

Comparison of Research Topics on Transportation Decarbonization Between Asian and Non-Asian Regions:

Using Topic Modeling and Machine Learning Algorithms*

Tsolmon Bayarsaikhan** · Tae-Hyoung Tommy Gim***

Abstract: The growing global interest to decarbonize the transportation industry has resulted in numerous scientific publications. This study reviews the rapidly expanding body of research and identifies the knowledge gaps in transport decarbonization between regions. This study employs a hybrid approach combining topic modeling and machine learning to identify research topics and their knowledge structures, and then compares the main debated topics between Asia and non-Asian regions. A dataset of 777 articles, including 410 Asian and 367 non-Asian articles, published between 1990 and 2022 was extracted from the Scopus database. The latent Dirichlet allocation topic modeling results showed that five potential topics were derived from Asia, while six were derived from non-Asian regions, and the knowledge structure of each topic differed between the two regions. The K-nearest neighbor machine learning algorithm results indicated a 92% accuracy for Asian topics and an 89% accuracy for non-Asian topics. The findings suggest that the Asian studies focused on “energy use in transportation” and “drivers of CO₂ emissions in transportation,” while the non-Asian studies focused on “electric vehicles” and “fuel consumption.” This paper will keep academics and practitioners updated on the paradigm shift in the research on transportation decarbonization.

Key Words: Decarbonization, Transportation, Asia, Topic Modeling, Machine Learning

* This research was supported by the Research Grants for Asian Studies funded by Seoul National University Asia Center (SNUAC) in 2021 (0448A-20210073)

** First Author, Ph.D. candidate, Graduate School of Environmental Studies, Seoul National University

*** Corresponding Author, Associate Professor, Graduate School of Environmental Studies, Interdisciplinary Program in Landscape Architecture, and Environmental Planning Institute, Seoul National University

I. Introduction

The challenge of decarbonizing transportation is one of the most pressing issues in the transportation sector. Transport decarbonization is defined as initiatives to reduce unnecessary travel, shift necessary travel to sustainable modes, and promote carbon-neutral mobility and energy solutions for reducing transportation carbon emissions (Organisation for Economic Co-operation and Development [OECD], 2021). In other words, reducing carbon emissions from transport is critical to accelerating the decarbonization trend (Meyer, 2020) and meeting the Paris Agreement goal of limiting global warming to under 1.5 degrees Celsius (United Nations [UN], 2021). According to the International Energy Agency (IEA), transportation is the second largest carbon dioxide producer by sector, accounting for 37% of global carbon dioxide (CO₂) emissions from end-use sectors in 2021 (IEA, 2022). Despite global efforts to decarbonize transportation, carbon emissions in the transportation sector have continued to rise, reaching 8% to about 7.7 Gt CO₂ in 2021. That increased level is significantly different by region.

In particular, Asia is responsible for more than half (41% in 2021) of global carbon emissions (IEA, 2022). Asia recorded the most significant increase in transportation CO₂ emissions of any region, along with the highest reliance on fossil fuels (Foster et al., 2021). Thus, Asia's influence on global CO₂ emissions from the transportation sector cannot be neglected. In this sense, the current state of transportation decarbonization differs substantially depending on the disparity of each region. As a result, decarbonizing transportation requires an approach from a regional standpoint to

close the gap between regions.

Despite the fact that Asia is at the forefront of transportation decarbonization research and practice (Meyer, 2020), few studies have investigated the academic knowledge structure characteristics and crucial research areas in the Asian region. Existing bibliometric research (Tian et al., 2018; Meyer, 2020) found that Asian publications dominated in the areas of transportation carbon emissions and decarbonization. They have noted the importance of the Asian region in research on carbon emissions in the transportation sector, but there has been insufficient coverage of major areas and topics.

By contrast, global interest in transportation decarbonization is growing, as carbon emissions in the transportation sector must be reduced from 20% to less than 6 Gt by 2030 to meet the net-zero goal (IEA, 2022). As a result, scientific publications in this field are increasing dramatically. Existing research has been undertaken in a variety of disciplines and perspectives, including an empirical analysis of transportation emissions drivers (e.g., Gim, 2022; Wang and Su, 2020), scenario analysis of future emissions (e.g., Mohamed, 2016), and conceptual and policy analysis of transportation decarbonization (e.g., Cleophas et al., 2019).

Notwithstanding these academic efforts, technological solutions or carbon-cutting measures remain unclear. In light of this, some review studies have reviewed existing research on transport carbon emissions to gain insight into existing areas of academic research using different approaches. As for methods, they were primarily conducted through content analysis, bibliometric and network analysis (Tian et al., 2018; Meyer, 2020), meta-analysis (Chiaramonti et al., 2020), and systematic review (Emodi et al., 2022).

