

A Review of Sustainable Supplier Selection and Evaluation Using Topic Modelling

Napat Srinual¹, Jörn Mehnen², and Nantawan Boonraksa³

^{1,2}Department of Design, Manufacturing and Engineering Management, Faculty of Engineering,
University of Strathclyde, Glasgow, UK

³Department of Logistics and Supply Chain Management, Faculty of Business Administration,
Huachiew Chalermprakiet University, Samutprakarn, Thailand
Email: napat.srinual@strath.ac.uk Tel. +44 759-786-8890

Abstract

The study provides a review of sustainable supplier selection (SSS), which has proceeded to emerge and expand along with greater speed over the last decades. This research aims to identify and explore the use of machine learning and text mining methods to automatically identify relevant core topics in sustainable supplier selection. The traditional approach by personal judgments with predetermined categories, cannot adequately take latent topics from large volumes of research data. Therefore, this study selects the topic modelling approach, which automatically identify topics that extend a large and unstructured collection of documents to uncover research topics in sustainable supplier selection research. The papers for the literature review were collected from SCOPUS database. The model with 20 topics was selected through the Latent Dirichlet Allocation (LDA) model from selected articles published from 2010 to 2020 in associated with various journals, and the top 5 most popular topics in sustainable supplier selection research are reviewed in a nutshell. We then explore topic trends by considering the transformation over different topics of sustainable supplier selection research over the last few decades. For each of the top 10 journals, the areas of subspecialty and the effects of editor changes on the topic portfolios are also investigated. The findings of this study are expected to provide implications for researchers, journal editors, and policy makers in the field of sustainable supplier selection.

Keywords: *Sustainability; Supplier selection and evaluation; Latent Dirichlet Allocation; Topic modelling; Unstructured documents*

1. Introduction

Corporate sustainability management is not limited to a single company. The importance of the influence of suppliers and the supply chain on the sustainability of products and processes is now undisputed. (Mani, Gunasekaran, & Delgado, 2018). Considering supplier practices and characteristics, determining the best sustainability match for those suppliers, and then selecting those suppliers are both important and accepted by practitioners and researchers to achieve truly sustainable outcomes (Shahryari et al. 2016).

In addition, both the literature review and the review papers on the problem of sustainable supplier selection (SSSP) have grown at a rapid pace and have become an academic discipline in their own right (Bai & Sarkis 2014). The SSSP literature has expanded significantly with more than a dozen academic journals devoted to the field. The substantial growth of the research has spurred bibliometric studies on itself because it is common for scholars to turn their attention to the literature itself once an academic discipline reaches a certain level of maturity (Chai, Liu, & Ngai, 2013). For example, 143 papers were selected for a literature review on supplier selection in sustainability. However, this is a time-consuming process with limited processing power, resulting in a small number of papers analysed. Researchers, especially early career researchers, often need to find, organise, and understand new and unexplored areas of research. Since a literature review often involves a large number of papers at the initial stage, the option for a researcher is either to limit the number of papers to be reviewed a priori or to review papers using other methods. It is interesting to note that there is another literature review related to sustainable supplier selection, e.g. Genovese et al. (2013) reviewed quantitative models to support green supplier selection. Moreover, these studies use the conventional discrete task approach, where each article is classified into a single category of predetermined topics based on subjective judgement. However, such discrete tasks cannot effectively hold latent themes from large scientific data. For one, manual classification based on reading abstracts or author keywords always carries the risk of classification error. Moreover, it costs too much time, especially when the number of articles to be assigned is very large. Second, the given categories are by no means exhaustive; relatively new and emerging topics are likely to be ignored. Convergence topics are also difficult to handle. Third, an article usually contains two or more topics. So far, handling large collections of articles has been structured into topics or categories by using coding sheets (Antons, Kleer, & Salge, 2016), dictionaries, or supervised learning methods. After considering some weaknesses in the existing literature reviews, we attempt to build a comprehensive literature review on supplier selection for sustainability from the database SCOPUS. This work applies topic modelling (Blei, Ng, & Jordan, 2003) called Latent Dirichlet Allocation (LDA). The period of data collection is 2010 to 2020.

2. Objectives

To review on topic trends, differing thematic fields and emerging research gap areas in the topic in sustainable supplier selection and evaluation by applying LDA topic modelling technique.

