Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.039104
Article

ESG Discourse Analysis Through BERTopic: Comparing News Articles and Academic Papers

Haein Lee[1], Seon Hong Lee[1], Kyeo Re Lee[2] and Jang Hyun Kim[3,*]

[1]Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, 03063, Korea
[2]Center for SW Education, Hanyang University, Ansan, 15588, Korea
[3]Department of Interaction Science/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, 03063, Korea

*Corresponding Author: Jang Hyun Kim. Email: alohakim@skku.edu
Received: 10 January 2023; Accepted: 10 April 2023

Abstract: Environmental, social, and governance (ESG) factors are critical in achieving sustainability in business management and are used as values aiming to enhance corporate value. Recently, non-financial indicators have been considered as important for the actual valuation of corporations, thus analyzing natural language data related to ESG is essential. Several previous studies limited their focus to specific countries or have not used big data. Past methodologies are insufficient for obtaining potential insights into the best practices to leverage ESG. To address this problem, in this study, the authors used data from two platforms: LexisNexis, a platform that provides media monitoring, and Web of Science, a platform that provides scientific papers. These big data were analyzed by topic modeling. Topic modeling can derive hidden semantic structures within the text. Through this process, it is possible to collect information on public and academic sentiment. The authors explored data from a text-mining perspective using bidirectional encoder representations from transformers topic (BERTopic)—a state-of-the-art topic-modeling technique. In addition, changes in subject patterns over time were considered using dynamic topic modeling. As a result, concepts proposed in an international organization such as the United Nations (UN) have been discussed in academia, and the media have formed a variety of agendas.

Keywords: ESG; BERTopic; natural language processing; topic modeling

1 Introduction

The term environmental, social, and governance (ESG) factors were first used officially in a report called "Who Cares Wins" published by the United Nations Global Compact (UNGC) in 2004 [1]. Further, in 2006, the United Nations Principles for Responsible Investment (UNPRI) emphasized ESG as a financial investment principle and presented the cornerstone of the ESG framework [2].

Subsequently, interest in ESG issues that include climate change, environmental protection, and public health has increased owing to the COVID-19 pandemic. In this context, in 2020, the importance of ESG from a long-term investment perspective was highlighted [3]. In addition, ESG can provide evaluations of corporate economic processes to regional and global positions [4].

In previous studies on ESG, the relationship between ESG concepts and investors was analyzed using data from specific countries [5]. Alternatively, the effects of ESG-related natural language data within financial markets were investigated [6]. Additionally, conceptions related to ESG were examined using data on corporate social responsibility (CSR) reports [7]. However, these studies did not use comprehensive data. Therefore, their findings do not suffice to gain insights into the utilization of ESG.

In this study, we used LexisNexis, a platform that provides news from major media outlets worldwide, and Web of Science, a platform providing academic papers, to extract unstructured text data and understand the views of the general public and academia. In particular, we employed topic modeling, which automatically clusters words based on machine learning and statistics, to consider the essential meanings of each item. This methodology assumes that each document is a function of latent variables and groups text documents by automatically identifying topics. Subsequently, we explored text mining related to ESG through a bidirectional encoder representations from transformers topic (BERTopic) model, which has been recognized as a state-of-the-art (SOTA) approach among topic modeling technologies. Based on the various topics in the text, we analyzed the media and academic discourse on ESG and discovered some notable differences. The present work contributes to global awareness of ESG across discourse sources. Specifically, based on changes in topics discussed in academia and media over time, we found that the value of utility was first debated in academia according to proposals of the United Nations organization, and then the media focused on these items.

The remainder of this study is arranged as follows. In the next section, we explore the related literature. In the Methods section, we outline the data collection and analysis procedures used in this work. Our findings are reported in the Results section, and the implications of this study are presented in the Discussion and Conclusion section.

2 Related Works

2.1 Research on ESG Using Natural Language Processing

As of the 2020s, research on ESG and natural language processing (NLP) has been actively conducted. Serafeim et al. [8] examined how investors react to various disclosures related to ESG by using a unique dataset containing ESG news sentiment scores acquired by tracking daily news on thousands of companies. They found that investors have a greater response to ESG news that is financially significant, positive, and related to social capital issues than to other news. Sokolov et al. [9] applied the latest NLP technology to automatically change unstructured natural language data into ESG scores with a Bidirectional Encoder Representations from Transformers (BERT) model and social media data. With this approach, they identified ESG risks using NLP technology. Moreover, they developed NLP models that could generate ESG scores. Bapat et al. [6] studied the impact of ESG on social media and news articles data on stock market outcomes. To understand the effects of articles on ESG, they selected four stocks from well-known companies, including HSBC Bank, Tesla, Goldman Sachs, and Amazon. By summarizing real-time data from newspaper articles and Twitter, they created a sentiment index to compare a company's overall sentiment with the percentage change over a specified time period. Consequently, from a sample of ESG-related articles in the financial sector and online retail, they concluded that news about ESG exhibited a positive impact on stock prices. Seo et al. [5] presented policy implications for the introduction and diffusion of ESG through an analysis of news data. They collected news data related to ESG management and analyzed text keywords, using latent Dirichlet allocation (LDA) for topic modeling. Subsequently, they showed a strategic direction for the successful ESG management of companies in a rapidly changing business environment. Goloshchapova et al. [10] used topic modeling with CSR reports, which are indicators closely related to ESG. Data were collected from publicly available CSR reports from 1999–2016 for all constituent companies of the stock market in 15 developed countries in MSCI Europe. Subsequently, common topics reported by companies in the UK and Europe emerged, such as "employee safety," "employees training support," "human rights," and "healthcare and medicine." Amin et al. [7] studied companies' social media accounts for CSR disclosure and identified their determinants. To analyze tweets about CSR, they used topic modeling and regressions. The results indicated that Twitter's popularity as a CSR disclosure platform had increased remarkably over the past few years.

