



Tracing the Trends of General Construction and Demolition Waste Research Using LDA Modeling Combined With Topic Intensity

Ze Zhou Wu¹, Peiying Xie¹, Jinming Zhang^{2*}, Baojian Zhan¹ and Qiufeng He¹

¹ College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, China, ² School of Political Studies, Nanjing Agricultural University, Nanjing, China

OPEN ACCESS

Edited by:

Linchuan Yang,
Southwest Jiaotong University, China

Reviewed by:

Zhikang BAO,
The University of Hong Kong, Hong
Kong SAR, China
Jingkuang Liu,
Guangzhou University, China

*Correspondence:

Jinming Zhang
zjming@njau.edu.cn

Specialty section:

This article was submitted to
Environmental Health and Exposome,
a section of the journal
Frontiers in Public Health

Received: 19 March 2022

Accepted: 12 April 2022

Published: 25 May 2022

Citation:

Wu Z, Xie P, Zhang J, Zhan B and
He Q (2022) Tracing the Trends of
General Construction and Demolition
Waste Research Using LDA Modeling
Combined With Topic Intensity.
Front. Public Health 10:899705.
doi: 10.3389/fpubh.2022.899705

The study of construction and demolition waste (CDW) has attracted more and more attentions with the increasing CDW pollution caused by the large-scale infrastructure construction. This study used the Latent Dirichlet Allocation (LDA) combined with topic intensity to discover hot topics and development trends in the study area of CDW. First, the LDA was used for topic modeling to extract the existing topics from textual data. Second, the topic intensity was calculated for the extracted topics and the numerical values of the topic intensity represented the popularity of the topics. In this study, 4 topics were extracted from 1,849 relevant articles through the LDA modeling and topic intensity calculation. The results showed that the topic of “CDW management” had an upward trend. Topics such as “recycled aggregate,” “environmental impact,” and “study of CDW on soil” all showed a downward trend. The methods of this study can dig into the topics of CDW study and help scholars to engage in this field for better understanding the prevalence and evolution trends of these topics.

Keywords: construction and demolition waste, research trend, linear discriminant analysis, topic modeling, topic intensity

INTRODUCTION

With the constant advancement of the construction industry, the problems that construction brings to the environment are becoming more and more obvious (1–3) and the daily construction and demolition waste (CDW) output of CDW is increasing daily (4–6). To improve the level of CDW, various countries are implementing a series of CDW measures, such as Sweden’s implementation of the “Ordinance on Landfilling of Waste,” the Netherlands’s “Environmental Protection Act,” and Japan’s “Construction Material Recycling Law”(7). In this context, researchers and policymakers are required to accurately grasp the development trend of the field to support the formulation of national CDW-related policies (8, 9). Scientific measurement and text mining methods are used to sort out the study context from the massive articles and predict the research frontier that plays an extremely critical role in grasping the latest study trends of CDW.

In recent years, to grasp the research trend of CDW, relevant scholars have discussed the development trend of CDW from the perspective of the whole or partial field of CDW (10, 11). The majority of previous study used the method of literature measurement or the keywords combined with the method of scientific mapping analysis to predict the topic. In 2011, Yuan (12) selected

87 articles from 8 journals of CDW management during 2000–2009 as a sample library. The descriptive analysis, simulation/modeling, statistical analysis, and cost-benefit analysis were used to mine and analyze study topics and trends of CDW management. In 2017, Liu (13) adopted bibliometric analysis and imagined investigations were likewise applied in this article to assess the status and patterns of CDW study in view of articles from 2000 to 2016. In 2020, Li (14) used journal articles concerning the study and development of information technology in CDW management from 2000 to 2019 as the data source and used the method of scientific mapping analysis to analyze the study trend in the past 20 years. In addition, Ding (15) and Ranjbari (16) used topic modeling to disclose the study topic of CDW management. These existing review studies, however, only surveyed articles in the field of CDW management and also failed to investigate study trends in the field of CDW. As a result, the existing literature lacks an analysis of study hotspots in CDW. According to the above analysis, study reviews can be achieved through bibliometric analysis or topic modeling. However, most such study reviews mainly analyze journal, time, country, and author information and rarely analyze the content of the study, which limits the analysis of relevant study content. Among them, Latent Dirichlet Allocation (LDA) is an unsupervised topic model that can be used to analyze the topic importance of textual documents. To fill the study gap, this article used the LDA modeling combined with topic intensity to conduct an in-depth analysis of the main study topics of CDW and explored the development trend of each topic according to the average weight and topic intensity of different periods.

The LDA model is a generative probabilistic topic model. Given that the potential subject is the likelihood appropriation of a progression of words in the corpus, then, at that point, the record is the likelihood dispersion of a progression of expected themes. Unlike the overall bunching technique, the LDA permits a record to contain different themes simultaneously, so it is more suitable to extract study topics from scientific articles (17). At the earliest, Griffiths (18) used the LDA to extract the subject and subject change trend of the Proceedings of the National Academy of Sciences (PNAS) journal abstracts and used the Gibbs sampling algorithm to infer the LDA. Currently, the LDA has been used for the analysis of computer linguistics (19), library and information management (20), economics (21), and other fields (22–25), to understand the study status of a particular field by automatically mining study topics of a large number of articles. Hence, in this article, the LDA was used as the strategy for theme display. First, essential issues can be examined in CDW by determining the extent to which they have been addressed in CDW study. Second, information about the study trend of CDW can be extracted by studying how people's interest in the subject changes over time. Third, relatively neglected or marginalized topics can be explored and used to identify future study topics. To investigate the trend of CDW study, this study is based on the articles distributed within the Web of Science (WoS), trying to map the development of CDW study in the past 10 years (2011–2020). According to the average weight and topic intensity of different periods, it profoundly analyzes the main study topics of CDW and the development trend of each topic. The purpose

of evaluating these topics is to study the research topics discussed by researchers in CDW. The text and abstract of the article are used as analysis objects by discussing its objectives, methods, and results. This article analyzes the use of point demonstration to extricate the principal subjects of the exploration and the distinctions of themes distributed in different periods to improve the accuracy of topic selection.

To sum up, this study extracts the topics of CDW articles with the help of the LDA classic model. It makes a detailed analysis of the content and intensity of each topic to deeply understand the study status of CDW in China from 2011 to 2020.

The next part of this article is organized as follows. Section 2 introduces the previous study work on the data source and implementation of the LDA-based topic modeling. Section 3 describes the analysis results. The last part is the conclusion.

