The Bibliometric Analysis on Finite Mixture Model

Seuk Yen Phoong¹, Shi Ling Khek¹, and Seuk Wai Phoong²

Abstract
A finite mixture model is well-known in statistics due to its versatility and is being actively researched. This paper reviews finite mixture model literature via bibliometric analysis, focusing on the trend and link between finite mixture model studies. The bibliometric analysis consists of four main phases; formulating research questions, locating research, research selection, evaluation, and analyzing and synthesizing selected papers. There are 667 journal articles extracted from the Web of Science (WoS) database from publications within 1988 to 2020. The Biblioshiny with R packages and VOSViewer were used as analytical tools. The findings show that there is an increasing trend of annual publication on the finite mixture model study. The results also outline key journals and the highest cited articles. Network analysis was also conducted and explored in scientific cooperation in the finite mixture model study. This study proposed a research agenda in the finite mixture model by identifying its current state and population trends.

Keywords
bibliometric analysis, finite mixture model, mixture model, three-field plot

Introduction
Mixture models have been used in statistics since the late 19th century when Pearson (1894) proposed studying the morphometry of crabs. Previously, mixture model applications faced a considerable challenge, especially before the advent of computers (Lindsay, 1995). The 21st century witnessed the prodigious flow of data along with technological advancements. The rapid development of computer technology provided statisticians with the idea of exploring probabilistic models for analyzing and handling complex properties of massive data. The mixture model is ubiquitous in statistics as it has been proven to be an excellent framework for modeling clustered data (Simola et al., 2021). Subsequently, a considerable shift from a single distribution to a mixture distribution occurred due to increased methodological complications (Makov, 2001). Mixture models can specify the number of different component distributions, making it viable for handling complex systems (Marin et al., 2005).

A mixture model is defined as a probabilistic model representing subpopulations within a population without needing to undergo the subpopulation identification process where an individual observation belongs to observed datasets. In other words, the mixture model contains datasets that are assumed to possess more than one distribution combined in different proportions (Makov, 2001). A mixture model is linked to mixture distribution, denoted as a probability distribution of observations within a population. A mixture distribution consists of finite or infinite components (Phoong et al., 2016), which can describe datasets with distinct features (Marin et al., 2005). However, the purpose of a mixture model is to build a statistical inference about subpopulations’ properties with observations provided on a population without subpopulations identities’ information. Mixture distributions are applied to illustrate the properties of the overall population based on subpopulations (McLachlan & Basford, 1988).

The mixture model offers distinct ways for interpretation, direct and indirect. For the former, the mixture model assumes that the underlying population with \( k \) subpopulations and observation \( X \) belongs to one of these subpopulations without knowledge regarding the subpopulations’ identity. Therefore, it can be assumed that the mixture model is a direct representation of an underlying phenomenon. The mixture model is applied and used as an approximation to an unknown distribution in the indirect...
interpretation, where these subpopulations lack physical and direct interpretations.

Technological evolution leads to tremendous data production, which attracted many studies to devise the most plausible statistical method for estimating a model’s parameters (Makov, 2001). The finite mixture model is an extensive statistical model used in a broad range of scientific fields, such as machine learning, data mining, estimation of density, pattern recognition, and image analysis (McLachlan & Peel, 2000) due to its appealing interpretability (Zhu & Melnykov, 2018). Articles reported studying finite mixture model in various real-life situations, such as power disaggregation and criminal studies. Zhou et al. (2019) used the finite mixture model to disaggregate commercial buildings’ power consumption, which allows the transformation of the power disaggregating process to behavioral analysis. Zhu and Melnykov (2018) studied crime trends using the finite mixture model due to its remarkable modeling flexibility in heterogeneous data. Literature on the subject reviews the model’s usefulness in handling real-life situations. Although the finite mixture model is becoming prominent in statistical modeling over the past few years, the finite mixture model’s bibliometric analysis remains dismal. Therefore, it is essential to conduct a thorough analysis to identify the most popular research trends or directions of the finite mixture model analysis. A comprehensive quantitative evaluation of the finite mixture model literature is presented using bibliometric analysis in this study.

The main objectives of this paper are:

i. To analyze trends of finite mixture model including publication performance, the most productive journals, critical articles, and the most influential keywords.

ii. To provide a comprehensive view toward the connections of finite mixture model study using network mapping.

iii. To analyze the relations among journals, keywords, countries, and authors to apply three field plot.

Finite mixture model is a getting attention in analyzing and modeling datasets, especially in the finance and economics area. This motivate us to investigate the trend of using this statistical method in publications. Therefore, bibliometric analysis of finite mixture model can attract the researchers’ attention, and thus popularize this mixture model. Furthermore, the contribution of this study identifies the applications of finite mixture model in various fields, including medicine, geoscience, and remote sensing. From this study, the collaboration between Asia countries can be further improved to increase the publication volume of finite mixture model among Asia nations. Other than that, the analysis of journals allows researchers to explore more about the leading sources used to discuss the usability of finite mixture model. Besides, the keywords analysis enable researchers to understand about the relatedness of this model with other popular methods, such as maximum likelihood estimation and EM algorithm. With this, it is believed that the contribution of this study can gain the attractions in research field especially in this technology era.