2.2 Topic Modeling

Topic modeling is used in many domains and is one of the most widely used methodologies for extracting meaningful text information from text-mining technologies. Representative models include LDA and Dirichlet multinomial regression (DMR). Zhao et al. [11] showed that global hashtags of social network services (SNS) and the relationship between topics and hashtags could be found through an approach called Hashtag-LDA; meaningful topics were identified using this model. In addition, based on experimental results on a confirmed Twitter dataset, they proved that Hashtag-LDA is superior to other recommendation methods. Yun et al. [12] used LDA to create an automated patent classification system and increase class prediction performance. Tonidandel et al. [13] developed an approach to leadership study by using structural topic models (STM). In this way, STM was applied with a huge corpus of unstructured text responses from various samples of leaders to create a classification structure for leadership problems. Li et al. [14] proposed a new methodology to evaluate bicycle accessibility for various travel motivations in Shanghai, China. They used the DMR to determine the various reasons for travel by considering both the departure and arrival times of individuals. These studies demonstrate the applications of topic modeling in various domains.

3 Method

The overall flow of the experiment is detailed in this section and illustrated in Fig. 1. First, data were collected from LexisNexis and the Web of Science for analysis. Second, preprocessing was performed to use the necessary text data. Third, dimensionality reduction and a Class-based Term Frequency-Inverse Document Frequency (c-TF-IDF) were performed through the BERTopic model using the data. Thus, topic clusters were formed from the data. The model automatically found an optimal number of topic pairs with a minimum similarity of 0.9 or more in the cluster.

3.1 Data Collection

To investigate the perception of ESG, we utilized data from LexisNexis, which comprised news data, and the Web of Science database, which offers academic literature. In the process of data collection, the data queries "ESG" and "esg" were used, and totally of 15,688 and 1,747 data points, respectively, were collected for the duration of January 1, 2016–October 14, 2022.

3.2 Data Preprocessing

In the collected data, texts with the same characters as 'esg' but different meanings were excluded from the analysis (e.g., *Ephedra sinica* granules, Electrical shift gearbox). In addition, the article in the

LexisNexis and the abstract part in the Web of Science were used for study. In the data, 15,688 data from LexisNexis and 1,706 data from the Web of Science were utilized, excluding cases where text did not exist in the data. To know the distribution of the word count in each data set, we showed Figs. 2 and 3. Subsequently, we adapted the lemmatization of the spaCy library to extract only nouns from the data. Unnecessary words were removed using the stopword list in spaCy [15,16].

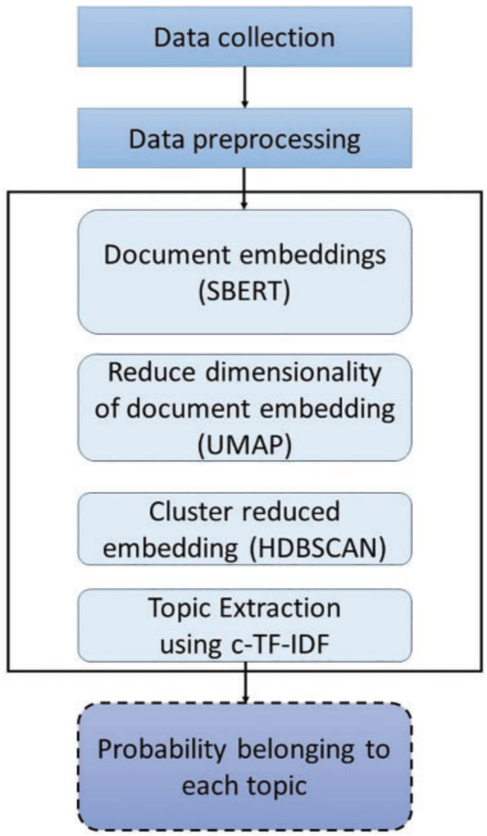

Figure 1: Diagram of the entire process

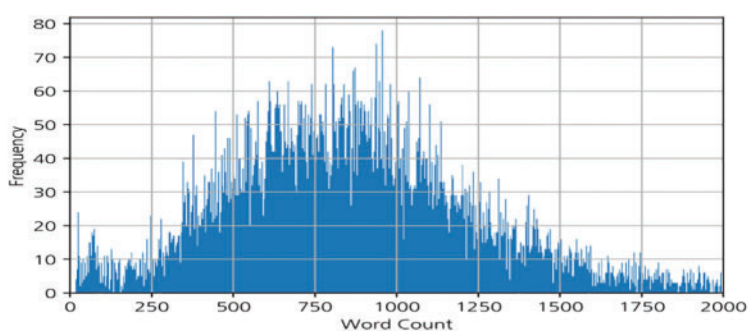

Figure 2: Word count from the LexisNexis data

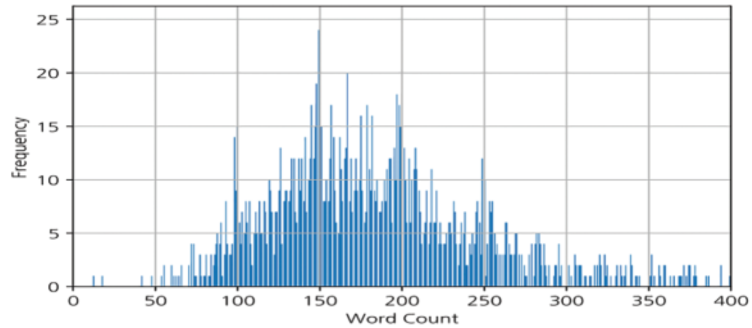

Figure 3: Word count from Web of Science data

3.3 BERTopic

3.3.1 Generation of Embedding and Clusters

Topic modeling is a methodology for identifying latent topics from a group of documents and is a part of clustering that uses unsupervised learning. LDA and non-negative matrix factorization (NMF) models, which describe a document as a latent mixture of topics through a collection of words, were used as the traditional topic modeling methods. However, these methodologies ignore the semantic relationship between words and do not describe the context of the words in a sentence. A collection of words, which is an input value, cannot represent a document accurately. To solve this problem, we adopted a model based on BERT. BERTopic generates contextual words and sentence vector representations, and then considers the semantic properties of vector representations such that similar texts are located close to the vector space [17]. BERTopic uses sentence BERT (SBERT) to perform documentary embedding to create dense vectors for sentences and paragraphs. Subsequently, the dimensionality of the data was reduced through uniform manifold approximation and projection (UMAP), and clusters were created using hierarchical density-based spatial clustering of applications with noise (HDBSCAN).

