Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SYSTEMATIC LITERATURE REVIEW ON ENHANCING RECOMMENDATION SYSTEM BY ELIMINATING DATA SPARSITY

DAPHNE BUNGA DWIPUTRIANE[1], ZURAIDA ABAL ABAS[2], NANNA SURYANA HERMAN[3]

[1]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malacca, Malaysia

[2] Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malacca, Malaysia

[3] Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia

E-mail: [1]daphnebunga1727@gmail.com, [2]zuraidaa@utem.edu.my, [3]nsuryana.herman@mmu.edu.my

ABSTRACT

The aim of this project is to develop an approach using machine learning and matrix factorization to improve recommendation system. Nowadays, recommendation system has become an important part of our lives. It has helped us to make our decision-making process easier and faster as it could recommend us products that are similar with our taste. These systems can be seen everywhere such as online shopping or browsing through film catalogues. Unfortunately, the system still has its weakness where it faced difficulty in recommending products if there are insufficient reviews left by the users on products. It is difficult for the system to recommend said products because it is difficult to pinpoint what kind of users would be interested in the products. Research studies have used matrix factorization as the standard to solve this issue but lately, machine learning has come up as a good alternative to solve data sparsity. This project compares results of the recommendation system using RMSE to see how each proposed methods performs using three different datasets from MovieLens. We have selected two models – matrix factorization with SVD and deep learning-based model to evaluate these approaches and understand why they are popular solution to data sparsity. We have found that SVD brought in a lower RMSE as compared to deep learning. The reason behind this was discussed in the latter chapter of this thesis. We have also found possible research in capitalising categorical variables in recommendation system and the experiment achieved a lower RMSE score as compared to SVD and deep learning, showing the many possibilities of the future directions of the research in recommendation system.

Keywords: *Recommendation System, SVD, Matrix Factorization, Deep Learning, Data Sparsity*

1. INTRODUCTION

E-commerce has been a booming industry ever since the age of Internet has transcended. It has become one of the most important services in the world with millions of people using them every day. E-commerce service provides a platform for internet users to browse through a wide variety of choices such as electronic devices, clothes, artist's merchandise to groceries. Based on Statista, a widely known website on statistics, there were 1.8 billion global online buyers in 2018. A little reminder that there are approximately 7.7 billion people in the world. This means the amount of global online buyers' accounts for 23.4% of the global population. In 2021, it is expected that these numbers will rise to over 2.14 billion. This indicates a huge number of potential buyers in the near future. E-commerce has also proven itself to be a reliable supporter in global economic growth. In 2018, global retail e-commerce sales reach $2.84 Billion USD and the number is set to rise to $4.87 Billion USD by 2021.

The growth of e-commerce sales worldwide is undoubtedly influenced by the rising number of internet users, fuelled by the expansion of internet access across the world. This also indicates that more people are willing to shop online rather than personally attending the physical stores, which is more convenient and less time consuming for people nowadays. Hence, businesses should consider

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

readjusting their business priorities and start making plans to shift their business model to a model that puts online shopping as their priority.

Hence, we know that the growth will be a combination of few factors – one that includes old users preferring to use e-commerce, which lead us to a new set of question: 'What is so good in an e-commerce to the point people are using them repeatedly?'. The answer might lie on the fact that e-commerce companies used a technology called 'recommendation system' that becomes a fundamental backbone of their structure. Recommendation system is a rather compulsory item to have in a modern e-commerce platform in order for businesses to achieve the furthest audience as well as their targeted sales.

Recommendation system is a type of information filtering system that filters the available information and later generates them to the audience (user) based on their preferences, interest or in the case of e-commerce, their past transaction. The generated information can also be based on the user's behaviour whenever they are on any website. The system has the ability in predicting items that a certain user would like based on their profile.

There is no better time to discuss on recommendation systems due to how prevalent it is in our lives, particularly on our online journeys. A lot of companies have applied this system in their applications such as Netflix [1] , Youtube [2], Twitter [3], and even Google [4]. E-commerce is one of the businesses where we can see recommendation systems as it is used to generate sales via their website. According to Schafer, Konstan, and Riedl [5], businesses who use recommendation systems scored 60 percent higher growth compared to traditional businesses. The system can also improve customer satisfaction as it could cut down their decision-making process by recommending products related to their taste. Maintaining a high customer satisfaction helps businesses to continuously generate sales, which helps them in the long run.

There have been many discussions on defining the term 'recommendation system'. Recommendation system defined as a decision-making process for users under complicated environment[6]. From the perspective of e-commerce, recommendation system is defined as a tool that can help users to go through records that is related to their interest. According to Sinha, recommendation system is commonly adopted in an e-commerce environment [7] because it was developed to improve the sales of certain product, allowing the platform to achieve their marketing target of the said item. It is frequently defined as a way to help users to make a choice before purchasing an item that they have no sufficient knowledge on item or even having the experience of buying similar items. For Xiao and Benbasat, the authors agreed that a recommendation system is part of information systems[8]. A recommendation system is part of a decision-making step that are implemented to help people in deciding which items they should purchase that are similar with their preferences.

Businesses use a few strategies in their recommendation systems to maximise the potential of the system. There are a four models that have been used by businesses to generate recommendations [9], [10] such as:

- **Collaborative Filtering**: This method groups people based on how similar their taste is.
- **Content-Based Filtering**: This method assumes that people will buy products that are similar with the products that they have bought.
- **Demographic Filtering**: This method uses data related to the demographic of the user such as age and gender to recommend potential items for the users.
- **Knowledge-Based Filtering**: This method capitalises on the knowledge that the system has on their users and recommend items that suits the existing knowledge of the user.

The most common method in many companies is collaborative filtering. Collaborative filtering recommends items based on the behaviors of the user in the past. For example, if person A and person B bought similar items in the past, it is very likely that person A would buy items that person B has also bought rather than buying an item that a random person has bought. Koren, Bell, and Volinsky, deemed that this method is found to be more accurate than content-based[11], the second most common method. This is because content-based filtering unable to make accurate recommendation if not enough information is provided to differentiate items, which means content-based filtering requires intensive domain knowledge.

While collaborative filtering does show its potential in aiding the development of the business world, there are still a few challenges that needs to be solved before the system could reach its full

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

potential use. Collaborative Filtering, the most popular recommendation model suffers from a “cold start” problem, where models would face problems due to the sparsity of the data. Data sparsity describes the difficulty in having enough data. Cold-start, meanwhile, refer to the difficulty for the models in obtaining accurate recommendations. In this project, we would be focusing on the former which is data sparsity.

Data sparsity is resulted from users only leaving a few ratings to items in the catalogue. There is a lack of incentives for users to rate items. Hence, the recommendation system faced a difficulty in recommending accurate items to users that does not provide any feedback or ratings towards an item. Few studies have mentioned how data sparsity has a negative impact in the system. Hanafi, Suryana and Sammad agreed with the claim that cold-start is an issue and also added that the accuracy of the final results could be heavily impacted from the amount of information that the model has stored [10]. They also believe that no recommendation could also be produced if the sparsity of data is high. Guo, meanwhile, mentioned that the insufficient ratings in the collection will prevent the model from producing accurate observations on the users’ preference and this also slows down their ability in searching for users with the similar taste [12].

