*November-December 2019*
*ISSN: 0193-4120 Page No. 1869 - 1875*

# A Study on Hybrid Recommender System with Deep Learning and Deployment in Big Data

[1]Srivats.S.Ramanujam, [2]Sivaneshwar. P, [3]J.Naren, [4]Madhumitha.S, [5]Dr.G.Vithya

*[1]B. Tech Computer Science and Engineering, School of Computing, SASTRA Deemed University, Thanjavur, India.*
*srivatsramanujam@sastra.ac.in*

*[2]B. Tech Information and Communication Technology School of Computing, SASTRA Deemed University, Thanjavur, India.*
*sivaneshwar.p@gmail.com*

*[3]Asst Professor, School of Computing, SASTRA Deemed University, Tirumalaisamudram, Thanjavur, Tamil Nadu, India. naren.jeeva3@gmail.com*

*[4]B. Tech Electronics and Communication Engineering, School of Computing, SASTRA Deemed University, Thanjavur, India.*
*Madhumitha.mohan98@gmail.com*

*[5] Professor, School of Computing*
*KL University, Vijayawada*
*. vithyamtech@gmail.com*

***Article Info***
***Volume 81***
***Page Number: 1869 - 1875***
***Publication Issue:***
***November-December 2019***

***Abstract***
A recommender system is a filtering tool that provides customized suggestions of products to users by using various techniques. A hybrid recommender system combines the approaches of two or more techniques, primarily focusing on Content-based recommender systems and Collaborative recommender systems, thus proving better suggestions by exploiting the advantages of both the strategies. This paper provides a detailed review of the various ways to implement a Hybrid Recommender System, enabling the reader to get a bird's eye view on the concept. The study allows the user to understand the relationship between data characteristics and the relative performance of various approaches to a Hybrid Recommender System. The flow of the research starts with a description of the initial models of Recommender Systems, followed by a study on the application of Machine Learning and Deep Learning algorithms to the Hybrid Recommenders. The contributions to the study are dichotomous: 1) to identify and discuss the various Techniques for Hybrid Recommendation Systems via a Systematic literature review and 2) to compare the various techniques to get a clear view of the efficiency and the accuracy of each technique. At the end of the study, the reader understands the development flow of Hybrid Recommender Systems and gets a clear idea of how various Recommender Systems work. It is observed that the technique which incorporates Deep Learning provides better results.

***Keywords****: Hybrid Recommender Systems, Big data, Machine learning, Neural networks, Ontology*

***Article History***
***Article Received:*** *5 March 2019*
***Revised:*** *18 May 2019*
***Accepted:*** *24 September 2019*
***Publication:*** *10 December 2019*

## I. INTRODUCTION

When the amount of information fed to a System exceeds the threshold of processing capability of a system, information overload occurs. Technology has been the primary reason for the overload problem in today's world. Information overload also occurs when

*November-December 2019*
*ISSN: 0193-4120 Page No. 1869 - 1875*

a variety of data is loaded to the System, affecting the efficiency of the decisions of the Systems. In the current generation of smart devices, data is generated from a variety of sources like Facebook, Twitter, WhatsApp and email services. This makes it difficult for users to find the relevant required information, thus resulting in errors in the decision-making process. But, the vast information available cannot be neglected. So users require Intelligent Systems that filter information and provide efficient results to the users.