The remaining sections of this paper are organized as follows. Section 2 highlights the finite mixture model overview, while Section 3 discusses the methodology utilized in extracting relevant papers to accomplish research objectives. Section 4 focuses on the extraction method of finite mixture literature, while Section 5 discusses the analysis results of scientific production, and Section 6 states the network analysis results and explanation. Finally, Section 7 concludes the work and elucidates the limitations of the study.

An Overview of Finite Mixture Model

Since there is an enormous amount of data collected over the years due to computer technology’s rapid development, different models have been introduced and used for modeling. In the 19th century, the mixture model with finite dimension, called the finite mixture model, first appeared in a modern statistical literature article (Newcomb, 1886) for displaying outliers. Although the finite mixture model often faces technical challenges from population inference due to its non-regularity (Chen, 2017; Chung et al., 2004) it remains popular. Nowadays, a finite mixture model is utilized in various fields to model the distributions of a wide variety of random phenomena. The finite mixture model’s work has proliferated in various fields, including agriculture, economics, and psychiatry (McLachlan & Peel, 2000).

The finite mixture model is a powerful probabilistic model used to model univariate and multivariate datasets. It offers a natural representation of heterogeneity in a finite number of latent classes and is also known as the latent class model (Masyn, 2013). The finite mixture model also acts as an unsupervised learning model, which gives a probabilistic model-based method to unsupervised learning (Bishop, 2007).

The finite mixture model is hierarchical (Li & Yamanishi, 2003); hence, there are $m$-components with $N$ random variables in the model. Each of the $m$-component is derived from the same parametric distribution family with different parameters. For instance, all of the components might come from the Gaussian mixture distribution but with different mean and variance. Moreover, the finite mixture model’s flexible properties allow it to be fitted with different types of distributions, such as beta, Weibull, Gaussian, gamma, and Poisson (McLachlan & Peel, 2000), while every distribution has its respective criteria and properties.

Since finite mixture model is a great, well-known approach utilized to model a variety of datasets, a bibliometric analysis is introduced to analyze the trend of this model. The bibliometric analysis regarded as a part of scientometrics, where mathematics and statistics are utilized for data analysis (Aparicio, 2019) to elucidate the research trends in areas such as agroecosystem (Liu et al., 2019), disease (Lou, 2003), etc.
This review technique can be conducted via analysis of bibliometric feature and document system of selected papers (Van Eck & Waltman, 2010; Van Leeuwen et al., 2013) by evaluating and identifying scientific publications in a relevant field (Carrion-Mero, 2020). This approach makes bibliometric analysis different from other reviews methods, as it is adequate for quantitative analysis (Mao et al., 2015) within a broad range of research papers.

**Methodology**

Methodology refers to the strategy proposed to realize the research objectives. This study’s main contribution is the provision of a deep understanding of the finite mixture model literature through bibliometric analysis. Bibliometric analysis’s systematic, rigorous, and transparent procedures can help researchers get high-quality reviews studies, as it involves stringent stages such as illustration, evaluation, and monitoring of relevant studies (Lamboglia et al., 2020). Therefore, this analysis can help researchers have a comprehensive view of the leading research characteristics and, hence, identify the trends by analyzing an enormous number of relevant papers.

The research hypotheses of this research are as follows:

- **H₁**: There is a significant relationship between keywords, journals, and countries.
- **H₂**: There is a significant relationship between authors, keywords, and sources.

**Bibliometric Analysis of Finite Mixture Model**

Over the years, literature reviews have been conducted to provide an in-depth description, including determining potential research gaps and highlighting the threshold of knowledge (Tranfield et al., 2003) in a given field by undergoing stringent procedures that can minimize bias to generate reliable results (Moher et al., 2009). Concurrently, bibliometrics analysis, which contemplates almost the same systematic methodology, is suggested by Rowley and Slack (2004) to structure literature reviews. The term “bibliometrics” was first coined by Pritchard (1969), who defines bibliometrics as “the implementation of mathematics and statistics to books and other communication media.” However, the research of bibliometrics analysis can be traced back to the 20th century. Bibliometrics analysis pioneer is Cole and Eales (1917), who studied comparative anatomy literature by calculating the relevant books and articles and classifying these documents according to the country from 1543 until 1860. The bibliometrics analysis’s primary significance is its ability to synthesize vast amounts of bibliographical data to identify research trends and characteristics (Zhang & Liang, 2020). The review technique is the primary key to ensure that synthesizing all research is always rigorous and transparent. Reviews studies not conducted systematically can result in serious errors (Snyder, 2019). Therefore, a step-by-step process is required for bibliometrics studies to achieve rigor and transparency in the analysis process. In this study, the methodology of bibliometrics analysis is referred to and modified from the steps of systematic literature reviews introduced by Denyer and Tranfield (2009) because it described every critical stage required for conducting reviews studies in order for high-quality analysis. Subsequently, four essential stages are involved in this study’s methodology, presented in Figure 1. Generally, it is an analytical approach that involves critically evaluating, selecting, interpreting, and synthesizing relevant studies on the finite mixture model analysis.