These reviews generally discussed transportation research, but they were either too broad (e.g., sustainable transport) or not directly related to transportation decarbonization (e.g., low-carbon fuels). Among them, Meyer (2020) and Emodi et al. (2022) are the only reviews that address the research on transport decarbonization. However, Meyer (2020) analyzed general trends in journals, institutions, and authors with related papers, whereas Emodi et al. (2022) systematically reviewed the evolution of a small number of articles on the decarbonization of transport in developing countries.

In this regard, we intend to go one step beyond their bibliometric findings by focusing on key areas and knowledge structures in transportation decarbonization research. Besides, these studies primarily focused on bibliometric analysis on an international or developing-country scale, so the results did not reflect the state of the literature on a regional scale. None of the previous attempts have been made region-wise, and comparisons between Asia and other regions were not even considered.

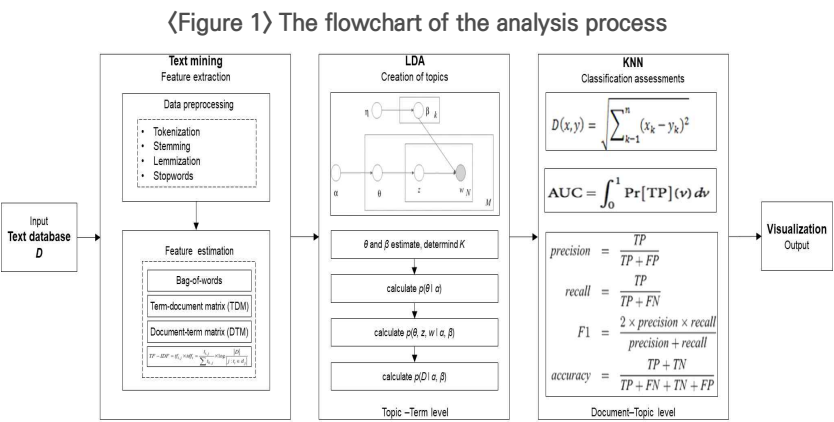
Against this backdrop, this present study aims to analyze the literature on the decarbonization of the transportation sector in Asia and non-Asian regions to identify core research topics and their knowledge structures and compare their similarities and differences on key topics discussed in these regions. To this aim, we performed the analysis using a hybrid approach combining topic modeling and machine learning algorithms based on the transport decarbonization-related articles published from 1990 to 2022.

Our study results can help identify future research gaps by revealing key areas that need further attention in each region and structuring a large body of existing knowledge. The rest of our work

is organized as follows. Section 2 describes the data collection process and research methodology; Section 3 presents the results of topic modeling and machine learning analysis; and Section 4 discusses major findings and outlines a future research direction.

II. Materials and methods

We employed a hybrid approach combining text mining, topic modeling, and machine learning to identify potential topics and their characteristics of knowledge structure in the literature on transportation decarbonization in Asian and non-Asian countries. Our analysis process consisted of the following steps: data collection, data preprocessing, text mining analysis, topic modeling, machine learning, and visualization. The flowchart of the analysis process is shown in <Figure 1>.



1. Data source

We first collected existing academic papers related to the transportation decarbonization from the SCOPUS database. The search string process was based on a two-level assembly structure, with level one containing transportation-related keywords (e.g., “transportation” OR “transport”) and level two containing decarbonization- related keywords (e.g., “decarboniz(s)ation” OR “decoupling” OR “decarboni*” OR “carbon emission”). The publication period was set to indefinite for all related articles published by August 1, 2022. Only articles written in English were selected.

The initial set included 1,342 academic papers that matched at least one of the keywords either in the title, abstract, or keywords. The screening process excluded non-peer-reviewed publications (e.g., conference papers, letters, book chapters), insufficient papers (e.g., articles without abstracts or keywords), irrelevant papers (e.g., articles unrelated to transportation), and duplicated papers. After the screening process, a total of 779 articles were included in the final dataset.

2. Data description

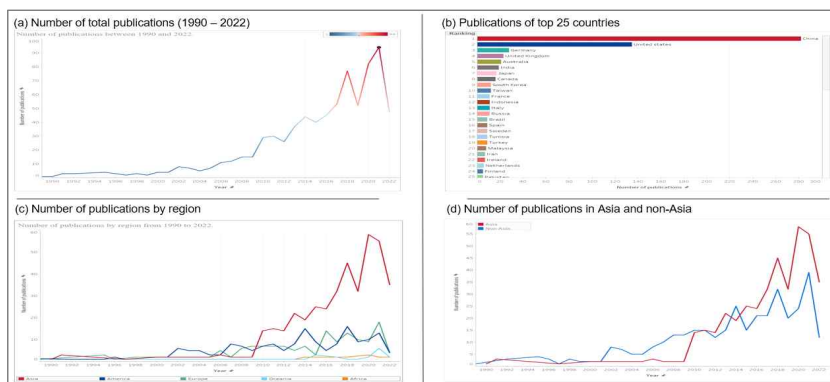
We divided the final dataset into two regions to compare the transport decarbonization articles from Asian and non-Asian countries. To do this, we first divided the article (classification of the collaborative article was based on the first author's country) by country according to the author's affiliation and country table provided by SCOPUS. During this process, the final dataset contained 777 articles, except for two that missed country information.