3. Materials and methods

This section clarifies the methods and procedure for distinguishing themes or topics in sustainable supplier selection research through LDA topic modelling. The

procedure involves three steps: (1) target journal selection and data collection, (2) text pre-processing, (3) topic modelling, as shown in Figure 1

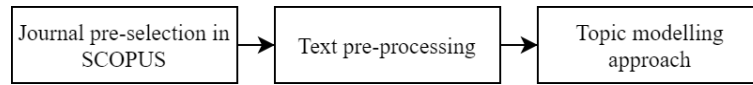


Figure 1 Procedure for identifying topics and applying topic modelling

3.1 Journal pre-selection in SCOPUS

The first step is to select relevant and leading journals for the article collection of the research on sustainable supplier selection (SSS) problems. Figure 2 shows that the number of articles has increased sharply from 2010 to 2020. This is the significant evidence based on the frequency of distribution by year of publication in sustainable supplier selection.

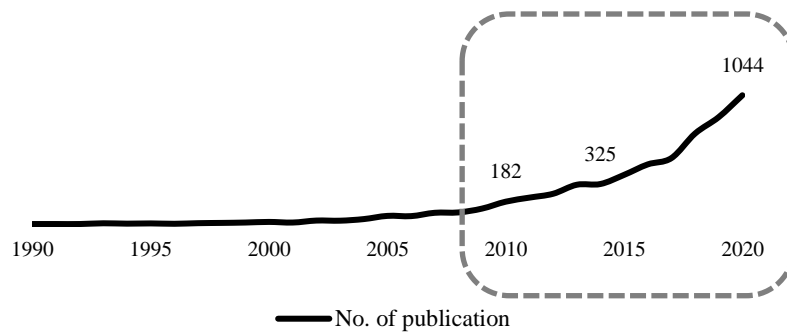


Figure 2 The numbers of published paper in sustainable from 2010 to 2020

The selection of leading journals in the field of sustainable supplier selection (SSS) is influenced by many aspects of SSS, but recent studies looking at SSS literature provide evidence on journal selection. Some recent studies confirmed that this journal can be considered as one of the leading journals due to its history and relatively high impact compared to other journals.

Table 1 lists the selected 10 journals and the corresponding number of articles included in this study.

Table 1 Lists of Sustainable Supplier Selection (SSS) Journals

Journal and conference	Year started	Impact Factor	Number
Journal Of Cleaner Production	1993	7.246	325
Sustainability Switzerland	2009	2.576	307
Iop Conference Series Earth And Environmental Science	2008	0.45	65
Energy	1976	6.082	58
Science Of The Total Environment	1972	6.551	55
Energies	2008	2.702	48
Advances In Intelligent Systems And Computing	2012	0.57	42
Journal Of Environmental Management	1999	4.175	41
Applied Energy	1999	8.848	37
Iop Conference Series Materials Science And Engineering	2009	0.53	37

For each journal, the in-depth process is carried out to gather the raw data for the targeted articles. First, documents that are original articles and reviews are examined in the database SCOPUS using different keywords for each journal starting from 2010 to 2020 as shown in Table 2.

Table 2 Identifying keywords in source database

Keywords	
Sustainable	Supplier selection and evaluation
sustainabl*	supplier, selecti*, evaluate*
	vendor, selecti*, evaluat*

Then, five meta-variables, namely title, year, DOI, abstract and keywords are selectively downloaded for the selected articles. In total, we collected 1,015 articles published in the top ten journals from 2010 to 2020. Table 3 shows the first three rows of my selected article data below.

Table 3 Sample of first three rows from selected articles

Title	Year	DOI	Abstract	Keywords
An Approach to Develop a Sustainable Preference Index for Self Compacting Concrete	2020	10.1088/1757-899X/998/1/012058	In construction industry, selecting a right material from the various alternatives is a critical task in contributing to promote sustainability. For evaluating performance of the material different conflicting attributes are considered which is very ...	-
Decision-making information support methodology for protective measures implementation in the AIC	2020	10.1088/1755-1315/613/1/012073	The article describes the methodology of decision-making information support for protective measures implementation in the agro-industrial complex (AIC). The aim is to determine the rational area of treated land plots for protective measures implementation as ...	Analytic hierarchy process; Application programs; Biotechnology; Crops; Engineering education; Environmental management; Environmental technology ...
An integrated model for selecting suppliers on the basis of sustainability innovation	2020	10.1016/j.jclepro.2020.123261	In today's competitive business environment, corporations attempt to achieve sustainability through innovation. Innovation is considered by researchers and scholars to be a key driver for achieving sustainability...	BWM; PROMETHEE; Supplier selection; Sustainable innovation; Sustainable supply chain management ...