**Formulating Research Questions**

The research questions are:

i. What are the finite mixture model trends in publication performance, the most productive journals, critical articles, and the most influential keywords?

ii. What are the connections of the finite mixture model study using network mapping?

iii. What are the relations between journals, keywords, countries, and authors to implement the three-field plot?

**Locating Research**

The Web of Science (WoS) has been selected as the primary database because it is the most extensive database that allows users to access the world’s leading academic documents in
the sciences, arts, and social sciences. WoS is also recognized as one of the primary databases with high convergence of peer-review journals (Meho & Yang, 2007). However, only reviewed English journal articles published in the WoS are used in this study. The primary reason for restricting the articles’ selection to journal articles is for quality control purposes (David & Han, 2004). In other words, this study does not include other types of documents, such as book chapter, proceedings, and reviews.

This study emphasizes the analysis of the finite mixture model. Researchers used multiple techniques to search for relevant articles involving keywords and Boolean logic. For instance, the keywords used in systematic search are “mixture model” and “finite mixture model.” Via the Boolean logic application, the papers consisting of the exact phrases of “mixture model” and “finite mixture model” are extracted.

**Research Selection and Evaluation**

Due to the vast number of finite mixture model studies over the past decade, selecting and evaluating studies is critical to ensure that selected ones can address the research questions. The articles selection phase is initiated via WoS. The time was chosen from 1988 to 2020, as the finite mixture model studies started appearing in 1988 and rapidly increased over the years. The search keyword was initiated using the keywords of “mixture model” followed by “mixture model” and “finite mixture model” to narrow the results.

Subsequently, the inclusion and exclusion criteria of existing articles are provided to filter out irrelevant articles from the selected list. There are only journal articles in English to avoid any misinterpretation. Also, the articles selected must be published in peer-reviewed journals in the database in the past 51 years, or 1988 to 2020. Studies issued in peer-reviewed journals confer legitimacy because these articles are always involved in the necessary and vital peer-review process (Goldbeck-Wood, 1999).

A detailed articles extraction method is displayed in Figure 2. Based on the criteria above, the preliminary results furnished 19,635 documents emphasizing a mixture model from 1988 until 2020. Next, among these articles, the results showed 896 documents on the finite mixture model within the same time. Subsequently, there are peer-reviewed articles in this study; hence, the documents were filtered to remove other documents such as early access paper, proceeding papers, and review papers. As a result, a total of 679 articles on the finite mixture model are presented in WoS. Among these articles, 677 of them were in English. After that, the authors conducted the time filtration to ensure all the relevant articles were published within the designated time. In short, the total number of related peer-reviewed journal articles selected was 665, successfully published over the past 51 years. These articles were then analyzed and synthesized critically using the Biblioshiny with R packages and VOSViewer. Biblioshiny is an application for non-coders for bibliometric analysis, such as computing general statistics and presenting the link among selected items using the Three-FieldsPlot (Van Eck & Waltman, 2014). VOSViewer is one of the popular bibliometric software that can be used to construct and explore the co-occurrence networks of authors, countries, and journals (Montalvan-Burbano et al., 2020) using the 2D image mapping technique (Shah et al., 2019; Van Eck & Waltman, 2010). In short, these
bibliometric analysis tools have different strengths, where Biblioshiny can be applied for fundamental bibliometric analysis, while VOSViewer is an excellent tool that can be used for networks visualization (Moral-Munoz et al., 2020).

Analysis and synthesis of selected studies. In this stage, selected articles’ publication trends were analyzed via various publications such as authors, institutions, journals, most cited articles, and author keywords.

Results and Discussion

Publication Performance

The publication performance of relevant journal articles was evaluated from 1988 until 2020, as shown in Figure 3, to provide a comprehensive view of finite mixture model publication trends. The publication of the finite mixture model study first appeared in the late 1980s; it grew in the 1990s until the early 2000s, then dramatically increased after 2004. Generally, the finite mixture model study shows a rapid annual growth trend, especially with the advances of computer technology.

Countries Scientific Production

Over the years, the scientific production of finite mixture model articles was contributed by 61 countries. The top 12 leading countries in the production of finite mixture model literature are listed in Table 1, where it can be seen that the USA ranks first, with 257 publications within 1988 to 2020, followed by China, Canada, England, and Italy with 116, 64, 47, and 39 publications, respectively.