As shown in <figure 2(a)>, the total number of publications has significantly increased from 1990 to 2022. Particularly, academic interest in transport decarbonization has grown sharply since 2015; it seems to positively reflect the Paris Agreement's establishment of the goal of achieving net zero emissions. Between 1990 and 2022, 65 different countries (Asia=21, non-Asia=44) published articles related to transportation decarbonization. The top five publishing countries were China (32%), the United States (19%), Germany (9%), the United Kingdom (7%), and Australia (5%) (see <Figure 2(b)>).

Next, we classified the articles into five regions based on the United Nations (UN) region classification criteria: America (22%), Asia (53%), Europe (20%), Africa (3%), and Oceania (2%) (see <Figure 2(c)>). As a result of regional classification, articles from Asian countries accounted for 53% of the total articles compared to other regions, so we combined all other regions into non-Asian countries.

Finally, 410 articles (53%) from Asian countries and 367 articles (47%) from non-Asian countries were included in the final dataset. A comparison of publication trends in the two regions is shown in <Figure 2(d)>. Before 2009, non-Asian countries published more than Asian countries, but Asian publications increased dramatically after 2010 and began to dominate in 2015. Overall, the number of publications in Asia and non-Asian increased significantly between 1990 and 2022.

〈Figure 2〉 Publications trends in research of transport decarbonization



3. Analytical methods

(1) Data pre-processing: The procedures for cleaning and preprocessing textual data using Natural Language Processing (NLP) are as follows. First, Tokenization process was used to split the textual data into tokens (simple units). Next, Parts of speech (POS) tagging was set to label each token with a tag (NN, VB, DT, etc.) based on the word type (noun, verb, adverb, etc.), and we extracted only nouns or noun phrases for analysis. Then, Stemming was set to process several words with similar meanings into a single word (e.g., transportation, transport > transportation). Next, Lemmization was set to that complex words (e.g., green, house, gas > greenhouse gas) were not extracted separately. Lastly, Stopwords was set to exclude unnecessary words (e.g., study, purpose) frequently occurring in the document corpus.

(2) Text mining: Following data pre-processing, we used the bag of words (BOWs) to extract the keywords with the high frequency of common occurrence in the documents of each region. Then, in preparation for topic modeling analysis, we generated a term-document

matrix (TDM) and a document-term matrix (DTM). We also estimate the term frequency-inverse document frequency (TF-IDF) and word distance weights for weight settings. The TF-IDF is a combination of two indicators: term frequency and inverse document frequency, and is computed by multiplying the term frequency by the inverse document frequency (Bai et al., 2021). In the TF-IDF equation shown in <Figure 1>, $TF(t,d)$ represents the term frequency of term t within document d , and IDF represents the inverse document frequency of term t .

(3) Topic modeling: The Latent Dirichlet Allocation (LDA) topic model was employed to reveal key latent topics and determine their knowledge structure in the literature corpus by each region. LDA is a probabilistic-based classification technique that extracts content from unsupervised data (Boyer et al., 2017). That is, it is an appropriate method for exploring unobserved information and identifying latent themes and their semantic structures in large volumes of textual data (Bayarsaikhan et al., 2022; Cao et al., 2009).

For inference of the LDA, we set the TF-IDF threshold (λ) = 0.05 (mean value), iterations = 1000, and hyperparameters alpha (α) = 0.1 and beta (β) = 0.01. Both α and β are parameters for determining the topic modeling probability (Griffiths and Steyvers, 2004). In order to determine the optimal number of topics and to derive the probability distribution of the terms in each topic, we used the perplexity of the Gibbs sampling algorithm (Griffiths and Steyvers, 2004). The perplexity of Gibbs Sampling is a sampling method from a joint distribution when conditional distributions of topics and words can be computed efficiently (Boyer et al., 2017). The generating process of LDA is presented in <Figure 1>.

(4) Machine learning: The K-Nearest Neighbor (KNN) algorithm was

utilized to evaluate the accuracy of latent topics extracted from LDA and to predict the fit of the topic classifier by supervised data. In the process of KNN, we first applied each dataset from Asia and non-Asian countries by dividing them into training datasets (i.e., classified data) and test datasets (i.e., non-classified data). We set training data to 80% and text data to 20% for our machine-learning experiments. That is, the standard for the training-to-test ratio is 80% to 20%, as it is more effective when the training dataset ratio is higher than the test dataset ratio (Casillo et al., 2021).