3.3.2 Finding Topics Using Class-Based Term Frequency-Inverse Document Frequency (c-TF-IDF)

With the clusters, the subject representation was extracted from each topic through a class-based term frequency-inverse document frequency (c-TF-IDF). In particular, the maximal marginal relevance (MMR) algorithm adjusts various keywords representing a topic and calculates the cosine similarity between the document and word embedding vectors to produce a list of keywords.

$$W_{t,c} = tf_{t,c} \cdot \left(1 + \frac{A}{tf_t}\right) \quad (1)$$

The topic representation deals with each clustered document as a single document, and the IDF is replaced by an inverse class frequency to measure the amount of information the word provides for the class. Specifically, the average number of words A per class (i.e., topic) was divided by the frequency of word t in all classes and takes the logarithm value. In addition, to consider only positive values, it was necessary to add one within the logarithm.

The methodology for finding these topic representations is an implementation of the importance of words in a cluster and can create a topic-word distribution for each cluster in the document. Therefore, it can be used to acquire words that represent the corresponding topics in the cluster.

4 Results

4.1 Topic Analysis from LexisNexis Data

The topics from the LexisNexis data have various meanings for ESG. Each topic representation was retrieved with a c-TF-IDF score derived within the HDBSCAN cluster (Table 1). In particular, the documents in the cluster were treated as a single document. Because the BERTopic model generated the c-TF-IDF for each class aggregated by HDBSCAN, this technique produces consistent topics. The hierarchical clustering root of the entire topic began with the industrial changes over time. In particular, it was divided into ESG value topics used as asset indicators and concentrated on Asian companies (Fig. 4).

Table 1: Topics related to ESG from LexisNexis

Number	Summary	Topics (c-TF-IDF score)
0	Industry change over time	(‘company’, 0.0290)(‘say’, 0.0262)(‘business’, 0.0249)(‘year’, 0.0188)(‘esg’, 0.0180)(‘time’, 0.0172)(‘work’, 0.0165)(‘change’, 0.0153)(‘people’, 0.0148)(‘industry’, 0.0142)
1	Environmental energy industry	(‘climate’, 0.0464)(‘energy’, 0.0464)(‘carbon’, 0.0415)(‘company’, 0.0327)(‘change’, 0.0323)(‘emission’, 0.0291)(‘say’, 0.0261)(‘esg’, 0.0239)(‘industry’, 0.0236)(‘business’, 0.0215)
2	Asian green finance hub	(‘singapore’, 0.0800)(‘business’, 0.0373)(‘cent’, 0.0357)(‘say’, 0.0309)(‘company’, 0.0309)(‘year’, 0.0290)(‘trade’, 0.0230)(‘industry’, 0.0224)(‘asia’, 0.0216)(‘esg’, 0.0212)
3	Asset indicators and ESG	(‘statement’, 0.0364)(‘company’, 0.0362)(‘production’, 0.0314)(‘cost’, 0.0311)(‘information’, 0.0287)(‘result’, 0.0281)(‘include’, 0.0278)(‘look’, 0.0276)(‘cash’, 0.0271)(‘gold’, 0.0251)
4	ESG business of Korean companies	(‘Korea’, 0.0525)(‘esg’, 0.0512)(‘company’, 0.0410)(‘group’, 0.0386)(‘management’, 0.0307)(‘business’, 0.0285)(‘say’, 0.0266)(‘industry’, 0.0253)(‘governance’, 0.0244)(‘factors’, 0.0240)
5	Corporate investment in ESG	(‘fund’, 0.0873)(‘investment’, 0.0558)(‘esg’, 0.049)(‘investing’, 0.0389)(‘index’, 0.0385)(‘company’, 0.0351)(‘investor’, 0.0340)(‘say’, 0.0302)(‘invest’, 0.0292)(‘market’, 0.0275)

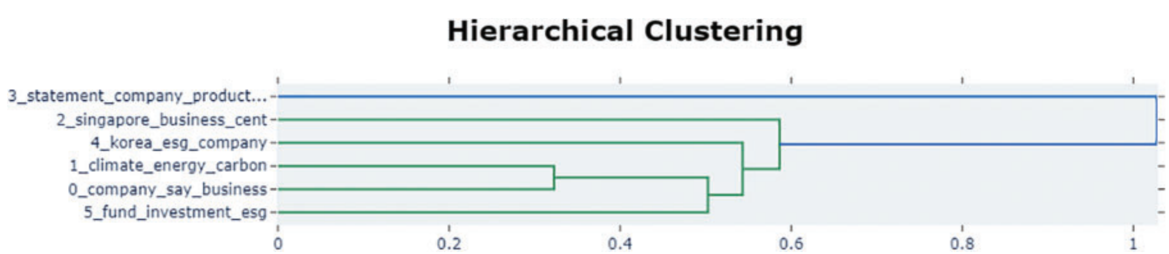

Figure 4: Hierarchical clustering from the LexisNexis data

Topic 0 was named “Industry change over time” through “company,” “industry,” “change,” “year,” “time,” and “esg.” Crespi et al. [18] surveyed 727 financial companies in 22 countries and found that ESG scores increased over time. In particular, companies applied corporate social performance (CSP) and ESG ratings to management. This tendency is related to company size and profitability. Ionescu et al. [19] investigated the relationship between market value and ESG factors in the travel industry based on listed companies. Consequently, they found that among the ESG factors, governance plays the most important role in the market value of companies. Chevrollier et al. [20] argued the strategic direction and effect of a company on its ESG performance through an evaluation of listed companies for a period of seven years. Based on this, the strategy for corporate ESG performance and actual behavior was considered to be positively related and effective in pursuing long-term sustainability. Therefore, companies from various domains have been striving to advance their ESG achievements for continued growth over many years.