This research has several purposes on why it is being conducted. The purposes are described as follows:

- To study the impact of data sparsity in recommendation system.
- To evaluate the popular techniques implemented in recommendation system

Recommendation system posed an incredibly exciting venture for many businesses. Hence, any kind of weaknesses that the system posed must be solved quickly to ensure high quality implementation and could be easily used by e-commerce platforms around the world. Data sparsity, in general, has been one of the main roadblocks for recommendation system to achieve their full potential. The lack of information available has caused inaccurate predictions for the system recommending items to their users.

To solve this issue, many algorithms has been proposed by researchers, showing the flexibility of the recommendation system when it comes to adapting itself with new algorithms. E-commerce businesses need to find the suitable algorithms that could enhance the performance of their recommendation system and subsequently, improve the chances of the system to accurately recommend the correct items to the users.

This research claims to embark several contributions which are described as follows:

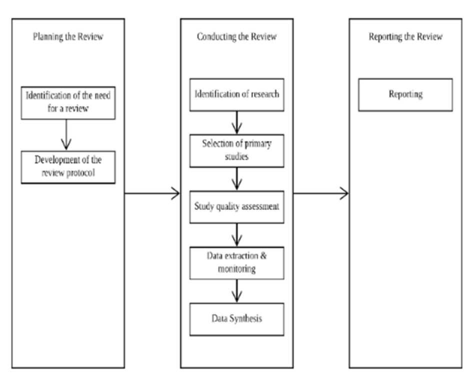
- To conduct a comprehensive literature review on the impact of data sparsity in a recommendation system, investigating how it could reduce the accuracy of recommended items to the users.
- To evaluate how popular approaches used to reduce data sparsity in a recommendation system are able to minimize the effect of sparsity, investigating its key features on why it is frequently used by other research studies.

This paper includes also on the discussion for the trends of research in data sparsity, particularly on its effect on recommendation system and the type of approaches that are proposed in research paper over the past few years. However, it will not include our proposed solution to this issue and rather just the approaches that have been discussed in previous research studies.

2. SYSTEMATIC LITERATURE REVIEW

2.1 Overview

The Systematic Literature Review (SLR) method was chosen as the method to perform literature review as it is found to have the capability of performing a fair and non-bias evaluation on a particular topic due to its predefined and well-executed strategy when conducting a research [13]. By using the SLR method in conducting the literature review, the method is expected to produce answers for the project questions that have been explained in the previous chapter.



Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 1: Stages of Systematic Literature Review

The stages underway in a typical SLR process is shown in Figure 1. SLR consists of three main stages: Planning the Review stage, conducting the Review stage, and Reporting the Review stage. These stages served as a guideline in performing a systematic review. In the first stage of SLR, where the planning process is evaluated, project questions and project objectives are curated in order to determine a general conclusion based on the studies that have been collected.

The second stage of SLR, where reviews are conducted, has been deemed as the most critical part of the literature review. This particular stage is divided into five sub-stages, in which procedures to conduct the studies are decided. All review protocol is developed in this stage to allow an accumulation of high quality and highly relevant results. The final stage of SLR is performed to evaluate the findings based on the research studies that has been collected.

2.2 Research Question

In the Planning the Review stage, the reason on the need for a literature review is formulated as there is a need to summarize available information related to the project, allowing a preliminary conclusion to be written out. To keep the project on the right track, Research Questions (RQ) are formulated. However, the RQ for the literature review is different with the RQ for this final year project. The RQ for the literature review is formulated based on the needs and requirements of the project. The list for the RQ is shown as follows:

Table 1: Research Questions for SLR

RQ	Questions	Motivation
RQ1	What is the trend of research related to recommendation system from 2016 until 2021?	To identify the research trends related to data sparsity in recommendation system.
RQ2	What is the most prevalent issue in recommendation system?	To identify the most serious issue in recommendation system that hampers them from achieving full potential.
RQ3	What approach are often used to solve the issue in recommendation system?	To identify possible approaches used to solve the issues in recommendation system.

Based on Table 2, there are three RQ that have been defined to help the development of the project in the right track. RQ1 is formulated to evaluate research trends related to recommendation system. Meanwhile, RQ2 and RQ3 were defined to narrow down the scope of the project, allowing a better focus on topics that are closely related with the objective of the project. The Research Questions can be illustrated in a mind map diagram. The mind map is shown as follows:

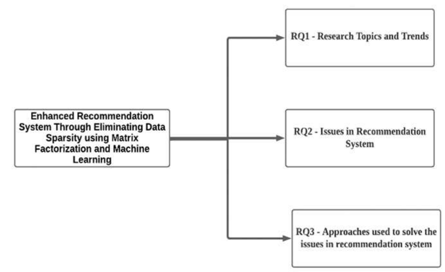


Figure 2: Mind Map for Research Questions

2.3 Search Identification

Search identification is the first sub-stage of the second SLR stage, namely the 'Conducting the Review' stage. This particular stage aims to collect as many relevant research studies as possible. By using the SLR process, research studies collected is expected to be fair and unbiased to the RQ. Generating a good search strategy is the key in finding research studies relevant to the topics of the project.

A list of potentially relevant studies can be compiled once the search strategy is applied in relevant research databases. Before applying the right search strategy, a list of research database is produced. Relevant studies related to this project can be found from these following databases. The list of database includes:

1. ACM Digital Library (http://dl.acm.org)
2. IEEEXplore (http://ieeexplore.ieee.org)
3. ScienceDirect (http://www.sciencedirect.com)
4. Google Scholar (https://scholar.google.com)

After the list of databases is determined, the next step in the search strategy is to come up with the accurate search string in order to find relevant studies to the projects. Based on RQ1-RQ3, the search string is extracted using the relevant words.

The Table 2as below showed the search string that was used to find relevant research studies.

Table 2: Search String Strategy

Search String		
Topic	**Activities**	**Methods**
Recommendation System	Identification	Matrix Factorization
Data Sparsity		Machine Learning

According to Table 2, the search string is formulated as follows: ("Recommendation System" OR "Data Sparsity") AND ("Identification") AND ("Matrix Factorization" OR "Machine Learning").

2.4 Selection of Research Studies

After generating a search strategy, the inclusion and exclusion criteria must be decided. It has to be formulated based on the RQ. This sub-stage aims to compliment the search string strategy by filtering out irrelevant studies in order to collect more accurate findings. There are five steps that was performed in order to find relevant studies related to this project:

1. If ('the search result is a general article') then discard, else go to step 2
2. If ('multiple publication of the same study') then discard, else go to step 3
3. If ('written in English') then go to step 4, else discard
4. If ('written as a thesis or dissertation') then discard
5. If ('full pdf is unavailable') then discard

The overall process starts by selecting one of the digital libraries listed, followed by applying the search string formulated as well as other appropriate strings into the database and then narrowing down the range of publication years. If these steps failed to show relevant research studies, the search string is then redefined. Figure 3. shows the flowchart of the complete search strategy.