We also applied the Euclidean distance parameter to calculate document-topic level similarity. The Euclidean distance indicates a higher degree of similarity between the documents based on their distance value (Gallego et al., 2018). We set the topic similarity threshold (λ) = 0.05, iterations = 10, and the number of topics in the trained model of each region to the same as in the LDA result. In KNN, the confusion matrix can present the proportion of topics between the predicted and actual as well as the correct and incorrect classification of the classified topics (Casillo et al., 2021). Therefore, the confusion matrix is used to check the accuracy of the classified topics based on the document-topic matrix generated by LDA. The standard parameters AUC-ROC, classification accuracy (CA), F-criteria (F1), precision, and recall (Casillo et al., 2021) are applied to evaluate the accuracy of classified topics.

III. Results

1. LDA topic modeling results

As a result of the perplexity of the Gibbs Sampling iterations by setting the number of topics in the range of 2 to 15, the most optimal number of topics in Asia was six (perplexity $\log=66.735$), while the most suitable number of topics in non-Asia was five (perplexity $\log=57.797$). As shown in <Table 1> and <Figure 3>, six topics were derived from the Asian dataset.

Topic 1 accounted for about 16% of the total dataset and includes core terms (i.e., sub-theme) such as “effect,” “carbon neutrality,” “infrastructure,” and “urban.” Among these core terms, the topic probability value of the term “effect” is the highest, indicating that this includes research on the impact of carbon emissions in the transportation sector. Thus, the label of topic 1 was set as “drivers of transport CO₂ emissions.” Topic 2 indicated the second most covered topic in Asian literature, which included approximately 18% of all papers. Topic 2 consists of core terms such as “energy saving,” “reduction,” “urban,” “strategy,” and “public transport,” indicating that it mainly deals with topics related to reducing transportation carbon emissions. Topic 3 accounted for 14% of the total dataset, and the topic probability of terms such as “freight,” “road,” “trade,” “cost,” and “material flow” was the highest.

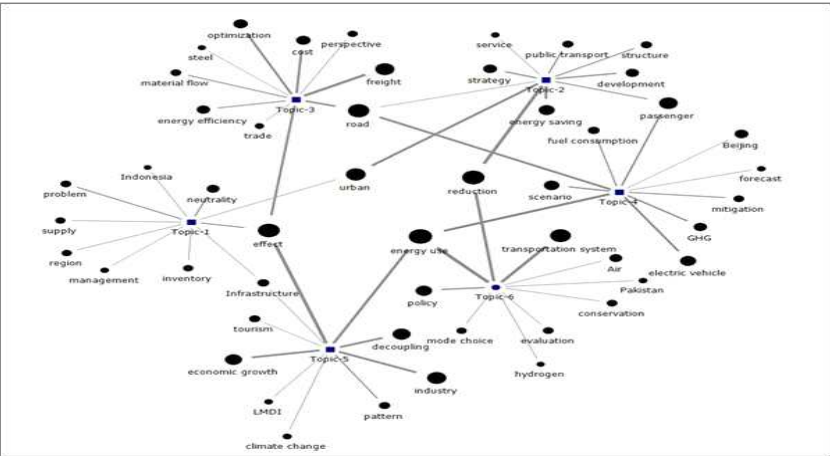
Therefore, it can be inferred that this topic deals with decarbonization in terms of road freight transport. Topic 4 accounted for 13%, the lowest proportion compared to other topics, which focused on electric vehicles. It shows that in this topic, key terms such as “fuel consumption”, “Greenhouse Gas (GHG),” “energy use,”

and “mitigation” are connected to electric vehicles. Topic 5 accounted for about 15%, and the topic probability value of “economic growth,” “decoupling,” “industry,” and “climate change” terms was the highest. Finally, topic 6 accounted for roughly 18% of the dataset, indicating that it is the most discussed topic compared to other topics. This topic consists of major related terms such as “energy use,” “transportation system,” “policy,” and “conservation” and includes studies dealing with energy consumption in the transport sector.

〈Table 1〉 LDA topic modeling results of Asian dataset

No.	Sub-themes (T/P)	Topic label	Freq. (%)
Topic 1	Effect (0.040), neutrality (0.033), infrastructure (0.032), inventory (0.026), supply (0.024), region (0.022), urban (0.015), management (0.015), problem (0.013), Indonesia (0.013)	Drivers of transport CO ₂ emissions	66 (16%)
Topic 2	Energy saving (0.127), reduction (0.076), urban (0.065), strategy (0.037), public transport (0.035), passenger (0.031), travel behavior (0.031), road (0.030), development (0.018), structure (0.015)	Reduction of transport CO ₂	75 (18%)
Topic 3	Freight (0.091), road (0.061), trade (0.053), cost (0.051), energy efficiency (0.051), optimization (0.032), material flow (0.026), perspective (0.020), steel (0.019), effect (0.015)	Road freight	57 (14%)
Topic 4	Electric vehicle (0.153), fuel consumption (0.066), passenger (0.046), GHG (0.044), energy use (0.043), mitigation (0.039), Beijing (0.031), gas (0.026), scenario (0.022), road (0.015)	Electric vehicles	51 (13%)
Topic 5	Economic growth (0.098), decoupling (0.086), industry (0.072), climate change (0.070), pattern (0.070), tourism (0.031), energy use (0.022), energy efficiency (0.018), infrastructure (0.018), LMDI (0.017)	Decoupling of transport CO ₂	61 (15%)
Topic 6	Energy use (0.194), transportation system (0.183), reduction (0.069), policy (0.044), conservation (0.025), air (0.022), evaluation (0.022), mode choice (0.019), Pakistan (0.014), hydrogen (0.013)	Energy use of transportation	100 (24%)