Topic 1 was labeled “Environmental energy industry” through “climate,” “energy,” “carbon,” “esg,” and “change.” According to Zainullin et al. [21], the energy industry is the most affected by ESG investments. Thus, to raise the attractiveness of the energy business for ESG investment, the best practices of energy corporations were employed, and corporate culture development was proposed. Naeem et al. [22] studied the impact of ESG on the financial outcome of global power and energy companies. The findings of the statistical research that combined financial and ESG data showed a significant correlation between financial achievement and ESG performance. In the face of global systemic risks from climate change, Chen et al. [23] analyzed the impact on corporate profits using manufacturers’ responses to low-carbon transitions, ESG performance, and financial data. Ultimately, the increasing crisis awareness about global climate change and eco-friendly energy business is continuously attracting attention.

Topic 2 was named “Asian green finance hub” through “Singapore,” “Asia,” “esg,” “industrial,” and “trade.” Singapore, the center of the Asian financial hub, aims for Asian green finance. Thus, the Singapore Exchange requested that member companies comply strictly with ESG disclosure. Specifically, the Association of Banks in Singapore issued guidelines on responsible finance to facilitate and support ESG disclosure [24]. Likewise, regarding green finance, which considers environmental impacts and promotes sustainability in Asia, Volz [25] suggested that lending decisions and investments should be based on risk assessment and environmental screening. Dikau et al. [26] proposed that central banks are responsible for financial and macroeconomic stability, addressing climate change-related risks at a systemic level. Overall, financial institutions in Asian countries, centered on Singapore, are continuously making efforts to realize green finance.

Topic 3 was summarized as “Asset indicators and ESG” through “gold,” “cash,” “production,” and “information.” Piserà et al. [27] examined safe assets by comparing cryptocurrencies, gold, and ESG investments. Their results indicated that ESG indices were validated even amid economic and financial crises caused by pandemics such as Covid-19. Andersson et al. [28] discussed the causal relationship between the ESG portfolio and stock exchange rates, prices, and commodity prices included in traditional portfolios. Consequently, ESG portfolio returns lead to currency and commodities returns. Rubbaniy et al. [29] noted that incorporating ESG stocks into portfolios was effective for equity investors and asset managers in enhancing portfolio performance. Along with the growing interest in ESG worldwide, the addition of ESG indicators to individual and corporate portfolios is drawing attention.

Topic 4 was summarized as “ESG business of Korean companies” through “Korea,” “esg,” “business,” and “governance.” Yoon et al. [30] examined the impact of companies on market and stock prices by using ESG scores for corporate social responsibility in the Korean market. The results suggested that governance reform should be carried out conditionally, because the impact on stock prices varies depending on the characteristics of the company. Yoon et al. [31] considered the relationship between corporate participation in social responsibility activities, measured by ESG scores and corporate tax avoidance tendencies in the Korean financial market. Similar to corporate culture theory, the higher the CSR management performance, the less taxable profit is manipulated. During the heightened uncertainty with the spread of Covid-19, the impact of ESG action on financial outcomes was investigated [32]. While most companies suffered financial difficulties in Korea in the first quarter of 2020, it was seen that the higher achievement of ESG activities, the smaller the reduction in profit. These results indicate that the performance of a company's ESG activities leads to positive financial outcomes in uncertain situations. Accordingly, the performance of many companies in Korea, which adjusts ESG activity indicators, is attracting attention.

Topic 5 accounted for "Corporate investment in ESG" through "fund," "investment," "esg," and "market." Eccles et al. [33] examined the perceived barriers that companies face for ESG metrics. Investors were surveyed to consider this, and it was found that standards for measuring ESG performance were lacking, and efforts should be made to incorporate them. Folqué et al. [34] explored the importance of ESG portfolio management for sustainable investment funds. Ultimately, integrating the evaluation system is essential for applying ESG indicators to the portfolio. In addition, ESG funds increased along with the rise in socially responsible investment funds. Furthermore, Raghunandan et al. [35] pointed out that ESG scores do not correlate with actual carbon emissions, and that the concerns of real funding stakeholders are poorly implemented. Consequently, portfolios employing ESG indicators have increased, but integrated systems are lacking, and companies have not fulfilled their social responsibilities, such as controlling carbon emissions [36].

4.2 *Topic Analysis from Web of Science Data*

Topics included in the Web of Science were considered hierarchically, starting with topics related to the impact on bond markets. In addition, clustering was separated into the energy industry and the company's performance in ESG (Fig. 5, Table 2).

Table 2: Topics related to ESG from Web of Science

Number	Summary	Topics (c-TF-IDF score)
0	ESG performance and investment	('esg', 0.0676)('performance', 0.0378)('study', 0.0356)('investment', 0.0345)('firm', 0.0313)('sustainability', 0.0293)('governance', 0.0287)('csr', 0.0286)('risk', 0.0279)('market', 0.0265)
1	Carbon emission and portfolio	('carbon', 0.1584)('climate', 0.1316)('emission', 0.1111)('risk', 0.0520)('change', 0.0513)('portfolio', 0.0430)('company', 0.0367)('policy', 0.0362)('reduction', 0.0342)('investor', 0.0330)
2	Bond market of ESG	('bond', 0.4223)('market', 0.0845)('issuance', 0.0780)('spread', 0.0712)('yield', 0.0710)('esg', 0.0627)('government', 0.0585)('issuer', 0.0573)('country', 0.0478)('credit', 0.0434)

(Continued)