An Intelligent Recommender System is an application that provide efficient suggestions on information to users by dealing with the obstacles caused by information overloading. The recommendations generated by these systems are customized for every user, thus making them very efficient, accounting for their extensive usage in many applications. The System provides the users with various information on what is available, how it has made an impact in the society etc. Recommender System is in the trend as it can deal with the information overload problems efficiently by providing personalized recommendation to the users. Some examples of these intelligent Recommendation Systems are product recommendations by Amazon, movies by Netflix, profile suggestions by Facebook, news articles, financial services, Twitter, etc. The infrastructure of various organizations incorporate these systems, which has resulted in the betterment of sales in organizations, the satisfaction of users by a better understanding of their needs effectively. The major advantages comprises of improvement in sales of the organization, satisfaction of user and understanding the needs of the user effectively. [1]

In the following survey, we will go through various techniques so that we can get a clear view of the efficiencies based on the dataset used or the time taken. We review major techniques like Ontology, usage of Big Data, and usage of Neural Networks (Deep Learning) to clearly distinguish our needs and the techniques used base on it. We analyze each model based on the log loss and utilize the weight of the parameters used in these models to write the following review. We also assess the efficiency of a Hybrid Recommender System by analyzing the efficiency, technique, and accuracy of the observations used in each model.

## II. LITERATURE REVIEW

### Hybrid Recommender Systems

There are two broad categories of Recommender System algorithms: Item-based and User-Based Algorithms. These algorithms are implemented by using two filtering techniques called collaborative filtering and content-based filtering.

(a) **Collaborative Filtering**: Collaborative filtering generates recommendations by collecting preferences and information about users, and using this information to find similarities among the various users. Recommendations are then made based on the matching profiles. It is based on the assumption that, if two users share a similar view on a few issues, their views on other issues are most likely similar.

(b) **Content-based Filtering**: Content-based filtering approaches make use of discrete properties of an item, to recommend additional items with similar characteristics.

Early recommender systems applied either Collaborative filtering (CF) approach or a Content-Based Filtering (CBF) approach to provide suggestions. Eventually, researchers started to combine the two techniques which came to be referred to as Hybrid Recommendation/Recommender Systems, which integrated the advantages of both approaches and reduced the shortcomings, providing better results. Initially, a content-based analysis is done on user-profiles and the information gained is used to build an effective recommendation. The above approach gives the application advantage of providing efficient and effective recommendations to the user based on the search history and interests. A pseudo-rating System is created for the user and an efficient Recommendation System is used to predict the ratings that are missing for the user. Now, the Pearson correlation coefficient is used for creating a user pair similarity. A final content-boosted collaborative filtering prediction is generated for the user after computing Harmonic Mean weighting for incorporating better correlations.

An example of an application that uses this technique is an e-learning Recommender System that suggests learning materials using a system that recognizes a pattern and similarities among user actions which can be used to give suggestions. The two stages of the

November-December 2019
ISSN: 0193-4120 Page No. 1869 - 1875

System are (1) Finding related resources using collaborative filtering, and (2) Filtering items from common learning using sequential pattern mining.

**Ontology based Recommender Systems**

Ontology is a set of concepts in a domain that shows the relationship between them. Ontology contains major concepts like the attributes and entities and the relationships between them. Ontologies reuse the domain knowledge and improves efficiency. Reusing ontologies saves time and is more efficient as it has been tested a number of times with various data. Ontologies are widely adopted by many researchers and organizations for retrieval of information and to build an intelligent System.

Ontology is the specification of concepts in a domain. It includes concepts like attributes, entities, characteristics and establishes the relationships among them. Ontologies can be constructed using Ontology languages like Operational Conceptual Modelling Language (OCML), Web Ontology Language (OWL) and Gellish. Ontologies allow the reuse of domain knowledge, thus improving efficiency by saving time and ensuring the usage of pre-tested and used components. Hence ontologies are widely adopted by many researchers and organizations for retrieval of information and building Intelligent Systems.