〈Figure 3〉 Latent topic structure of Asia



Meanwhile, the optimal number of topics derived from non-Asian datasets is five (see 〈Table 2〉 and 〈Figure 4〉). Topic 1 accounted for about 16% of the total dataset and includes core terms such as “road,” “energy efficiency,” “climate change,” and “passenger.” The topic probability value of the term “road” is the highest, indicating that includes research related to road transport decarbonization. Topic 2 indicated the second most covered topic in Asian literature, which included approximately 24%. Topic 2 consists of core terms such as “GHG,” “alternative fuels,” and “technology;” it implies that it mainly deals with topics related to energy consumption in the transport sector.

Topic 3 accounted for 11% of the total dataset, the lowest proportion compared to other topics, focused on reducing transport CO₂. The topic probability of terms such as reduction, “Environmental Kuznets Curve (EKC),” “decoupling,” “mode choice,” “energy efficiency,” “passenger,” and “economic growth” was the highest. Topic 4 accounted for 22%, and the topic probability of terms such as “fuel consumption,” “transportation

system,” “gas,” “public transport,” “hydrogen,” and “oil” was the highest. It shows that this topic deals with fuel consumption issues in transport decarbonization. Finally, topic 5 accounted for about 25% of the dataset, which was the most covered topic compared to the other topics. This topic addressed electric vehicles related to the terms “carbon capture and storage (CCS),” “pollution,” and “supply.”

〈Table 2〉 LDA topic modeling results of non-Asian dataset

No.	Sub-themes by terms (T/P)	Topic label	Freq. (%)
Topic 1	Road (0.150), energy efficiency (0.123), infrastructure (0.097), climate change (0.054), policy (0.045), passenger (0.044), reduction (0.030) rail (0.029), mitigation (0.027), United Kingdom (0.022)	Road transport	59 (16%)
Topic 2	GHG (0.162), fuel consumption (0.151), impact (0.147) technology (0.135), alternative fuels (0.104), freight (0.077), mobility (0.075), policy (0.045), cost (0.043), road (0.042), material flow (0.035), saving (0.032), evolution (0.031), zero-emissions (0.031)	Energy use of transport	89 (24%)
Topic 3	Reduction (0.198), urban (0.181), transportation sector (0.121), Environmental Kuznets Curve (0.075), Decoupling (0.065), mode choice (0.063), energy efficiency (0.062), passenger (0.052), economic growth (0.040), neutrality (0.033)	Reduction of transport CO ₂	43 (11%)
Topic 4	Fuel consumption (0.128), GHG (0.100), transportation system (0.099), electric vehicle (0.095), gas (0.078), public transport (0.054), life cycle assessment (0.049), hydrogen (0.042), oil (0.039), conversion (0.038)	Fuel consumption	83 (22%)
Topic 5	Electric vehicle (0.158), United States (0.120), CCS (0.094), pollution (0.083), plan (0.070), travel (0.062), air (0.042), monoxide (0.040), supply (0.038), source (0.035)	Electric vehicles	93 (25%)

at an appropriate level.¹⁾

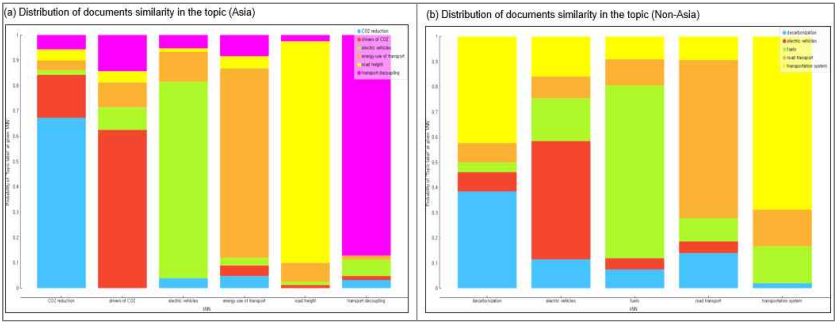
〈Figure 5〉 shows the distribution of documents by topic extracted from the LDA. In Asia, similarities between documents accounted for the highest proportion of topics as “decoupling of transport carbon emissions,” “freight road transport,” and “energy use of transport” in the order of 89%, 82%, and 76%, respectively. On the other hand, the similarity between documents distributed in the rest topics of “transport CO₂ drivers,” “transport CO₂ reduction,” and “electric vehicles” was slightly lower, which seems to be mixed with documents from all other topics. In non-Asia, similarities between documents were high in topics of “fuel consumption,” “road transport,” and “transportation energy use,” and the distribution of documents by topic accounted for the highest proportion in the order of 82%, 79%, and 72%, respectively.

