

FA Study on the NLP Approach to Detect Greenwashing Signs of Companies in Korea

2022.10.27

Yongjik Jay Lee

Contents





Overview of Our Research

- The recognition of 'Greenwashing' is not easy because there are limitations of 'information asymmetry' due to it being an undisclosed corporate's internal information area's activities.
- This study aims to find a way to detect the authenticity of the company's green activities or the possibility of greenwashing through comprehensive and multi-faceted tracking of public information about corporate activities that anyone can easily acquire.
- To this end, considering the characteristics of greenwashing information composed of large-scale unstructured data, natural language processing (NLP) techniques based on big data and machine learning can be a meaningful solution.
- We are developing the Korean Language based Climate Focused NLP Language Model (K-ClimateBERT), which can detect the authenticity of the company's green activities through tracking of public text information about corporate's green activities
- Our Recent Ver.0.3 of K-ClimateBERT is targeting all 2,541 listed Korean corporates through NLP analysis of 0.7 million articles of media text information from the five major daily newspapers (Chosun, JoongAng, Donga, Hankyoreh, Kyunghyang), two economic newspapers (Hankyung, Meakyung), and five broadcasting media (KBS, MBC, SBS, OBS, YTN) for the past five years (2017-09-01 ~ 2022-09-01)

Key Hypothesis on Our Research

(Definition of Greenwashing)

Defined as a state of relative inconsistency between symbolic actions and actual actions related to the corporate's green image (identity) Materiality



- (Problem) The authenticity of the company's green image cannot be quickly confirmed by the ordinary observation of the general public, who do not have specific internal information about the company.
- (Hypothesis)
 - By measuring the weight between the symbolic mention ^{Cheap Talk, Implicit} and actual action^{Explicit} that companies use/expose daily in external communication media (news/media) and comparing it with the social standard on Green talk behavior of industries, the degree and possibility of Greenwashing can be measured.
 - In addition, by measuring the quantifiable greenwashing position of companies, it is expected that the practical effect of social activities of greenwashing monitoring can be increased by limiting/selecting subjects that require intensive observation of greenwashing.

ClimateBERT: A Pretrained Language Model for Climate-Related Text

Summary

By <u>Nicolas Webersinke</u>, <u>Mathias Kraus</u>(FAU Erlangen-Nuremberg, Germany), <u>Julia Anna Bingler</u> (ETH Zurich, Switzerland), <u>Markus</u> <u>Leippold</u>(University of Zurich, Switzerland)

Subjects: Computation and Language (cs.CL) Cite as: arXiv:2110.12010 [cs.CL] or arXiv:2110.12010v1 [cs.CL] https://doi.org/10.48550/arXiv.2110.12010 Submission history From: Markus Leippold [v1] Fri, 22 Oct 2021 18:47:34 UTC (3,545 KB)

Abstract

Over the recent years, large pretrained language models (LM) have revolutionized the field of natural language processing (NLP). However, while pretraining on general language has been shown to work very well for common language, it has been observed that niche language poses problems. In particular, climate-related texts include specific language that common LMs can not represent accurately. We argue that this shortcoming of today's LMs limits the applicability of modern NLP to the broad field of text processing of climate-related texts. As a remedy, we propose ClimateBert, a transformer-based language model that is further pretrained on over 1.6 million paragraphs of climate-related texts, crawled from various sources such as common news, research articles, and climate reporting of companies. We find that ClimateBert leads to a 46% improvement on a masked language model objective which, in turn, leads to lowering error rates by 3.57% to 35.71% for various climate-related downstream tasks like text classification, sentiment analysis, and fact-checking.

Approaches

- Many studies based on traditional NLP approaches in Green Monitoring face considerable limitations, since climate related wording could vary substantially by source (Kim and Kang, 2018).
- Deep learning techniques promise higher accuracy. (e.g., Kölbel et al., 2020; Luccioni et al., 2020; Bingler et al., 2021; Callaghan et al., 2021; Wang et al., 2021; Varini et al., 2020).
- One of the most prominent NLP models is called BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018)



- O Domain Specific pre-training : As climate change is a complex, fast-moving, and often ambiguous topic with scarce resources, extracting climate-related information from textual sources is challenging
 - **Climate Text Corpus :** Pretrainned 1,662,206 climate-related paragraphs of text from <u>news articles, research abstracts, and corporate climate reports</u>
- ② Downstream Task : Text classification, Sentiment analysis, Factchecking → Cheap talk in corporate climate commitments: The role of active institutional ownership, signaling, materiality, and sentiment (Nicolas et al., Jan. 2022)