3.2 Text pre-processing

Since we load the selected papers into Google Colab (python programming) , we applied 'title , 'abstract' and 'keywords' column of each paper to handle them as a single document to form a new column called 'Article text'. Python code and 'Article text' column can be seen in Figure 3 and Figure 4 below.

```

1 cols_article_text = ['Title', 'Abstract', 'Author Keywords', 'Index Keywords']
2 df_dataset["Artcile text"] = df_dataset[cols].apply(lambda row: '_'
3             .join(row.values.astype(str)), axis=1)
4 cols_keywords = ['Author Keywords', 'Index Keywords']
5 df_dataset["Keywords"] = df_dataset[cols_keywords].apply(lambda row: '_'
6             .join(row.values.astype(str)), axis=1)
7 dataset_drop = df_dataset.drop(columns = ['Abstract', 'Link', 'Authors', 'Source',
8             'Author Keywords', 'Index Keywords',
9             'Source title', 'Cited by', 'DOI',
10            'Document Type'], axis = 1)
11 dataset_drop.tail()

```

	Title	Year	Artcile text	Keywords
1010	Integration of artificial neural network and M...	2010	Integration of artificial neural network and M...	Analytic network process; Artificial neural ne...
1011	Assessing the sustainability of forest managem...	2010	Assessing the sustainability of forest managem...	Criteria and indicators; Ethiopia; Exclosures;...
1012	An analytical method for the measurement of en...	2010	An analytical method for the measurement of en...	Energy indicators; Multi-criteria decision met...
1013	Environmental performance evaluation of agro-i...	2010	Environmental performance evaluation of agro-i...	Environmental impact; Methodology; Multi-crite...
1014	Multicriteria analysis to evaluate the energet...	2010	Multicriteria analysis to evaluate the energet...	Bio-energy; Life Cycle Assessment (LCA); Multi...

Figure 3 Load and merge a raw text data in Google Colab

Article text

Decision-making information support methodology for protective measures implementation in the AIC. The article describes the methodology of decision-making information support for protective measures implementation in the agro-industrial complex (AIC). The aim is to determine the rational area of treated land plots for protective measures implementation as components of cultivated crop area-based industrial technology. The methodology is based on the method of analytic hierarchy process (AHP) constructing and analyzing of multicriteria decision-making task solving in the selection of a rational land plot area for protection treatment. The following selection criteria are used: land plots route survey results giving the information about crops diseases throughout the area under study; affected plant damage degree; diseases harmfulness; environmental conditions (air temperature, precipitation (humidity)); probable losses. A hierarchical estimation of the used criteria importance is obtained, and an information model is formed. The model made it possible to establish a rational land plot option recommended for treatment. The novelty of the proposed methodology for solving the task is the following: the rationale for information technology (IT) application for a new class of tasks dealt with by the AIC and the technology practical focus on environmental friendliness ensuring and resources saving; on the basis of the methodology, the scientifically-based method, that does not require the involvement of a large number of experts, is used; the software implementation of the methodology on a PC does not impose special requirements for computing resources. Analytic hierarchy process; Application programs; Biotechnology; Crops; Engineering education; Environmental management; Environmental technology; Personal computers; Surveys; Sustainable development; Agro-industrial complex; Analytic hierarchy process (ahp); Environmental conditions; Environmental friendliness; Hierarchical estimation; Industrial technology; Multi criteria decision making; Software implementation; Decision making

→ Title

→ Abstract

→ Keywords

Figure 4 A sample of “Article text” data

In the text pre-processing method, a course of cleaning processes (tokenization, lowercase, etc.) to exceptional results of documents is then presented in Figure 5.