It can be seen that the higher the number of articles published in a country, the higher the number of citations gained by said country, referring to Table 1. For example, countries such as the USA, China, Canada, and England, the top 4 leading countries in finite mixture model articles production, also are countries with the highest citations. Nevertheless, exceptions are evident for Italy and Germany, where the citation number is much lower than the countries with lower publications, such as France, Australia, and Taiwan. The average citation for the top 12 countries seen in Table 1 shows that the USA is the highly cited country with 33.93 citations per document, while China is the least cited country with 10.82 citations per document.

The Most Productive Journals

In this study, a total of 667 articles published in 370 journals were collected. Table 2 shows the top 12 most influential
Table 2. Scientific Production of the Top 12 Journals.

| Journals                                             | TP | TP% | TC    | AC   | IF    |
|------------------------------------------------------|----|-----|-------|------|-------|
| Computational Statistics & Data Analysis             | 20 | 3.00| 193   | 9.65 | 1.186 |
| Statistics in Medicine                               | 15 | 2.25| 165   | 11.0 | 1.783 |
| Biometrics                                           | 10 | 1.50| 1,182 | 118.2| 1.711 |
| Journal of the American Statistical Association      | 10 | 1.50| 542   | 54.2 | 3.989 |
| Pattern Recognition                                  | 9  | 1.35| 164   | 18.22| 7.196 |
| Quaternary Geochronology                             | 9  | 1.35| 405   | 45.0 | 3.079 |
| Statistics and Computing                             | 9  | 1.35| 201   | 22.33| 3.035 |
| Canadian Journal of Statistics-Revue Canadienne De Statistique | 8  | 1.20| 112   | 14.0 | 0.656 |
| IEEE Transactions on Geoscience and Remote Sensing   | 8  | 1.20| 232   | 29.0 | 5.855 |
| Communications in Statistics-Simulation and Computation | 7  | 1.05| 32    | 4.57 | 0.651 |
| Journal of the Royal Statistical Society Series C-Applied Statistics | 7  | 1.05| 87    | 12.43| 1.590 |
| Neurocomputing                                       | 7  | 1.05| 83    | 11.86| 4.438 |

Note. TP = total publications; TC = total citations; AC = average citation; IF = impact factors from 2019 Journal Citation Reports.

The total number of citations for an article is always a topic of concern for researchers due to its vital role in elucidating an article’s academic impact. Other researchers often cite an article with excellent content. Table 3 shows the top 10 highly cited articles in the field of finite mixture models within the study period. The article with the highest citations is “Unsupervised learning of finite mixture models” by Figueiredo and Jain (2002), published in IEEE Transactions on Pattern Analysis and Machine Intelligence, with a total of 1,301 citations. The article “Unsupervised learning of finite mixture models” not only is the highest cited article but also ranks first place in an average citation (65.05). Subsequently, the “Finite mixture modeling with mixture outcomes using the EM algorithm” by Muthen and Shedden (1999) is the second most highly cited paper (886), followed by Carlin and Chib (1995), with “Bayesian model choice via Markov-Chain Monte-Carlo methods” (490).

Analysis of The Most-Cited Articles

The total number of citations for an article is always a topic of concern for researchers due to its vital role in elucidating an article’s academic impact. Other researchers often cite an article with excellent content. Table 3 shows the top 10 highly cited articles in the field of finite mixture models for the past 51 years. Computational Statistics & Data Analysis (20), Statistics in Medicine (15), Biometrics (10), and Journal of the American Statistical Association (10) are ranked as the top 4 most productive journals, accounting for 3.00%, 2.25%, 1.50%, and 1.50% respectively. Since the publication percentage for these top journals is not high, accounting for 8.25% among all publications indicates that there are many journals available related to the research in the field.

The total citation cited by Web of Science core for the top 10 key journals was evaluated and listed in Table 2. Biometrics showed the best performance in a total citation (1,182), followed by the Journal of the American Statistical Association (542), Quaternary Geochronology (405), IEEE Transactions on Geoscience and Remote Sensing (232), and Statistics and Computing (201). In terms of average citation, Biometrics (118.2), Journal of the American Statistical Association (54.2), Quaternary Geochronology (45.0), IEEE Transactions on Geoscience and Remote Sensing (29.0), and Statistics and Computing (22.33) were ranked as the top 5 journals with the highest average citation. In other words, in this study, it can be seen that the journal ranking for total citation and average citation are similar. Moreover, the impact factor of 2019 of the top 10 productive journals is shown in Table 2 to illustrate the journal’s importance. A higher impact factor means a more influential journal. Pattern Recognition gained the highest impact factor of 7.196, followed by the IEEE Transactions on Geoscience and Remote Sensing (5.855) and Neurocomputing (4.438).

Author’s Keywords

The author’s keyword is an essential item in bibliometric analysis, as researchers can identify the main trends in the literature. The most frequently used of the author’s keywords were analyzed and presented in Table 4. The keywords of “finite mixture model” is the most common keywords used by authors, with 237 times, accounting for 35.53%, followed by “EM algorithm” (51), “mixture model” (40), “finite mixture models” (38), and “clustering” (30). These keywords occur 30 or more times, accounting for 59.37% of the total of 667 articles.