Ontology-based Recommender Systems are information-based Recommender Systems that use Ontology and occasionally Semantics for representing the required knowledge and information. Ontologies are used to deduce user interests and efficiently enhance user profiles. In the Hybrid Ontology-based Recommender System, the ratings of each user are used along with the ontological knowledge to map the similarities efficiently. Once the mapping of concepts is completed, the regular recommendation concepts can be deployed. In the application which uses Recommender Systems for e-learning suggestions, ontologies are designed that will combine and relate the knowledge between the user/learner and the learning resources. A major advantage of using this technique is that the Intelligent Ontology-based Recommenders Systems do not face the problems generally encountered in traditional Recommenders Systems like Data Sparsity, over- specialization and cold-start. The recommendations generated closely match the end user's preferences through the personalized learner profiles created using Ontology-based Recommender Systems. Ontologies play an important role in the field of e-learning for personalized Recommendations, where user profiles are created alongside the profiles created for learning resources, providing additional information about the characteristics of the user like learning style and level of knowledge. But, if the entire ontology is stored in a profile, including the relationships, the problem of data redundancy and inconsistency arises. To overcome this inconsistency, the profiles can eliminate the storage of concept relationships, which can be obtained by queries in the domain ontology. [2]

**Big Data Recommendation Systems**

With the heightened use of the internet, the amount of data that is generated and available for processing is multiplying rapidly, demanding the usage of a robust system to handle the huge volumes. Hadoop is an open-source framework for storing and processing huge amounts of data. Hadoop Distributed File System (HDFS) is a structured file system that is built to run on various platforms. HDFS can be deployed on any low-cost hardware and is fault-tolerant. Only minor modifications are required for working with different ecological systems. Large volumes and varieties of data are processed parallelly on clusters of commodity hardware using a programming model called Map Reduce. A warehousing solution to the data is the Hive, which consists of a query language called HiveQL that can be used to manipulate the software with ease. Apache Mahout is a library primarily used for creating scalable machine learning algorithms, which runs on the Hadoop platform. A Recommender System built on a Big Data platform is much more efficient than the Ontology-Based Recommender System for large datasets as the system is equipped to handle huge volumes and varieties of data. [3]

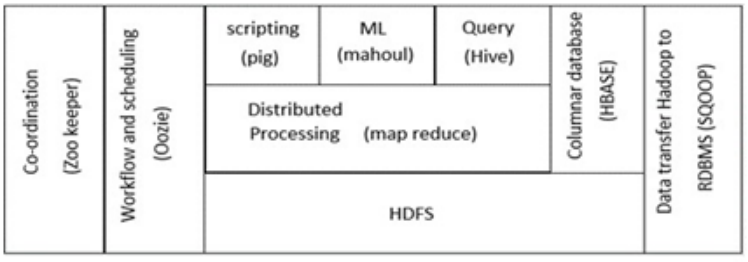


**Fig 1.1** Hadoop Eco System

November-December 2019
ISSN: 0193-4120 Page No. 1869 - 1875

**Machine Learning Recommendation Systems**

Machine Learning (ML) simulates human learning with computers, where information is fed from the real world which is then used by the computer used to acquire knowledge and generate inferences that increase the performance of tasks based on this knowledge. Recommender Systems are being used in many highly reputed information-based organizations such as Google, Twitter, LinkedIn, and Netflix. These systems can incorporate Artificial Intelligence methods to improve efficiency and provide better recommendations. For example, an online bookstore can classify books based on genre and these attributes can then be used for recommendations. The field of Machine Learning is ever-growing and is the future of technology. Recommendation Systems and Machine Learning fields together combined can benefit Software Engineering principles and definitions in the Intelligent Systems. However, since there are several approaches of ML algorithms without a clear scheme, it becomes difficult and confusing to choose an ML algorithm that fits one's need when developing a Recommender System. Thus a systematic analysis of real-time recommender systems implementing ML algorithms is required, and their results might aid in choosing the right algorithm.