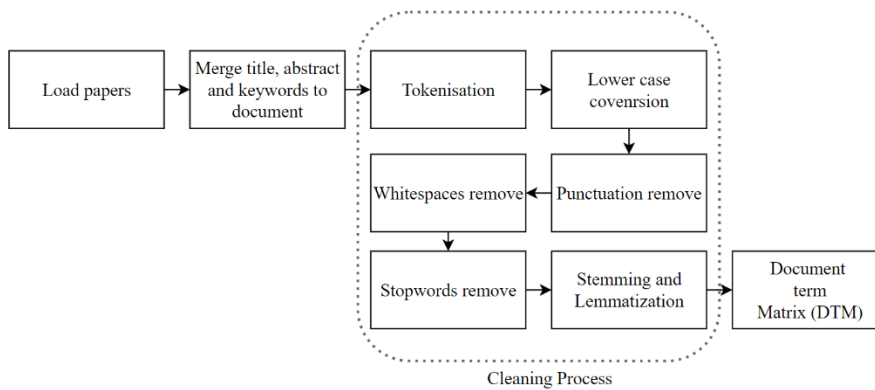


Figure 5 Overview of cleaning process in text pre-processing

The cleaning process is usually a repetitive one, as it can be challenging to locate the entire misinterpreted and meaningless words. The corpora of several papers consist of different words, which means that an accurate cleaning process cannot be guaranteed when a new review is performed. Therefore, developing features and applying text mining packages for cleaning raw text can contribute to a better cleaning process. Figure 6 shows the process of text cleaning in an example article.

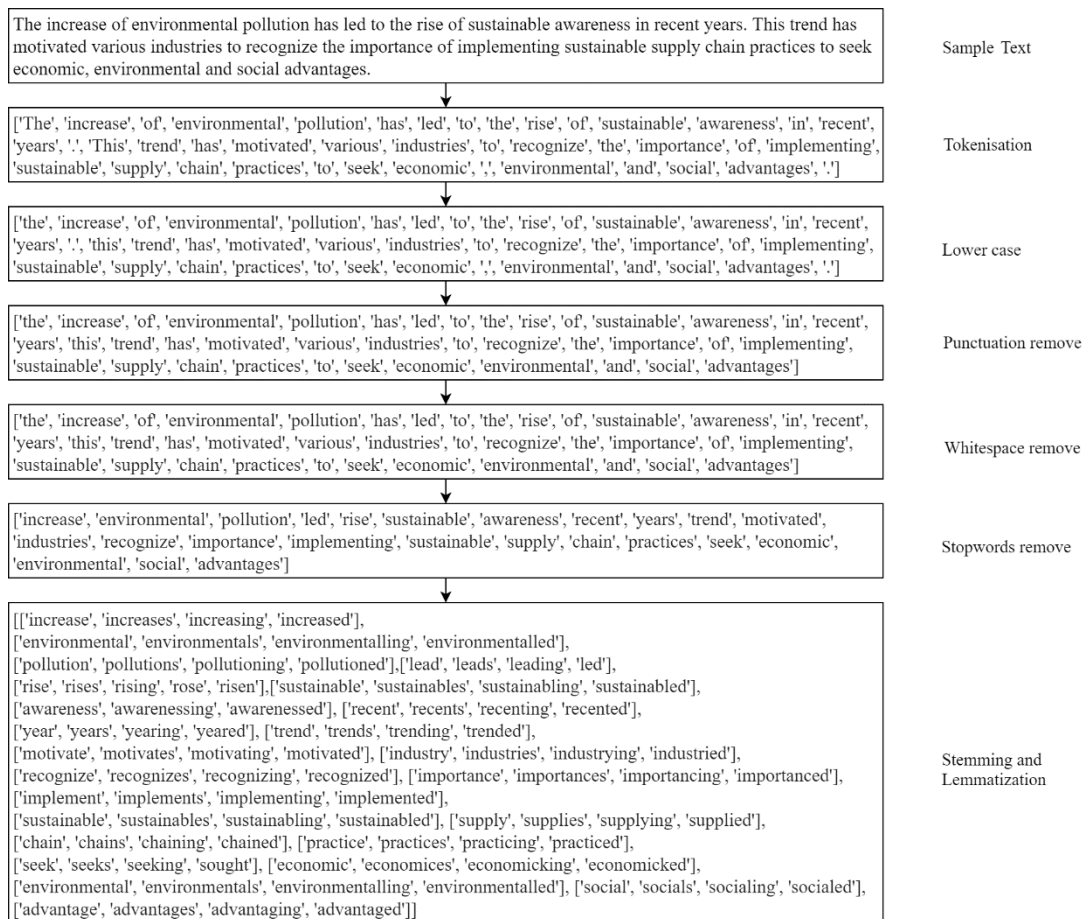


Figure 6 Text-cleaning process sample

After text preprocessing, the cleaned and preprocessed text data was inserted into a new column in the data frame labelled "Clean Article Text" and then de-tokenized. The example from the "Clean Article Text" column is shown in Table 4.