Network Analysis

Keywords Co-Occurrence Analysis

In this study, the mapping analysis was used to identify the research hotspots and intellectual structure of the related field, with the presentation of visual maps via contents analysis of relevant articles (Dong & Chen, 2015), which are usually represented by distinct units of analysis such as countries, authors, and journals (Cobo et al., 2014). Table 5
shows the top 10 author’s keywords ranked by total link strength. According to Waltman et al. (2010), the article has been linked with others many times if the total link strength is high. Based on Table 5, the “finite mixture model” is the keyword with the best link strength (254), followed by “EM algorithm” (81) and “clustering” (51). Apart from that, these keywords also act as the top keywords in link numbers with 93 (“finite mixture model”), 40 (“EM algorithm”), and 33 (“clustering”).

The co-occurrence of the author’s keywords is presented in the form of a network shown in Figure 4 to better visualize the research hotspots in the field of finite mixture models. In this phase, the minimum number of keywords was set to 3 and 115 author’s keywords out of 2,151 keywords were classified and applied as visualization items in network analysis. The size of circles represents the total number of occurrences of related author’s keywords. Figure 4 reveals that the larger size of the circles, the higher frequency co-occurrence of the author’s keywords. Meanwhile, the distance between every circle’s elements is used to illustrate the topic’s similarities and its strength. Moreover, clusters can be found on the network analysis where different circles represent different keywords clusters. Based on Figure 4, 13 distinct clusters exist in representing individual subareas of the research in finite mixture models’ studies.

There are 13 clusters of author’s keywords found in the articles on finite mixture models. Clusters with red, green, blue, greenish-yellow, and purple colors, the top 5 main sub-fields are listed below.

- The red cluster consists of keywords such as “finite mixture models,” “data mining,” “latent class,” “finite mixture model (fmm),” “clustering,” “unsupervised learning,” and “expectation maximization.”
- The green cluster group together keywords such as “asymptotically normality,” “Bayesian,” “clustering analysis,” “density estimation,” “finite mixture,” “latent classes,” and “unobserved heterogeneity.”
- The blue cluster contains keywords such as “Bayesian analysis,” “clustering,” “latent class analysis,” “latent variables,” “missing data,” and “mixtures models.”
- The greenish-yellow cluster consists of the keywords “cluster analysis,” “EM-algorithm,” “latent variable,” “MCMC,” “dimension reduction,” and “ordinal data.”
- The purple cluster group together keywords such as “expectation maximization (EM),” “Gaussian mixture,” “Gaussian mixture model,” “student’s t-distribution,” “gradient descent,” and “image segmentation.”

From the results, the co-occurrence of the authors’ keywords showed diverse trends in the publication of finite mixture model papers. There is, therefore, a great chance of exploring the finite mixture model due to its multi-dimensional character in research.
Regional collaboration was analyzed and presented via network analysis of countries’ co-authorship. Table 6 shows the top 12 leading countries with the highest total link strength. The total link strength of a country represents the total number of papers where authors from different countries collaborated. The connection line between countries represents the total link strength. The link number denotes the number of countries linked to another. According to the values of total link strength seen in Table 6, the USA leads the group, with total link strength of 104 within the study period, followed by China (64) and England (61). Network analysis of countries co-authorship is highlighted in Figure 5. At this stage, the minimum number of documents was set to be one. The size of the circles illustrates the number of documents occurrence of countries. The larger the circle’s size, the higher the number of occurrence of documents in that country. The thicker the connection line, the greater the collaboration between both countries.

From the network analysis of countries co-authorship, it can be seen that a total of 61 countries represented by authors collaborating in finite mixture model literature. The USA is the top country with the highest number of publications and the best total link strength. The primary partner countries of the USA are China, Canada, and England. Also, China comes in second with the highest publications and strongest cooperation relations with its major partners such as the USA, Canada, and Germany. By referring to Figure 5, it can be seen that the co-authorship in the analysis of the finite mixture model is mainly focused on the developed countries, while the collaboration relations in developing countries remain slow. For instance, the top 9 countries with the highest link strength are monopolized by developed countries, while developing countries such as South Africa only ranks 10th in cooperation with other countries.

Three-Field Plots

Three-field plots were employed to elucidate the relationship between three distinct pieces of information. The three-field plots were created to visualize the proportion of selected items during the study period. In the plot, every item lies along with a rectangle. The height of the rectangles plays a role in illustrating the relations between elements in a different row. The stronger the relations between the elements, the higher the height of the rectangle. There are two types of three-field plots presented in the present study, as shown in Figures 6 and 7.

