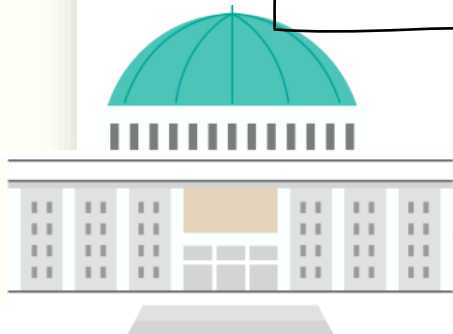
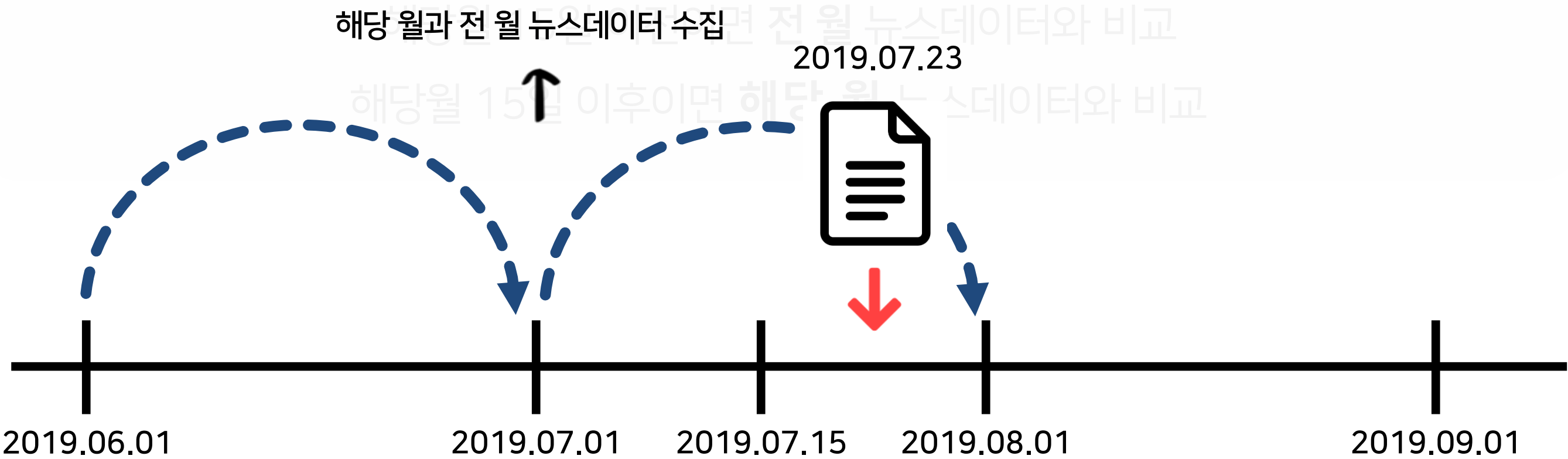


4. 데이터 활용

(2) 뉴스 데이터 활용 | 데이터 수집

| 수집범위 설정 | 기간

② 발의안이 해당 월 15일 이후이면 전 월과 해당 월 뉴스데이터와 비교



4. 데이터 활용

(2) 뉴스 데이터 활용 | 데이터 수집

■ 수집범위 설정 | 언론사

경향신문

국민일보

동아일보

매일경제

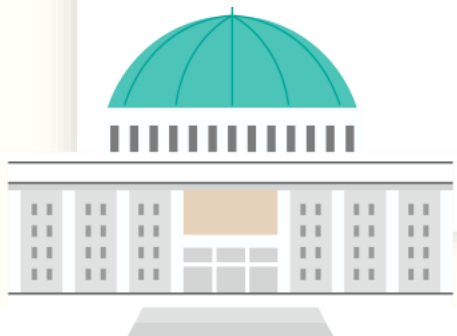
문화일보

중앙일보

한겨레

⋮

정치성향, 주요분야 등을 고려한 주요 언론사 7개로 필터링하여 뉴스데이터 수집



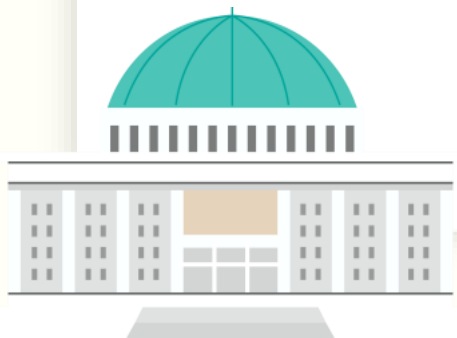
4. 데이터 활용

(2) 뉴스 데이터 활용 | 데이터 수집

최종 데이터 수집

시점, 언론사, 주제를 고려한 최종 뉴스 키워드 데이터 수집

정치>행정_자치	<행정_자치19_4>					
행정_자치19_4	일자	언론사	통합분류1	통합분류2	...	특성추출(상위 50개)
행정_자치19_5	20190401	중앙일보	정치 > 행정자치	NULL	...	후보자, 국토부, 행정
행정_자치19_6	20190401	동아일보	정치 > 행정자치	정치 > 국회	...	탄력근무제, 입법, 국회
⋮	20190401	한겨레	사회 > 노동복지	정치 > 행정자치	...	노동이사제, 공공
행정_자치20_5	⋮	⋮	⋮	⋮	...	⋮
	20190430	매일경제	정치 > 행정자치	정치 > 청와대	...	본회의, 공수처법



4. 데이터 활용

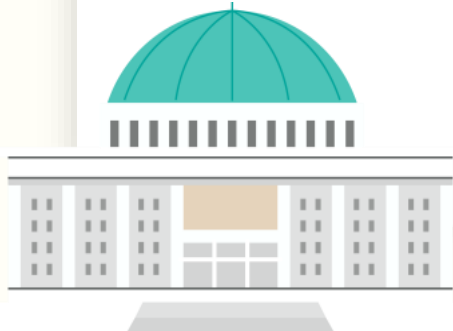
(2) 뉴스 데이터 활용 | 데이터 전처리

소주제 전처리

통합분류1, 통합분류2, 통합분류3 중에
하나만 소분류에 해당해도 해당 소분류로 추출됨
→ 통합분류1이 해당 소주제인 것만 필터링

행정_자치19_4

일자	언론사	통합분류1	통합분류2	...	특성추출(상위 50개)
20190401	중앙일보	정치 > 행정자치	NULL	...	후보자, 국토부, 행정
20190401	동아일보	정치 > 국회	정치 > 행정자치	...	탄력근무제, 입법, 국회
20190401	한겨레	사회 > 노동복지	정치 > 행정자치	...	노동이사제, 공공
⋮	⋮	⋮	⋮	...	⋮
20190430	매일경제	정치 > 행정자치	정치 > 청와대	...	본회의, 공수처법



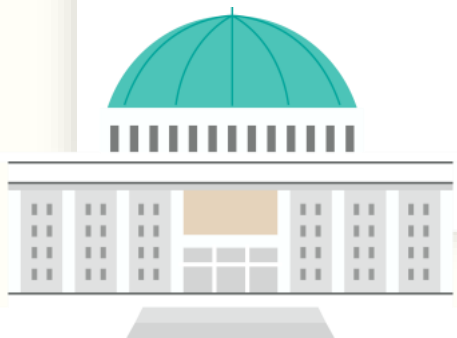
4. 데이터 활용

(2) 뉴스 데이터 활용 | 데이터 전처리

소주제 별 키워드 전처리

월별 소주제에 불필요한 키워드가 많음을 확인
→ ChatGPT API를 이용하여 상위 10개 키워드 추출

소주제_날짜	불필요 키워드 예시
행정_자치19_4	'엑소', '엉덩이춤', 'id', 'party', '복씨', '500조', '홈페이지', '윤씨', '만큼' ...



4. 데이터 활용

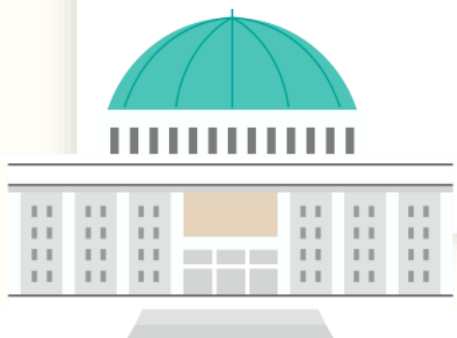
(2) 뉴스 데이터 활용 | 데이터 전처리

■ 소주제 별 키워드 전처리

너는 지금부터 법안과 사회 이슈를 비교하는 일을 맡은 전문가야.
지금부터 너가 맡을 일은 주어진 단어들 중 사람 이름과 정당명을 제외하고,
아래 키워드들에 대해, 처음에 제시된 단어와 관련된 사회이슈를 가장 잘 나타내는
중요한 키워드를 10개로 추려주는 거야.
정당명에는 '민주당', '자유한국당', ...과 같은 것이 있어.
이름에는 '김정은', '나경원', '유시민', '문재인', '황교안', ... 이 있고,
추려주는 키워드단어에서 제외해줘



이름, 정당명 제외하고 사회이슈를 잘 나타내는 상위 10개 키워드 추출



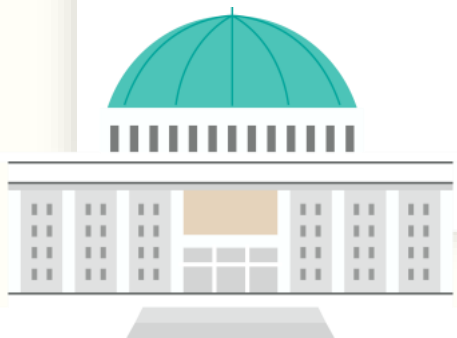
4. 데이터 활용

(2) 뉴스 데이터 활용 | 데이터 전처리

소주제 별 키워드 전처리

상위 10개 핵심 키워드 추출 후 확인
→ 소주제에 맞는 핵심 키워드가 잘 추출되었다고 판단

소주제_날짜	상위10개 키워드
국회정당19_6	'국회의원', '공공주택지구', '패스트트랙', '국회의원', '국민소환제', '한국당', '윤석열', '공공주택지구', '청와대', 국민소환제



4. 데이터 활용

(3) 발의안의 사회동향 반영 정도

발의 시점을 고려하여

발의안 키워드와 소주제 뉴스 키워드를 매칭해주는 것이 목표!