By contrast, CO₂ reduction and electric vehicle topics were mixed in with others, as were Asian topics. The similarity between the documents distributed on the topics of “reduction of transport CO₂” and “electric vehicles” was slightly lower, and shows that they are mixed with documents of others similar to the results of the Asian topic distribution. Therefore, it can be concluded that these topics, which consist of mixed-documents, appear to be dominated by articles dealing with incorporated combination research, or are areas of research that have not yet been sufficiently addressed as core topics in this field.

1) Asia: Topic 1 (d)=0.014, Topic 2 (d)=0.039, Topic 3 (d)=0.006, Topic 4 (d)=0.022, Topic 5 (d)=0.075, Topic 6 (d)=0.023

Non-Asia: Topic 1 (d)=0.026, Topic 2 (d)=0.031, Topic 3 (d)=0.033, Topic 4 (d)=0.057, Topic 5 (d)=0.015

〈Figure 5〉 The distribution of documents similarity in the topic



2) Evaluation of the accuracy of classified topics extracted from LDA.

We evaluated the accuracy of classified topics extracted from LDA using standard parameters of machine learning algorithms. The score range of performance parameters is 0 to 1, which means that the closer the value is to 1, the higher the classification accuracy (Casillo et al., 2021). 〈Table 3〉 and 〈Figure 6〉 present the classification evaluation results for the proposed topic model by LDA. The AUC was 0.925 in Asia and 0.892 in non-Asia.

The classification accuracy (CA) was 0.888 in Asia and 0.660 in non-Asian regions. The F1, precision, and recall values also achieved high accuracy for the topic classification of each region. In other words, these results indicated that the topic classifications of both Asian and non-Asian datasets were accurately classified. Thus, it implies that the proposed model with latent topics can significantly improve the performance of unsupervised textual data classification.

The confusion matrix can present the proportion of topics between the predicted and actual as well as the correct and incorrect classification of the classified topics (Kulkarni et al., 2020). Therefore, the confusion matrix is used to observe which specific topics were

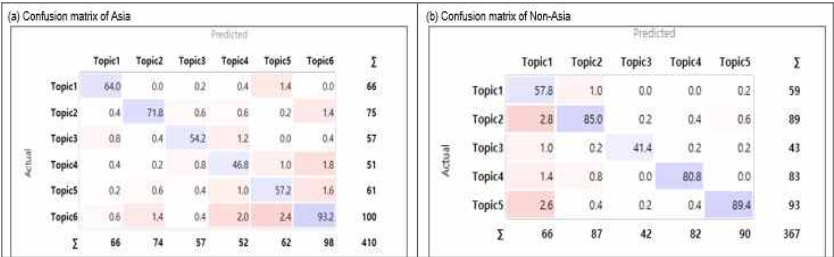
misclassified and/or how they were classified. <Figure 6> presents the confusion matrix obtained from each dataset (i.e., Asia = 6×6, non-Asia = 5×5).

In the confusion matrix, the rows represent the actual numbers of the document samples, while the columns reflect the estimates of the model. The sum of the total probabilities of each document shows a slight difference between the actual and predicted. The purple diagonal also displays the correct classification ratio, and the rest shows the ratio of incorrect classification displayed in pink. Among Asian topics, 64%~93% were correctly classified, whereas 57%~89% of non-Asian topics were correctly classified. Regarding misclassification, however, Topic 1 in Asia and Topic 2 in non-Asia had slightly higher error classifications.

<Table 3> The classification evaluation on the proposed topic model

KNN model	N	AUC	CA	F1	Precision	Recall
Asia	410	0.925	0.888	0.884	0.889	0.888
Non-Asia	367	0.892	0.660	0.608	0.743	0.660

<Figure 6> The confusion matrix of KNN



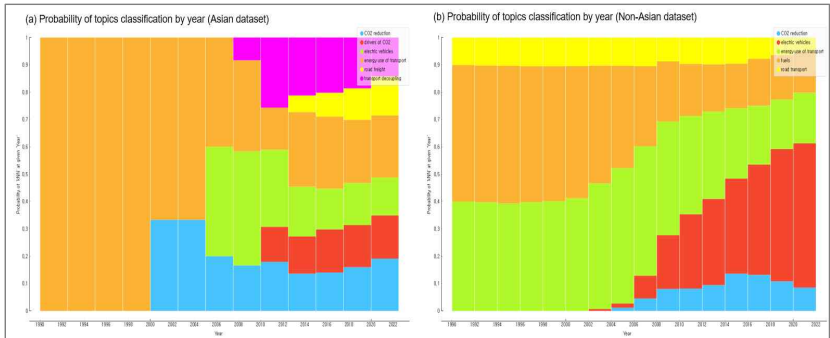
3) Histogram results for document–topic probability

<Figure 7> depicts the histogram results of the document–topic

probability. Asian document flows show that topics related to transportation decarbonization were primarily addressed in terms of energy in the transportation sector between 1990 and 2000. Then, in the early 2000s, investigations about reducing transportation carbon emissions began. Since 2010, transportation decarbonization issues have been debated in research areas such as “transportation CO₂ emission drivers,” “transportation decoupling,” and “electric vehicles.”