Table 4 Comparison of documents current-former cleaning process

No.	Article text (former)	Clean article text (current)
1	An Approach to Develop a Sustainable Preference Index for Self Compacting Concrete in construction industry, selecting a right material from the various alternatives is a critical task in contributing to promote sustainability...	approach develop sustainable preference index self compact concrete in construction industry select right material various alternative critical task contribute promote sustainability ...
2	The aim is to determine the rational area of treated land plots for protective measures implementation as components of cultivated crop area-based industrial technology ...	determine rational area treat land plot protective measure implementation component cultivate crop area based industrial technology ...

The final step of preprocessing is to transform the documents into the document term matrix representation to generate the input for LDA. As input to LDA, the preprocessed corpus was aligned to a document term matrix (DTM). A document term matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of documents. Figure 7 shows how the usage of some user specified words changes over time. It also represents the number of each term per document in Table 5. To do this, the Gensim package can generate a unique ID for each word in the document.

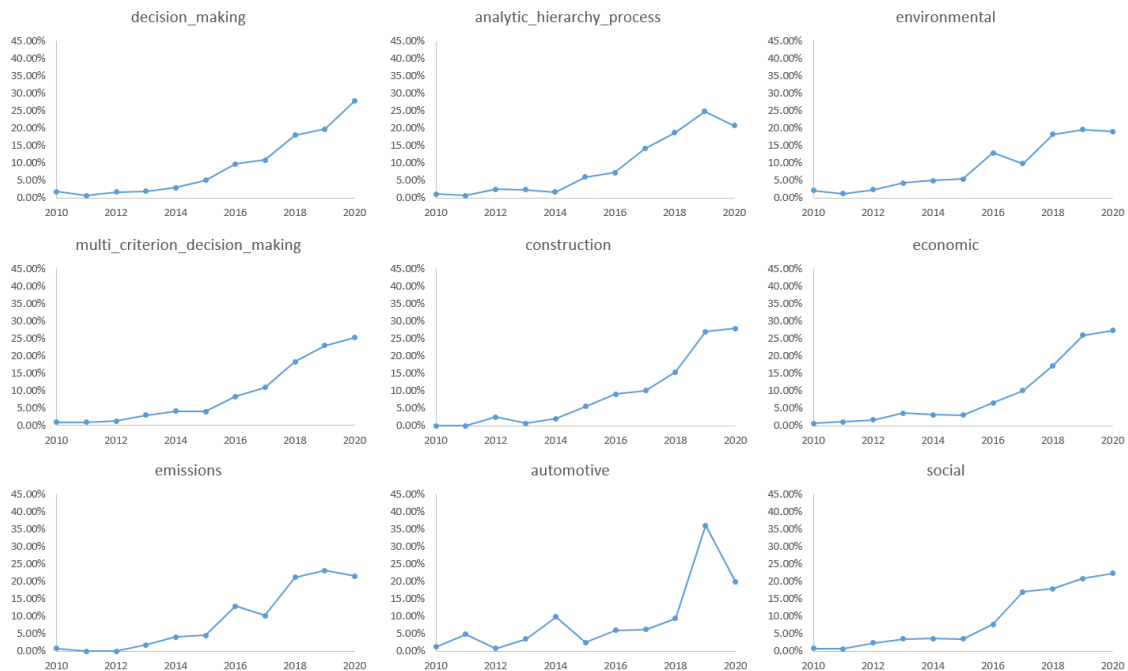


Figure 7 Trend of % frequency of sample word from 2010 to 2020

Table 5 lists the 20 words with the highest frequencies, as well as words taken from the Clean Article Text column.

Table 5 Top 20 no. of occurrences from ‘Clean article text’ in 1,150 papers

Rank	Words	Frequency	Rank	Words	Frequency
1	sustainability	1704	11	fuzzy	524
2	method	1599	12	policy	444
3	analysis	992	13	stakeholder	433
4	energy	937	14	cost	431
5	criterion	751	15	optimization	424
6	environmental	718	16	construction	398
7	green	654	17	ranking	352
8	decision making	595	18	manufacturing	237
9	design	586	19	ahp	227
10	evaluation	557	20	water	209

Then, we create a wordcloud to see the most common words used in papers from 2010 to 2020, as shown in Figure 8.



Figure 8 Summary a frequent word in Wordcloud

3.3 Topic modelling approach

Topic models have several advantages, both from an algorithmic and a practical point of view. First, they have principled mathematical foundations that help us identify the mechanism of document structure. Second, they do not require prior labeling of documents, so documents can be viewed without the help of expert knowledge (Blei et al. 2003). Finally, they can shape and organize documents without any doubt (Griffiths and Steyvers 2004). Due to these advantages, topic models have gained new importance and have been successfully applied to a wide range of text mining tasks. LDA, proposed by Blei (2012), is the most widely used topic modeling algorithm and has several practical advantages over other topic modeling algorithms. It can classify the combination of existing topics in new documents without updating the current model. The key idea of LDA is that documents are specified as random mixtures over latent topics, each of which is defined by a distribution over words.