Relationship Between Keywords, Journals, and Countries

Figure 6 shows there is a relationship among keywords, journals, and countries. The top 10 most frequently used keywords are displayed on the figure’s left side to explore the finite mixture model studies’ research hotspots. The top 10 most productive journals using these keywords are represented in the middle part of the figure, while the most productive countries with the highest publications of finite mixture model papers are seen on the right side of the figure. The analysis depicts that research topics of “finite mixture model,” “mixture model,” and “EM algorithm” mainly from five productive countries such as the USA, China, Canada, Italy, and Australia are published in Computational Statistics & Data Analysis, Statistics in Medicine, Journal of American Statistical Association, Quaternary Geochronology, and IEEE Transaction on Geoscience and Remote Sensing.

Relationship Between Authors, Keywords, and Sources

The relation plot (Figure 7) shows there is relationship between the author’s name, keywords, and sources. The 10 most productive authors are listed on the left side of the figure, the top 10 keywords used by these authors are presented in the middle, while the top 10 influential journals
are displayed on the right side of Figure 7. As per Figure 7, there are five authors (i.e., Chen, Maiboroda, Lord, Zou, and Melnyov) and five sources (i.e., Computational Statistics & Data Analysis, Biometrics, Pattern Recognition, Journal of American Statistical Association, and Statistics in Medicine) have strong bond or relation with the keywords of “finite mixture model,” “mixture model,” and “EM algorithm.” The three keywords are also the most recurring keywords used in analyzing the finite mixture model among 667 journal articles.

**Conclusion**

This study presents a review on the finite mixture model within 1988 to 2020 via several dimensions, such as annual production trends, scientific production of journals and articles, author’s keywords, and country scientific production. After data extraction, a total of 667 publications (61 countries, 2,151 author’s keywords, and 370 journals) were analyzed using Biblioshiny with R packages and VOSViewer.

The bibliometric analysis results indicated that the number of publications on the finite mixture model studies increased, especially after 2004. However, most of the publications were mainly contributed by the USA (257), China (116), Canada (64), England (47), and Italy (39). Therefore, it can be concluded that the publications of finite mixture model papers are shaped mainly by developed countries.

The statistical computing of bibliometric analysis indicated that the most productive journals on finite mixture model are Computational Statistics & Data Analysis, Statistics in Medicine, Biometrics, and the Journal of the American Statistical Association and the Pattern Recognition. Furthermore, the highly cited articles in the finite mixture model area were identified to uncover the most influential articles in this field. Based on the analysis of the author’s keywords, the central research hotspots were examined, which are the finite mixture model, EM algorithm, mixture model, finite mixture models, and clustering.
Subsequently, network analysis was presented to provide better visualization of the finite mixture model literature. The analysis of keywords co-occurrence reveals that research in finite mixture model highly concerned the keywords of “finite mixture model,” where this keyword recorded the highest occurrence and link strength. Simultaneously,
diversity of research topics co-occurrence was determined via clusters identified from the keywords co-occurrence network analysis. For instance, “finite mixture model,” the most popular keyword, was found to co-occur with “Akaike Information Criterion,” “China,” “maximum likelihood estimation,” and “optically stimulated luminescence.” The “finite mixture models” always co-occurred with keywords such as “data mining,” “latent class,” “customer segmentation,” “expectation maximization,” “Markov random field,” and “unsupervised learning.” This scenario shows a multidimensional view on the field of finite mixture model since there are plenty of keywords bonded with the word of a finite mixture model.

Additionally, co-authorship among countries was displayed in the form of a network map. The countries co-authorship analysis depicts that the USA and China are leaders of collaboration in the publication of finite mixture model studies. Concurrently, both countries were also the top countries with the highest document publications.

The relations between keywords, journals, countries, and authors were displayed in three-field plots to provide a different perspective on finite mixture model literature analysis. By analyzing the most popular author’s keywords, researchers can explore related research topics by understanding the finite mixture model research trends. This analysis furnishes the information of the most productive countries and authors. By exploring the critical journals with the highest publications, researchers can analyze and then implement a technique to increase publication chance.

There are few limitations faced in this study, such as the database issue. This study emphasized published journal articles from the Web of Science only. However, further study should involve articles from other well-known databases such as Scopus, Google Scholar, and Science Direct in the bibliometric analysis, as the higher the documents analyzed can produce robust and reliable findings on the finite mixture model analysis. Furthermore, the bibliometric analysis should be extended to explore more items such as institutions and authors to provide a complete and comprehensive view on finite mixture model papers.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research has been carried out under Fundamental Research Grants Scheme (FRGS/1/2019/STG06/UPSI/02/2) provided by Ministry of Education of Malaysia. The authors would like to extend their gratitude to Universiti Pendidikan Sultan Idris (UPSI) that helped manage the grants.

**ORCID iDs**

Seuk Yen Phoong [https://orcid.org/0000-0002-9978-9771](https://orcid.org/0000-0002-9978-9771)
Seuk Wai Phoong [https://orcid.org/0000-0002-9925-0901](https://orcid.org/0000-0002-9925-0901)
References

Aparicio, G., Iturralde, T., & Maseda, A. (2019). Conceptual structure and perspectives on entrepreneurship education research: A bibliometric review. European Research on Management and Business Economics, 25(3), 105–113. https://doi.org/10.1016/j.errebe.2019.04.003

Arnold, L. J., & Roberts, R. G. (2009). Stochastic modelling of multi-grain equivalent dose (D-e) distributions: Implications for OSL dating of sediment mixtures. Quaternary Geochronology, 4(3), 204–230. https://doi.org/10.1016/j.quageo.2008.12.001

Bishop, C. M. (2007). Pattern recognition and machine learning. Springer.