**HYBRID RECOMMENDER SYSTEMS**

Kavinkumar et.al illustrated a Hybrid Recommendation System using user-based and item-based collaborative filtering. Additionally, a feedback mechanism is included to improve the efficiency of the suggestions. A two-level feedback system is proposed - (i)external feedback and (ii)internal feedback. External feedback involves gathering general comments and blogs from public platforms like websites and social media. Internal feedback is obtained from the users currently using the recommended products. This proposed system is applied to a dataset consisting of ratings for 40 cars by 50 users. Reviews are then collected from feedback forms and Twitter for feedback analysis, which is then classified and used for opinion extraction through sentiment analysis. Further research is encouraged for better sentiment analysis, and inclusion of other relevant geological, psychological and behavioral patterns of users are suggested for better results. [4]

Xia Wu et.al proposed two hybrid models for recommending mobile apps for various users. These models make use of RFD (Recency, Frequency, Duration) analysis, where customers are ranked based on how recently, how frequently and how consistently their purchases have been. This avoids the inaccuracy which results because of judging customer value based on a single parameter. The two models proposed are Improved Item-Oriented Collaborative Filtering (IIOCF) and Hybrid Latent Filtering (HLF) approach. The former improves the efficiency and performance of the system by making use of the latent factor model while the latter combines the latent factor model with the item-based approach. Experimental results showed that the HLF approach was better than the individual models. [5] Kazuyoshi et.al proposed a Hybrid music Recommender System. It uses a probabilistic generative model combines the collaborative and content-based approach efficiently. The primary goal was to provide an accurate recommendation for a huge variety of artists. Traditional techniques like Collaborative filtering could not recommend unranked music and could provide only a limited range of artists. On the other hand, content-based filtering could not develop an accurate picture of user preferences since the method looks for similar music content whereas user preferences vary across genres. To overcome the problem, a model based on Bayesian networks was developed that could effectively integrate the observed rating scores and acoustic features of the music. The System provided a high level of accuracy when new users and rating scored were added and learn incrementally. [6]

Toon De Pessemier et.al proposed a Hybrid Recommender System that gives personalized recommendations for various travel destinations to people. Collaborative, knowledge and content-based algorithms are combined in the Hybrid approach of the Recommender System. A group of users was used to evaluate the recommendation System using a web application that was created as a prototype. All the inputs were based on ratings, personal interests and specific demands of the user in the process of recommendation. A Sentimental Analysis and inclusion of more places can be done as a future work of the research paper. Machine Learning can also be deployed into the project. [7]

Intekhab Naser et.al proposed a Hybrid music recommendation System which uses context and

interests of the user to provide an enhanced recommendation. Firstly, the user's contextual information like location, motion, time etc., were collected and the users were classified into 2 categories based on the context and behavior of the user from the "context log" and "rating matrix" respectively. The recommended song list was provided by integrating the Refined User Sub-Matrix and Relevant Song List. The combined approach effectively recommended music to the users based on their interests and moods. [8]

**ONTOLOGY BASED HYBRID RECOMMENDER SYSTEMS**

Omar Portilla et.al presented a semantic personalized Hybrid recommended System for Biomedical Ontologies. A Semantic Repository of Biomedical Ontologies were designed and implemented to build the recommended System. The proposed recommended System also considers Hybridization, considering the calibration of the impact generated by two aspects: customization (adaptability) for each user, and the quality evaluation of each ontology to recommend. The knowledge stored in the metadata of the semantic repository is used for the Recommender System, in order to give prioritized recommendations of the different biomedical ontologies that meet certain search criteria. [9]

John K. Taurus et.al proposed a Hybrid Recommended System that uses Ontology and Sequential Pattern Mining (SPM) to recommend e-learning resources for students based on interests in various platforms. SPM algorithm has been deployed to discover the Sequential Pattern of the learner. A PageRank algorithm has been used to rank the recommended links for the user to choose the best one. The proposed Hybrid approach alleviates data sparsity problem and the efficiency has been improved. A data set from 50 people has been taken into account to evaluate the System. Big Data and Machine Learning can be deployed as a future work to improve the Recommended System[10].