Therefore, once the papers have been cleaned, the LDA method can be operate. To build the LDA model, these settings consist of three key parameters alpha (α), beta (β) and the number of topics (K). Alpha (α) estimated the topic distribution over documents, Beta (β) determines the specificity of topics. To build LDA model, we employed alpha = 50/K, beta = 0.1 and the number of topics (K) = 20.

4. Results

The LDA results from Google Colab in Table 6 show 20 topics, each having 10 keywords, in 1,015 papers.

Table 6 20 topics in sustainable supplier selection (1,015 papers)

No. of Keywords	Topic # 00	Topic # 01	Topic # 02	...	Topic # 17	Topic # 18	Topic # 19
0	model	supplier	energy	...	china	food	concrete
1	method	supplier selection	renewable energy	...	power	food waste	construction
2	analysis	industry	electricity	...	power generation	local	aggregate
3	development	suppliers	biomass	...	performance evaluation	production	material
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	approach	procurement	energy efficiency	...	life cycle	evaluate	chain
9	criterion	criterion	electricity generation	...	materials	operation	chains

Then, the LDA results from 1,015 papers are viewed in Figure 9. The closer the distance among the circles, the more related topics with words among the circles.

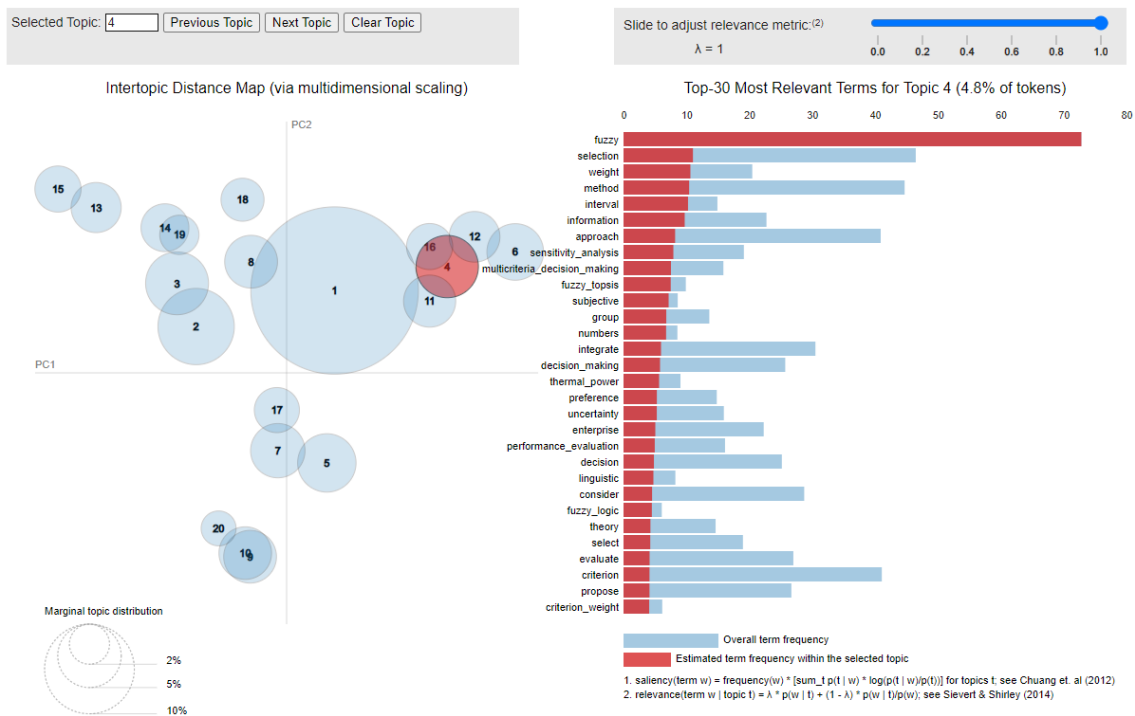


Figure 9 20 topics of papers by LDAvis results

Figure 9 shows the two major groups with a small distance between the circles. The first group consists of topic 1, 4, 6, 8, 11, 12 and 16. There is a relationship in decision making analysis with linguistics, fuzzy logic, which has increased in the context of sustainable supplier selection techniques. In addition, China is still leading the research on sustainable supplier selection. The second group is topic 2, 3, 19 and 14, which focuses