Carlin, B. P., & Chib, S. (1995). Bayesian model choice via markov-chain monte-carlo methods. Journal of the Royal Statistical Society: Series B-Statistical Methodological, 57(3), 473–484.

Carrion-Mero, P., Montalvan-Burbano, N., Paz-Salas, N., & Morante-Carballe, F. (2020). Volcanic geomorphology: A review of worldwide research. Geosciences, 10(9), 347. https://doi.org/10.3390/geosciences10090347

Chen, J. (2017). On finite mixture models. Statistical Theory and Related Fields, 1(1), 15–27. https://doi.org/10.1007/24754269-017.321883

Chung, H., Lokken, E., & Schafer, J. L. (2004). Difficulties in drawing inferences with finite mixture models: A simple example with a simple solution. The American Statistician, 58(2), 152–158. https://doi.org/10.1198/0003130043286

Cobo, M. J., Chiclana, F., Collop, A., De Oña, J., & Herrera-Viedma, E. (2014). A bibliometric analysis of the intelligent transportation systems research based on science mapping. IEEE Transactions on Intelligent Transportation System, 15(2), 901–908. https://doi.org/10.1109/TITS.2013.2284756

Cole, F. J., & Eales, N. B. (1917). The history of comparative anatomy: Part 1—A statistical analysis of the literature. Science Progress, 11(44), 578–596.

David, R. J., & Han, S. K. (2004). A systematic assessment of the empirical support for transaction cost economics. Strategic Management Journal, 25(1), 39–58. https://doi.org/10.1002/smj.359

Deb, P., & Trivedi, P. K. (1997). Demand for medical care by the elderly: A finite mixture approach. Journal of Applied Econometrics, 12(3), 313–336.

Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In D. A. Buchanan & A. Bryman (Eds.), The Sage handbook of organizational research methods (pp. 671–689). SAGE.

Dong, D., & Chen, M. L. (2015). Publication trends and co-citation mapping of translation studies between 2000 and 2015. Scientometrics, 105(2), 1111–1128.

Figueiredo, M. A. T., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3), 381–396. https://doi.org/10.1109/34.990138

Goldbeck-Wood, S. (1999). Evidence on peer review: Scientific quality control or smokescreen? British Medical Journal, 318(7175), 44–45. https://doi.org/10.1136/bmj.318.7175.44

Gu, B., Sun, X. M., & Sheng, V. S. (2017). Structural minimax probability machine. IEEE Transactions on Neural Networks and Learning Systems, 28(7), 1646–1656. https://doi.org/10.1109/TNNLS.2016.2544779

Jedidi, K., Jagal, H. S., & DESarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. Marketing Science, 16(1), 39–59. https://doi.org/10.1287/mksc.16.1.39

Lamboglia, R., Lavorato, D., Scornavacca, E., & Za, S. (2020). Exploring the relationship between audit and technology. A bibliometric analysis. Meditari Accountancy Research. Advance online publication. https://doi.org/10.1108/MEDAR-03-2020-0836

Li, H., & Yamanishi, K. (2003). Topic analysis using a finite mixture model. Information Processing and Management, 39(4), 521–541. https://doi.org/10.1016/S0306-4573(02)00035-3

Lindsay, B. G. (1995). Mixture models: Theory, geometry and applications. NSF-CBMS regional conference series in probability and statistics (Vol. 5, pp. 1–163). Institute of Mathematical Statistics. https://doi.org/10.1214/cbms/1462106015

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. Journal of Statistical Software, 42(10), 1–29.

Liu, W., Wang, J., Li, C., Chen, B., & Sun, Y. (2019). Using bibliometric analysis to understand the recent progress in agro-ecosystem services research. Ecological Economics, 156, 293–305. https://doi.org/10.1016/j.ecolecen.2018.09.001

Lou, J., Tian, S. J., Niu, S. M., Kang, X. Q., Liu, H. X., Zhang, L. X., & Zhang, J. J. (2020). Coronavirus disease 2019: A bibliometric analysis and review. European Review for Medical and Pharmacological Sciences, 24(6), 3411–3421. https://doi.org/10.26555/eurrev_202003_2071

Makov, U. E. (2001). Mixture models in statistics. In N. J. Smelser & P.B. Baltes (Eds.), International Encyclopedia of the social & behavioral sciences (pp. 9910–9915). Elsevier. https://doi.org/10.1016/%0200-843076-7/00464-2

Mao, G. Z., Zou, H. Y., Chen, G. Y., Du, H. B., & Zuo, J. (2015). Past, present and future of biomass energy research: A bibliometric analysis. Renewable and Sustainable Energy Reviews, 52, 1823–1833. https://doi.org/10.1016/j.rser.2015.07.141

Marin, J. M., Mengersen, K., & Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. In D. Dey & C. Rao (Eds.), Handbook of statistics (Vol. 25, pp. 459–507). Elsevier. https://doi.org/10.1016/S0169-7161(05)25016-2

Masy, K. N. (2013). Latent class analysis and finite mixture modeling. In T. Little (Ed.), Oxford handbook of quantitative methods (pp. 551–611). Oxford University Press.