DATA SOURCE AND IMPLEMENTATION OF THE LATENT DIRICHLET ALLOCATION-BASED TOPIC MODELING

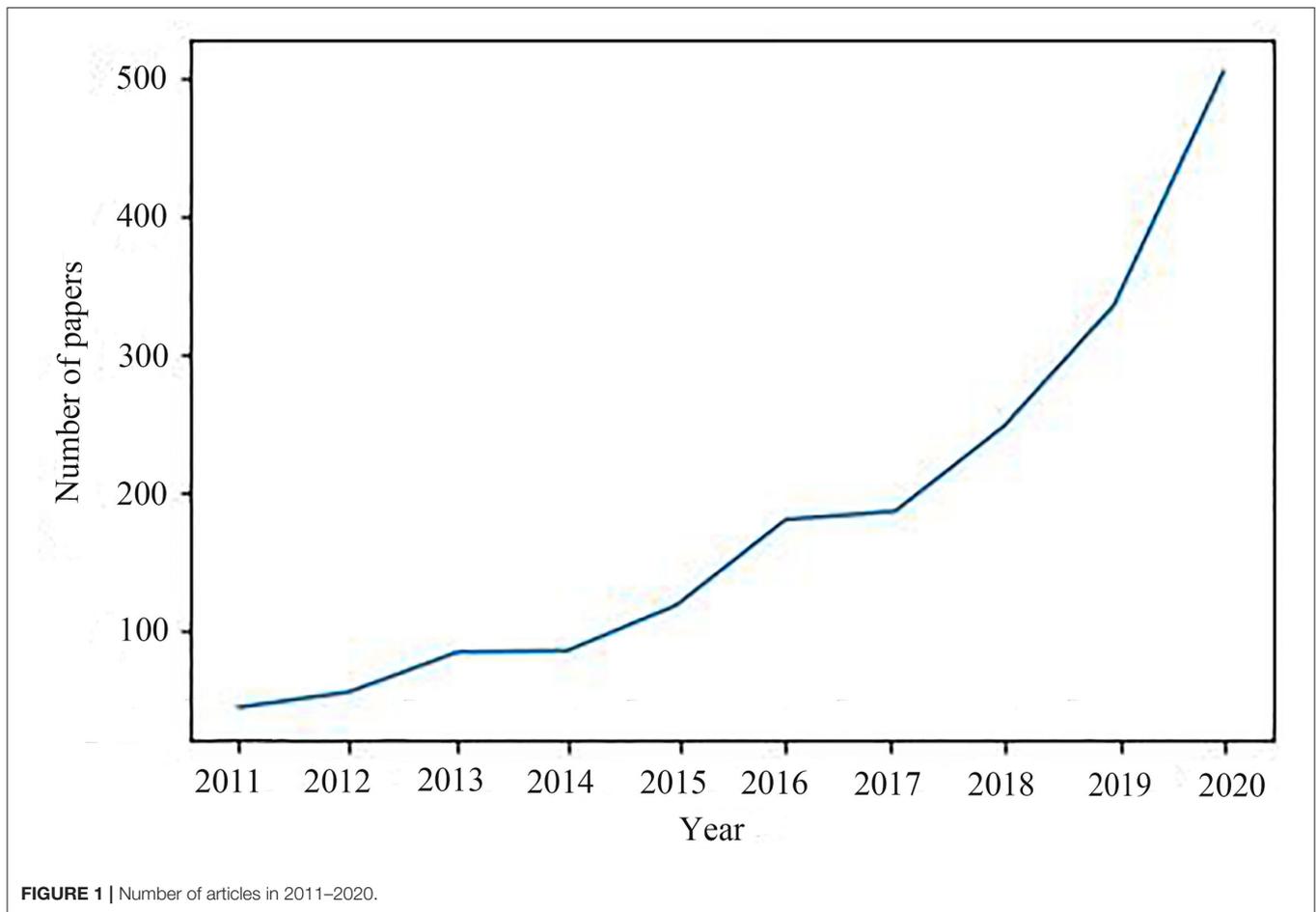
Data Source

When retrieving from the WoS database, TS = “demolition and construction waste” or “C&D waste” or “C&D wastes” or “CDW” or “demolition waste” or “construction waste” (as the search type). At the same time, conference papers were excluded given that conference papers provide less data than journal articles. Finally, 1,849 articles related to the English language from 2011 to 2020 were collected. The annual distribution of 1,849 articles is shown in **Figure 1**, indicating that the overall number of articles in CDW study is steadily increasing.

Data Preprocessing

The stage of preprocessing is exceptionally important to move forward the quality investigation of text data that is unstructured and information preprocessing is an activity that should be embraced, particularly within the investigation of web-based unstructured literary materials (26). The processing of data is mainly for identification dispensation, stop word removal, and morphological restoration of the above data. In this study, the processing of data is implemented and programmed in Python. In data processing, the smallest processing unit understood by the computer is a word, so text information must be identified. The identification here is to divide the sentence into a series of meaningful word segmentation.

Stop words removal is to remove meaningless words such as conjunctions, articles, and pronouns. The vocabulary made up of meaningless words is called the stop word list. This article removes the English stop words and adopts the stop word list of the Natural Language Toolkit (NLTK). The stop word list of the NLTK is relatively standard and complete, covering the stop word list in 22 languages. By passing language name parameters to the library's stop words constructor, stop words in different languages can be removed to improve the reliability of results. The lemmatization operation uses the setting and morphology to decide the variety of related words and uses different normalization rules to obtain connected roots based



on the grammatical feature (27). In this study, the titles and abstracts of texts were extracted from 1,849 articles. All the text characters are changed into lowercase and any exceptional characters or stop words are deleted. Any grammatical marker, such as the plural form of a noun or the past form of a verb, can be converted to its general form, removing punctuation and unnecessary characters altogether. Some articles have two-letter combinations of words, such as focus group, requiring two-letter combinations and lexical recovery, as well as calling them sequentially.

Latent Dirichlet Allocation-Based Topic Modeling Implementation

The choice of the number of topics is the first step in building a model. The topic of the LDA model is an abstract concept. According to different granularities, the topics in the corpus can be divided into various quantities of themes. The quantity of subjects in the LDA needs to be given in advance. In general, larger corpora represent more topics.

Furthermore, the LDA gives a bunch of words compared to themes based on the number of points. Hence, it is essential to investigate the appropriate number of themes. If not, it will affect the advantages and disadvantages of the entire

topic model. Suppose that the selected value of the number of topics is too small. In this case, the ability of the model to describe the data will be limited. Instead, it will prolong the training time of the model and hinder the effect of data analysis. Perplexity is often used to determine the number of topics (28).