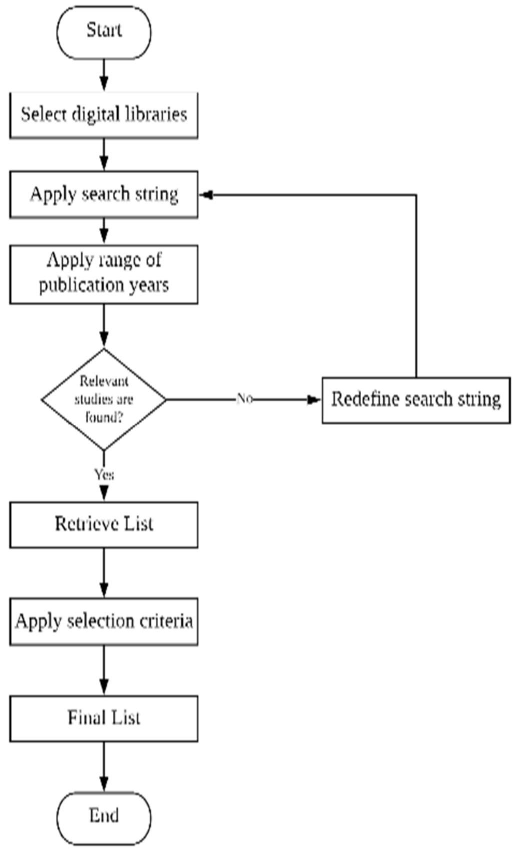


Figure 3: Flowchart for Search Strategy

2.5 Study Quality Assessment

In order to collect high quality data extraction results, an assessment must be performed regarding the quality of the studies that have been compiled. The studies that appeared on the final list need to have a certain degree of relevance and contained factual information that is essential for this project. A quality assessment Table 3 has been summarised from a research studies done by Barbara Kitchenham for software engineering domain [13]. The table has been readjusted in order to be relevant to the project and the list of quality assessment.

Table 3: Quality Assessment Strategy

Number	Quality Assessment
1	Did the aim of the research clearly stated?
2	Did the research studies fit with the project objectives?
3	Are the research relevant to the project questions?
4	Does the research studies have a clear process from introduction until conclusion?
5	Did the research studies mention the relationship of recommendation system and data sparsity?
6	Did the research studies describe any possible future works?

2.6 Data Extraction and Monitoring Process

Data Extraction and Monitoring Progress is the fourth sub-stage in the second stage of the SLR method. This stage is conducted to ensure the collection of information is accurate and relevant to the project. The data that were extracted from the research studies were the title of the publication, name of the authors, year of publication, problem statement, methodology, goals and proposed future works.

The mapping of Data Extraction is shown in Table 4. Data Extraction is collected in three parts. Research trends answer RQ1. Identification answer Research Questions in RQ2 while Matrix Factorization and Machine Learning will answer RQ3.

Table 4: Mapping Data Extraction to Research Questions

Data Extraction Property	Project Questions
Research Trends	RQ1
Identification	RQ2
Matrix Factorization and Machine Learning	RQ3

2.7 Data Synthesis

In the final sub-stage of the 'Conducting the Review' stage, a data synthesis method was performed to summarise the findings. There were 30 studies found between the four digital libraries that are relevant with the project. 9 of the research studies were found on ScienceDirect, 5 research studies were found on ACM Digital Library, 3 research studies were found on Google Scholar and the remaining 13 research studies were collected from IEEE Xplore.

2.8 Publication Frequency

There were 30 research studies found after the execution of the SLR process, targeted publications published from 2016 to 2021. Table 5 describes the classification of research studies from the selected digital libraries.

Table 5: Publication Frequency on Recommendation System

No	Search Database	Publications	Total
1	ACM Digital Library	[14], [15], [16], [17], [18]	5
2	Google Scholar	[9], [19], [10]	3
3	IEEE Xplore	[20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32]	13
4	ScienceDirect	[33], [34], [35], [36], [37], [38], [39], [40], [41]	9

2.9 Research Trends

To find the research trends on Recommendation System, the research studies were grouped in two categories: the issues in recommendation system and the research trends on recommendation system. The data extraction from the research studies are mapped based on the year of publication in Table 6 below.

Table 6: Research Trends on Recommendation System

Publication Year	Research Studies	Total
2016	[20], [17]	2
2017	[15], [16], [9], [21], [22], [33]	6
2018	[19], [10], [23], [24], [25], [34]	6
2019	[14], [18], [26], [27], [28], [29], [30], [31], [35], [36]	10
2020	[37], [38]	2
2021	[32], [39], [40], [41]	4

The Table 6 shows the increasing interests in research for recommendation system between the year of 2016 and 2021. There was a slight decrease from 2019 to 2020 but the number of publications has increased steadily, showing sign of researchers gaining an interest in researching on recommendation system. The decrease in terms of publications from 2019 until 2020 might be due to the filtering results or inability to find articles that are suitable to be studied for this research.

The distribution of research domain, meanwhile, can be seen in Figure 4. The project domain is separated into three focus areas: identification component, matrix factorization method, and machine learning method. For the identification component, we would like to find out what are the popular issues in recommendation system.

Journal of Theoretical and Applied Information Technology
15th April 2022. Vol.100. No 7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

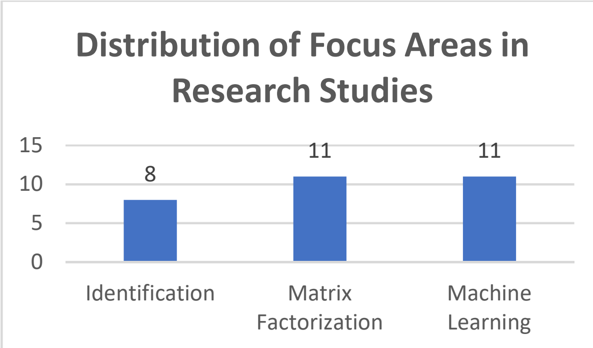


Figure 4: Distribution of Focus Areas in Research Studies

3. IDENTIFICATION

3.1 Overview

In recommendation system, there has been several issues that has been deemed as a hinderance to the development of the system. These issues have stopped recommendation system from achieving its full potential. Based on the SLR findings and Table 7 below, there are 8 research studies that specifically discussed on the issues in recommendation system. These researchers mostly discussed on data sparsity and cold start problem as the biggest issues in recommendation system nowadays.

Table 7: Mapping Data Extraction to Identification

Identification and Analysis	Research Studies	Total
Data Sparsity	[33], [9], [10], [19], [24], [37], [38], [39]	8
Cold Start	[9], [10], [37]	3

Several research studies put emphasis on both data sparsity and cold start in order to have a result that could solve both issues in the same approach. Çano and Morisio placed attention to not only data sparsity but also cold start and subsequently, to a certain extent, the accuracy of recommendation system resulting from not solving both issues mentioned [9].

Data sparsity refers to the fact that users would only leave a few ratings over a large catalogue of items [24]. The catalogue, in general, is referring to the dataset, and the items in the catalogue tends to not be rated frequently enough, which eventually caused the recommendation system facing difficulties in predicting items that accurately recommends products that fits their taste [9], [24]. Hwangbo and Kim also noted that these issues persisted in both user-based collaborative filtering and item-based collaborative filtering, which made data sparsity as an unavoidable issue in improving recommendation system [33]. As data sparsity has a huge influence in the eventual recommended items, the quality of a recommendation system highly depends on how researchers are able to minimise data sparsity in their dataset [37].

Hanafi, Suryana, and Basari, meanwhile, mentioned both problems but placed an emphasis on cold start in order to improve the functionality of a recommendation system [10]. Cold start refers to the difficulty for recommendation system to recommend items to new users, as opposed to data sparsity, which focus on how the lack of reviews or ratings by users stops the system from recommending a more accurate item. Lastly, Natarajan found both data sparsity and cold start is an equally serious issue and tried to solve both issues with the same method [37].