Table 2: Continued

Number	Summary	Topics (c-TF-IDF score)
3	Energy industry	('mining', 0.1267)('copper', 0.0744)('supply', 0.0732)('metal', 0.0716)('mineral', 0.0627)('energy', 0.0610)('cobalt', 0.0510)('industry', 0.0483)('battery', 0.0460)('demand', 0.0452)

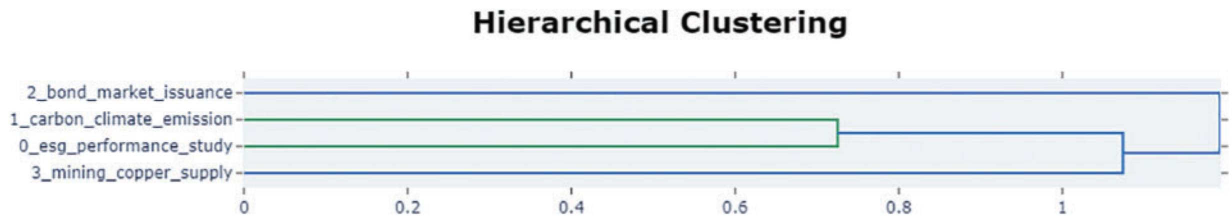

Figure 5: Hierarchical clustering from Web of Science data

Topic 0 was named "ESG performance and investment" through "performance," "investment," "firm," "governance," and "csr." Giese et al. [37] suggested the possibility of using ESG as a financial indicator by identifying the relationship between a firm's ESG information and the capital profile value. In other words, ESG ratings are suitable for incorporating policy benchmarks and financial analyses. Additionally, the rise in CSR worldwide has highlighted the relationship between corporate financial profiles and ESG performance [38]. In particular, the relationship between the impact of ESG disclosure, financial performance, and ESG performance is demonstrated through correlation. Van Duuren et al. [39] investigated how asset managers utilize ESG factors in their investments. Managers identify the use of ESG information for responsible investment in their processes, and then employ it for risk warning and management. Therefore, corporations have integrated ESG into their capital profiles and financial metrics for sustainable growth and risk management.

Topic 1 was summarized as "Carbon emission and portfolio" through "carbon," "climate," "emission," and "portfolio." Bressan et al. [40] studied rebalancing corporate bond portfolios to reduce exposure to climate change risks. Consequently, portfolio rebalancing that reflects climate awareness was necessary for an ongoing investment strategy. Likewise, oil and gas companies account for 56% of global carbon emissions. Shakil [41] described the adverse effects of ESG performance on financial risk in oil and gas companies. Furthermore, they found that gender diversity and ESG exhibited significant moderating effects on financial risk. Bolton et al. [42] found that carbon dioxide emissions control return predictors, such as stock size, leading to higher returns. Ultimately, a company's unexpected profitability or risk factors can be explained through the carbon premium. Thus, carbon emission variables are closely related to the composition of financial portfolios and are highly correlated with ESG.

Topic 2 represented a bond market of ESG "Bond market of ESG" through "bond," "market," and "esg." Amiraslani et al. [43] investigated the impact of bond spreads as a substitute for social capital using a firm's ESG performance. Consequently, companies with high social capital benefited from low bond spreads during the 2008–2009 financial crisis period. Gerard [44] examined the association between ESG policies, bond prices, risk, and returns. In this regard, an extremely high ESG performance does not unconditionally cause good bonds. Apergis et al. [45] studied the relationship between debt costs and ESG scores. The authors discovered that higher ESG ratings lead to lower debt costs in the bond market. Hence, ESG achievement corresponding to social capital is necessary to protect corporations from financial risk.

In the case of Topic 3, words such as “copper,” “mining,” “metal,” “energy,” and “industry” were included to summarize “Energy industry.” Behl et al. [46] tested causality using ESG disclosures and energy sector company data. As a result, corporate values did not have a bidirectional relationship with ESG and each factor, and the authors proposed that executives assess corporate value and investment advantages in terms of relationships. Zhao et al. [47] surveyed Chinese groups to explore the relationship between financial indicators in the energy power market and ESG performance. Their analysis showed that good ESG ability could improve financial accomplishment. In summary, the number of companies in the energy industry considering ESG factors to improve their financial factors has increased.

4.3 Dynamic Topic Model (DTM)

Dynamic Topic Modeling (DTM) is a technique for analyzing changes in topics over time [48]. In the BERTopic, DTM is performed by calculating the c-TF-IDF for each subject and time. Fig. 6 visualizes the shift in the subject according to the change in the year through data obtained from LexisNexis. The frequency of various topics with ESG as keywords noticeably increased by 2020. In particular, Topic 0 (Industry change over time) has the largest portion every year, as more companies focus on ESG. Topic 3 (Asset indicators and ESG) received much attention in 2021, as the value of ESG as a stable asset was highly discussed during the heightened uncertainty of Covid-19. In addition, interest in topic 1 (Environmental energy industry) increased as crisis awareness about climate change has grown along with the Covid-19 pandemic. Fig. 7 presents the topic change over time based on Web of Science data. In academia, discussions have centered on ESG performance and investment derived from topic 0 rather than dealing with various topics. In addition, interest in ESG-related topics has increased since 2018. Therefore, the ESG agenda has drawn attention from exploratory research through academic papers to the media.