⋮

발의안명	제안이유
발의안1	제안이유1
발의안2	제안이유2
발의안3	제안이유3
⋮	⋮



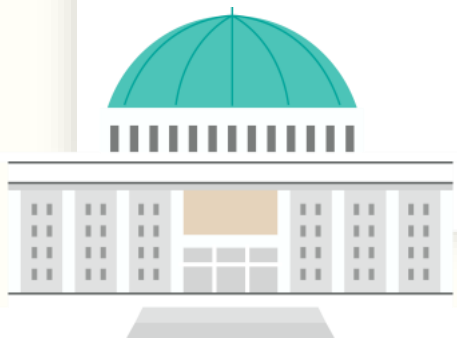
뉴스 대주제
정치
경제
사회
문화
IT과학



뉴스 소주제
정치>행정_자치
정치>북한
⋮
IT과학>IT과학일반
IT과학>보안



(소주제별)	
날짜	주요 키워드
19년 4월	키워드1
19년 5월	키워드2
⋮	⋮
20년 5월	키워드14



4. 데이터 활용

(3) 발의안의 사회동향 반영 정도

발의 시점을 고려하여

발의안 키워드와 소주제 뉴스 키워드를 매칭해주는 것이 목표!

⋮

발의안명	제안이유
발의안1	제안이유1
발의안2	제안이유2
발의안3	제안이유3
⋮	⋮

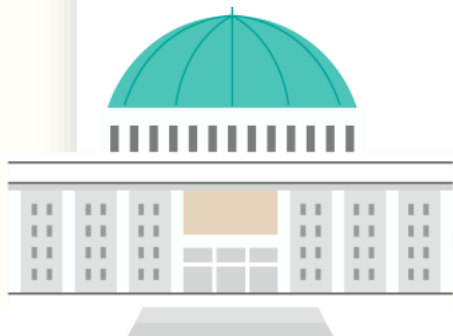


뉴스 대주제
정치
경제
사회
문화
IT과학



뉴스 소주제	(소주제별)	날짜	주요 키워드
정치>행정_자치		19년 5월	키워드1
정치>북한		19년 5월	키워드2
⋮		⋮	⋮
IT과학>IT과학일반		20년 5월	키워드14
IT과학>보안			

전이학습을 이용해
KoBERT 다중분류 시도



4. 데이터 활용



(3) 발의안의 사회동향 반영 정도

발의 시점을 고려하여

발의안을 바로 뉴스 소주제로 분류하지 않고,

두 번에 나눠서 분류하는 이유

발의안 키워드와 소주제 뉴스 키워드를 매칭해주는 것이 목표!

사용하기로 한 소주제는 총 48개

발의안명	주요 키워드
발의안1	키워드1
발의안2	키워드2
발의안3	키워드3
⋮	⋮

뉴스 대주제
정치

뉴스 소주제
정치>행정_자치

(소주제별)	
날짜	주요 키워드
19년 4월	키워드1
19년 5월	키워드2
⋮	⋮
20년 5월	키워드14

48개의 소주제에 대해 바로 분류하게 되면,
분류해야 할 클래스가 많아 계산 비용이 커지고
성능도 좋지 않을 것이라 판단했기 때문!



4. 데이터 활용

(3) 발의안의 사회동향 반영 정도 | 전이학습

사전 학습된 모델 선택

모델의 상위층 동결

이미 학습된 특징들이 손상되는 것을 방지!

새로운 출력층 추가

새로운 데이터셋으로 훈련

새로운 층 + 일부 기존 층으로 학습

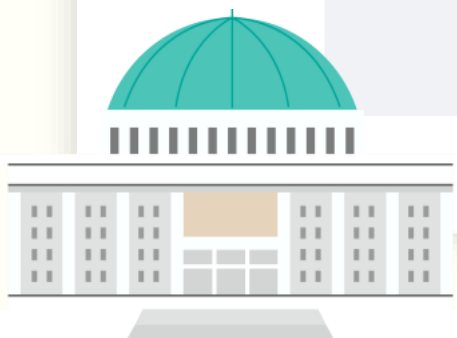
파인튜닝

우리의 목표

발의안을 뉴스 대주제로 분류하는 것

⋮

모든 발의안 데이터를 수집하여
학습시키는 데에는 한계가 있으므로,
전이학습을 써서 분류해보자!



4. 데이터 활용

(3) 발의안의 사회동향 반영 정도 | 전이학습

사전 학습된 모델 선택



사전 학습된 모델인 **KoBERT** 선택

모델의 상위층 동결

이미 학습된 특징들이 손상되는 것을 방지!

새로운 출력층 추가

새로운 데이터셋으로 훈련

새로운 층 + 일부 기존 층으로 학습



19년 1월 ~ 19년 5월
발의안의 대주제를 직접 라벨링하여
새로운 학습데이터셋 생성

파인튜닝



4. 데이터 활용

(3) 발의안의 사회동향 반영 정도 | 전이학습

미세 조정 학습(Fine tuning)

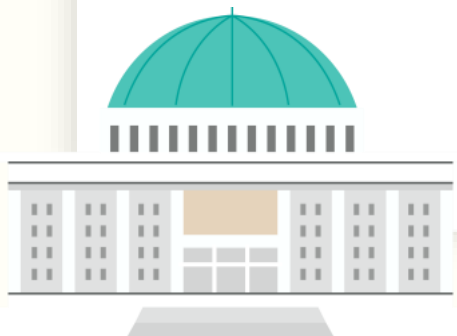
딥 러닝에서 사전 훈련된 모델의 가중치가
새로운 데이터에 대해 훈련되는 전이학습에 대한 접근 방식