Sheng-Yuan Yang et.al proposed a Hybrid System to aid the scholars using an ontology based model. The System extracted important information from specific domain documents and by using Hybrid filtering technique performed ranking to the recommendations and integrates the available information. Firstly, the System invoked the web crawler to collect web pages. The paged were classified and the necessary information was extracted. The information Recommender made resulted in an integration in the recommendation. Artificial Intelligence, ARN (Artificial Neural Networks), Fuzzy Logic were chosen as application for the scholar information domain in the proposed System. [11]

**BIG DATA BASED HYBRID RECOMMENDER SYSTEMS**

Girish Prasad et.al employed Big Data in Hybrid Recommended Systems to recommend the exact need for a user from a huge database of products. An experiment on Movie Data Sets have been conducted for the System. A Map Reduce framework has been deployed into the Hadoop framework and the difference in time utilized by serial and parallel Hybrid recommendation algorithm by using various clusters has been measured and calculated. A collaborative System Combined with Parallel K-Mean Clustering algorithm and slope-one algorithm on Hadoop Framework has been used to improve the efficiency of the Recommended System. [12]

Ludovico Boratto et.al investigated a prediction task in Granularity based Recommender System for predicting ratings in Big Data. A granularity-based group Recommender System is a System that can detect K groups of users from a set of n users and generate recommendations for the K groups. In the Big data scenario in businesses, providing personalized advertisement for every individual is complex and practically not affordable. The proposed System avoids both data sparsity in clustering task and the issues due to curse of dimensionality. [13]

**HYBRID RECOMMENDER SYSTEMS USING DEEP NEURAL NETWORKS**

Tulasi K.Paradarami et.al illustrated a Hybrid Recommendation System which uses deep learning. Product and service reviews were considered along with rating. A neural network model was developed from a group of collaborative and content features minimizing the loss due to logarithms and misclassification of error using stochastic gradient descent optimization algorithm. The model's performance was evaluated from Yelp for a dataset of

November-December 2019
ISSN: 0193-4120 Page No. 1869 - 1875

restaurants and resulted in a loss of 0.142 and an accuracy of 91.35%. The optimized learning capabilities of an ANN framework is the reason for the overall recommendation of the Hybrid System being improved. [14]

Ankit Kanojia et.al illustrated the usage of neural networks to build a personalized Hybrid Book Recommender System. Demographic information like age, sex, geographic location etc., were collected about the customer from user activity, cookies, and IP address. Each of these parameters improved the personalization factor in the recommendation factor. The System provided a rating from 1 to 5 using a MLP-neural network. A Hybrid Recommender System was proposed combining collaborative filtering and content based recommendation technique which overcame the limitations of an individual approach when used separately. The System was evaluated in Matlab. [15]

## IMPLEMENTATION DETAILS

The "Articles sharing and reading from CI&T Desktop" dataset from Kaggle was used to build collaborative, content-based and hybrid recommender system models. The dataset contains a real sample of **12 months logs** (Mar. 2016 - Feb. 2017) from CI&T's Internal Communication platform. It contains about **73000 logged users interactions** on more than **3000 public articles shared** in the platform. The Recall@N metric was used to evaluate the performance of the models. It evaluates whether the interacted item is among the top N items in the ranked list of recommendations for a user. For the content-based recommender model, the text corpus is preprocessed using TF-IDF which is used to represent each sentence in a vector form so that Cosine similarity can be used to find the similarity between each text sample. For the collaborative recommender model, Singular Value Decomposition is used to represent the user-item matrix into low dimensional representation in terms of latent factors. Hybrid recommender systems are a cross of content-based and collaborative approaches. The Hybrid recommender model's scores are obtained by multiplying the scores of content-based and collaborative models.