3.2 Eliminating Data Sparsity in Collaborative Filtering

The fastest answer one could give to the "what are the impact of sparse dataset?" question would be inaccurate recommendation. No item could be recommended accurately if the number of ratings left by users on the products are less than four percent of the total number of users and items. Collaborative filtering relies on ratings of items from the users. Unfortunately, users tend to not leave any recommendation, reviews or even ratings after purchasing an item, making it complicated for the system to recommend other items that would be interesting to the user. To measure sparsity level, the following formula can be used:

Figure 5: Sparsity Level Measurement

$$1 - \left(\frac{number\ of\ ratings}{number\ of\ users * number\ of\ items}\right)$$

According to Lee, Sun and Guy, in order to conduct a sparsity-related experiment, the density of the dataset should be in the range of 1% until 5% [42]. This low density has led to difficulty in improving the accuracy of recommendation. A low accuracy will lead to inaccurate recommendation, which may lead to users shifting to another platform to find items similar with their preference using the help of a more accurate recommendations.

An effective and accurate collaborative filtering is the key to have a successful e-commerce business. The challenge to achieve those goal would be eliminating, or minimizing, the sparse rating matrix, contributed by the sparsity of the dataset. The main concept of collaborative filtering method would be showcasing the relationship between users and items that have been rated by the users as ratings would show a strong correlation between the users

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and items. Previously, collaborative filtering has used the memory-based approach in recommending items, but many researchers have developed a better solution by using model-based approach, which has been found to alleviate scalability issue and high computational cost in memory-based approach.

4. MATRIX FACTORIZATION

4.1 Overview

As shown in Table 8, research studies on matrix factorization method were extracted and mapped. In order to come up with potential solutions to the issues raised in the identification section, researchers have mostly chosen matrix factorization as their preferred solution. Most researchers have opted to place their attention to one type of matrix factorization method as their approach in solving data sparsity.

Table 8: Mapping Data Extraction to Matrix Factorization method

Matrix Factorization Method	Description	Research Studies	Total
General Matrix Factorization (General MF)	Discussion on the general structure of matrix factorization, focusing on the basic component before touching on more complex structure.	[21]	1
Deep Matrix Factorization (Deep MF)	This architecture is inspired by deep learning, taking in the concept and applying them in matrix factorization.	[16], [30], [36]	3
Distributed Matrix Factorization (Distributed MF)	Designed based on Apache Spark in order to perform matrix factorization	[35]	1
Matrix Factorization with Implicit Feedback (MF with Implicit Feedback)	In additional to the traditional matrix factorization approach, this method put implicit feedback from users into the consideration.	[29], [31], [40]	3
Matrix Factorization with Rating Completion (MF with Rating Completion)	This is a new matrix factorization model that incorporates ratings completion inspired by active learning.	[22]	1
Preliminary Data-Based Matrix Factorization (Preliminary Data-Based MF)	This method generates preliminary data from the algorithm's closest neighbour and later factorize both the preliminary and original matrix.	[41]	1
Non-Negative Matrix Factorization (Non-Negative MF)	This approach tries to predict with whatever they have and proceed to predict further items based on the extracted latent features.	[27]	1

Matrix factorization is a type of model that is commonly used for recommendation system to predict items for users. W. Zhang have used matrix factorization to discover latent factors from rating matrix and later map the items and users against those factors [36]. In a more technical explanation, it reduces a matrix into its smaller parts, simplifying matrix operations rather than a complex one. For example, given a M x N user-item rating, the matrix first learns about the latent representation of the users as well as items, projecting a low-dimensional approximation, then later generate ratings by predicting a score from the multiplication of the user and item latent vector [43]. This methodology has made matrix factorization as an integral component of a recommendation system over the years, even with the rising popularity of other technologies.

This step has highlighted the fact that deep matrix factorization and matrix factorization with implicit feedback are the most popular ones among the research studies that we have collected. Deep matrix factorization, for example, may have gained their popularity due to the advancement of deep learning, the architecture that it was mainly inspired from, and the ability for many researchers to support the factorization process with adequate tools. Deep matrix factorization utilize deep-learning architecture [30]. This model captures latent features of the users and items and later generate ratings for other items as part of this approach [36].

After learning the latent features, the system later project the results after going through multiple layers. In addition to the deep learning concept, this model has also been said as being partly

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

inspired by deep structured semantic models, primarily used for web search [16].

4.2 Requirement for Matrix Factorization with SVD

SVD was briefly mentioned in the previous section. It was deemed as a popular or benchmark solution in enhancing the performance of collaborative filtering and subsequently, the recommendation system, in recommending accurate item to users[44], [45]. SVD has the ability of decomposing a matrix into a smaller estimation of the original matrix. In other words, it decomposes a matrix into two unitary matrices and one diagonal matrix, as shown in the figure below:

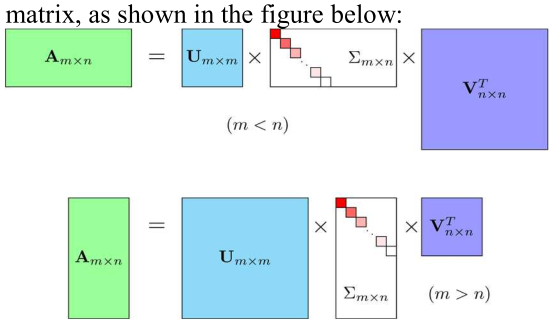


Figure 6: Structure of a SVD Matrix

A act as the input matrix (user's ratings), U is the left singular vectors (user "features" matrix), ∑ is the diagonal matrix of singular values and V^T would be the right singular vectors (movie "features" matrix). U and V^T represent different items. In order to receive the approximation of the lower ranks,this computation would be implemented and only the top k features would be retained afterwards. The top K features specifically refers to the user's taste and preferences vectors.

The SVD function is then added to ensure the calculation would include the number of latent factors wanted by the analyst in order to estimate the original matrix. As the next step would usually require matrix multiplication to obtain predictions, the ∑ values are converted into diagonal matrix form.

5. MACHINE LEARNING

5.1 Overview

Machine learning is a subset of artificial intelligence and refers to computer systems that learns automatically from the data with minimal help from humans. It has been implemented in many industry as more people have access to adequate resources to run machine learning algorithms [32]. This method of analysis allows computer to draw observations based on the pattern of the data that they have collected [34]. This allows the recommendation system to predict items even with the presence of a high amount of empty data as the machine will learn underlying patterns from existing data. Hence, implementing machine learning in a recommendation system is a good option as it allows the system to draw inferences based on the ratings that has previously existed. Machine learning is also known for greatly improving accuracy of many applications and this includes recommendation system, as proven by several experiments done by Alfarhood and Cheng [26].

Portugal, Alencar and Cowan noted that there are increasing interests among researchers to implement machine learning algorithms in recommendation system using supervised learning, unsupervised learning as well as deep learning [34]. As the field of machine learning continues to develop, it is worth to look at machine learning algorithms to see whether the performance of a recommendation system can be improved when the algorithms are being implemented in the system.

Table 9 below shows how research studies regarding machine learning algorithms in recommendation system are mapped after data extraction and a SLR process. As shown in Table 9 below, various mechanism has been proposed to allow researchers to focus on different components of the architecture.