수동으로 라벨을 분류한 발의안 데이터 학습



국회 도메인과 분류 작업에 맞게 파인 튜닝



4. 데이터 활용

(3) 발의안의 사회동향 반영 정도 | 전이학습

■ 성능 평가 - 혼동 행렬 시각화

학습 데이터 내에서 다시 train test split 후
검증 데이터의 라벨을 제거한 뒤 예측

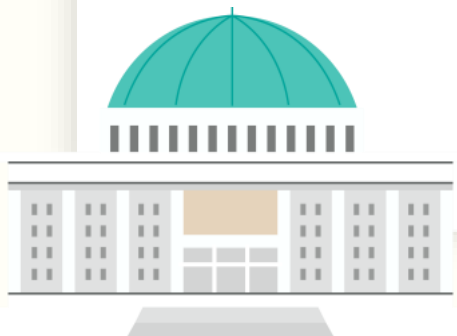
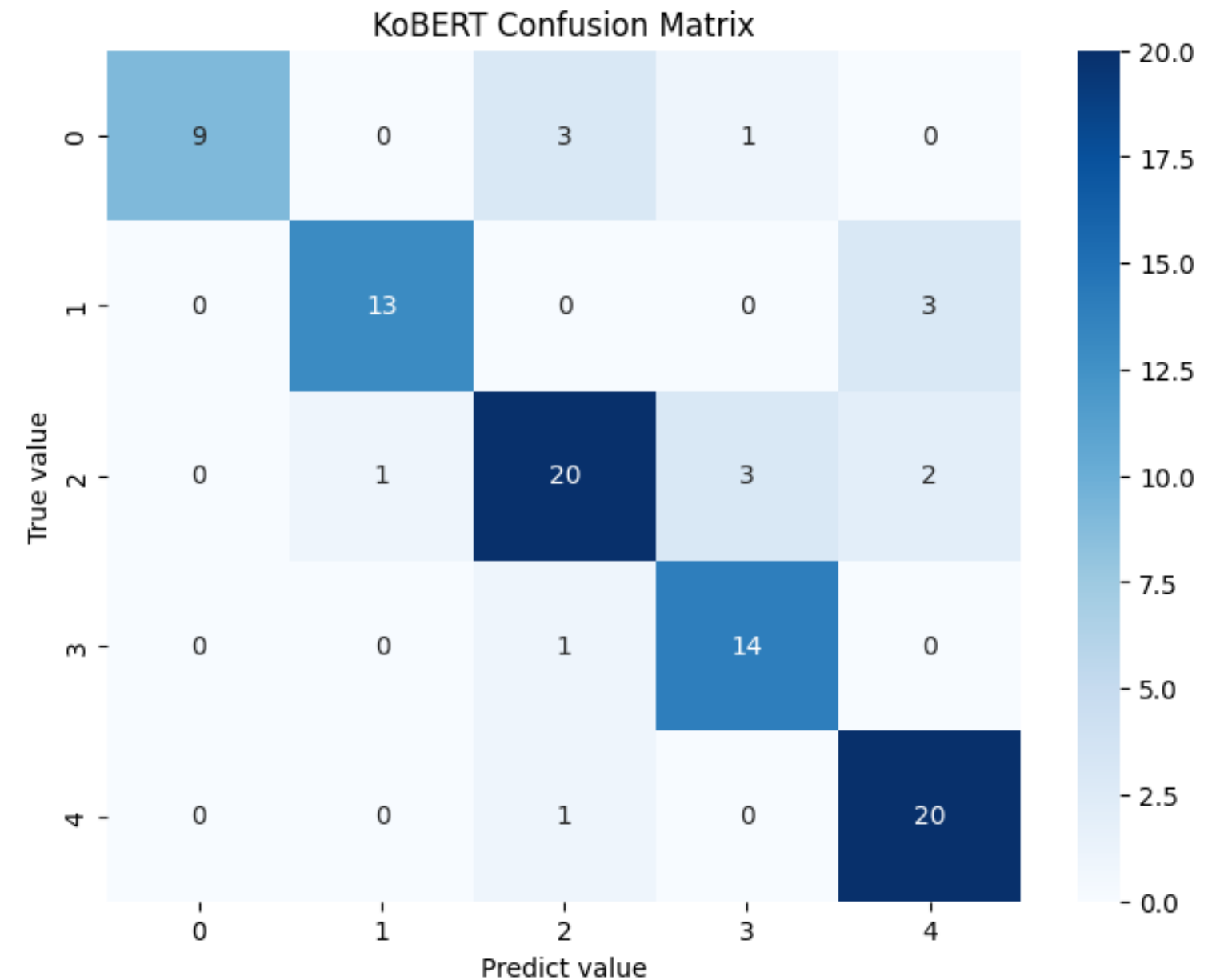
⋮

[정밀도 : 0.8612 재현율 : 0.8319]

→ 학습이 잘 되었음을 확인



라벨링되지 않은 새로운 데이터에 대해 예측 시도!



1. 발의안과 뉴스 소주제 매칭

(2) KPF-BERT

■ 학습 과정

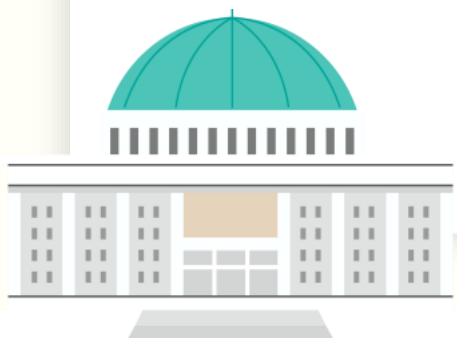
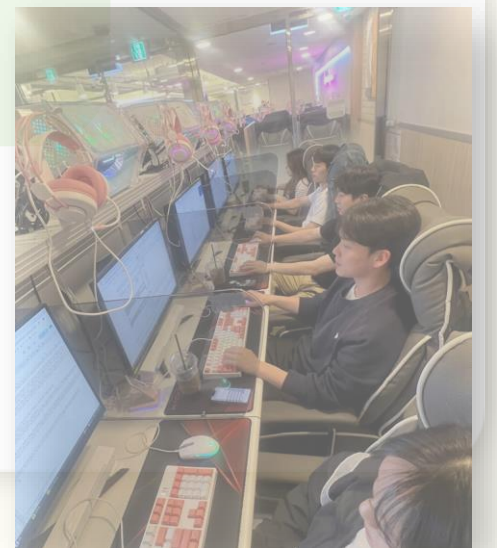
19대 발의안을 소관위별로 30개씩 직접 라벨링하여 학습 데이터셋 구성

1123개 발의안 손 라벨링



⋮

대분류별로 학습 데이터셋을 5개로 나누고,
`BertForSequenceClassification.from_pretrained("kpfbert-base")`
사용하여 데이터셋 각각 소주제 학습 진행!



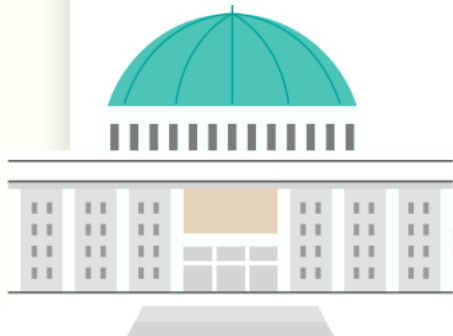
1. 발의안과 뉴스 소주제 매칭

(2) KPF-BERT

■ 학습 과정

KPF-BERT [정치] 학습 결과

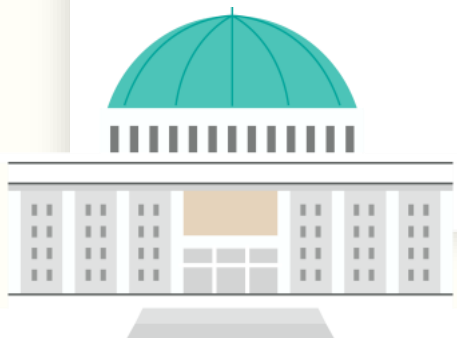
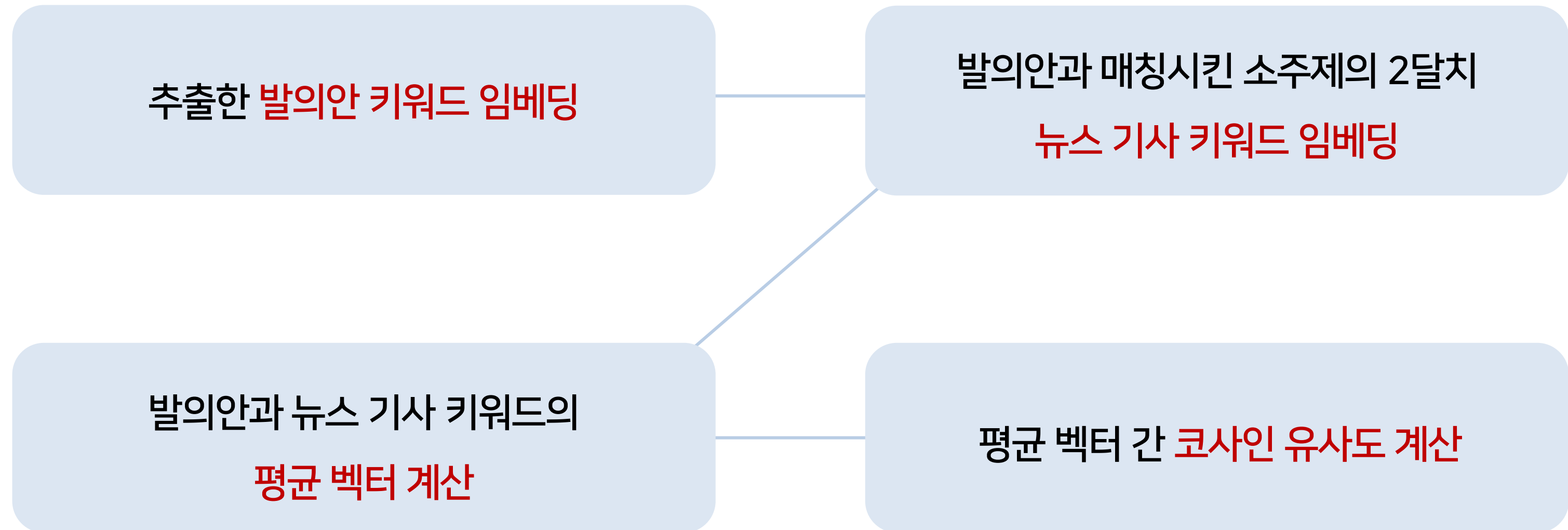
제안이유	분류한 소주제
지방자치단체의 재정은 지방자치단체의 장이 제반 여건을 고려하여 가장 효율적으로 운용할 수 있도록 해야 함에도 불구하고 현행 특별회계나 기금 등에 관한 은 해당 목적 재원이 충분히 조성되어 있는지 여부와 무관하게 재원의 융통을 제한함에 따라 ...	행정_자치
대한민국 정부는 대한민국과 아메리카합중국 간의 상호방위조약 제4조에 의한 시설과 구역 대한민국에서의 합중국군대의 지위에 관한 협정 제5조에도 불구하고 주한미군 주둔비용의 공평한 분담과 한미동맹의 발전을 위해 미국 정부와 1991년부터 대한민국과 아메리카합...	외교
⋮	⋮
현행법은 누구든지 선거에 관한 여론조사 결과를 공표보도하려는 때에는 여론조사와 관련 있는 자료일체를 해당 선거의 선거일 후 6개월까지 보관하도록 하고 있을 뿐 표본구성의 절차 입증을 위한 원자료RAW DATA의 제출의무가 배제되어 선거여론조사의 작위적 표본...	선거