**Table 1.1:** A Comparative Study of various techniques used in Hybrid Recommender System

| Title of the Paper | Data Set | Techniques Used | Output |
|---|---|---|---|
| An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model | Audio signals from commercial CD's and its ratings from e-commerce sites | Probabilistic generative model: - a Bayesian network model called a three-way aspect model which were incrementally trained online with new users and ratings | Highly accurate ranking of music across different genres and large variety of artists as per user preference. Handled non-rated musical pieces as well and made appropriate suggestions |
| A Hybrid knowledge-based Recommender System for e-learning based on ontology and sequential pattern mining | Ratings of learning resources collected from a public university. | 1. Ontology for learning sources and learners. 2. Computing similarities and prediction rating from ontology. 3. Generation of top N learning resources by the CF Recommender. 4. Application of GSP algorithm to top N learning resources. | Recommended list of suitable learning resources for a learner Established the superiority of (CF + Onto + SPM) Hybrid System over (CF only) and (CF + Onto only) Systems. Accuracy was 94% with the final recommendations Tackled issues with cold start and data sparsity |
| A Hybrid Recommender System using artificial neural networks | Yelp (a company publishing crowd sourced reviews about local businesses) academic dataset (2015) was used. | Artificial neural network – a multi categorical classification model that predicts the class of a rating. LogLoss is the cost function minimized by applying stochastic gradient descent. | Interprets user reviews of services from Yelp dataset and recommends appropriate service for the user. The ANN trained Hybrid RS demonstrated a 75% improvement over conventional user based CF algorithm. Overall accuracy of 91.35% |
| Building Hybrid Recommendation System Based on Hadoop Framework | Movie Lens Dataset from DEC Research. | K-mean Clustering Algorithm and Slope One on Hadoop Framework Distance measure and initial centroids method based on MapReduce | Comparison between Time consumption of serial and parallel Hybrid RS. The results showed that with increase in Hadoop nodes there was decrease in time consumption in Parallel Hybrid RS. Distance measure accuracy: 94.73%, Initial centroids method accuracy: 97.33% |

## III. RESULTS AND CONCLUSION

Hybrid recommender model has higher Recall@5 and Recall@10 scores than collaborative and content-based models. The content-based model has a higher Recall@5 and Recall@10 scores than the collaborative model.

The above work analyzes in depth the various Hybrid Recommended Systems and its efficiencies in various technologies and methods available. It starts with a Systematic literature review regarding related work. The literature review shows that most work is performed on Collaborative Filtering, with a reduced number of datasets, algorithms and evaluation measures in the base-level. However, recent works have improved such dimensions, effectively contributing to advances in the field. Integrating technologies like Machine Learning and Data Mining into the Recommender System will enhance the efficiency of the System and this should be the focus if researchers in the future. Nevertheless, the knowledge gathered in the document allows to positively direct future work in this research topic. [16]

November-December 2019
ISSN: 0193-4120 Page No. 1869 - 1875

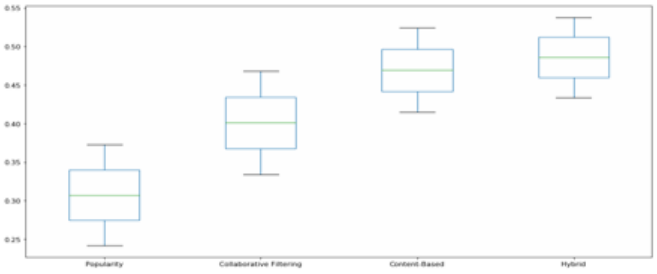

**Fig 1.2:** BoxPlot to compare the various types of Recommender Systems

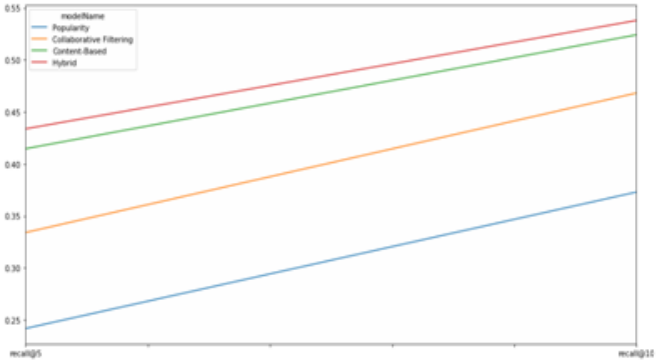

**Fig 1.3:** LinePlot to compare the various types of Recommender Systems

## REFERENCES

[1] Jose Aguilar*, Priscila Valdiviezo-Dı´az, Guido Riofrio. (2017), "A general framework for intelligent Recommender Systems", Applied Computing and Informatics, pp.147-160.