Moreover, although perplexity is widely used, it is not the best way to evaluate the number of topics because it does not consider the semantic association between word contexts. This study selects the ideal number of themes in light of the coherence score. It uses the method provided by Mallet to calculate a coherence score, depending on the level of co-event between words removed from the subject to work out 10 expressions of every point. Topic modeling gives a bunch of 10 words to encapsulate a subject and the theme is created by subjectively examining the collection of words introduced in the topic displaying results. Then, the most noteworthy top articles of each topic are analyzed and referred to make the theme title.

The topic coherence score compels the number of topics K and when the coherence score is the highest, K takes the optimal value. This article employs the assessment index coherence score in the statistical language model show to decide the ideal number of topics (29).

The formula for the coherence score is described as follows:

$$\text{coherence}(V) = \sum_{(v_i, v_j) \in V} \text{score}(v_i, v_j, \varepsilon)$$

$$\text{score}(v_i, v_j, \varepsilon) = \log \frac{p(v_i, v_j) + \varepsilon}{p(v_j)}$$

In the formula, V is a series of words that describe the topic. ε is the smoothing factor, which is taken as 1 according to experience. v_i and v_j are any two words belonging to V , respectively. $p(v_i, v_j)$ represents the co-occurrence of v_i and v_j probability. The consistency score is positively correlated with the sentence similarity, as obtained by calculating the co-event recurrence of words in the sentence. Therefore, a higher consistency score could give a better result (30).

Topic Intensity

The center element of the exploration front is “topic intensity”(31). Topic intensity addresses how much specialists focus on the study output. Higher topic attention represents higher scientific study output. From a bibliometric perspective, the number of documents in a particular field of study frontier topics is vast compared with other study topics. When identifying frontiers based on the LDA model in this article, topic intensity is recommended to reflect the great attention of the study frontier.

Topic intensity is the extent of the point in the archive. It is a quantitative pointer to quantify whether the study theme is a state-of-the-art subject. Topic intensity can be communicated as a proportion of the number of weights for exploring topics in all the scientific reports to the aggregate sum of archives. The formula is:

$$\theta_j = \sum_d \theta_j^{(d)} / M$$

In the formula, M represents the complete number of reports. θ_j indicates the point force value of the j subject. $\theta_j^{(d)}$ refers to the weight of the j topic on document d . Topic intensity can reflect the significance of the topic within the document set. The larger the topic intensity value (θ_j) of the topic j , the more significant the proportion of the topic j in the document set and the more critical the topic j . The study frontier has the characteristic of great attention, hence the frontier topic has a relatively high proportion in the document set, with a large topic intensity value. The average topic intensity value of each topic in the document set can be calculated according to the formula. Topics with a greater topic intensity value than the average can be filtered out to obtain popular topics.

When individual topics extracted from topic modeling, the corresponding study trends of these topics over time are investigated. The weight upsides of the multitude of investigated articles were acquired and the year-to-year mean weight was charted. The exploration of preparation can be partitioned into five steps:

(1) Obtain the article data of CDW from the past 10 years.

- (2) Preprocess the collected article data set to form a document set for running the LDA algorithm.
- (3) Select the appropriate topic to run the LDA and get the topic and subject terms.
- (4) Establish the topic intensity model, calculate the intensity of each hot topic, and analyze hot topics.
- (5) According to the weight of each topic, analyze the study trend of each hot topic.

ANALYSIS RESULTS

Suitable Number of Topics

Figure 2 illustrates how the number of topics varies from 2 to 40 and how coherence scores are calculated by the number of subjects. In general, a greater coherence value indicates that there are more themes that can adequately represent all of the data. As shown in Figure 2, the coherence score is the highest based on four topics. Therefore, four themes could be sufficient for this investigation.

Making the Title of Each Topic

Python language was used to implement the LDA modeling, extract the potential topic distribution, and present it in the topic-word and document-topic distributions. After estimating the consistency score, the ideal number of points is 4. Table 1 displays the subject vocabulary distribution as a function of topic extraction. Because of spatial consideration, only the top 10 terms with the highest likelihood in each category are given in Table 1.

Table 1 presents the topics in the form of a topic-word probability distribution table. The specific topic name for each topic was summarized qualitatively according to the keywords. The final four topics are: “recycled aggregate” (topic 0), “environmental impact” (topic 1), “construction and demolition waste management (CDWM)” (topic 2), and “study of CDW on soil” (topic 3).

The LDA topic recognition generated document-topic probability matrix that is based on the probability of the $4 \times 1,849$ matrix. This study visualizes each document in the subject probability of topic information and intuitively investigates each document's relevant information. Table 2 presents the keywords comprising topics and the top articles of each topic and provides detailed information about each topic for easy understanding and analysis, where the Perc_Contribution column is the percent contribution of the topic in a given article.

Four study topics reflect the current study status in the field of CDW over the past decade. Based on the words in the topic-word matrix and the top articles of each topic in the document-topic distribution, the study content in the field is analyzed and summarized, mainly focusing on the following areas:

Category I: Study on material recycling technology. The study related to material recycling technology focuses on recycled aggregate (topic 0). This category highlights material-related stability studies, which are essential for technical study on various recycled materials in CDW. In this topic, Sasanipour et al. (32) with the highest contribution rate studied the replacement of sand aggregate with recycled brick aggregate. The results showed that when using recycled brick aggregate instead of sand