2. 발의안의 사회동향 반영정도

(1) 발의안 - 뉴스키워드 유사도

유사도 도출 과정



2. 발의안의 사회동향 반영정도

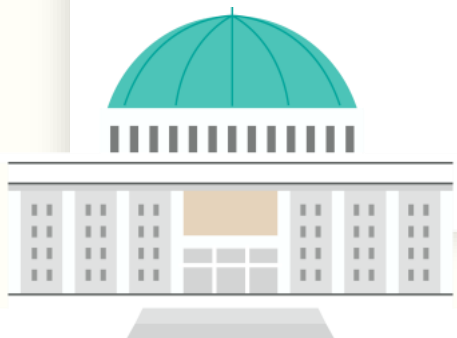
(1) 발의안 - 뉴스키워드 유사도

■ 문서 임베딩: Doc2Vec

Doc2Vec

문장 이상의 단위의 문서를 임베딩하는 자연어 처리 기법으로,
문서의 의미를 벡터 공간에서 나타내는 것을 목표로 함

키워드를 추출했던 발의안의 '제안이유'를 학습시켜
모델이 키워드가 추출된 **맥락**을 고려하게 함



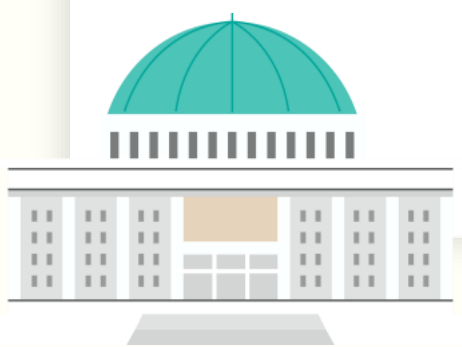
2. 발의안의 사회동향 반영정도

(1) 발의안 - 뉴스키워드 유사도

Doc2Vec 유사도 계산 결과

발의안 키워드	뉴스기사 키워드	유사도
대기 먼지 환경부 주입 물질 수도권 미세먼지 계획 환경 기본	미세먼지 국가기후환경회의 전기요금 환경회의 ... 폐기물 환경 농업용수 한국농어촌공사 오염원	0.61335266
피해자 범죄 성폭력 성희롱 징계 강조 권리 보장 반영 보호	여성 강남구 주거침입 서울 강간미수 사망자 ... 성폭행 성희롱 성범죄 성평등 성폭력	0.51590323
자녀 개정 법률안 반영 의결 민법 파양 신고 일부 가족	건설 빗물 노동자 반려인 감리사 공사장 ... 사법연감 외국인 대법원 변호인단 불법대부업 공정위	-0.20881182
검사 시험 대상 대표 신뢰도 시행 확보 분석 여부 주체	공개매각 바이오 조선업 중소기업 ... 과학기술 컴퓨터 공정거래 중공업 통신 대한상의 경쟁력	-0.15062407
⋮	⋮	⋮

키워드들 간 명확하고 직관적인 유사도 차이가 나타나는 것을 볼 수 있음!



2. 발의안의 사회동향 반영정도

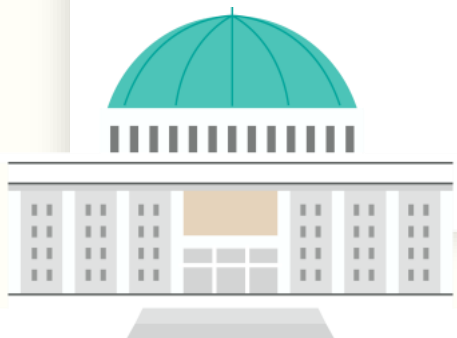
(2) 발의안 관련 논문 수

논문 수 크롤링

발의안 키워드 상위 3개를 성균관대학교 학술정보관 홈페이지에 검색

검색 결과 중 '학술 논문' 카테고리만 필터링

제안일로부터 6개월 이내에 발행된 논문으로 한 번 더 필터링!

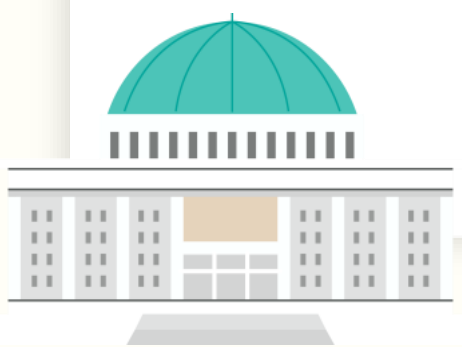


2. 발의안의 사회동향 반영정도

(2) 발의안 관련 논문 수

논문 수 크롤링

의안번호	법률안명	대표발의자	...	제안일	키워드	관련 논문 수
2024924	전기통신사업법 일부개정법률안	이원욱	...	2020-05-04	정보 통신 의무 통신망 유통 촬영 ...	53
2024923	정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안	이원욱	...	2020-05-04	해외 통신 서비스 정보 피해 접속 ...	22
2024856	정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안	한정애	...	2020-04-21	음란 판매 영상물 부호 피해자 화상 ...	0
2024830	성폭력범죄의 처벌 등에 관한 특례법 일부개정법률안	박광온	...	2020-03-31	복제물 징역 소지 이하 시청 천만 ...	0
2024095	정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안	박대출	...	2020-03-31	딥페이크 판별 지능 여부 이용자 해당 ...	2
						...



3. 최종 데이터셋 (1)

(2) 발의안 점수 산출

■ 비음수행렬분해 (NMF)

차원 축소와 특성 추출을 목적으로
주어진 비음수 행렬을 두 개의 비음수 행렬로 분해하는 비지도 학습의 일종

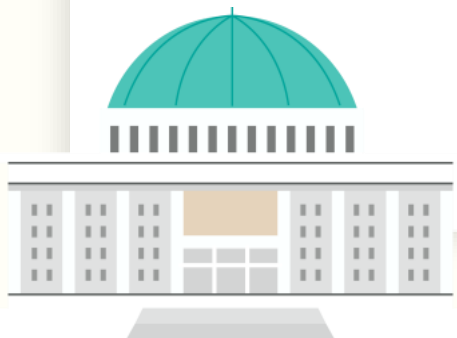
⋮

$$X = WH$$

X: 원본 데이터셋

W: 기저 행렬 또는 특징행렬 즉, 원본 데이터의 패턴이나 구성요소

H: 계수 행렬 즉, 각 패턴과 구성요소의 가중치

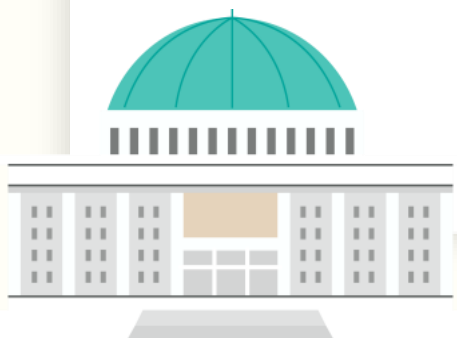
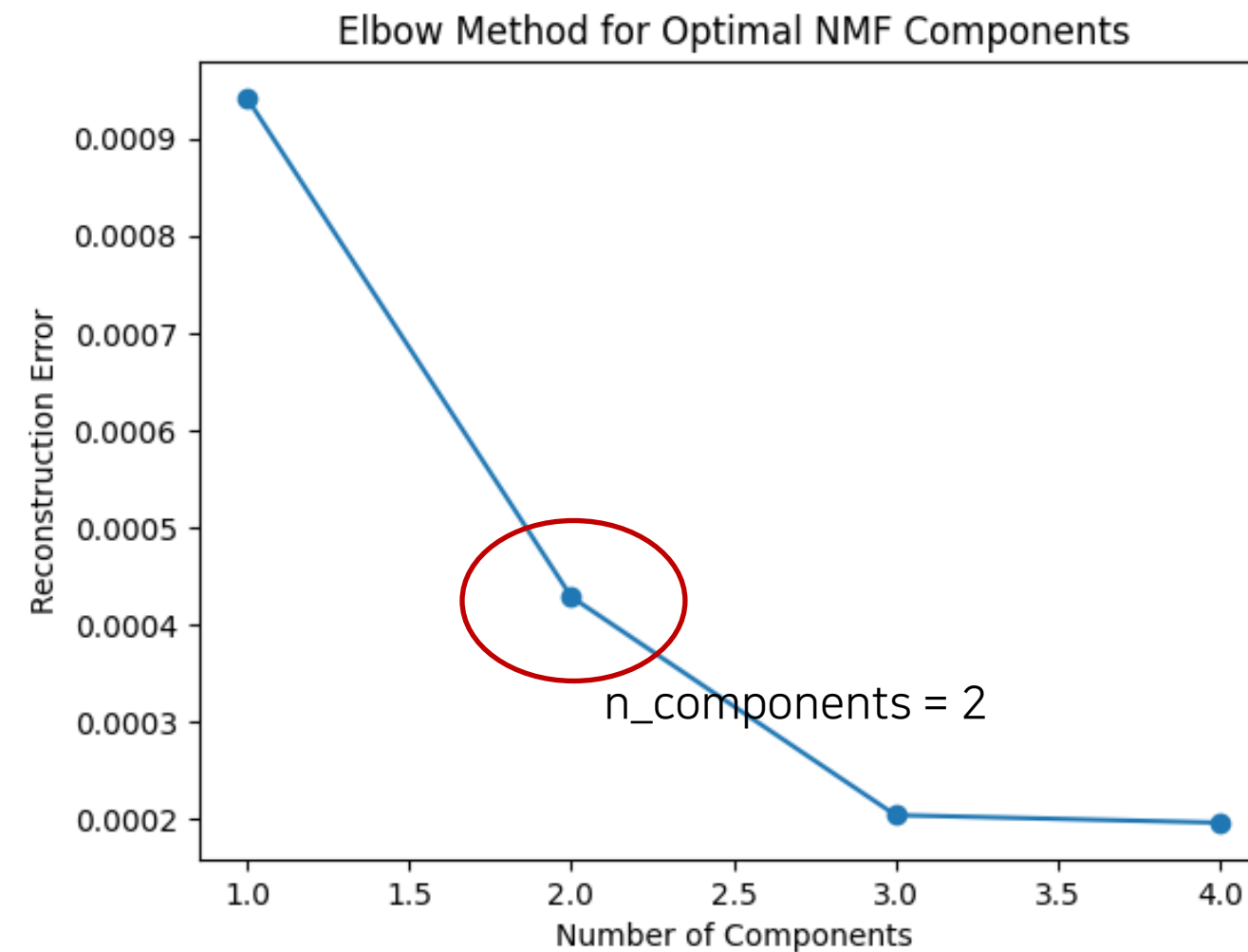