By contrast, non-Asian literature has discussed transport decarbonization topics in “fuel,” “road transport,” and “transport energy use” research since 1990, but the probability of distributing this body of literature has decreased since 2008. It implies that this is due to the emergence of new research areas rather than a decrease in relevant research topics. Since the mid-2000s, there has been an increase in both Asian and non-Asian literature on “transport CO₂ reduction” and “electric vehicles.” Notably, the topic of transportation decarbonization has been overgrown in the non-Asian electric vehicle-related literature since 2008. It may reflect the rapid expansion of international academic interest in this transportation decarbonization research area.

〈Figure 7〉 Trend histograms for document–topic probabilities



IV. Discussions and Conclusion

This study examined the main research topics and their knowledge structures and compared the key topics discussed in the literature on the decarbonization of the transportation sector in Asian and non-Asian regions. For this purpose, we analyzed based on 777 articles (Asia=410, non-Asia=367) on transportation decarbonization published from 1990 to 2022 using a hybrid approach that combines LDA topic modeling and KNN machine learning algorithms.

Our analysis yielded the following key findings. First, the growth rate of publications related to transportation decarbonization increased by 97.1% in Asia and 85.7% in non-Asia between 1990 and 2022. Particularly, Asian publications have increased dramatically since 2010. As reported in the previous work (Tian et al., 2018, Meyer, 2020), the pace of publication in Asia is expected to grow. However, China leads in Asia, while the United States and Europe lead in non-Asian countries, suggesting that further contributions from other countries will be required.

Second, LDA results show that topics such as “energy use of transportation sector,” “reduction of transport CO₂ emissions,” “electric vehicles,” and “road transport” are commonly addressed for transport decarbonization in the two regions. The topic differences between the two regions are that Asian research has focused on “transport energy use” and “transport CO₂ emissions drivers,” whereas non-Asian studies have focused on “electric vehicles” and “fuel consumption.”

Furthermore, while the topics derived from the transportation decarbonization study in Asia and non-Asia appeared to cover similar scopes, the structure of each topic clearly differed between the two

regions. Notably, the topic structure of “road transport decarbonization” revealed that non-Asian studies had led by “passenger” road-related themes, whereas Asian studies had led by “freight” road-related themes. In the structure of “electric vehicle” topics, non-Asia has a strong connection with “fuel consumption” topics and consists of “alternative fuels,” “technologies,” and “CCS”-related sub-themes, whereas Asia has a strong connection with “road transport” and “energy use” topics and consists of “mitigation,” “GHG,” and “scenarios”-related sub-themes.

Finally, the KNN results showed that the similarity of documents distributed under the topics of “reduction of transport CO₂” and “electric vehicles” was low in both regions. It can be concluded that these mixed-document topics appear to be dominated by articles dealing with incorporated research or that they are areas of research that have not yet been fully addressed as core topics in this field. In addition, topic probability began to be evenly distributed in both Asia and non-Asia from the mid-2000s, implying that the topic of transportation decarbonization became common at the end of 2000. the KNN results revealed a 92% accuracy for Asian topics and an 89% accuracy for non-Asian topics.

Our findings can help capture a comprehensive picture of the transport decarbonization-related research activities that are exploding in the scientific community. This study can also contribute to expanding the existing work area by tackling the regional scale. Furthermore, the LDA-KNN integrated model can improve and supplement existing analysis methods to generate additional important insights. It is the first attempt to explore topics in transportation decarbonization research using the LDA-KNN integrated model, including relevant research fields to the author's knowledge.

Meanwhile, the current work contributes to keeping academics and practitioners informed of the paradigm shift and characteristics of transport decarbonization research from a regional perspective. The outcomes of this study can also pave the way for formulating and developing regional-level strategies and policies to decarbonize the transportation sector. Academics and practitioners are kept informed about the current paradigm shift and state of transport decarbonization research by current work.

This study only used data from the Scopus database; further research needs to include data from other sources such as WoS and Google Scholar. We also collected only academic publications, which are insufficient for evaluating the actual implementation of transportation decarbonization practices and policies. Future research can apply our approach to address official documents such as transportation decarbonization plans and policies. Future studies also can improve our LDA-KNN integration model by including other text mining or machine-learning approaches.