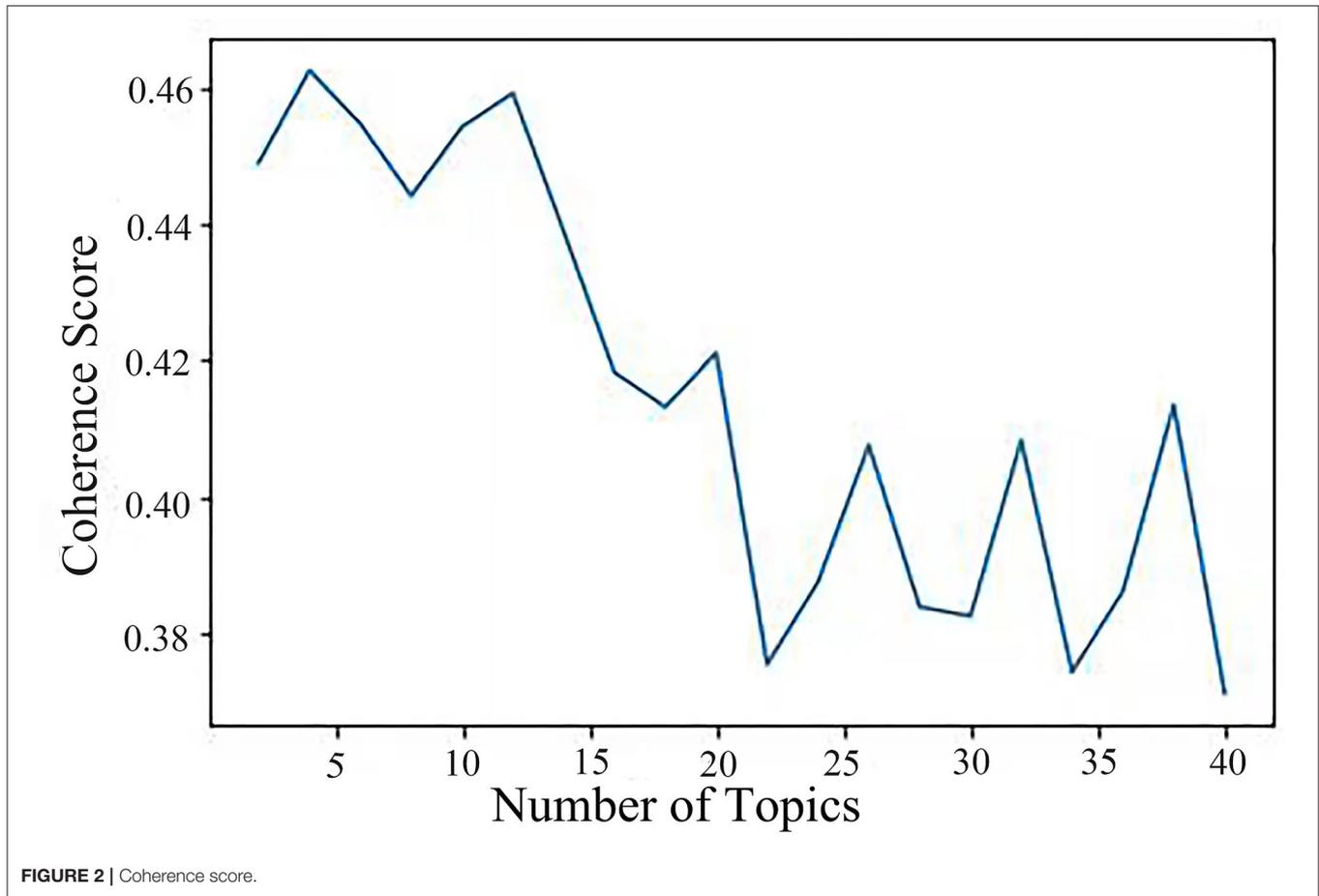


TABLE 1 | Topic-word probability distribution.

Topic 0	Topic 1	Topic 2	Topic 3
Aggregate (0.036)	Waste (0.059)	Construction (0.021)	Material (0.016)
Concrete (0.035)	Construction (0.041)	Model (0.018)	Soil (0.016)
Recycle (0.022)	Material (0.020)	Factor (0.015)	Test (0.010)
Strength (0.020)	Demolition (0.016)	Design (0.014)	High (0.009)
Cement (0.016)	Study (0.012)	Management (0.014)	Sample (0.008)
Property (0.016)	Environmental (0.012)	Project (0.010)	Increase (0.008)
Use (0.014)	Building (0.010)	Study (0.008)	Use (0.008)
Test (0.011)	Use (0.008)	Research (0.008)	Result (0.007)
Result (0.010)	Impact (0.007)	Policy (0.008)	Brick (0.007)
Recycled (0.010)	Recycling (0.007)	Practice (0.007)	Thermal (0.007)
Recycled aggregate	Environmental impact	CDWM	Study of CDW on Soil

aggregate, chloride migration was reduced, but water absorption, drying shrinkage, and carbonation were increased.

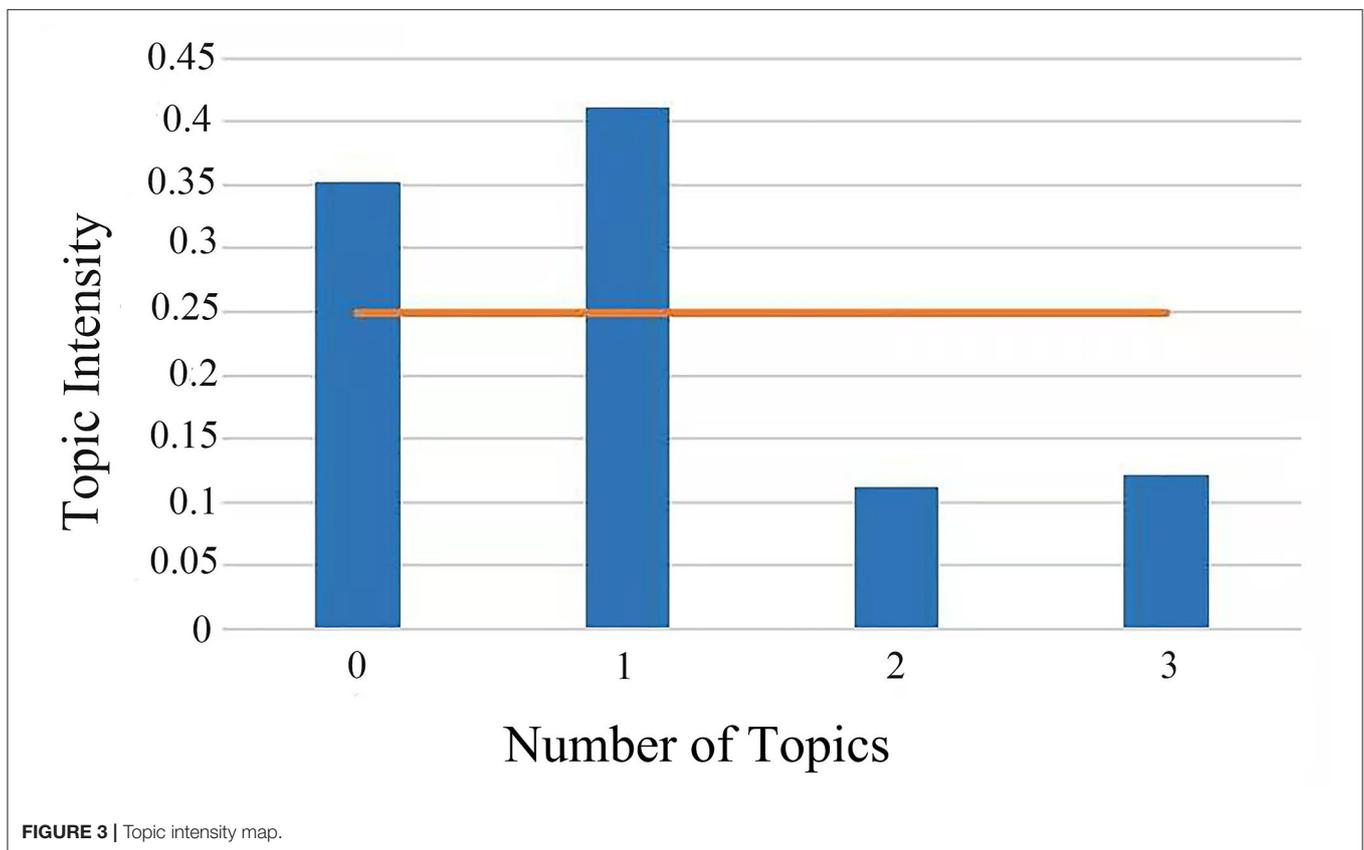
Category II: Study on the impact of CDW on the ecosystem is the environmental impact (topic 1). This category investigates the environmental factors of CDW and proposes some measures to reduce CDW in the environment. The highest contribution rate in this topic is that Dahlbo et al. (33) used a combination of different methods to study the performance of CDW

management system in Finland. In general, in most cases, the environmental benefits of waste recycling are greater than those of energy recycling.