3. 최종 데이터셋 (1)

(2) 발의안 점수 산출

가중치 산출 과정 | ① 잠재 요인 개수 정하기

컴포넌트 수를
1부터 특성 개수만큼 반복하여
재구성 오류가 점점 줄어드는 기점을
컴포넌트 수로 결정



3. 최종 데이터셋 (1)

(2) 발의안 점수 산출

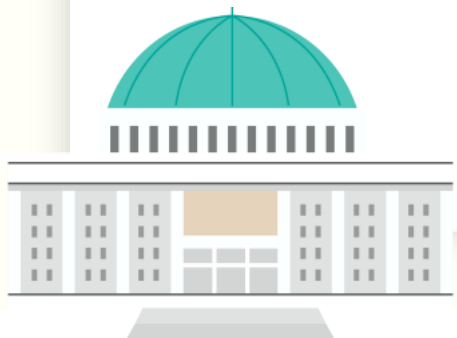
가중치 산출 과정 | ③ 비음수 행렬분해

계수 행렬 H 

	특성 1			
	0.07837227	0	0.461591	1.03436486
특성 별 가중치 ←	0.91804247	0.25010574	0.76119412	0
		⋮		

후
기

Ex) 특성 1 변수 가중치 = 특성 1 열의 합 / 전체 특성 열의 합



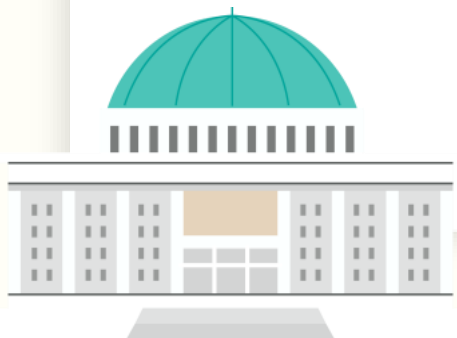
3. 최종 데이터셋 (1)

(2) 발의안 점수 산출

가중치 산출 과정 | ④ 가중치 산출

$$\text{발의안 점수} = 27 * \text{<신설>조항개수} + 7 * \text{<삭제>조항개수} + 36 * \text{뉴스유사도} + 30 * \text{관련논문수}$$

위와 같이 계산된 변수별 가중치 [28, 7, 35, 30]를
0-1 사이로 스케일링된 원본 데이터에 곱하여 100점 만점으로 평가



3. 최종 데이터셋 (1)

(2) 발의안 점수 산출

가중치 산출 과정 | ④ 가중치 산출

$$\text{발의안 점수} = 27 * \text{<신설>조항개수} + 7 * \text{<삭제>조항개수} + 36 * \text{뉴스유사도} + 30 * \text{관련논문수}$$

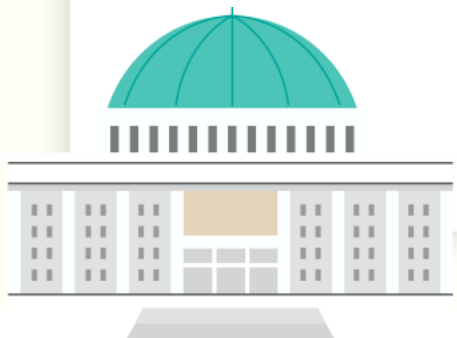
<삭제> 조항 개수의 중요도가 가장 낮고

뉴스 유사도의 중요도가 가장 낮으며

<신설> 조항 개수와 관련논문수가 그 다음으로 비등하게 높음



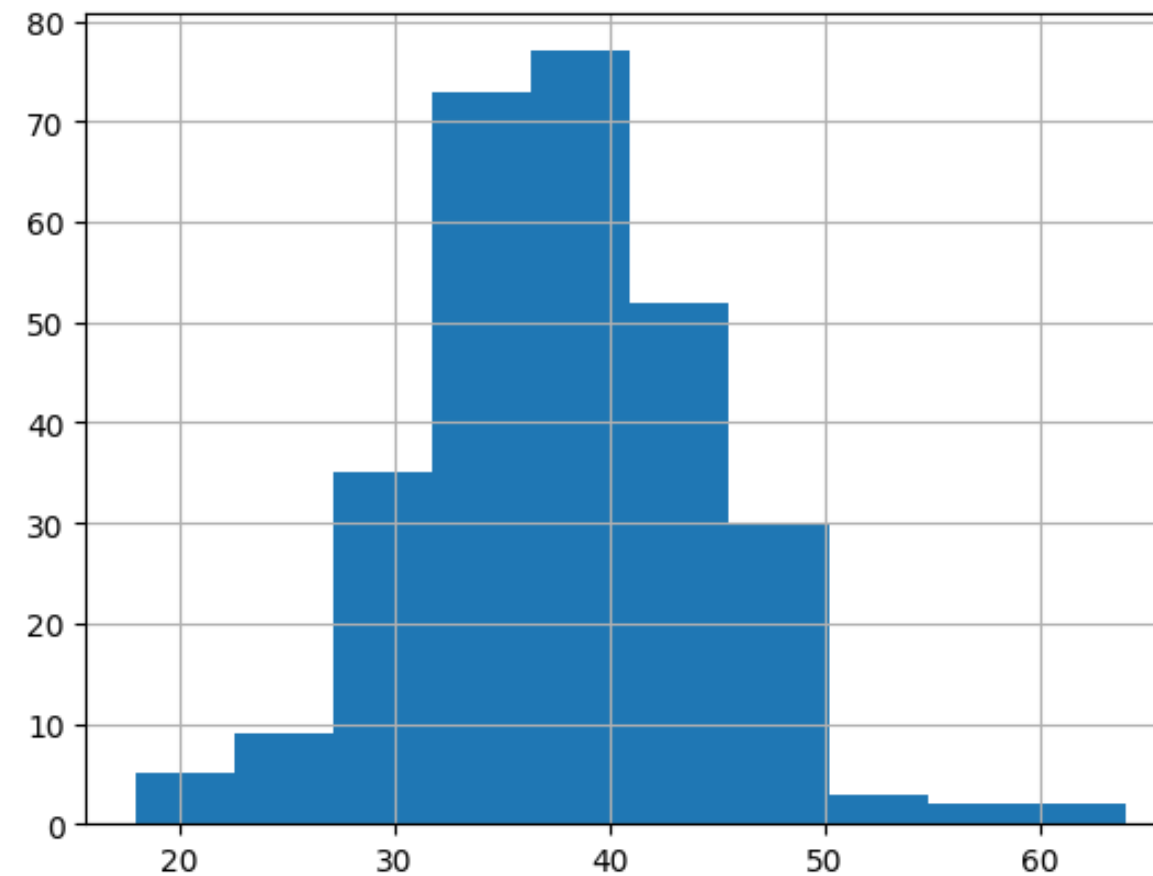
직관에 부합하므로 채택 !