Step 1: Segmentation & Sampling

	KOSPI	Kosdaq	Konex	SUM
Fossil fuel, Chemicals and material products	141	222	21	384
Electronics and Precision instrument manufacturing	46	300	16	362
Manufacture of electrical, mechanical and other products	62	260	17	339
Information/Communications/Broa dcasting and Media Services	34	237	21	292
Finance and Insurance	105	110	-	215
Logistics, Distribution and Transportation Industry	92	100	5	197
Manufacturing of automobiles and transportation equipment	61	67	5	133
Manufacturing of iron and metal products	53	48	2	103
Construction and Real estate	58	39	5	102
Education, R&D and Human Service	27	98	23	148
Textile and consumer goods manufacturing	38	29	4	71
Food and Beverage Manufacturing	26	32	2	60
Mining, mineral products and cement manufacturing	24	14	1	39
Wood product manufacturing	24	12	1	37
Natural resource production and processing	15	11	2	28
Environmental & Energy Services	12	9	-	21
Accommodation and leisure industry	5	5	-	10
	823	1,593	125	2,541

 (\mathfrak{B})

Approaches & Implication

At K-ClimateBERT v0.1, out of 2,499 listed companies in Korea, 549, about 20% + α (2%), were derived.

Our latest research for K-ClimateBERT V0.3 includes all of the listed companies in Korea.

195 out of 821 KOSPI
334 of 1,549 KOSDAQ
20 of KONEX 129

20% for each industry group, 20% + α ~ all sampling if the parameters are below the average for analysis of the characteristics of each industry group



K-ClimateBERT V0.3

6

Step 2: Crawling of news media information by company & pre-processing of data

	fx			4			11						
					∞ <th∞< th=""> ∞ <th∞< th=""> ∞ ∞ ∞ ∞</th∞<></th∞<>								
A		C	D	E	F	G	н	1	J	к	L	м	N
1	뉴스 식별자	일자	언론사	기고자	제목	통합 분류1	인물	위치	기관	키워드	(가중치순 심	본문	URL
2 1068	1100801.202	20200108	조선일보	박원익 기자	알토스벤처스 친환경 소비재 기업 제이케이인스퍼레이션에 투자 총 투자금 80억	경제>유통	김한준,이를	실리콘밸리	알토스벤치	알토스벤치	제이케이인:	실리콘밸리 벤처캐피털VC 알토스벤처스대표 김한준는 친환경 소비재 제품을 제조 유통하는 제이케이인스퍼레이션	https://bi
3 1087	1100801.202	20191211	조선일보	이진혁 기자	인도 태양광에 280억 투자 허윤홍 이끄는 GS건설 신사업 속도	경제>경제일	슈만트,마(우크라이니	IPP,신재생	태양광,투기	gs건설,태양	허윤홍 사장이 이끄는 GS건설의 신사업이 속도를 내고 있다 허윤홍 사장은 허창수 GS건설 명예회장의 장남으로 올	https://bi
4 1117	1100801.202	20190925	조선일보	이창환 기자	HDC현대EP EBPK 지분 인수로 친환경 소재 사업 진출	경제>자동차	강,강창균	일본,한국	마츠시다,0	HDC현대E	에코바이오	HDC현대EP가 친환경 플라스틱 제조 벤처기업 에코바이오플라스틱코리아EBPK의 지분을 부분 인수해 친환경 소재	http://biz
5 1322	1101001.202	20211206	한겨레	옥기원 기자	페트병 유니폼 입고 썩는 플라스틱 개발 식품업계 재활용 바람	경제>유통		고강	롯데칠성,	페트병,유니	거래처,페트	플라스틱 용기를 수거해 직원 유니폼을 제작하거나 썩는 플라스틱을 만들기 위한 기술개발도 한창인 식품기업들이	http://ww
6 1325	1100401.202	20211126	동아일보	윤우열	2023년까지 플라스틱 OUT 나뚜루 친환경 포장재 전환 작업 박차	경제>서비스	_쇼핑	바닐라,내	롯데제과,	2023년,플	나뚜루,포장	롯데제과의 디저트 브랜드 나뚜루가 Greens come true를 슬로건으로 친환경 포장 전환 작업에 박차를 가한다고 26	https://w
7 1355	1100901.202	20211029	중앙일보	백일현(baek.ilhy	Big Change 차세대 모빌리티 로봇 바이오 미래 산업 주도권 잡아라	경제>산업_	기업	페로브스키	UAM,CJ제	차세대,모!	차세대,친환	디지털 전환 친환경 경제 대응 위해 빅 체인지 나선 국내 기업들 차세대 모빌리티 로봇 바이오 등 미래 핵심 산업에	https://w
8 1376	1100401.202	20210827	동아일보	윤우열	롯데제과한솔제지 친환경 패키지 개발 맞손 제2의 카카오 판지 만든다	경제>경제일	민명기	서울,부산,	롯데제과,	롯데제과,	친환경,롯데	롯데제과가 한솔제지와 지속 가능한 친환경 패키지 개발을 위한 업무협약MOU을 맺었다고 27일 밝혔다 롯데제과는	https://w
9 1418	1100401.202	20210625	동아일보	윤우열	식품업계에 불어온 SES 바람 상반기 패키지 트렌드는	경제>유통		중국,월평	CJ제일제당	식품업계,9	소용량,친환	작년부터 올해 상반기까지 식품업계에서 패키지 관련 이슈가 이어지면서 패키지 트렌드도 변화하고 있다 식품업계(https://ww
10 1452	1100401.202	20210413	동아일보	이지윤	친구야 착한 소비 하자 비건 상품 선물 10배 껑충	경제>서비스	이은희	재구,캐슈,	롯데제과,:	친구,소비,	비건,아이스	채식주의자인 최서윤 씨24 여는 최근 동물성 원료를 쓰지 않은 비건vegan 핸드크림 세트를 카카오톡 선물하기 기능	https://w
11 1455	1100101.202	20210407	경향신문	김한술 기자 hans	플라스틱 줄여야 하긴 하지만 플라스틱 포장재 제거에 소극적인 기업들	경제>유통	김		엄마손파이	플라스틱,7	포장재,농심	4개 식풍제과기업들의 플라스틱 트레이 제거 계획 답변 보니 필요성 공감 한다지만 구체적 방법 시기 제시는 없어소	http://nev