■ References ■

- Bai, X., X. Zhang, K. X. Li, Y. Zhou, and K. F. Yuen, 2021, "Research topics and trends in the maritime transport: A structural topic model," *Transport Policy*, 102, pp.11-24.
- Bayarsaikhan, T., M. H. Kim, H. J. Oh, and T. H. T. Gim, 2022, "Toward sustainable development? Trend analysis of environmental policy in Korea from 1987 to 2040," *Journal of Environmental Planning and Management*, pp.1-15.
- Boyer, R. C., W. T. Scherer, and M. C. Smith, 2017, "Trends over two decades of transportation research: a machine learning approach," *Transportation research record*, 2614(1), pp.1-9.
- Casillo, M., F. Colace, B. B. Gupta, D. Santaniello, and C. Valentino, 2021, "Fake

- News Detection Using LDA Topic Modelling and K-Nearest Neighbor Classifier,” *Paper for International Conference on Computational Data and Social Networks*, Springer, Cham, pp.330–339.
- Chiaromonti, D., G. Talluri, N. Scarlat, and M. Prussi, 2021, “The challenge of forecasting the role of biofuel in EU transport decarbonisation at 2050: A meta-analysis review of published scenarios,” *Renewable and Sustainable Energy Reviews*, 139, 110715.
- Cleophas, C., C. Cottrill, J. F. Ehmke, and K. Tierney, 2019, “Collaborative urban transportation: Recent advances in theory and practice,” *European Journal of Operational Research*, 273(3), pp.801–816.
- Emodi, N. V., C. Okereke, F. I. Abam, O. E. Diemuodeke, K. Owebor, and U. A. Nnamani, 2022, “Transport sector decarbonisation in the Global South: A systematic literature review,” *Energy Strategy Reviews*, 43, 100925.
- Foster, V., J. U. Dim, S. Vollmer, and F. Zhang, 2021, “Understanding Drivers of Decoupling of Global Transport CO₂ Emissions from Economic Growth,” *Policy Research Working Paper*, World Bank, 9809.
- Gim, T.-H. T., 2022, “Analyzing the city-level effects of land use on travel time and CO₂ emissions: a global mediation study of travel time,” *International Journal of Sustainable Transportation*, 16(6), pp.496–513.
- Gallego, A. J., J. Calvo-Zaragoza, J. J. Valero-Mas, and J. R. Rico-Juan, 2018, “Clustering-based k-nearest neighbor classification for large-scale data with neural codes representation,” *Pattern Recognition*, 74, pp.531–543.
- IEA (2022), “Net Zero by 2050,” <https://www.iea.org/reports/net-zero-by-2050>, [Date Accessed: 7th December 2022].
- ITF, 2021, “ITF Transport Outlook 2021,” Paris: OECD Publishing, DOI: 10.1787/16826a30-en.
- Jin, C., J. D. Ampah, S. Afrane, Z. Yin, X. Liu, T. Sun, et al., 2022, “Low-carbon alcohol fuels for decarbonizing the road transportation industry: A bibliometric analysis 2000–2021,” *Environmental Science and Pollution Research*, 29(4), pp.5577–5604.
- Kulkarni, A., D. Chong, and F. A. Batarseh, 2020, “Foundations of data imbalance and solutions for a data democracy,” *Data democracy*, pp.83–106.
- Meyer, T., 2020, “Decarbonizing road freight transportation-A bibliometric and network analysis,” *Transportation Research Part D: Transport and Environment*, 89, 102619.
- Mohamed, M., C. Higgins, M. Ferguson, and P. Kanaroglou, 2016, “Identifying and

- characterizing potential electric vehicle adopters in Canada: A two-stage modelling approach," *Transport Policy*, 52, pp.100-112.
- Tian, X., Y. Geng, S. Zhong, J. Wilson, C. Gao, W. Chen et al., 2018, "A bibliometric analysis on trends and characters of carbon emissions from transport sector," *Transportation Research Part D: Transport and Environment*, 59, pp.1-10.
- Velazquez, L., N. E. Munguia, M. Will, A. G. Zavala, S. P. Verdugo, B. Delakowitz et al., 2015, "Sustainable transportation strategies for decoupling road vehicle transport and carbon dioxide emissions," *Management of Environmental Quality*, 26(3), pp.373-388
- Wang, Q., and M. Su, 2020, "Drivers of decoupling economic growth from carbon emission-an empirical analysis of 192 countries using decoupling model and decomposition method," *Environmental Impact Assessment Review*, 81, 106356.

Tsolmon Bayarsaikhan: Ph.D. candidate in environmental management at Seoul National University. Her current research interests include transportation energy consumption and carbon emissions in urban areas(stsweety@snu.ac.kr).

Tae-Hyoung Tommy Gim: Ph.D., is an associate professor in the Graduate School of Environmental Studies and jointly affiliated with the Interdisciplinary Program in Landscape Architecture and the Environmental Planning Institute at Seoul National University (SNU). He is also the director of the SNU Integrated Planning Lab. His fields of expertise include land use-transportation-environment interactions, and quantitative methods(taehyoung.gim@snu.ac.kr).

Received: 13 December 2022
 Revised: 18 December 2022
 Accepted: 22 December 2022