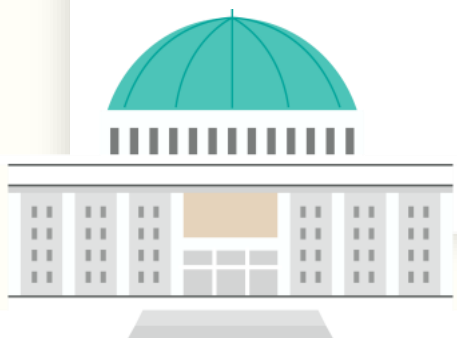
3. 최종 데이터셋 (1)

(2) 발의안 점수 산출

최종 발의안 점수 분포



정규화를 시도했음에도 전반적으로 낮은 값에 머무르고 있음
국회의원들이 질적발의를 하지 않았음을 반증하는 결과!



3. 최종 데이터셋 (2)

국회의원 데이터셋 구성

② 발의안 데이터에서 국회의원별 발의안 정보 병합

국회의원명	신설법안개수	대표발의비율	가결수	발의안점수	발의안평균분산도
강길부	1	0.03773...	1	9	
강병원	0	0.11564...	5	4.95	
강석진	2	0.15540...	15	4.65	
⋮	⋮	⋮	⋮	⋮	
황희	1	0.04255...	1	0	



3. 최종 데이터셋 (2)

국회의원 데이터셋 구성

③ 발의안점수

$$\text{발의안점수} = \frac{\text{대표발의안별 점수}}{\text{대표발의안 수}}$$

⋮

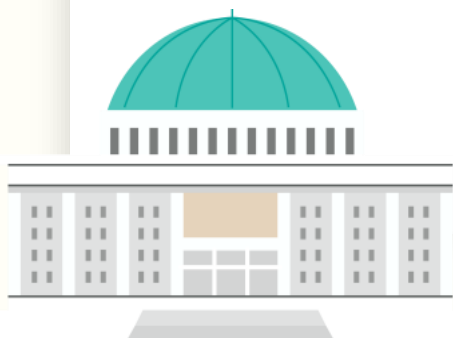
대표 발의안별 점수

= 발의안 데이터셋에서 구한 발의안 점수를
해당 국회의원에 대해 합산하여 구한 값



국회의원 데이터

국회의원명	발의안점수
강길부	9
강병원	4.95
강석진	4.65
⋮	⋮
황희	0



3. 최종 데이터셋 (2)

국회의원 데이터셋 구성

④ 발의안평균분산도 변수 생성

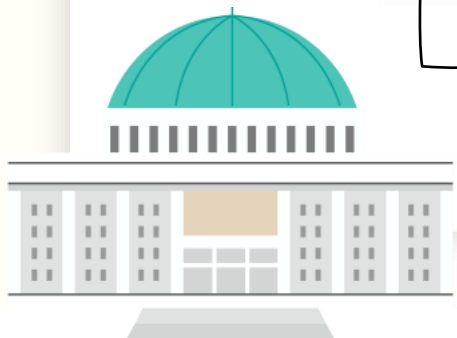
Inertia값

국회의원명	Inertia값
강길부	42.515...
⋮	⋮
이양수	59.349...
⋮	⋮
황희	38.432...

Inertia value: 각 클러스터의 중심에서 클러스터에 할당된
데이터 포인트간의 거리를 합산한 것
즉, 군집이 얼마나 잘 응집되었는지 보여주는 지표 !

K-means Clustering

Inertia 값 반환



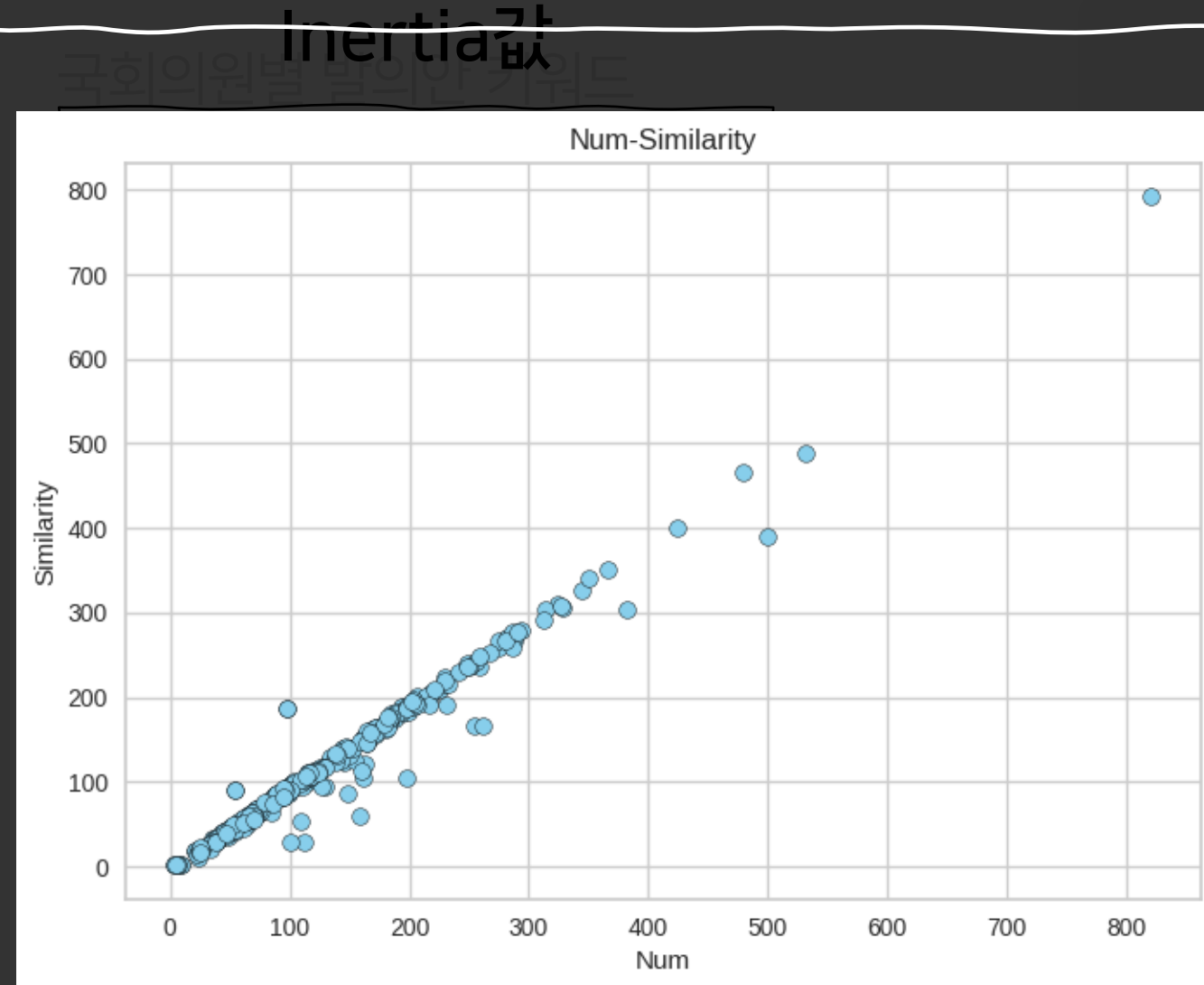
3. 최종 데이터셋 (2)



국회의원 데이터셋 구성

④ 발의안평균분산도 변수 하지만 Inertia 값을 바로 사용하면 ..

거리의 단순 합이기 때문 !



발의안 수와 Inertia값이
명확한 선형관계가 있는 것을 확인

데이터 포인트간의 거리를 합산한 것

즉, 군집이 얼마나 잘 응집되었는지 보여주는 지표 !

발의안수의 영향을 없애기 위해

Inertia값을 발의안 수로 나눔

$$\text{발의안평균분산도} = \frac{\text{Inertia값}}{\text{발의안수}}$$

Inertia 값 반환

3. 최종 데이터셋 (2)

국회의원 데이터셋 구성

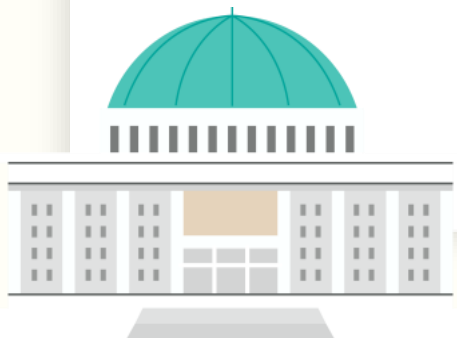
④ 발의안평균분산도 변수 생성 | 발의안평균분산도값 검증

하위 9등인

이양수 국회의원의 발의안 키워드가 매우 유사함을 알 수 있음

이양수 국회의원 발의안 키워드

'피의자 신상 공개 전자 정보 정면 흉악 기록 컬러 전신',
'한자어 국민 개정 확장 가능 축약 한글 정도 접근 신뢰',
'한자어 국민 개정 확장 가능 축약 한글 정도 접근 신뢰',
:
"국민 한자어 이해 기여 신뢰 한글 법률 접근 용어 축약',
'한자어 국민 개정 확장 가능 축약 한글 정도 접근 신뢰',
'한자어 국민 개정 확장 가능 축약 한글 정도 접근 신뢰',
'한자어 국민 개정 확장 가능 축약 한글 정도 접근 신뢰'

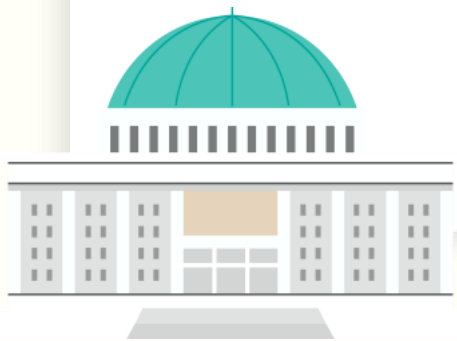


3. 최종 데이터셋 (2)

국회의원 데이터셋 구성

⑤ 최종 국회의원 데이터셋 완성

국회의원명	신설법안개수	대표발의비율	가결수	발의안점수	발의안평균분산도
강길부	1	0.03773...	1	9	0.9164490...
강병원	0	0.11564...	5	4.95	0.9718003...
강석진	2	0.15540...	15	4.65	0.6527361...
⋮	⋮	⋮	⋮	⋮	⋮
황희	1	0.04255...	1	0	0.7778446...