Category III: Study on CDW management (topic 2), mainly on policy support, reduction, and production calculation. This category favors CDW management and policy study, which is closely related to stakeholders. With the highest contribution rate in this topic, Kim et al. (34) used questionnaire survey and

TABLE 2 | Document-topic probability distribution.

Num	Perc_Contribution	Keywords	Title
0	0.996200025	aggregate, concrete, recycle, strength, cement, property, use, test, result, recycled	Durability properties evaluation of self-compacting concrete prepared with waste fine and coarse recycled concrete aggregates.
1	0.996200025	waste, construction, material, demolition, study, environmental, building, use, impact, recycling	Construction and demolition waste management-a holistic evaluation of environmental performance.
2	0.929400027	construction, model, factor, design, management, project, study, research, policy, practice	A performance evaluation framework for construction and demolition waste management: stakeholder perspectives.
3	0.965200007	material, soil, test, high, sample, increase, use, result, brick, thermal	Spatial Distribution Characteristics of Microorganisms in Constructed Wetland System with New Matrix and Its Effect on Sewage Purification.

**FIGURE 3** | Topic intensity map.

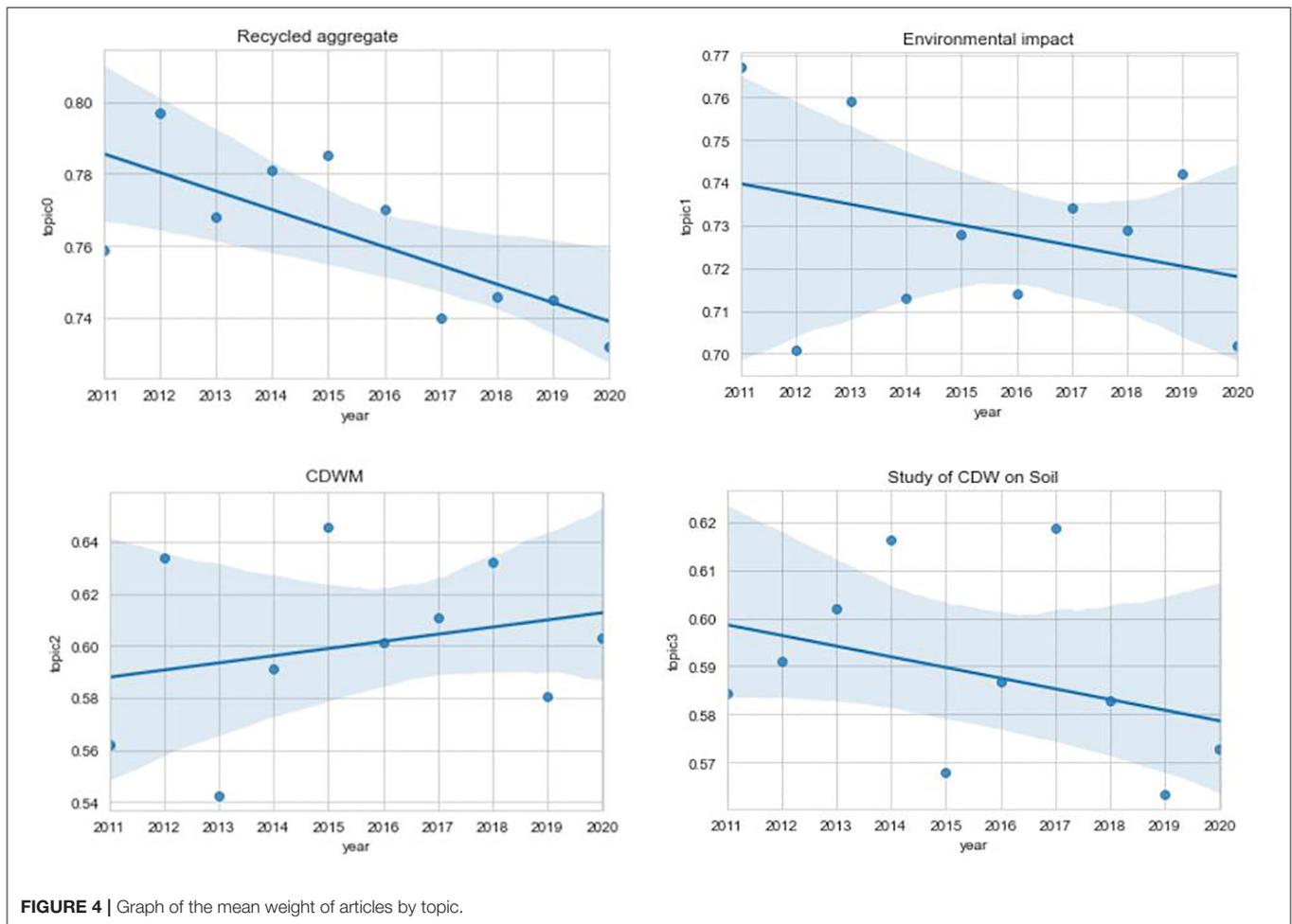
factor analysis study methods to analyze effective strategies for providing CDW by establishing a CDW management evaluation framework. The study showed that the conceptual performance evaluation framework for CDW management consists of six factors, the key factor of which is the attitudes of stakeholders toward CDW management.

Category IV: Study on CDW in other areas, such as the study of CDW on soil (topic 3). CDW is used as a new substrate to study the ability of microorganisms in the soil to purify sewage, increasing the recycling of CDW to help protect the environment. Therefore, study in other fields is another effective way to reuse CDW resources. With the highest contribution rate in this topic, Chen et al. (35) used CDW as one of the substrates to study the

spatial distribution of microorganisms and soil enzymes and the distribution characteristics of pollutants at different depths.