[2] Jai Prakash Verma, Bankim Patel, Atul Patel. (2015), "Big Data Analysis: Recommendation System with Hadoop Framework", Computational Intelligence & Communication Technology, pp. 92-97.

[3] Ivens Portugal, Paulo Alencar, Donald Cowan., (2018) "The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review", Expert Systems with Apploications, pp. 205-227.

[4] Kavinkumar.V, Rachamalla Rahul Reddy, Rohit Balasubramanian, Sridhar.M, Sridharan.K, Dr. D.Venkataraman. (2015), "A Hybrid Approach for Recommendation System with Added Feedback Component", Advances in Computing,Communications and Informatics (ICACCI), pp. 745-752.

[5] Xia Wu, Yanmin Zhu. (2016), "A Hybrid Approach based on Collaborative Filtering to Recommending Mobile Apps", Parallel and Distributed Systems, pp. 8-15.

[6] Kazuyoshi Yoshii, MasatakaGoto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno. (2008), An "Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model", Audio, Speech, And Language Processing, Feb, Vol.16, No.2, pp. 435-447.

[7] Toon De Pessemier, Jeroen Dhondt, Luc Martens. (2017), "Hybrid group recommendations for a travel service", Multimed Tools Appl, pp. 2787- 2811

[8] Intekhab Naser, Reena Pagare, NayanKumarWathap, Vinod Pingale. (2014), "Hybrid Music Recommendation System: Enhanced Collaborative Filtering Using Context and Interest Based Approach", IEEE (INDICON).

[9] Omar Portilla, Jose Aguilar*. (2016), "Hybrid Recommender System of Biomedical Ontologies", Computing Conference (CLEI), Jan, pp.1-12.

[10] John K. Tarus ,ZhendongNiua∗ , Abdallah Yousif, (2017), "A Hybrid knowledge-based Recommender System for e-learning based on ontology and sequential pattern mining", Future Generation Computer Systems , Feb, pp. 37-48

[11] Sheng-Yuan Yang, Chun-Liang Hs. (2010), "A New Ontology-Supported and Hybrid Recommending Information System for Scholars", Network- Based Information Systems, pp. 379-384.

[12] Girish Prasad, Mahendra Kumar Gourisaria, Lalit Kumar Vashishtha. (2016), "Building Hybrid Recommendation System Based on Hadoop Framework", Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 3493-3499.

[13] Ludovico Boratto∗ , Salvatore Carta, Gianni Fenu. (2017), "Investigating the role of the rating prediction task in granularity-based group Recommender Systems and big data scenarios", Information Sciences, pp. 424-443.

[14] Tulasi K. Paradarami∗ , Nathaniel D. Bastian, Jennifer L. Wightman. (2017), "A Hybrid Recommender System using artificial neural networks", Expert Systems with Applications, pp. 300-313

[15] Ankit Kanojia, Om Prakash Verma, Hitesh Nirwan. (2016), "Personalized Hybrid Book Recommender System using Neural Network, Computing for Sustainable Global Development (INDIACom), pp. 1281-1288.

[16] Tiago Cunhaa∗ , Carlos Soares, André C.P.L.F. de Carvalho. (2018), "Metalearning and Recommender Systems: A literature review and empirical study on the algorithm selection problem for Collaborative Filtering", Information Sciences, pp.128-144.