4. 입법활동 점수 산출

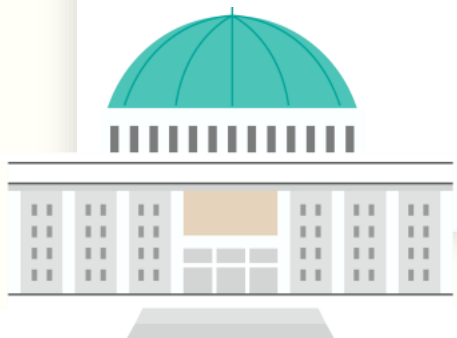
(2) 점수 산출 방법 | NMF

가중치 산출 과정

$$\begin{aligned} \text{최종 점수} = & 18 * \text{대표발의비율} + 15 * \text{가결수} + 22 * \text{발의안평균분산도} \\ & + 30 * \text{신설법안개수} + 15 * \text{발의안점수} \end{aligned}$$

⋮

질적 발의 평가 요소로 분류할 수 있는
신설법안개수와 발의안평균분산도의 중요도가 높고
양적 발의 평가 요소들의 중요도가 낮음



4. 입법활동 점수 산출

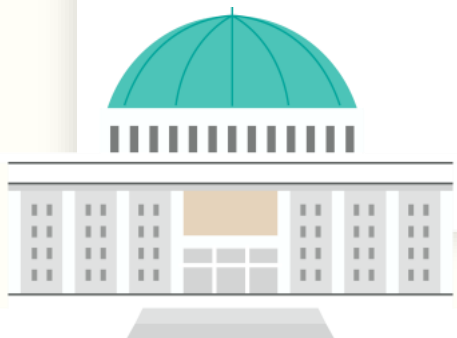
(2) 점수 산출 방법 | NMF

가중치 산출 과정

$$\begin{aligned} \text{최종 점수} = & 18 * \text{대표발의비율} + 15 * \text{가결수} + 22 * \text{발의안평균분산도} \\ & + 30 * \text{신설법안개수} + 15 * \text{발의안점수} \end{aligned}$$

⋮

질적 발의 평가 요소로 분류할 수 있는
신설법안과 발의안평균분산도의 중요도가 높고
양적 발의 평가 요소들의 중요도가 낮음

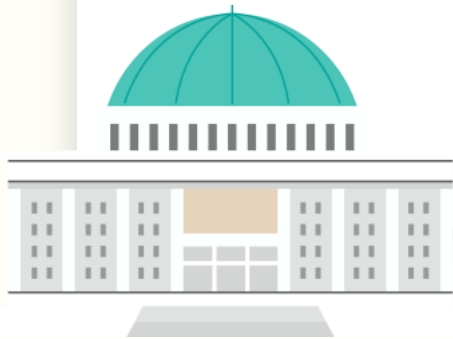


4. 입법활동 점수 산출

(4) 국회의원별 최종 입법활동 점수

5개의 변수의 총합이 최종점수

국회의원명	대표발의점수	가결점수	발의안분산점수	신설법안점수	발의안점수	최종점수
서영교	14.94	10.50	21.12	28.2	7.20	81.96
조응천	16.02	11.25	19.80	24.6	6.60	78.27
김정우	14.40	14.70	15.62	24.6	6.90	76.22
...
김현미	0.00	0.00	0.00	0.0	5.25	5.25
박영선	0.00	0.00	0.00	0.0	5.10	5.10



4. 입법활동 점수 산출

(4) 국회의원별 최종 입법활동 점수

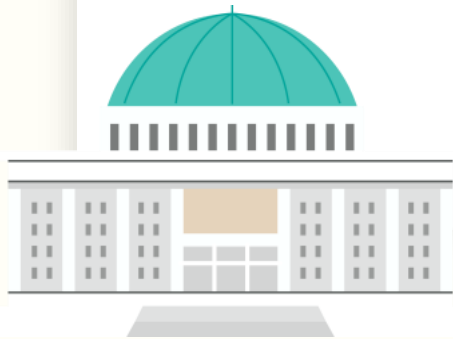
5개의 변수의 총합이 최종점수

국회의원명	대표발의점수	가결점수	발의안분산점수	신설법안점수	발의안점수	최종점수
서영교	14.94	10.50	21.12	28.2	7.20	81.96
조응천	16.02	11.25	19.80	24.6	6.60	78.27
김정우	14.40	14.70	15.62	24.6	6.90	76.22
...
김현미	0.00	0.00	0.00	0.0	5.25	5.25
박영선	0.00	0.00	0.00	0.0	5.10	5.10

<고교무상교육법>으로

'2019년도 입법 및 정책개발' 우수국회의원에 선정된 것을 확인

산출한 입법활동 최종 점수가 타당하다고 판단!



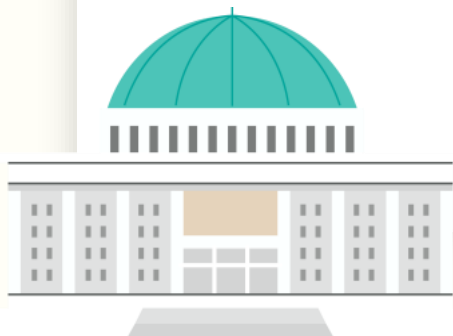
5. 의정활동 평가모델로의 확장

(4) 회의록 분석

① 질문 횟수

ex. 「외교통일위원회 회의록 中」

발언자	발언 내용
이석현 위원	그 부분에 대해서 지금 이상희 보건복지부 과장님 생각이 그런 건가요 또는 외교부와 의논도 좀 해 본 게 있나요?
진술인 이상희	지금 현재 사할린한인 영주귀국사업을 할 때도 대상자 선정과 그다음에 귀국자 선정하는 것 자체를 다 외교부에서 하고 있습니다. 물론 실질적인 것은 또 대한적 ‘?’가 발언 내용에 포함되는 경우 하는 것은 지금 현재도 외교부에서 지원을 해 줘야 가능한 사업입니다.
⋮	해당 발언자의 질문 횟수가 + 1 ⋮
이석현 위원	그런데 고려인들을 위해서, 고려인동포 합법적 체류자격 취득 및 정착 지원을 위한 특별법 이런 게 있어요. 그런 것 보면 외국에 체류 중인 그런 피해자들도 어떤 부류에 따라서, 어떤 특수성에 따라서 그런 법을 또 만들 수도 있는 것 아닌가, 고려인의 경우를 보면 사할린법도 만들 만한 것 아닌가. 그 점을 어떻게 생각해요? 말하자면 행안부에서 볼 때는 구태여 그런 법을 별도로 안 만들어도 지원할 수 있다 이런 입장인 것 같네요. 그런데 법을 만들어도 그런 고려인법의 전례를 보면 또 그럴 수도 있겠다는 생각이 들기도 하는데 어떻게 생각하세요?
진술인 박천수	행안부는 별도의 법을 만드는 것에 찬성하는 입장입니다.



5. 의정활동 평가모델로의 확장

(4) 회의록 분석

② 중간발언 (말끝음) 횟수

ex. 「환경노동위원회 회의록 中」

발언자	발언 내용
고용노동부차관 임서정	아닙니다. 다만 위원님께서 말씀하신 대로 개별 상품을, 몇백 개씩 되는 개별 상품에 대해서 하나하나씩 이게 어떤 정도의 이윤을 남기고 어느 정도.....
신보라 위원	아니, 그러니까 기금 투자의 기본적인 원칙이나 매뉴얼이 없으신 거예요?
:	'...' 말줄임표가 문장 끝에 들어가는 경우
신보라 위원	다음 발언자의 중간 발언 횟수가 + 1는 것 아닙니까? 이렇게 원금 손실을 이 큰데요? 그리고 아시겠지만 지금 이 DLS에 투자한 개인 투자자들도 많지만 개인 투자자들은 그 손실을 결국 개인이 감수합니다. 그런데 이 투자에 대한 손실은 지금 정부가 어떤 책임을 질 수 있는 거지요?
고용노동부차관 임서정	저희들이 이제 한국투자.....
신보라 위원	책임 못 지는 것 아닙니까?