on the industry in terms of environmental impact, including renewable energy, biomass, electricity and food waste from production to select sustainable suppliers. However, there is a positive trend in the automotive industry to apply supplier selection for sustainability. In addition to the two topic areas, it is interesting to note that topic 18 deals with risk assessment, which stays away from the other topic areas. This does not mean that research with such topics is not necessary or weak. Rather, it may be an emerging research topic. In particular, the research topic of findings could be triggered by them.

4.1 Clusters by “Clean article text”

Table 7 shows five clusters consisting of the following keywords among 1,015 articles associated with each cluster, as shown in Figure 10.

Table 7 Five clusters from LDAvis results

Cluster	Term for each topic
1	decision making, multicriteria analysis, fuzzy logic, linguistic, criterion, economic, China, ahp, anp
2	design, construction, building, material, eco-design, concrete, production, manufacturing
3	urban, recycling, waste management, social impact, automotive, delivery
4	green, environmental, renewable energy, energy consumption, optimization, electricity, food waste
5	soil, chemical, land use , agriculture, water management

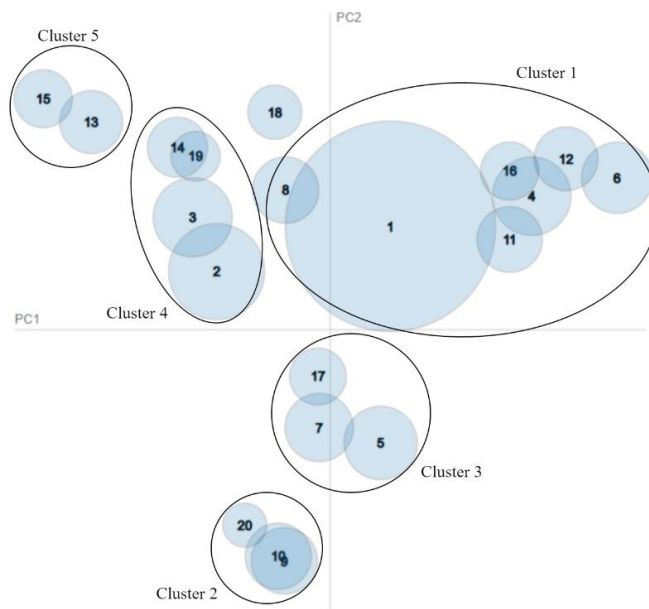


Figure 10 Cluster labelling from LDAvis

4.2 Publication papers on economy country conditions

Distribution of published papers based on country conditions is an interesting topic in our review paper. In this section, based on developed and developing countries we showed that, which economy region used decision making approaches in sustainable supplier selection in Figure 11. Still, Thailand, there is lack of research in this field.

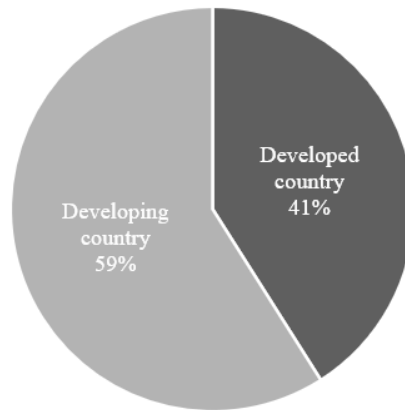


Figure 11 Distribution of papers based on economy country conditions

4.3 Publication by country

We also analyzed the distribution of reviewed papers by country of origin. Table 8 shows the countries in the field of sustainable supplier selection and evaluation. As can be seen, more than 70% of the publications (208 publications) are from four countries (China, Iran, India and United States), where China is the leading country in terms of number of publications with 73 publications, followed by Iran with 54 publications. On the other hand, there are only a few publications in Thailand that deal with sustainability in supplier selection.

Table 8 Distribution of papers based on country (Top 10)

Country	Number	Percentage (%)
China	73	22.26
Iran	54	16.46
India	45	13.72
United States	36	10.98
Taiwan	29	8.84
United Kingdom	23	7.01
Malaysia	21	6.40
Germany	20	6.10
Australia	14	4.27
Turkey	13	3.96

4.4 Publication by Multicriteria decision making approaches (MCDM)

In this study, we summarize two categories used for sustainable supplier evaluation and selection, namely the single approach and the hybrid approach.