After obtaining the topic through experiments, the focus is on evaluating the intensity of the topic and further analyzing its hot topic. Based on the LDA topic modeling, this study uses the topic-document probability matrix obtained by the model to calculate the topic intensity, reflecting the degree of attention paid to the topic and discovering hot topics.

As shown in **Figure 3**, the topic intensity of four topics can be calculated using the formula. The bar height represents the topic intensity value and the horizontal line in the middle is the average topic intensity value, which is 0.2491. At the same time, the topic intensity of topic 0 (recycled aggregate) is 0.3521,



the topic intensity of topic 1 (environmental impact) is 0.4112, the topic intensity of topic 2 (CDWM) is 0.1118, and the topic intensity of topic 3 (study of CDW on soil) is 0.1212. Among them, topic 1 has the highest topic intensity. Furthermore, the topic intensity of topic 0 and topic 1 is higher than the average topic intensity, indicating that topic 0 and topic 1 are hotspots in CDW study field.

Changing Interest in Each Topic Year by Year

The average annual weights of articles per topic from 2011 to 2020 are shown in **Figure 4**. A trend could not be discerned in this article because there were not enough data points. Consequently, the graph was redrawn with a regression line to show the trend of change.

As shown in **Figure 4**, only one topic (topic 2) is on the rise, representing the study on CDWM is on an upward trend. It can be seen that there is an increasing study in the field of CDW management, which may be related to the promulgation of CDW management policies in various regions in recent years. This is consistent with the findings of Li (36). In recent years, an

increasing amount of study activity has been devoted to CDWM to find solutions to reduce CDW, a challenging problem around the world.

Three topics (topic 0, topic 1, and topic 3) show a downward trend. Among them, topic 0 is declining, but accounting for a large proportion of all the topics, with a weight of 0.73–0.80. This indicates that the recycling of CDW has always been an important core technology or application field. The decline in study on recycled aggregates may be due to limited technical conditions, lacking study equipment, and high costs. Moreover, there are currently fewer CDW recycling companies and limited study subjects. From the topic intensity of CDW recycling, it can be seen that its value is relatively high, indicating that the weight of CDW recycling declines, but continues to develop, which is a study hotspot. Similarly, Liu (37) found that the promotion of recycled aggregate applications mainly depends on low-cost technology and low-cost technology will be a hotspot in the future. Topic 1 is in a downward trend, but accounts for a relatively large proportion of all the topics, with a weight of 0.71–0.75. Its topic intensity is also above the average, indicating that environmental impact study is also one of the necessary study fields in CDW. Its weight declined may be because the release of

CDW policies in various regions has certain particularities, study has certain limitations, and the lack of data on CDW disposal causes certain obstacles to study. At the same time, Chen (38) found that the study on the impact of CDW on the environment is still in the preliminary stage; however, the impact of CDW on the environment has gradually received more attention from researchers, which has promoted the environmental impact study of CDW. Topic 3 is also in a downward trend. It can be seen from the points that the study first rises and then falls, but its topic intensity is above the average, which indicates that the topic is still the study hotspot that year. The decline in the trend of this topic may be due to that scholars who are currently concerned about CDW are more scholars related to construction engineering. In addition, scholars in other fields pay less attention to the study on the field of CDW, which may easily lead to a decline in the study trend. Therefore, there is currently a lack of scholars who have reviewed this type of study.

Future Study

Future study is suggested to pay attention to the impact of CDW on the environment from the perspective of the whole life cycle. The governance of CDW should be seen as a sustainable system (39). Most of the existing study is limited to the study of CDW in the construction and demolition stages (40–42), while ignoring the generation of waste in all the stages of the project life cycle. Considering the generation of waste at the design stage of a construction project, it can effectively reduce the generation of waste. In addition, the transport process may cause a high environmental load (43). Future study should pay more attention to the impact of the entire CDW life cycle on the environment to improve environmental performance.

Another future study is suggested to conduct more extensive study on different kinds of CDW. There are many types of CDW, each with different recycling routes and process modes. A single CDW processing model may be too general and not suitable for all the types of CDW. Therefore, more in-depth study on different types of CDW should be carried out, such as study on recycled glass and recycled concrete, which can provide more valuable suggestions for CDW recycling.

Apply new technologies to study related to the entire field of CDW. By analyzing the literature and combining with the current development trend, it is found that there is still a relatively lack of study on new technologies. With the rapid development of information technology, more and more new technologies have been applied to engineering construction study, such as big data, artificial intelligence, building information model, geographic information system, and so on (44–47). For example, Lee used artificial neural networks and genetic algorithms to predict the amount and cost of waste in the early stages of construction. Wu developed an innovative method using embedded geographic information system (GIS) to quantify CDW from generation to final disposal. Therefore, future study can apply intelligent technology to the entire study field of CDW and can also use two or more technologies combined for study, which will help to overcome

the limitations of a single technology in application and improve the performance level.

Furthermore, CDW study can benefit from additional views, even after the findings of this study are considered. Study articles can be also expanded to incorporate additional databases [Scopus, China National Knowledge Infrastructure (CNKI), etc.]. A second option is to add earlier years to the list of publications. The history of CDW study trends is uncovered, which can shed light on the fads and features of CDW study evolving through time. In CDW field, these viewpoints provide patterns that can be used to anticipate future developments. Topic modeling is also feasible through the use of multiple algorithms, which allow for more refined or comparative identification of topics, with the potential to provide a test of the topics found. A corpus with additional metadata, such as author-offered keywords, abstracts, and titles, can be used for topic modeling. Finally, the extracted topics can also be utilized to understand and evaluate the progress of CDW study over time.

CONCLUSION

Tracking the study hotspots of CDW can help scholars to understand the current study status and the law of development over time. This study proposed a method to obtain short-text topics through the combination of the LDA and topic intensity. The topics and abstracts of articles in the WoS published from 2011 to 2020 were chosen as the study subject. First, we use the NLTK and spaCy to obtain the professional terminology of CDW articles. Second, the LDA model of Mallet, a generative probabilistic topic model, was used to capture potential topics in articles of CDW to obtain the topics and topic words. Third, the topic intensity was calculated to obtain hot topics. Finally, the periodic changes of topics to year-to-year mean weight were analyzed in different periods, discussing each topic's development trend and mapping the trend of CDW study over time.