Database of news articles generated by crawling code (MySQL)

		inyov	L Workbench						
news_db									
35 7 805	5 0 0 0				o 🗖 🗖				
Administration Schemas	🐔 news	F news							
CHEMAS 🚸		👰 🜔 🧙 📀 🕄 🗐 Don't Limit	A 4	🗸 🔍 🖺 🖃					
2 Filter objects	1 • SELECT * FROM news (b. news;								
	I . SELECT	<pre>FROM news_db.news;</pre>							
hscode									
, 📄 news_db									
🗸 🖻 Tables	100% 🗘 1:1								
> news	Result Grid 📋	Filter Rows: Q Search Export:	Fetch rows	⇒					
F Views	date_publish	company news_title	source_media	contents_summary	article				
Stored Procedures	2022.03.28.	현대중공… 현대重그룹, HD현대로 새출발정기선 지주사대표 선	~ 연합뉴스	정몽준 아산재단 이사장의 장남이자 현대가(家) 3세인	서울연합뉴스 김보경 기자 정몽준 아산재단 이사장의,				
Functions	2022.03.28.	현대중공… 정기선 현대중공업지주 사장, 주주총회에서 사내이사로	연합뉴스	28일 서울 종로구 계동 현대빌딩에서 열린 '현대중공업…					
	2022.03.28.	현대중공… 현대중공업지주 제5기 정기 주주총회	연합뉴스	28일 서울 종로구 계동 현대빌딩에서 '현대중공업지주					
sys	2022.03.24.	현대중공… 현대重그룹 권오갑 회장 "정주영 창조적 예지 따라 새 !		창립 50주년 맞아 임직원들에게 기념 메시지 권오갑 현…					
	2022.03.22.	현대중공··· 정기선, 한국조선해양 대표 선임가삼현 부회장과 각 현대중공··· 현대미포조선, 매출 2조8천872억원·영업손실 2천173		제48기 정기 주주총회 개최이사 선임 등 5개 안건 가… 이날 주총에선 전(前) 국민경제자문회의 자문위원으로	시절신입다고 남고경 가지 영웅은 이신체는 이지공처.				
	2022.03.22	현대중공… 참립 50주년 맞은 현대重그룹"100년 기업으로 도익		히글 구멍에는 단(히) 국민장제지문회의 지문위원으로 현대중공업그룹은 2018년 현대중공업지주를 출범시키…					
	2022.03.21.	현대중공… 좋은기업연구소 "카카오페이 신원근 CSO 사내이사 신							
	2022.03.20.	현대중공… 정주영 21주기 맞아 범현대가 청운동에 모여정몽준	~~ 연합뉴스	이후 정 명예회장의 아들인 정몽준 아산재단 이사장과	서울연합뉴스 김보경 기자 아산 정주영 현대그룹 명예				
	2022.03.28.	현대중공… 조선업계, 여성 사외이사 줄줄이 선임자본시장법-E		28일 조선업계에 따르면 현대중공업그룹의 지주사인	서울연합뉴스 김보경 기자 남성적 직장문화가 강한 것				
	2022.03.28.	현대중공… 현대重그룹, HD현대로 새출발정기선 지주사대표 선		정몽준 아산재단 이사장의 장남이자 현대가(家) 3세인	서울연합뉴스 김보경 기자 정몽준 아산재단 이사장의.				
	2022.03.28.	현대중공… 정기선 현대중공업지주 사장, 주주총회에서 사내이사로		28일 서울 종로구 계통 현대빌딩에서 열린 '현대중공업…					
	2022.03.28.	현대중공… 현대중공업지주 제5기 정기 주주총회	연합뉴스	28일 서울 종로구 계통 현대빌딩에서 '현대중공업지주	서울연합뉴스 김보경 기자 권오갑 현대중공업지주26. 51				
	2022.03.24.	현대중공… 현대重그룹 권오갑 회장 "정주영 창조적 예지 따라 새		창립 50주년 맞아 임직원들에게 기념 메시지 권오갑 현…					
	2022.03.22.	현대중공… 정기선, 한국조선해양 대표 선임가삼현 부회장과 각		제48기 정기 주주총회 개최이사 선임 등 5개 안건 가~					
	2022.03.22.	현대중공··· 현대미포조선, 매출 2조8천872억원·영업손실 2천173 현대중공··· 창립 50주년 맞은 현대重그름*100년 기업으로 도약		이날 주충에선 전(前) 국민경제자문회의 자문위원으로 현대중공업그룹은 2018년 현대중공업지주를 출범시키…					
	2022.03.21.	현대중공… 좋은기업연구소 "카카오페이 신원근 CSO 사내이사 신		수혜를 인을 뿐 아니라 두 회사의 이사를 경직해 이해	서울IR 제공 서울연합뉴스 이미령 기자 의경권 자문사 BIG				
	0000.00.00								
	news_db query				Read O				
	Action Output 0								
	Time	Action		Response	Duration / Fetch Time				
	0 7 15:38:3	SELECT * FROM news_db.news		3910 row(s) returned	0.00080 sec / 0.023				
ery Completed		ourset internet of the set		control (a) retained	0.00000 300 / 0.020				