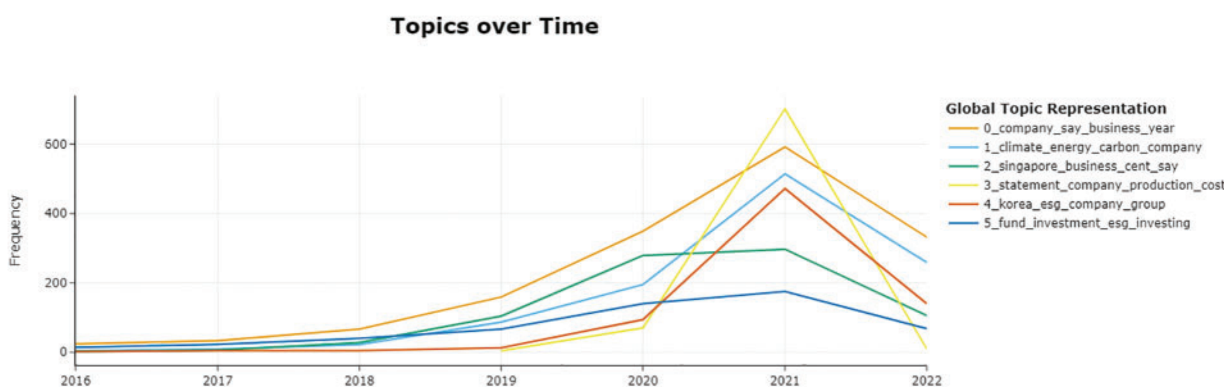

Figure 6: Visualization of topics over time from LexisNexis data

This point is associated with the UNPRI, which was created by major financial institutions worldwide and the United Nations in 2006 [49]. This includes the financial aspects of the company that investors want to invest in and in ESG, which is a non-financial indicator [50,51]. Since ESG was first mentioned from this viewpoint, academics have discussed the ESG performance of corporations. Likewise, while announcing the concept of Sustainable Development Goals (SDGs) reported by the United Nations in 2015, there was an official announcement asking major companies to play a significant role in achieving the SDGs [52]. Since then, major companies have been closely linked to ESG promotion and SDGs attainment in their sustainable management reports [53].

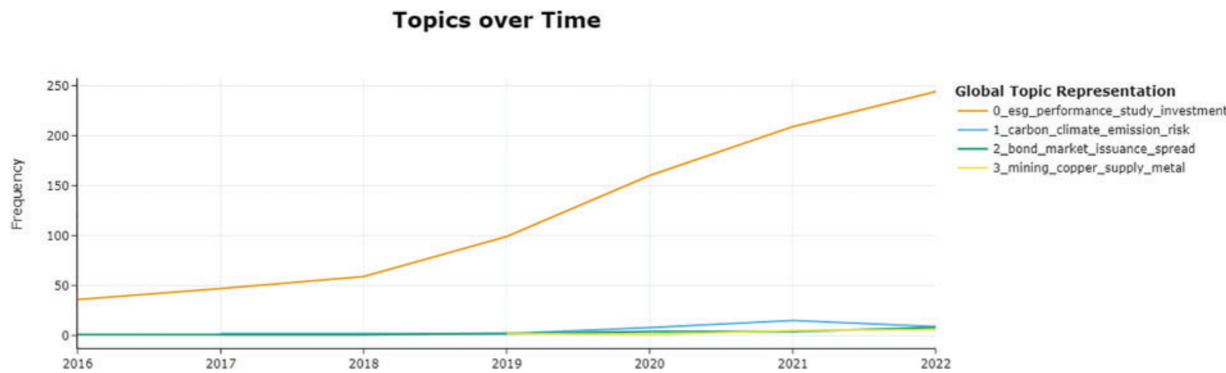

Figure 7: Visualization of topics over time from Web of Science data

5 Discussion

This section discusses the results of unsupervised learning algorithm (i.e., BERTopic) trained on LexisNexis dataset and Web of Science dataset. The ESG-related topics based on LexisNexis data ranged from industry changes over time to topics related to asset indicators and Asian companies. The topics included in the Web of Science data described the performance of the energy industry and companies, starting with topics related to the impact on the bond market. By analyzing the changes in the subject over time, we found that the value of efficiency was initially addressed in academia in response to the subjects discussed by international organizations, such as the United Nations Organization, and then the media focused on this content from various perspectives. As a result of the study, in the trend by international news reports, the topic of ESG and the asset market was greatly emphasized in the hierarchical topic composition, and this was prominent in the time series analysis. In addition, it was confirmed that topics on ESG and the bond market are hierarchically emphasized in trends in international academia. This is interpreted as the same importance on the economic value of ESG factors from the perspective of academic and international media trends, breaking away from the limitations of emphasizing that ESG is significant only at a specific national level in existing studies. In other words, social discussions on ESG should be interested in all internationally, and active social applications should be discussed.

6 Conclusion

Since ESG was first mentioned at the UNGC, it has gained considerable attention and become an important evaluation principle for corporations. In particular, with the outbreak of the Covid-19 pandemic, interest in ESG-related issues, such as climatic and environmental changes, has increased worldwide. Additionally, focusing on ESG from the perspective of long-term investments has become essential for companies. Therefore, understanding the public and academic perspectives on ESG is essential. Previous studies identified the relationship between ESG and investors using data from a specific country [5]. Additionally, natural language data were used to investigate the influence of ESG on financial markets [6]. However, these studies lacked comprehensive insights, because they employed limited data or focused on specific countries.

To overcome these limitations, in this study, we collected news articles and papers from major media companies worldwide, which were available on the LexisNexis and Web of Science platforms, respectively. Furthermore, we clustered and analyzed unstructured data from the perspective of text mining through BERTopic, which is considered the SOTA for topic modeling. Subsequently, based on various topics, we studied the media and academic discourse on ESG using the DTM to find differences.

Although this study studied the international relationship of ESG, our research methodology has several practical implications. Considering the results, BERTopic confirmed that there was little overlap between topics. Therefore, it would be suitable to analyze extensive content using large amounts of natural language data. Additionally, considering the change in time through the DTM, the temporal causal relationship of each platform data can be identified, and how each agenda was formed can be interpreted.

Theoretically, the results of this work suggest that applying news data and papers to understand public awareness and academic interest is appropriate. Moreover, ESG did not immediately receive media attention after it was presented by global institutions. In other words, after exploratory studies in academia, the media concentrates on it from various perspectives. Therefore, if exploration through papers is preceded after a new concept comes out, it can be helpful to figure out the public's perception.