Table 9 the classification of MCDM techniques

Single approach	Percentages	Hybrid approach	Percentages
AHP	24.29	AHP, TOPSIS	36.04
ANP	21.43	AHP, ANP	12.61
TOPSIS	34.29	ANP, ELECTRE	10.81
VIKOR	4.29	AHP, PROMETHEE	10.81
ELECTRE	4.29	Fuzzy AHP, Fuzzy TOPSIS	22.52
PROMETHEE	7.14	Fuzzy ANP, Fuzzy TOPSIS	3.60
BWM	2.86	Fuzzy ANP, DEMATEL	1.80
MAUT	1.43	Fuzzy DEMATEL, Fuzzy VIKOR	1.80

Table 9 shows that about 80% of the single or individual approaches refer to AHP, ANP and TOPIS. As a hybrid or combined approach, AHP, ANP and TOPSIS are introduced to integrate between these techniques. Nowadays, most authors focus on modelling approaches that support combined multi-criteria decision making rather than single approaches. However, in sustainable supplier evaluation and selection, there is a gap between using single and hybrid approaches to solve this problem. Therefore, integrating fuzzy logic to create a novel approach in MCDM techniques is still a space for research and overcome the shortcomings of traditional methods.

5. Conclusion

This study identified 20 sustainable supplier selection research topics and examined their changes and relationships by applying the LDA theme modelling approach to 1,015 articles published in the SCOPUS database over the past decade. The findings on the research trends in the articles during the period of 2010-2020 are summarized in five key trend themes. First, the majority of research topics in sustainable supplier selection deal with a multi-criteria decision-making approach. A single approach has a negative trend, while a hybrid MCDM approach has a positive trend over the past 10 years. Second, although the construction industry mostly applies sustainable supplier selection, the automotive industry has increased the trend to focus on sustainable supplier selection. Third, carbon emissions and renewable energy still play an important role, not only in economic and environmental terms, but also in terms of social impact for the welfare of people in the future generation. The fourth theme shows that circular economy trends have a significant impact on sustainability, such as recycling and waste management. Finally, the impact of land use in agriculture has also increased the trends in sustainability. Nevertheless, this study also has some limitations that could serve as areas for future research. The primary limitation is that we show the themes for the entire 10-year period. We cannot discover the evolution of the themes using such a cumulative approach. Next, we applied the basic LDA model to discover topics and their popularity trends over time. While there are LDA extensions, these have been evolved. Finally, we put only 20 topics through the majority of topic modelling approach. For further research, expanding the scope of research in this way would support explicit consideration for sustainability transformation.

6. References

- Antons, D., Kleer, R., & Salge, T. O. (2016). Mapping the topic landscape of JPIM, 1984–2013. In search of hidden structures and development trajectories. *Journal of Product Innovation Management*, 33(6), 726–749.
- Bai, C., & Sarkis, J. (2014). Determining and Applying Sustainable Supplier Key Performance Indicators. *Supply Chain Management: An International Journal*, 19(3), 275–291.
- Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Blei, D.M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84.
- Chai, J., Liu N.K., & Ngai, W. T. (2013). Application of Decision-making Techniques in Supplier Selection: A Systematic Review of Literature. *Expert Systems with Applications*, 40(10), 3872–3885.

- Genovese, A., Lenny Koh, S., Bruno, G., & Esposito, E. (2013). Greener supplier selection: state of the art and some empirical evidence. *International Journal of Production Research*, 51(10), 2868-2886.
- Griffiths, T.L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101, 5228–5235.
- Mani, V., Gunasekaran, A., & Delgado, C. (2018). Enhancing supply chain performance through supplier social sustainability: an emerging economy perspective. *International Journal of Production Economics*, 195, 259-272.
- Shahryari Nia, A., Olfat, L., Esmaeili, A., Rostamzadeh, R., & Antuchevičienė, J. (2016). Using fuzzy Choquet Integral operator for supplier selection with environmental considerations. *Journal of Business Economics and Management*, 17(4), 503-526
- Tracey, M., & Leng Tan, C. (2001). Empirical analysis of supplier selection and involvement, customer satisfaction, and firm performance. *Supply Chain Management: An International Journal*, 6(4), 174-188.
- Zhou, X., & Xu, Z. (2018). An Integrated Sustainable Supplier Selection Approach Based on Hybrid Information Aggregation. *Sustainability*, 10(7), 2543.