Table 9: Mapping Data Extraction to Machine Learning Method

Machine Learning Method	Description	Research Studies	Total
General Machine Learning (General MF)	The studies mentioned here discussed on the general applications of machine learning in recommendation system	[34], [25]	2
Unsupervised Learning - K-Means	Clustering can be used to find distances between two items in a hope to predict things that are similar between them.	[28], [32]	2
Deep Learning	This approach applies the concept of multi-layers neural networks that allows feature extraction, in which recommendation system will highlight distinct patterns	[20], [17], [15], [23], [14], [26], [18], [32]	8

The mechanism shown in Table 9 are proposed mechanisms that has been suggested by the researchers. Based on the amount of research studies collected, deep learning seemed to be the favourite among researchers to solve data sparsity in recommendation system. As mentioned by Alfarhood and Cheng, researchers are capitalising on their latest tools and software to test deep learning in many applications including recommendation system [26]. However, there are less works on researching the use of deep learning in recommendation system as compared to matrix factorization, making this field relatively new. This made the prospect of researching on the implementation of machine learning, specifically deep learning, in recommendation system a higher and brighter prospect. While the field is relatively new, Ong, Haw, and Ng, found that deep learning approaches are able to show a great sign of improvement as compared to traditional approaches in recommendation system [14].

5.2 Requirement for Deep Learning Recommendation System

The popularity of a deep learning-based recommendation system has been steadily increasing over the past years[18]. Deep Learning has proven its effectiveness in adapting to various domains in the field of computer science and researchers are eager to implement a model inspired by the same concept in recommendation system, a field where mainly one approach called matrix factorization has been dominating for a long time[14], [18].

The concept of implementing deep learning in recommendation system is similar with model-based matrix factorization[23]. For matrix factorization, we would be decomposing the original sparse matrix into a product of low rank orthogonal matrices. Deep learning implementation, meanwhile, allows us to learn the values of the embedded matrix on its own[46]d. The embedding matrices would allow the user latent features and movie latent features to be looked up to for a specific combination between a movie and a user.

6. RECOMMENDATION RESEARCH SCOPE

6.1 Overview

Machine learning is a subset of artificial intelligence and refers to computer systems that learns automatically from the data with minimal help from humans. It has been implemented in

In order to be clear with the intention of our research, the scope of this research has been determined. Based on the scope we have set in the previous section, we will now be even more specific with our scope, aligning ourselves with a clearer scope after conducting literature review. The boundaries for this research have been determined by the following characteristics:

- The mapping of e-commerce recommendation system research area
- The mapping of research topic and sub-topics
- The mapping of contribution characteristics
- The mapping of dataset characteristics
- The mapping of methodology approach
- The scope of the qualitative evaluation method
- The scope of evaluation metrics and statistical approach

Based on the Figure 7 below, a chart has been designed in order to visualise the filtering of the research scope along with its design. The diagram represents the structure and organization of mapping and issues that belongs to the recommendation system research area, the research methodologies designed, the detail output estimated and the adjusted parameters.

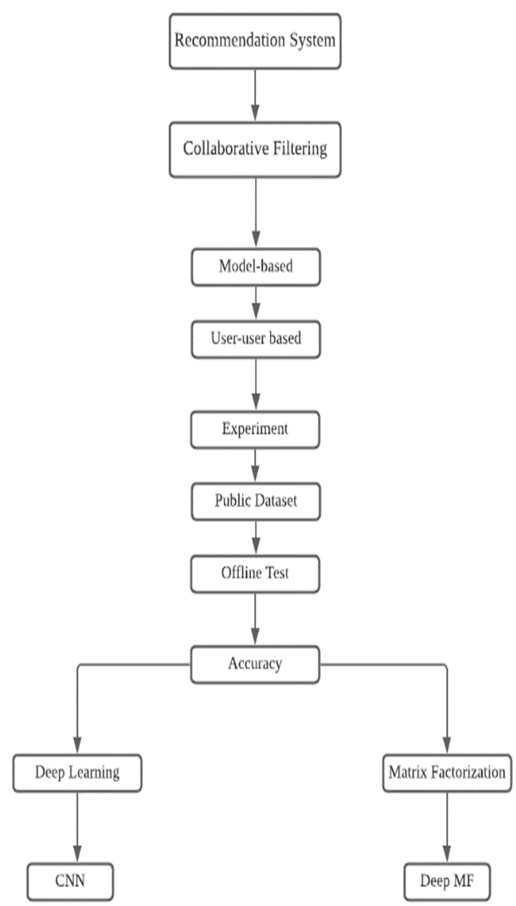


Figure 7: Chart of a Research Scope

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6.2 Research Issues and Analytics

The recommendation system has been a vital component to the business field, particularly the e-commerce industry. It has a critical role in the industry, ensuring that users are being recommended items that suits their past behaviour or in general, their personality.

The research objectives for studies on recommendation system mainly focuses on the following criteria [9]:

- Eliminate data sparsity for new items as well as new users.
- Solve data sparsity problems that are caused by minimum ratings on certain products.
- Improve the accuracy of products recommended to users.
- Able to be scalable as new information are expected to flood the system most of the time.

Over the years, researchers have made commendable attempts in solving data sparsity issues to solve the issues explained above. Those attempts have come in many forms. As researchers notice the information overload draining a lot of system in the past decade, they have tried to improve the capability of a recommendation system by testing them with high amount of information[35]. While they insert a lot of new information to the system, they have decided to capitalise on the rapid development of artificial intelligence to extract relevant information and test whether the recommendation system can produce good results despite being given a truckload more of information[18]s.

From the literature review, it becomes apparent that whenever it comes to information related to recommendation system, it consists of two sides – item side and user side. The item side owns information related to the abstract, product reviews, ratings, product overview and basically any information concerning to a product. The user side, meanwhile, belongs to the users that uses the systems, and it has information related to the demographics of the users such as age, gender, address, occupation, and marital status.

Due to the rising popularity of hybrid recommendation system, this research has proposed to implement this method to solve data sparsity issues. The hybrid recommendation system hypothetically is expected to increase the effectiveness of matrix factorization, the most popular solution to data sparsity. This research aims to adopt the concept of deep learning to improve the overall system.

Deep learning has quickly gained an influential role in the development of artificial intelligence and has quickly become an integral part of many systems due to its ability in handling a high number of data and effectively provide results at a fast rate, given a suitable support are provided to the system[47]. Deep learning approach has made important contributions into multiple fields of computer sciences such as speech recognition, computer vision, natural language processing as well as sentiment analysis.

Data sparsity issues mainly stems from the lack of information available for the system to process relevant results related to the user. This problem has led to inaccurate results of predicted recommendations, causing the accuracy level to be at a questionable level[36]. According to a study by Bobadilla et al., the solution to this issue is by adding more information from side information and affirmed that it is a promising research field in the near future[48]. The side information that Bobadilla et al. refers to are explicit feedback as well as implicit feedback[48].

The categories of feedback are described below:

- **Explicit –** The system solely relies on product rating and it is considered to be the simpler approach for collaborative filtering as it only accounts ratings in order to calculate product recommendation.
- **Implicit –** The system refers to the act of inferring a similar product that the user would likely give their ratings based on the observations of the user's behaviour on the platform. Certain products would be recommended based on the way the user behave and this does not involve any kind of ratings or specific actions.

Based on several research studies, it has been found that experiments involving secondary information such as implicit and explicit feedback have successfully increased the effectiveness of rating predictions. This breakthrough was definitely boosted by the increasing availability of side information shared by organisations, researchers and even communities in social media.

From literature studies shows that collaborative filtering mainly exploits explicit feedback in order to recommend products to the users accurately. While explicit feedback still remains the main option for collaborative filtering in order to build a recommendation system, classical recommendation system still remains a strict structure that only relies on product ratings by the users. The inability for the existing method to find their way to capitalise existing structures to the maximum in order to recommend items as accurately

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as possible has become a major roadblock in the development of recommendation system.