Approaches & Implication

- Crawling News media exposure information and data collection
- Data Source: Online news articles from the five major daily newspapers (Joseon, JoongAng, Donga, Hankyoreh, Kyunghyang), two daily economic newspapers (한경,매경), and five broadcasting media KBS, MBC, SBS, OBS, YTN for the past five years (2017-09-01 ~ 2022-09-01)
 - **Method:** Using the Korea Press Foundation's Big Keys API, extract a list of news articles on selected sample companies, and analyze the metadata (media company, contributor, title, etc.) and entity name (person, institution, place, etc.) of the newsText data
- preprocessing
 - Hangul Natural Language Processing (NLP) Precedent: Converting Input Strings to Morphological Columns
 - * Morphological analysis refers to the process of separating words, which are words with spaces, into morphemes, which are the minor units of meaning that can no longer be analyzed.
 - This study deleted stopwords such as special characters and punctuation marks from article data, and tokenization was performed.
 - To further increase analysis accuracy, only significant parts of speech, such as nouns, verbs, adjectives, and adverbs, are extracted and applied.
 - Bot creation unrelated to Green Talk among news articles Article articles in the form of stock market conditions (e.g., Chosun Ilbo C-Biz bot), duplicate reports, and personnel transfer/obituary-related messages are excluded through exception handling.

Step 3 : GreenTalk Identification by K-ClimateBERT and Relation Analysis Between GreenTalk Behavior & Scientific Results by Industries



Key Findings & Implication

Rank	Industry	GreenTalk Ratio	CO2 Reduction Rate*
1/17	Environmental & Energy Services	7.69%	-11.2%
2/17	Manufacturing of automobiles and transportation equipment	3.09%	-2.1%
3/17	Fossil fuel, Chemicals and material products	2.70%	-2.0%
5/17	Logistics, Distribution and Transportation Industry	2.04%	-5.3%
6/17	Wood product manufacturing	1.92%	-1.9%
15/17	Construction and Real estate	0.56%	2.5%
16/17	Information/Communications/Broad casting and Media Services	0.52%	2.1%

* CO2 Reduction Rate (2017 ~ 2020, CAGR)

More GreenTalker make Better Performance on climate contribution !

Step 4 : Green Talk Classification (N_{Explicit} / N_{explicit+Implicit})



Does Better Quality of GreenTalker make Better Performance of Contributing to the climate?

Performance Analysis

"S" Company vs. "L" Company

(Fossil fuel, Chemicals and material products Manufacturing



^{·······} 선형 ("S" Company) ······· 선형 ("L" Company)