5. 의정활동 평가모델로의 확장

(4) 회의록 분석

③ 감성분석



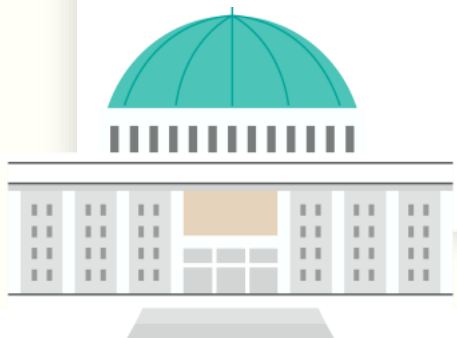
서울특별시의회 교육행정감사 회의록 감성분석 데이터



KoBERT 사전학습 모델에 전이학습 시도



긍정, 부정, 중립 세 가지 감정으로 예측



5. 의정활동 평가모델로의 확장

(4) 회의록 분석

③ 감성분석 | 분류 결과 예시

그러니까 그것에 대해서 조사를 하라는 거잖아요 왜 사실이 아니라고 얘기를 해요
교육위원회끼리 무슨 경고예요 경고는

...

부정

시간 남았습니다 시간 주신 것만큼 하겠습니다
지금 전희경 위원님이 발언 준비하고 있습니다

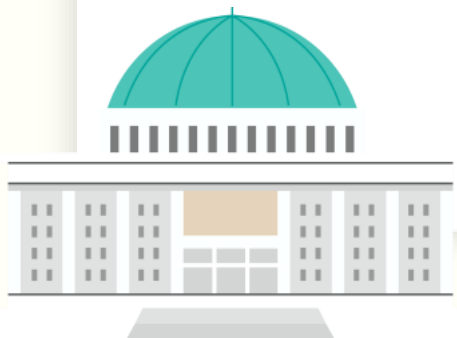
...

중립

재정 당국과 또 지자체와 교육청까지 사실 교육감님들이 굉장히 노력을 많이 해 주셨고 어려웠지만
고교 무상교육을 실시하는 게 국가 중에 마지막 우리나라만 남아서 늦었다는 판단들이 있었고요...

...


긍정



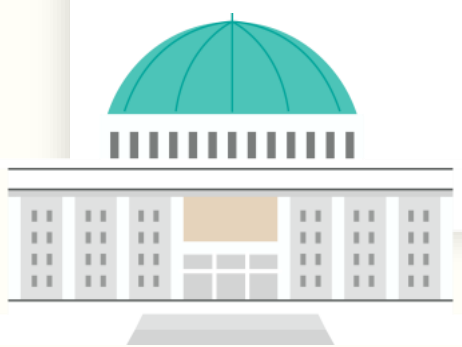
5. 의정활동 평가모델로의 확장

(4) 회의록 분석

③ 감성분석 | 통합 결과

 $\frac{\text{긍정 발언 수}}{\text{전체 발언 수}}$

발언자	긍정	부정	중립	감성지수
강길부	47	5	67	0.39
강병원	24	23	34	0.3
강석진	24	42	39	0.23
...
황주홍	87	30	347	0.19
황희	2	2	3	0.29



5. 의정활동 평가모델로의 확장

(5) 최종 회의록 데이터셋

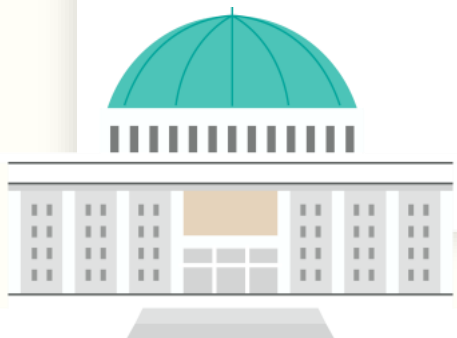
■ NMF를 통한 가중치 산출

비음수여야 하므로 역수를 취함!
(0인 경우 아예 발언이 없는 경우가 많았으므로 그대로 0으로 처리)

$$\text{의정활동 점수} = 29 * \text{중간발언_역수} + 23 * \text{질문} + 20 * \text{발언횟수} + 28 * \text{감정지수}$$

⋮

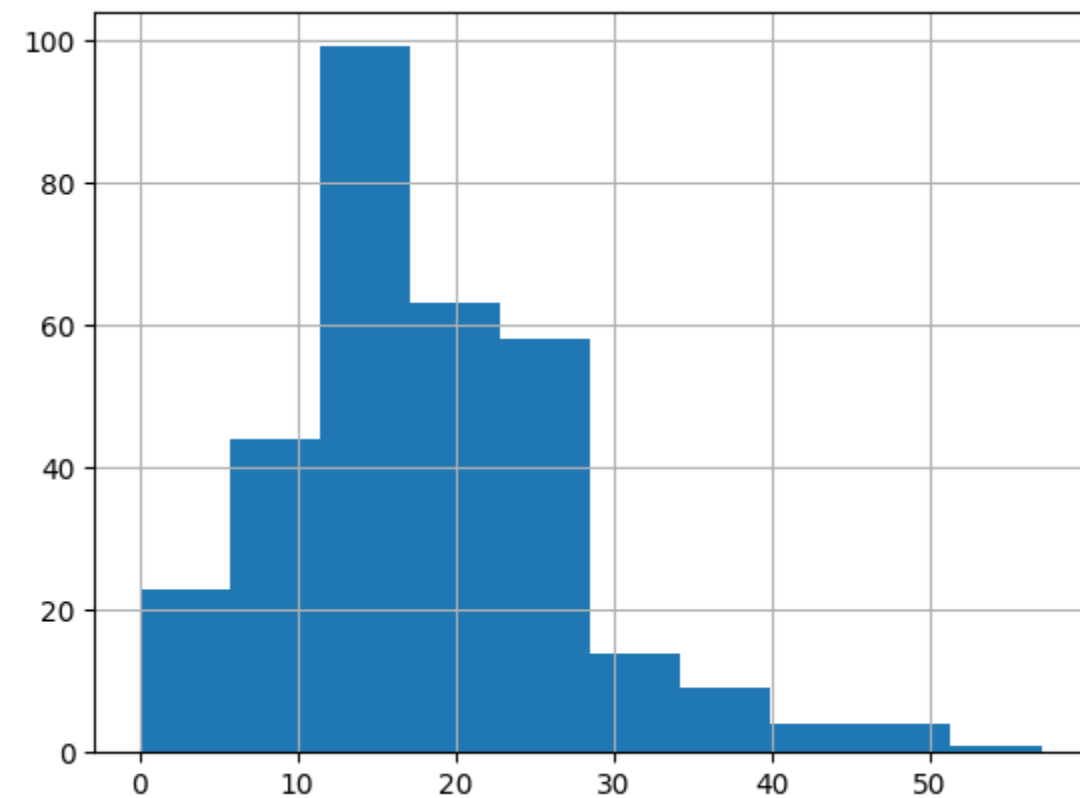
앞의 발의안 점수/ 최종 점수의 산출 과정과 동일하게 진행,
계수 행렬 H의 열 평균을 가중치로 하여 최종 점수 산출



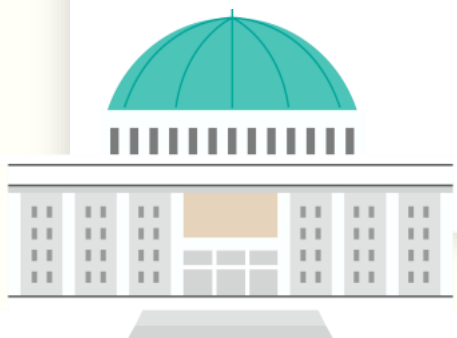
5. 의정활동 평가모델로의 확장

(5) 최종 회의록 데이터셋

의정활동 최종 점수



국회의원명	중간발언_역수	질문	발언횟수	감성지수	최종점수
강길부	29	2.30	1.2	16.24	48.74
강병원	4.93	2.07	0.6	12.60	20.20
강석진	3.19	2.30	1.0	9.52	16.01
강석호	2.32	2.30	0.6	13.72	18.94
강창일	2.03	3.22	1.4	8.40	15.05
...



6. 의의 및 한계

(1) 의의

1.

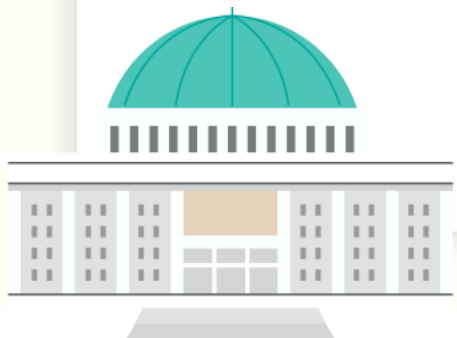
가결과 대표발의만 고려하는 실적 위주의 국회의원의 입법 평가지표를 재설정함

2.

입법활동의 양이 아닌 질을 평가하여 무분별한 발의의 감소를 기대해볼 수 있음

3.

발의안의 사회동향 반영 정도를 키워드 유사도 비교를 통해 수치화하려는 시도



6. 의의 및 한계

(1) 의의

4.

자동화된 입법활동 평가모델로 확장해볼 수 있음

5.

기존의 평가체계 대신 PCA같은 비지도학습을 통해 항목별 가중치를 구하려는 시도