6.3 Implementation and Benchmark Method

Based on the findings in the previous section, several approaches have been discovered that could possibly improve the performance of a recommendation system despite having the presence of data sparsity by capitalising on auxiliary information gained from product reviews. The other popular approaches found also relates on the implementation of deep learning, specifically neural networks, in recommendation system. There are several models that has proven to be successful in implementing both concept and subsequently improve the performance of the recommendation system. In this research the word of model has been defined as a processing model and it can be understood as the simplification of the real-world process. Four research studies have been selected and placed in table below. The said table shows research studies that have either implement a model that relies on implicit feedback and deep learning in order to generate product recommendation. However, the implicit feedback in this current research was not treated extensively.

Table 10: State-of-the-Art Collaborative Filtering Method

Number	Research Title	Description	References
1	Recommendation algorithm based on Explicit and Implicit feedback Matrix factorization	An approach that combines both explicit models and implicit data by assigning weights to the two models.	[29]
2	Collaborative Filtering and Deep Learning Based Hybrid Recommendation For Cold Start Problem	A hybrid recommendation model is suggested where it combines SVD++ model with a deep learning architecture that aims to address the data sparsity problem.	[20]
3	Convolutional Matrix Factorization for Document Context-Aware Recommendation	A version of collaborative filtering that combines convolutional neural networks and matrix factorization which aims to improve product recommendation by understanding the context of product document.	[49]
4	Deep Matrix Factorization With Implicit Feedback Embedding for Recommendation System	An approach that proposes deep learning based collaborative filtering structure that aims to combine any kind of side information effectively and handily.	[31]

Collaborative filtering is a one of the many types of recommendation system and is considered to be widely popular among the e-commerce industry. It implements a latent factor model using matrix factorization in order to recommend products to the users. The reason why matrix factorization is one of the most popular approaches in collaborative filtering is due to its adaptability in integrating multiple information into one single system. However, the system is still vulnerable and has issues in achieving optimal accuracy due to data sparsity, caused mainly by the minimum number of ratings given by the users or the fact that there are insufficient ratings on the products. Researchers have tried to come up with solutions to tackle this problem and many of them found that in order to minimise the effect of data sparsity in a recommendation system, they suggested to use additional information to obtain a better performance in producing product recommendation ranking.

This approach aims to support latent factor by adding product document information as a new latent factor while another approach suggest that user information representation is integrated into a latent factor. According to several research studies, these approaches have successfully improved the performance of recommendation system based on evaluation metrices. Most of these research studies have used RMSE and MAE to evaluate the performance of the rating prediction[14].

However, there are still challenges in extracting additional information related to user and item information representation and later incorporating them into the recommendation system. For example, a model proposed by Wei (2016) considered to adopt SDAE so they could extract user information correctly [20] . Wang (2015), meanwhile, proposed a hybrid model that combines both auto encoder and matrix factorization [50]. Kim (2016) came up with an improved model based on the works by Wei (2016), in which the proposed model adopted convolutional neural networks and matrix factorization using the same concept of deep learning model [49]. Most of these research studies aims to solve data sparsity that are caused by minimum rating and their experiment has shown that their proposed model could successfully improve latent factor model that subsequently helped the recommendation system to achieve a better performance.

6.4 Improvement Efforts

There has been a number of methods being proposed in order to improve the accuracy of the predictions by a recommendation system. Due to the increasing popularity of implementing product opinion in the business field, many are expecting the research area for recommendation system will also rapidly grow to ensure the system structure could keep up with the dynamics of the industry and avoid a potential collapse due to the high amount of information exchange happening at once. As data sparsity remains an issue that plagues the recommendation system for a long time, matrix factorization has risen to be a popular solution proposed by researchers. A variation of matrix factorization has been proposed and the ones that certainly caught many interests is the incorporation of implicit feedback into the matrix factorization, making the model consider the user's behaviour on the platform to predict the right products to recommend. From a non-scientific point of view, this will make a lot of sense if we reflect on our own behaviour on e-commerce platforms or any platforms that recommends us certain items. Some of us rarely leave a rating after using a certain product but at one point, did try to use them[51].

For example, if you are on Netflix, there is high chance you probably did not finish one of their original series. This act should be put under the consideration when Netflix tries to recommend us a new series to watch. If a user watches a series until ¾ of the series ended, they should consider that there is a high chance that the user enjoys the series but if they quit the series barely halfway through, Netflix should know that the series might not be liked enough by the user to be bothered enough to finish the series. Hence, the factor of auxiliary information should be seriously considered before proposing an improvement of a recommendation system as not only it can solve data sparsity, it can also provide more insights that allows more products to be accurately recommended. Implementing a hybrid approach that involves deep learning also seemed to be an effective solution. These two solutions proposed, which are providing auxiliary information and hybrid approach, is still a rather new discussion in the field of recommendation system [9]. According to Zheng (2017), utilizing deep learning has proven to be able to significantly improved the recommendation system [52].

This research proposes a matrix factorization approach that considers implicit feedback to generate effective rating predictions. This research also proposes supplementary models such as a support vector decomposition (SVD) model, a well-known matrix factorization method, and a deep learning-inspired model, in order to compare which models, have the best performance well within the dataset given to give a clearer idea on the performance of each models given a situation of data sparsity. The major contribution of this research is providing a model comparison of three popular approaches in solving data sparsity, matrix factorization with implicit feedback, a SVD matrix factorization method and deep convolutional neural networks model. This proposed model uses the foundation of a collaborative filtering method to produce rating predictions in order to develop a recommendation system.

6.5 Implementation on real datasets and pre-processing

There are two kinds of datasets in the research field of recommendation system. The first kind of dataset would be private datasets. Private datasets stemmed from the permission only being held by dataset owner and others who would like access to them need to gain permission from the owner, making those datasets less accessible. Public dataset, meanwhile, is designed to be accessible by many other people and is the more popular kind of dataset in the research field. Public datasets are easily accessible through websites such as Kaggle, GitHub and GroupLens.

This research would implement dataset from MovieLens movie dataset. This decision was made after considering the hardware availability and how the dataset is easily understood due to the author's interest in researching more of movie recommendation system. It is also a widely popular dataset and has been tested for a number of applications so using dataset from MovieLens would mean that there are high chances that this method would be highly compatible in other domains as well.

In this research project, implementing an algorithm based on a MovieLens dataset means we would be seeing information related to the users, movies and ratings. While referring to the user latent factor point of view, this research would consider user information in terms of demographics information and movie ratings from users.

6.6 Critical Assessment

Based on these research studies, there are multiple suggestions to improve recommendation system though most of the discussion revolves around machine learning and matrix factorization. This is probably due to how matrix factorization has been frequently used in recommendation system since this system have been pushed to be applied in many real-life settings. Machine learning, meanwhile, have been capitalised due to its rising popularity in the field of computer science due to the many advantages that it could give to an application. It is widely believed that both these technologies could tremendously improve recommendation system, making the application to reach its full potential.

However, despite the advantages that recommendation system has given to many industries, there are still many gaps that needs to be addressed in research studies. For example, some studies have used deep learning but unfortunately the equipment required to run this technology is far more costly than the standard computers running. There is also concern that deep learning requires more storage to run their processes. There is a need to consider cost and storage restraints while conducting the experiment to ensure more parties could run similar tests with recommendation system. Next, certain research studies conducted their data sparsity research using music-based dataset. While there is nothing wrong with this, it tough to gauge the effectiveness of the research as music-based dataset might give different result when the same approaches are used for text content such as movies. This is because music-based recommendation system learned about the content features of items such as the spectrum of music. There is no standard result for approaches proposed in reducing data sparsity in recommendation system.