The results show that the topic effect is the best when the number of topics is four. One topic is trending up and the other three topics are trending down. Results based on the LDA have revealed these topics. Only one topic is on the rise (e.g., “CDWM”). The rise is relatively straightforward, indicating that CDWM is receiving an increasing attention in society, while others are decreasing gradually (e.g., “recycled aggregate,” “environmental impact,” and “study of CDW on soil”). Such decline may be affected by the economy, technology, and policy to a certain extent, but according to the topic intensity, it can be seen that topic 0 and topic 1 are both the hot topics. The results show the trends of CDW study and provide more valuable information for real-time, detailed, and scientific tracking of its study. The LDA can more accurately extract the study topics of CDW articles, which is helpful for researchers to have a preliminary understanding of the development status of the field, while grasping the future study direction and looking for emerging topics. Compared with the existing similar studies, previous studies have found that the study in the field of CDWM has attracted more and more scholars' attention.

However, in addition to finding that CDWM has attracted much attention, this study also found that recycled aggregate and environmental impact are also hotspots in CDW study. In addition, some scholars have studied the impact of CDW on soil. It is important to consider the study's limitations when interpreting findings. First, the titles and abstracts provided by authors are studied. Previous study has focused on the article's title, abstract, keywords, and text. Titles and abstracts may not sufficiently convey the study, despite that authors are proficient in their field and in a position to do so. Second, the sample size of 1,849 articles is large, which has specific practical significance. However, only CDW articles of the WoS are considered. CDW articles published by scholars in comprehensive excellent domestic journals are not considered. Future study can consider increasing the example size for subject analysis to completely comprehend the current development status of CDW field. Third, the analysis is based on the LDA algorithm, which does not look at different calculations for point demonstration (LSI/LSA). The themes and findings of this study may vary slightly, if other topic modeling techniques are used. Fourth, the LDA model assumes that topics are independent of each other. There are always inseparable links between topics in the same discipline. Therefore, the related topic model can be added to the model in future study. Finally, this study only considers the external characteristics of the publication time. In future study, the structural topic model can be used to add external information, including authors and journal categories to understand the study status more accurately.

REFERENCES

1. Wang J, Wu X, Wang R, He D, Li D, Yang L, et al. Review of associations between built environment characteristics and severe acute respiratory syndrome coronavirus 2 infection risk. *Int J Environ Res Public Health*. (2021) 18:7561. doi: 10.3390/ijerph18147561
2. Wei Y, Li Y, Liu X, Wu M. Sustainable development and green gross domestic product assessments in megacities based on the emergy analysis method—a case study of Wuhan. *Sustain Dev*. (2020) 28:294–307. doi: 10.1002/sd.2003
3. Wei Y, Huang C, Lam PTI, Yuan Z. Sustainable urban development: a review on urban carrying capacity assessment. *Habitat Int*. (2015) 46:64–71. doi: 10.1016/j.habitatint.2014.10.015
4. Liu J, Chen Y, Wang X. Factors driving waste sorting in construction projects in China. *J Clean Prod*. (2022) 336:130397. doi: 10.1016/j.jclepro.2022.130397
5. Bao ZK, Lu WS, Hao JL. Tackling the “Last Mile” problem in renovation waste management: a case study in China. *Sci Total Environ*. (2021) 790. doi: 10.1016/j.scitotenv.2021.148261
6. Lin ZQ, Xie Q, Feng YB, Zhang P, Yang Y. An optimal C&D waste logistics network design from contractors' perspective. *J Green Building*. (2020) 15:121–37.
7. Wu Z, Yu ATW, Poon CS. Promoting effective construction and demolition waste management towards sustainable development: a case study of Hong Kong. *Sustain Dev*. (2020) 28:1713–24.
8. Wu Z, Yu ATW, Wang H, Wei Y, Huo X. Driving factors for construction waste minimization: empirical studies in Hong Kong and Shenzhen. *J Green Building*. (2019) 14:155–67. doi: 10.3992/1943-4618.14.4.155
9. Wu Z, Yu ATW, Poon CS. An off-site snapshot methodology for estimating building construction waste composition - a case study of Hong Kong. *Environ Impact Assess Rev*. (2019) 77:128–35. doi: 10.1016/j.eiar.2019.03.006
10. Bao ZK, Lu WS, Chi B, Yuan HP, Hao JL. Procurement innovation for a circular economy of construction and demolition waste:

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

PX and ZW contributed to conception and design of the study. QH organized the database. BZ performed the statistical analysis. PX wrote the first draft of the manuscript. ZW and JZ wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version of the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 72001148), the Foundation for Basic and Applied Basic Research of Guangdong Province (No. 2019A1515110247), the Research Start-up Funding in Shenzhen (No. 000376), and the Natural Science Foundation of Shenzhen University (SZU).

ACKNOWLEDGMENTS

The authors thank the editors and reviewers for their valuable comments.

lessons learnt from Suzhou, China. *Waste Manage*. (2019) 99:12–21. doi: 10.1016/j.wasman.2019.08.031

11. Wu Z, Yang K, Xue H, Zuo J, Li S. Major barriers to information sharing in reverse logistics of construction and demolition waste. *J Clean Prod*. (2022) 350:131331. doi: 10.1016/j.jclepro.2022.131331
12. Yuan H, Shen L. Trend of the research on construction and demolition waste management. *Waste Management*. (2011) 31:670–9. doi: 10.1016/j.wasman.2010.10.030
13. Liu Y, Sun T, Yang L. Evaluating the performance and intellectual structure of construction and demolition waste research during 2000–2016. *Environ Sci Pollut Res*. (2017) 24:19259–66. doi: 10.1007/s11356-017-9598-9
14. Li CZ, Zhao Y, Xiao B, Yu B, Tam VWY, Chen Z, et al. Research trend of the application of information technologies in construction and demolition waste management. *J Clean Prod*. (2020) 263:121458. doi: 10.1016/j.jclepro.2020.121458
15. Ding ZK, Liu RS, Yuan HP. A text mining-based thematic model for analyzing construction and demolition waste management studies. *Environ Sci Pollut Res*. (2021) 28:30499–527. doi: 10.1007/s11356-021-13989-1
16. Ranjbari M, Saidani M, Shams Esfandabadi Z, Peng W, Lam SS, Aghbashlo M, et al. Two decades of research on waste management in the circular economy: insights from bibliometric, text mining, and content analyses. *J Clean Prod*. (2021) 314. doi: 10.1016/j.jclepro.2021.128009
17. Kavvadias S, Drosatos G, Kaldoudi E. Supporting topic modeling and trends analysis in biomedical literature. *J Biomed Informat*. (2020) 110:103574. doi: 10.1016/j.jbi.2020.103574
18. Griffiths TL, Steyvers M. Finding scientific topics. *Proc Natl Acad Sci U S A*. (2004) 101:5228–35. doi: 10.1073/pnas.0307752101
19. Wang J, Fan Y, Zhang H, Feng L. Technology hotspot tracking: topic discovery and evolution of China's blockchain patents based on a dynamic lda model. *Symmetry-Basel*. (2021) 13:415. doi: 10.3390/sym13030415