McLachlan, G. J., & Basford, K. E. (1988). Mixture models: Inference and applications to clustering. Marcel Dekker.

McLachlan, G. J., & Peel, D. (2000). Finite mixture models. Wiley.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and google scholar. Journal of the American Society for Information Science and Technology, 58(3), 2105–2125. https://doi.org/10.1002/asi.20677

Melnikov, V., & Zhu, X. (2020). Coronavirus disease 2019: A systematic review and meta-analysis to understand the recent progress in agro-ecosystem services research. Ecological Economics, 156, 293–305. https://doi.org/10.1016/j.ecolecen.2018.09.001

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Annals of Internal Medicine, 151(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Montalvan-Burbano, N., Perez-Valls, M., & Plaza-Ubeda, J. (2020). Analysis of scientific production on organizational innovation.
Moral-Munoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *El profesional de la información, 29*(1), e290103. https://doi.org/10.3145/epi.2020.ene.03

Muthen, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463–469. https://doi.org/10.1111/j.0006-341X.1999.00463.x

Newcomb, S. (1886). A generalized theory of the combination of observations so as to obtain the best result. *American Journal of Mathematics, 8*(4), 343–366.

Pearson, K. (1894). Contributions to the theory of mathematics evolution. *Philosophical Transactions of the Royal Society of London, A185*, 71–110. https://doi.org/10.1098/rsta.1894.0003

Phoong, S. Y., Ismail, M. T., Phoong, S. W., & Rahman, R. A. (2016). Finite mixture model: A comparison method for nonlinear time series data. *International Journal of Computing Science and Mathematics, 7*(4), 381–393. https://doi.org/10.1504/IJCSM.2016.078684

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation, 25*(4), 348–349.

Pyne, S., Hu, X. L., Wang, K., Rossin, E., Lin, T. I., Maier, L. M., Baecher-Allan, C., McLachlan, G. J., Tamayo, P., Hafler, D. A., De Jager, P. L., & Mesirov, J. P. (2009). Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences of the United States of America, 106*(21), 8519–8524. https://doi.org/10.1073/pnas.0912012106

Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management Research News, 27*(6), 31–39. https://doi.org/10.1108/01409170410784185

Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology, 87*(4), 835–841.

Shah, S. H. H., Lei, S., Ali, M., Doronin, D., & Hussain, S. T. (2019). Presumptions: Bibliometric analysis using HistCite and VOSviewer. *Kybernetes, 49*(3), 1020–1045. https://doi.org/10.1108/K-12-2018-0696

Simola, U., Cisewski-Kehe, J., & Wolpert, R. L. (2021). Approximate Bayesian computation for finite mixture models. *Journal of Statistical Computation and Simulation, 91*, 1155–1174. https://doi.org/10.1080/00949655.2020.1843169

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence informed management knowledge by means of systematic review. *British Journal of Management, 14*, 207–222. https://doi.org/10.1111/1467-8551.00375

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSViewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 285–320). Springer.

Van Leeuwen, T., Costas, R., Calero-Medina, C., & Visser, M. (2013). The role of editorial material in bibliometric research performance assessments. *Scientometrics, 95*(2), 817–828. https://doi.org/10.1007/s11192-012-0904-5

Waltman, L., Van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics, 4*(4), 629–635. https://doi.org/10.1016/j.joi.2010.07.002

Xu, S., Zhang, X. T., Feng, L. P., & Yang, W. T. (2020). Disruption risks in supply chain management: A literature review based on bibliometric analysis. *International Journal of Production Research, 58*(11), 3508–3526. https://doi.org/10.1080/00207543.2020.1717011

Zhang, K., & Liang, Q. M. (2020). Recent progress of cooperation on climate mitigation: A bibliometric analysis. *Journal of Cleaner Production, 277*, 123495. https://doi.org/10.1016/j.jclepro.2020.123495

Zhou, Y., Shi, Z., Shi, Z., Gao, Q., & Wu, L. (2019). Disaggregating power consumption of commercial buildings based on the finite mixture model. *Applied Energy, 243*, 35–46. https://doi.org/10.1016/j.apenergy.2019.03.014

Zhu, X., & Melnykov, V. (2018). Manly transformation in finite mixture modelling. *Computational Statistics & Data Analysis, 121*, 190–208. https://doi.org/10.1016/j.csda.2016.01.015