This study involves several limitations. First, owing to the concentration of data on a few major countries, it was impossible to accurately assemble data from all over the world. Second, the data collection was limited to 2016–2022. Because the concept of ESG was introduced in 2006, future studies should expand the data collection to more extended periods.

Acknowledgement: We appreciate Editage (www.editage.co.kr) for their English editing service.

Funding Statement: This study was supported by a National Research Foundation of Korea (NRF) (http://nrf.re.kr/eng/index) grant funded by the Korean government (RS-2023-00208278).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] G. Hagart and I. Knoepfel, Communicating ESG value drivers at the company-investor interface: Who Cares Wins annual event 2006-conference report. Washington: World Bank Group, 2006.
[2] A. G. Hoepner, A. A. Majoch and X. Y. Zhou, "Does an asset owner's institutional setting influence its decision to sign the principles for responsible investment?," *Journal of Business Ethics*, vol. 168, no. 2, pp. 389–414, 2021.
[3] F. Ferriani and F. Natoli, "ESG risks in times of Covid-19," *Applied Economics Letters*, vol. 28, no. 18, pp. 1537–1541, 2021.
[4] M. Minkkinen, A. Niukkanen and M. Mäntymäki, "What about investors? ESG analyses as tools for ethics-based AI auditing," *AI & Society*, pp. 1–15, 2022.
[5] H. Seo, D. H. Jo and Z. Pan, "Big data analysis of 'ESG' news using topic modeling," in *Proc. Int. Conf. on Big Data, Cloud Computing, and Data Science (BCD)*, Danang, Vietnam, pp. 183–187, 2022.
[6] S. R. Bapat, S. Kothari and R. Bansal, "Sentiment analysis of ESG disclosures on stock market," 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2210.00731
[7] M. H. Amin, E. K. Mohamed and A. Elragal, "CSR disclosure on Twitter: Evidence from the UK," *International Journal of Accounting Information Systems*, vol. 40, pp. 100500, 2021.

[8] G. Serafeim and A. Yoon, "Which corporate ESG news does the market react to?," *Financial Analysts Journal*, vol. 78, no. 1, pp. 59–78, 2022.
[9] A. Sokolov, J. Mostovoy, J. Ding and L. Seco, "Building machine learning systems for automated ESG scoring," *The Journal of Impact and ESG Investing*, vol. 1, no. 3, pp. 39–50, 2021.
[10] I. Goloshchapova, S. H. Poon, M. Pritchard and P. Reed, "Corporate social responsibility reports: Topic analysis and big data approach," *The European Journal of Finance*, vol. 25, no. 17, pp. 1637–1654, 2019.
[11] F. Zhao, Y. Zhu, H. Jin and L. T. Yang, "A personalized hashtag recommendation approach using LDA-based topic model in microblog environment," *Future Generation Computer Systems*, vol. 65, pp. 196–206, 2016.
[12] J. Yun and Y. Geum, "Automated classification of patents: A topic modeling approach," *Computers & Industrial Engineering*, vol. 147, pp. 106636, 2020.
[13] S. Tonidandel, K. M. Summerville, W. A. Gentry and S. F. Young, "Using structural topic modeling to gain insight into challenges faced by leaders," *The Leadership Quarterly*, vol. 33, no. 5, pp. 101576, 2022.
[14] A. Li, Y. Huang and K. W. Axhausen, "An approach to imputing destination activities for inclusion in measures of bicycle accessibility," *Journal of Transport Geography*, vol. 82, pp. 102566, 2020.
[15] M. Honnibal and I. Montani, "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," *To Appear*, vol. 7, no. 1, pp. 411–420, 2017.
[16] H. Jang, E. Rempel, D. Roth, G. Carenini and N. Z. Janjua, "Tracking COVID-19 discourse on Twitter in North America: Infodemiology study using topic modeling and aspect-based sentiment analysis," *Journal of Medical Internet Research*, vol. 23, no. 2, pp. e25431, 2021.
[17] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2203.05794
[18] F. Crespi and M. Migliavacca, "The determinants of ESG rating in the financial industry: The same old story or a different tale?," *Sustainability*, vol. 12, no. 16, pp. 6398, 2020.
[19] G. H. Ionescu, D. Firoiu, R. Pirvu and R. D. Vilag, "The impact of ESG factors on market value of companies from travel and tourism industry," *Technological and Economic Development of Economy*, vol. 25, no. 5, pp. 820–849, 2019.
[20] N. Chevrollier, J. Zhang, T. van Leeuwen and A. Nijhof, "The predictive value of strategic orientation for ESG performance over time," *Corporate Governance: The International Journal of Business in Society*, vol. 20, pp. 123–142, 2020.
[21] S. Zainullin and O. Zainullina, "Scientific review digitalization of corporate culture as a factor influencing ESG investment in the energy sector," *International Review*, vol. 2021, no. 1–2, pp. 130–136, 2021.
[22] N. Naeem and S. Cankaya, "The impact of ESG performance over financial performance: A study on global energy and power generation companies," *International Journal of Commerce and Finance*, vol. 8, no. 1, pp. 1–25, 2022.
[23] H. M. Chen, T. C. Kuo and J. L. Chen, "Impacts on the ESG and financial performances of companies in the manufacturing industry based on the climate change related risks," *Journal of Cleaner Production*, vol. 380, pp. 134951, 2022.
[24] Y. Chang, "Green finance in Singapore: Barriers and solutions," *SSRN Electronic Journal*, 2019. https://doi.org/10.2139/ssrn.3326287
[25] U. Volz, "Fostering green finance for sustainable development in Asia," in *Routledge Handbook of Banking and Finance in Asia*, 1st ed., London, United Kingdom: Routledge, chapter. 27, pp. 488–504, 2018.
[26] S. Dikau and U. Volz, "Central banking, climate change and green finance," in *Handbook of Green Finance*. Singapore: Springer, 2018.
[27] S. Piserà and H. Chiappini, "Are ESG indexes a safe-haven or hedging asset? Evidence from the COVID-19 pandemic in China," *International Journal of Emerging Markets*, vol. 37, pp. 200, 2022.
[28] E. Andersson, M. Hoque, M. L. Rahman, G. S. Uddin and R. Jayasekera, "ESG investment: What do we learn from its interaction with stock, currency and commodity markets?," *International Journal of Finance & Economics*, vol. 27, no. 3, pp. 3623–3639, 2022.