20. Lin LL, Xiu-Feng MAJIS. The theme discovery and evolution analysis of domestic library and information science research based on lda. *Informat Sci.* (2019) 37:87–92. doi: 10.13833/j.issn.1007-7634.2019.12.013
21. Piepenbrink A, Nurmammadov E. Topics in the literature of transition economies and emerging markets. *Scientometrics.* (2015) 102:2107–30. doi: 10.1007/s11192-014-1513-2
22. Wu Z, He Q, Yang K, Zhang J, Xu K. Investigating the dynamics of China's green building policy development from 1986 to 2019. *Int J Environ Res Public Health.* (2021) 18:196. doi: 10.3390/ijerph18010196
23. Wu Z, Zhang Y, Chen Q, Wang H. Attitude of Chinese public towards municipal solid waste sorting policy: a text mining study. *Sci Total Environ.* (2021) 756:142674. doi: 10.1016/j.scitotenv.2020.142674
24. Fu H, Wang M, Li P, Jiang S, Hu W, Guo X, et al. Tracing knowledge development trajectories of the internet of things domain: a main path analysis. *IEEE Trans Industr Inform.* (2019) 15:6531–40. doi: 10.1109/TII.2019.2929414
25. Wu Z, He Q, Chen Q, Xue H, Li S, A. Topical network based analysis and visualization of global research trends on green building from 1990 to 2020. *J Clean Prod.* (2021) 320:128818. doi: 10.1016/j.jclepro.2021.128818
26. Kyriakopoulou A, Kalamoukis T. The impact of semi-supervised clustering on text classification. In: *Proceedings of the 17th Panhellenic Conference on Informatics*; Thessaloniki: Association for Computing Machinery (2013). p. 180–7.
27. Sizhu W, Qing Q, Tiejun H, Danya L, Junlian L, Na HJDA, et al. Contrast analysis of methods and tools for lemmatization. *Data Anal Knowl Discov.* (2012) 28:27–34. doi: 10.11925/infotech.1003-3513.2012.03.05
28. Foster C, Inglis M. Mathematics teacher professional journals: what topics appear and how has this changed over time? *Int J Sci Math Educ.* (2019) 17:1627–48. doi: 10.1007/s10763-018-9937-4
29. Xie T, Qin P, Zhu L. Study on the topic mining and dynamic visualization in view of lda model. *Mod Appl Sci.* (2018) 13:204–13.
30. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D, editors. Exploring topic coherence over many models and many topics. *Conference on Empirical Methods in Natural Language Processing.* Jeju Island (2012).
31. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res.* (2003) 3:993–1022. doi: 10.1162/jmlr.2003.3.4-5.993
32. Sasanipour H, Aslani F. Durability properties evaluation of self-compacting concrete prepared with waste fine and coarse recycled concrete aggregates. *Constr Build Mater.* (2020) 236. doi: 10.1016/j.conbuildmat.2019.117540
33. Dahlbo H, Bachoi: 10.1016/j.conbjouttij.2019.117540elf-compacting concrete prepared with waste fine and coarse recycled concrete aggregates. 2020. I. rom bibliometric, text miJ *Clean Prod.* (2015) 107:333–41. doi: 10.1016/j.conbjouttij.2019.117540elf
34. Kim SY, Nguyen MV, Luu VT. A performance evaluation framework for construction and demolition waste management: stakeholder perspectives. *Eng Constr Archit Manag.* (2020) 27:3189–213. doi: 10.1108/ECAM-12-2019-0683
35. Chen PZ, Ren TZ, Zheng XQ, Liu YM, Cheng WM, Sun J. Spatial distribution characteristics of microorganisms in constructed wetland system with new matrix and its effect on sewage purification. *Environ Eng Sci.* (2017) 34:828–34. doi: 10.1089/ees.2016.0250
36. Li Y, Li M, Sang P. A bibliometric review of studies on construction and demolition waste management by using citespace. *Energy Build.* (2022) 258:111822. doi: 10.1016/j.enbuild.2021.111822
37. Liu YL, Sun TT, Yang L. Evaluating the performance and intellectual structure of construction and demolition waste research during 2000–2016. *Environ Sci Pollut Res.* (2017) 24:19259–66.
38. Chen KY, Wang JY, Yu B, Wu HY, Zhang JR. Critical evaluation of construction and demolition waste and associated environmental impacts: a scientometric analysis. *J Clean Prod.* (2021) 287. doi: 10.1016/j.jclepro.2020.125071
39. Bao ZK, Laovisutthichai V, Tan T, Wang Q, Lu WS. Design for manufacture and assembly (Dfma) enablers for offsite interior design and construction. *Build Res Informat.* (2022) 50:325–38. doi: 10.1080/09613218.2021.1966734
40. Bao ZK, Lee WMW, Lu WS. Implementing on-site construction waste recycling in Hong Kong: barriers and facilitators. *Sci Total Environ.* (2020) 747. doi: 10.1016/j.scitotenv.2020.141091
41. Liu JK, Wu P, Jiang YH, Wang XT. Explore potential barriers of applying circular economy in construction and demolition waste recycling. *J Clean Prod.* (2021) 326. doi: 10.1016/j.jclepro.2021.129400
42. Bao ZK, Lu WS. A decision-support framework for planning construction waste recycling: a case study of Shenzhen, China. *J Clean Prod.* (2021) 309. doi: 10.1016/j.jclepro.2021.127449
43. Guo YY, Yang LC, Chen Y. Bike share usage and the built environment: a review. *Front Public Health.* (2022) 10:848169. doi: 10.3389/fpubh.2022.848169
44. Davidson J, Fowler J, Pantazis C, Sannino M, Walker J, Sheikhhoshkar M, et al. Integration of Vr with bim to facilitate real-time creation of bill of quantities during the design phase: a proof of concept study. *Front Eng Manag.* (2020) 7:396–403. doi: 10.1007/s42524-019-0039-y
45. Chalal ML, Medjdoub B, Bezaï N, Shrahily R. Big data to support sustainable urban energy planning: the evoenergy project. *Front Eng Manag.* (2020) 7:287–300. doi: 10.1007/s42524-019-0081-9
46. Aghimien D, Aigbavboa C, Aghimien L, Thwala W, Ndlovu L. 3d printing for sustainable low-income housing in South Africa: a case for the urban poor. *J Green Build.* (2021) 16:129–41.
47. Fu H, Manogaran G, Wu K, Cao M, Jiang S, Yang A. Intelligent decision-making of online shopping behavior based on internet of things. *Int J Inf Manage.* (2020) 50:515–25. doi: 10.1016/j.ijinfomgt.2019.03.010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wu, Xie, Zhang, Zhan and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.