Lastly, the research in recommendation system is still growing rapidly over the years. While there is a lot of new information on recommendation system available, no existing standard is available to measure the performance of a recommendation system, particularly when a study is trying to solve data sparsity issue in recommendation system. A more comprehensive guide to data sparsity in recommendation system should be produced and standardized to help researchers in solving data sparsity.

7. CONCLUSION

Data sparsity refers to the insufficient data in the dataset while cold start refers to the difficulty of the system to produce prediction due to the lack of information about a user. As this research focus on data sparsity, we noticed that data sparsity caused inaccuracy in predictions due to the lack of rated items in the dataset. This is a huge problem as inaccurate recommendation might make users to change their preferred e-commerce as it is unable to find products that closely resemble their taste.

The two selected models are Matrix Factorization with SVD and Deep Learning-based model. Both models were evaluated using RMSE and MAE and lower scores for both indicates better accuracy. The models processed datasets that were deemed to be sparse so having a lower accuracy after the experiment indicates that the influence of data sparsity have been successfully minimized.

On the surface, we have found that our SVD performed better than the deep learning-based model. There were, however, other issues that arise during the experiment such as the issues of finding the optimal number of latent factors as well as the issues of overfitting. The problem of overfitting seemed to be a common problem among data scientists while conducting deep learning-related experiments. In order to have a better accuracy for recommendation system in the future, more work need to be done to solve overfitting.

There are a lot of many potentials that the recommendation system field could head in the future, especially with the number of companies starting to use recommendation system in their business. One of the most interesting avenues would be graph-learning based recommendation system. Graph learning is related to the machine learning field that works on graph structure data[53]. Recommendation system is in a position to benefit the most as graph learning allows the model to learn relational data[53]. Recommendation system is essentially a graph structure where there is a relationship between a user and the things that like. Graph learning has the potential in deriving knowledge embedded in different kind of graphs[53]. If this can be capitalized in recommendation system, the quality of the recommendations might just improve.

The limitation of current knowledge on recommendation system is how data sparsity could cause the lack of content diversity. If only popular products are reviewed, the other products on the shelves will not be viewed by consumers as it is not recommended or be left with ratings from other consumers. In other context, the lack of content diversity might cause issue when recommending news. Assuming people are storming and pushing a false news to trend, this might cause false news to be recommended due to how popular it is. It is important to look at the negative side of recommendation system in order to improve the quality of applications. Improving recommendation system by reducing data sparsity will potentially help companies to improve their customer satisfaction which will lead to a higher revenue, an important highlight of research in recommendation system.

REFERENCES:

[1] H. Steck, "Evaluation of recommendations," pp. 213–220, 2013, doi: 10.1145/2507157.2507160.

[2] J. Davidson, B. Liebald, J. Liu, P. Nandy, and T. Van Vleet, "The YouTube video recommendation system," *RecSys'10 - Proc. 4th ACM Conf. Recomm. Syst.*, pp. 293–296, 2010, doi: 10.1145/1864708.1864770.

[3] A. Ahmed, B. Kanagal, S. Pandey, V. Josifovski, L. G. Pueyo, and J. Yuan, "Latent factor models with additive and hierarchically-smoothed user preferences," *WSDM 2013 - Proc. 6th ACM Int. Conf. Web Search Data Min.*, pp. 385–394, 2013, doi: 10.1145/2433396.2433445.

[4] Y. Wang and W. Shang, "Personalized news recommendation based on consumers' click behavior," *2015 12th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2015*, pp. 634–638, 2016, doi: 10.1109/FSKD.2015.7382016.

[5] J. Ben Schafer, J. A. Konstan, and J. Riedl, "E-commerce recommendation applications," *Data Min. Knowl. Discov.*, vol. 5, no. 1–2, pp. 115–153, 2001, doi: 10.1007/978-1-4615-1627-9_6.

[6] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*. 2015, doi: 10.1016/j.eij.2015.06.005.

[7] K. Swearingen and R. Sinha, "Interaction design for recommender systems," *Des. Interact. Syst.*, 2002.

[8] X. Bo and I. Benbasat, "e-commerce product recommendation agents: Use, characteristics, and impact," *MIS Q. Manag. Inf. Syst.*, 2007, doi: 10.2307/25148784.

[9] E. Çano and M. Morisio, "Hybrid recommender systems: A systematic literature review," *Intell. Data Anal.*, vol. 21, no. 6, pp. 1487–1524, 2017, doi: 10.3233/IDA-163209.

[10] Hanafi, N. Suryana, and A. S. B. H. Basari, "An understanding and approach solution for cold start problem associated with recommender system: A literature review," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 9, pp. 2677–2695, 2018.

[11] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer (Long. Beach. Calif).*, 2009, doi: 10.1109/MC.2009.263.

[12] G. Guo, "Improving the performance of recommender systems by alleviating the data sparsity and cold start problems," *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 3217–3218, 2013.

[13] B. Kitchenham, "Procedures for performing systematic reviews," *Keele Univ. Natl. ICT Aust.*, 2004, doi: 10.1.1.122.3308.

[14] K. Ong, S. C. Haw, and K. W. Ng, "Deep Learning Based-Recommendation System: An Overview on Models, Datasets, Evaluation Metrics, and Future Trends," *ACM Int. Conf. Proceeding Ser.*, pp. 6–11, 2019, doi: 10.1145/3372422.3372444.

[15] A. Karatzoglou and B. Hidasi, "Deep learning for recommender systems," *RecSys 2017 - Proc. 11th ACM Conf. Recomm. Syst.*, pp. 396–397, 2017, doi: 10.1145/3109859.3109933.

[16] H. J. Xue, X. Y. Dai, J. Zhang, S. Huang, and J. Chen, "Deep matrix factorization models for recommender systems," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 0, pp. 3203–3209, 2017, doi: 10.24963/ijcai.2017/447.

[17] P. Covington, J. Adams, and E. Sargin, "Deep Neural Networks for Youtube Recommendations," pp. 191–198, 2016, [Online]. Available: http://arxiv.org/abs/1611.00591.

[18] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–35, 2019, doi: 10.1145/3285029.

[19] N. F. Al-Bakri and S. H. Hashim, "Reducing Data Sparsity in Recommender Systems," *J. Al-Nahrain Univ. Sci.*, vol. 21, no. 2, pp. 138–147, 2018, doi: 10.22401/jnus.21.2.20.

[20] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative Filtering and Deep Learning Based Hybrid Recommendation

for Cold Start Problem," *Proc. - 2016 IEEE 14th Int. Conf. Dependable, Auton. Secur. Comput. DASC 2016, 2016 IEEE 14th Int. Conf. Pervasive Intell. Comput. PICom 2016, 2016 IEEE 2nd Int. Conf. Big Data*, pp. 874–877, 2016, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.149.

[21] R. Mehta and K. Rana, "A review on matrix factorization techniques in recommender systems," *2017 2nd Int. Conf. Commun. Syst. Comput. IT Appl. CSCITA 2017 - Proc.*, pp. 269–274, 2017, doi: 10.1109/CSCITA.2017.8066567.

[22] X. Guan, C. T. Li, and Y. Guan, "Matrix Factorization with Rating Completion: An Enhanced SVD Model for Collaborative Filtering Recommender Systems," *IEEE Access*, vol. 5, pp. 27668–27678, 2017, doi: 10.1109/ACCESS.2017.2772226.