[29] G. Rubbaniy, A. A. Khalid, M. F. Rizwan and S. Ali, "Are ESG stocks safe-haven during COVID-19?," *Studies in Economics and Finance*, vol. 39, pp. 239–255, 2021.
[30] B. Yoon, J. H. Lee and R. Byun, "Does ESG performance enhance firm value? Evidence from Korea," *Sustainability*, vol. 10, no. 10, pp. 3635, 2018.
[31] B. Yoon, J. H. Lee and J. H. Cho, "The effect of ESG performance on tax avoidance—Evidence from Korea," *Sustainability*, vol. 13, no. 12, pp. 6729, 2021.
[32] J. Hwang, H. Kim and D. Jung, "The Effect of ESG activities on financial performance during the COVID-19 pandemic—Evidence from Korea," *Sustainability*, vol. 13, no. 20, pp. 11362, 2021.
[33] R. G. Eccles, M. D. Kastrapeli and S. J. Potter, "How to integrate ESG into investment decision-making: Results of a global survey of institutional investors," *Journal of Applied Corporate Finance*, vol. 29, no. 4, pp. 125–133, 2017.
[34] M. Folqué, E. Escrig-Olmedo and T. Corzo Santamaría, "Sustainable development and financial system: Integrating ESG risks through sustainable investment strategies in a climate change context," *Sustainable Development*, vol. 29, no. 5, pp. 876–890, 2021.
[35] A. Raghunandan and S. Rajgopal, "Do ESG funds make stakeholder-friendly investments?," *Review of Accounting Studies*, vol. 27, no. 3, pp. 822–863, 2022.
[36] I. S. Popescu, C. Hitaj and E. Benetto, "Measuring the sustainability of investment funds: A critical review of methods and frameworks in sustainable finance," *Journal of Cleaner Production*, vol. 314, pp. 128016, 2021.
[37] G. Giese, L. E. Lee, D. Melas, Z. Nagy and L. Nishikawa, "Foundations of ESG investing: How ESG affects equity valuation, risk, and performance," *The Journal of Portfolio Management*, vol. 45, no. 5, pp. 69–83, 2019.
[38] A. S. Garcia, W. Mendes-Da-Silva and R. J. Orsato, "Sensitive industries produce better ESG performance: Evidence from emerging markets," *Journal of Cleaner Production*, vol. 150, pp. 135–147, 2017.
[39] E. Van Duuren, A. Plantinga and B. Scholtens, "ESG integration and the investment management process: Fundamental investing reinvented," *Journal of Business Ethics*, vol. 138, no. 3, pp. 525–533, 2016.
[40] G. Bressan, I. Monasterolo and S. Battiston, "Sustainable investing and climate transition risk: A portfolio rebalancing approach," *The Journal of Portfolio Management*, vol. 48, no. 10, pp. 165–192, 2022.
[41] M. H. Shakil, "Environmental, social and governance performance and financial risk: Moderating role of ESG controversies and board gender diversity," *Resources Policy*, vol. 72, pp. 102144, 2021.
[42] P. Bolton and M. Kacperczyk, "Do investors care about carbon risk?," *Journal of Financial Economics*, vol. 142, no. 2, pp. 517–549, 2021.
[43] H. Amiraslani, K. V. Lins, H. Servaes and A. Tamayo, "Trust, social capital, and the bond market benefits of ESG performance," *Review of Accounting Studies*, pp. 1–42, 2022.
[44] B. Gerard, "ESG and socially responsible investment: A critical review," *Beta*, vol. 33, no. 1, pp. 61–83, 2019.
[45] N. Apergis, T. Poufinas and A. Antonopoulos, "ESG scores and cost of debt," *Energy Economics*, vol. 112, pp. 106186, 2022.
[46] A. Behl, P. S. Kumari, H. Makhija and D. Sharma, "Exploring the relationship of ESG score and firm value using cross-lagged panel analyses: Case of the Indian energy sector," *Annals of Operations Research*, vol. 313, no. 1, pp. 231–256, 2022.
[47] C. Zhao, Y. Guo, J. Yuan, M. Wu, D. Li *et al.,* "ESG and corporate financial performance: Empirical evidence from China's listed power generation companies," *Sustainability*, vol. 10, no. 8, pp. 2607, 2018.
[48] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proc. ICML*, Pennsylvania, USA, pp. 113–120, 2006.
[49] N. S. Eccles, "UN principles for responsible investment signatories and the anti-apartheid SRI movement: A thought experiment," *Journal of Business Ethics*, vol. 95, no. 3, pp. 415–424, 2010.
[50] A. A. Majoch, A. G. Hoepner and T. Hebb, "Sources of stakeholder salience in the responsible investment movement: Why do investors sign the principles for responsible investment?," *Journal of Business Ethics*, vol. 140, no. 4, pp. 723–741, 2017.

[51] J. Diener and A. Habisch, “A plea for a stronger role of non-financial impact in the socially responsible investment discourse,” *Corporate Governance: The International Journal of Business in Society*, vol. 21, pp. 294–306, 2020.

[52] W. Leal Filho, “Accelerating the implementation of the SDGs,” *International Journal of Sustainability in Higher Education*, vol. 21, no. 3, pp. 507–511, 2020.

[53] V. Litvinenko, I. Bowbrik, I. Naumov and Z. Zaitseva, “Global guidelines and requirements for professional competencies of natural resource extraction engineers: Implications for ESG principles and sustainable development goals,” *Journal of Cleaner Production*, vol. 338, pp. 130530, 2022.