[23] L. Zhang, L. Tiejian, F. Zhang, and W. Yanjun, "A Recommendation Model Based on Deep Neural Network," vol. 6, 2018.

[24] J. F. G. Da Silva, N. N. De Moura, and L. P. Caloba, "Effects of Data Sparsity on Recommender Systems based on Collaborative Filtering," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2018-July, no. July, pp. 1–8, 2018, doi: 10.1109/IJCNN.2018.8489095.

[25] A. Nawrocka, A. Kot, and M. Nawrocki, "Application of machine learning in recommendation systems," 2018, doi: 10.1109/CarpathianCC.2018.8399650.

[26] M. Alfarhood and J. Cheng, "DeepHCF: A Deep Learning Based Hybrid Collaborative Filtering Approach for Recommendation Systems," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 89–96, 2019, doi: 10.1109/ICMLA.2018.00021.

[27] A. Slim, D. Hush, T. Ojha, C. Abdallah, G. Heileman, and G. El-Howayek, "An automated framework to recommend a suitable academic program, course and instructor," 2019, doi: 10.1109/BigDataService.2019.00026.

[28] R. Ahuja, A. Solanki, and A. Nayyar, "Movie recommender system using k-means clustering and k-nearest neighbor," *Proc. 9th Int. Conf. Cloud Comput. Data Sci. Eng. Conflu. 2019*, pp. 263–268, 2019, doi: 10.1109/CONFLUENCE.2019.8776969.

[29] X. Xiao, R. Yan, and D. Tan, "Recommendation algorithm based on explicit and implicit feedback matrix factorization," *Proc. - 2019 4th Int. Conf. Mech. Control Comput. Eng. ICMCCE 2019*, pp. 903–909, 2019, doi: 10.1109/ICMCCE48743.2019.00205.

[30] J. Zhou, J. Wen, S. Li, and W. Zhou, "From Content Text Encoding Perspective: A Hybrid Deep Matrix Factorization Approach for Recommender System," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, no. July, pp. 1–8, 2019, doi: 10.1109/IJCNN.2019.8852443.

[31] B. Yi *et al.*, "Deep Matrix Factorization With Implicit Feedback Embedding for Recommendation System," *IEEE Trans. Ind. Informatics*, vol. 15, no. 8, pp. 4591–4601, 2019, doi: 10.1109/tii.2019.2893714.

[32] A. Althbiti, R. Alshamrani, T. Alghamdi, S. Lee, and X. Ma, "Addressing Data Sparsity in Collaborative Filtering Based Recommender Systems Using Clustering and Artificial Neural Network," *2021 IEEE 11th Annu. Comput. Commun. Work. Conf. CCWC 2021*, pp. 218–227, 2021, doi: 10.1109/CCWC51732.2021.9376008.

[33] H. Hwangbo and Y. Kim, "An empirical study on the effect of data sparsity and data overlap on cross domain collaborative filtering performance," *Expert Syst. Appl.*, vol. 89, pp. 254–265, 2017, doi: 10.1016/j.eswa.2017.07.041.

[34] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Syst. Appl.*, vol. 97, pp. 205–227, 2018, doi: 10.1016/j.eswa.2017.12.020.

[35] B. Ait Hammou, A. Ait Lahcen, and S. Mouline, "An effective distributed predictive model with Matrix factorization and random forest for Big Data recommendation systems," *Expert Syst. Appl.*, vol. 137, pp. 253–265, 2019, doi: 10.1016/j.eswa.2019.06.046.

[36] W. Zhang, X. Zhang, H. Wang, and D. Chen, "A deep variational matrix factorization method for recommendation on large scale sparse dataset," *Neurocomputing*, vol. 334, pp. 206–218, 2019, doi: 10.1016/j.neucom.2019.01.028.

[37] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using Linked Open Data," *Expert Syst. Appl.*, vol. 149, 2020, doi:

10.1016/j.eswa.2020.113248.

[38] C. Feng, J. Liang, P. Song, and Z. Wang, "A fusion collaborative filtering method for sparse data in recommender systems," *Inf. Sci. (Ny).*, vol. 521, pp. 365–379, 2020, doi: 10.1016/j.ins.2020.02.052.

[39] N. Bhalse and R. Thakur, "Algorithm for movie recommendation system using collaborative filtering," *Mater. Today Proc.*, no. xxxx, pp. 1–6, 2021, doi: 10.1016/j.matpr.2021.01.235.

[40] F. Prathama, W. F. Senjaya, B. N. Yahya, and J. Z. Wu, "Personalized recommendation by matrix co-factorization with multiple implicit feedback on pairwise comparison," *Comput. Ind. Eng.*, vol. 152, no. December 2020, p. 107033, 2021, doi: 10.1016/j.cie.2020.107033.

[41] X. Yuan, L. Han, S. Qian, L. Zhu, J. Zhu, and H. Yan, "Preliminary data-based matrix factorization approach for recommendation," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102384, 2021, doi: 10.1016/j.ipm.2020.102384.

[42] L. Joonseok, S. Mingxuan, and L. Guy, "A Comparative Study of Collaborative Filtering Algorithms," *KDIR 2012 - Proc. Int. Conf. Knowl. Discov. Inf. Retr.*, pp. 132–137, 2012, doi: 10.5220/0004104001320137.

[43] T. Qian, Y. Liang, Q. Li, and H. Xiong, "Attribute Graph Neural Networks for Strict Cold Start Recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 8, 2020, doi: 10.1109/TKDE.2020.3038234.

[44] Z. Xian, Q. Li, G. Li, and L. Li, "New Collaborative Filtering Algorithms Based on SVD++ and Differential Privacy," *Math. Probl. Eng.*, 2017, doi: 10.1155/2017/1975719.

[45] M. Sunitha Reddy and T. Adilakshmi, "Music recommendation system based on matrix factorization technique-SVD," 2014, doi: 10.1109/ICCCI.2014.6921744.

[46] A. K. Sahoo, C. Pradhan, R. K. Barik, and H. Dubey, "DeepReco: Deep learning based health recommender system using collaborative filtering," *Computation*, 2019, doi: 10.3390/computation7020025.

[47] B. Ait Hammou and A. Ait Lahcen, "FRAIPA: A fast recommendation approach with improved prediction accuracy," *Expert Syst. Appl.*, 2017, doi: 10.1016/j.eswa.2017.06.001.

[48] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Syst.*, 2013, doi: 10.1016/j.knosys.2013.03.012.

[49] D. Kim, C. Park, J. Oh, S. Lee, and H. Yu, "Convolutional matrix factorization for document context-aware recommendation," 2016, doi: 10.1145/2959100.2959165.

[50] H. Wang, N. Wang, and D. Y. Yeung, "Collaborative deep learning for recommender systems," 2015, doi: 10.1145/2783258.2783273.

[51] X. Yang, S. Zhou, and M. Cao, "An Approach to Alleviate the Sparsity Problem of Hybrid Collaborative Filtering Based Recommendations: The Product-Attribute Perspective from User Reviews," *Mob. Networks Appl.*, 2020, doi: 10.1007/s11036-019-01246-2.

[52] L. Zheng, V. Noroozi, and P. S. Yu, "Joint deep modeling of users and items using reviews for recommendation," 2017, doi: 10.1145/3018661.3018665.

[53] S. Wang *et al.*, "Graph Learning based Recommender Systems: A Review," no. May, 2021, [Online]. Available: http://arxiv.org/abs/2105.06339.