IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 5 ISSUE : 6 Print / Issue Publication Date: 21-Dec-2020

ISSN : 2455-2143

DOI : 10.33564/IJEAST.2020.v05i06.036

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 242-248
Published Online October 2020 in IJEAST (http://www.ijeast.com)

COMPARISON OF DIFFERENT MUSIC RECOMMENDATION SYSTEM ALGORITHMS

Meghan Patil
Department of Computer Science
NMIMS University, Mukesh Patel School of Technology Management & Engineering
Mumbai, Maharashtra, India

Sainaya Brid
Department of Computer Science
NMIMS University, Mukesh Patel School of Technology Management and Engineering
Mumbai, Maharashtra, India

Stuti Dhebar
Department of Computer Science
NMIMS University, Mukesh Patel School of Technology Management and Engineering
Mumbai, Maharashtra, India

Abstract— **Recommendation systems have become very popular and an essential part of consumer-based systems such as shopping, music, movies, etc. New algorithms are being developed to improve performance and understand the needs of the consumers better. This paper reviews some of the most used machine learning and deep learning models for music recommendation systems and explains why deep learning algorithms are so efficient.**

Keywords— **Music recommendation system, machine learning, deep learning, neural network.**

I. INTRODUCTION

A recommendation system is based on two filtering methods:

Content Filtering – creates a profile for each user or item to characterize its nature. (Seo, 2018)

Collaborative Filtering – analyses relationships between users and inter-dependencies among items to identify new user-item pairs. Generally, more accurate than content filtering however, it suffers from a cold start problem. (Seo, 2018)

Cold start problem: If a new user exists and does not have any inter-dependencies among existing values, we are not able to recommend anything. (Seo, 2018)

Collaborative filtering has two methods:

Neighborhood Method – computing the relationships between items or between users. Example: K-Nearest Neighbors (KNN). (Seo, 2018)

Matrix Factorization Method – each row in the matrix represents each user, while each column represents a different item. Characteristic features of the matrix:

1) Matrix is sparse because not every user likes or uses every item. (Seo, 2018)

2) can incorporate implicit feedback, information that is not directly given but can be derived by analyzing user behavior. (Seo, 2018)

Example: Singular Value Decomposition (SVD), Probabilistic latent semantic analysis (PLSA). (Seo, 2018)

There are two types of data on which we can gather information about the likes and dislikes of users:

- Explicit Data: data where we have some sort of rating. Here we know how much a user likes or dislikes an item. (Victor, 2017)
- Implicit Data: data we gather from the user's behavior, with no ratings or specific actions needed like how many times a user played a song, or watched a movie, etc. (Victor, 2017)

Since users might not spend the time to rate items or your app might not work well with a rating approach in the first place most systems usually go for implicit data but the downside for this data is that it is noisier and not always apparent what it means (Victor, 2017). In our review, we are going to use implicit data from the lastfm dataset to compare the algorithms.

II. LITERATURE REVIEW

A. Singular value decomposition

SVD is a dimensionality reduction technique that reduces the features in a dataset. In the context of the recommender system, SVD is used as a collaborative filtering technique. Working of SVD in terms of a recommendation system is as follows:

- It uses a user-item-rating matrix where rows represent users, columns represent the items, and the users give a rating to the items which are used for predicting recommendations (Kumar, 2020).
- Perform factorization of the user-item-rating matrix to get the factors (Kumar, 2020).
- Assuming N - number of users, M - number of items, R - latent features, matrix A - N*R, matrix B - N*R (Kumar, 2020).

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 242-248
Published Online October 2020 in IJEAST (http://www.ijeast.com)

- S - R*R, which describes the strength of each latent factor and V - R*M, which indicates the similarity between items and latent factors (Kumar, 2020).
  Assuming that each item is represented by a vector (Xi) and each user is represented by a vector (Yu). The expected rating by a user on an item is represented as (Ui) (Kumar, 2020).
- The (Xi) and (Yu) can be obtained in a manner that the square error difference between their dot product and the expected rating in the user-item matrix is minimum (Kumar, 2020).
- To reduce the error between the value predicted by the model and the actual value, the algorithm uses a bias term (Kumar, 2020).

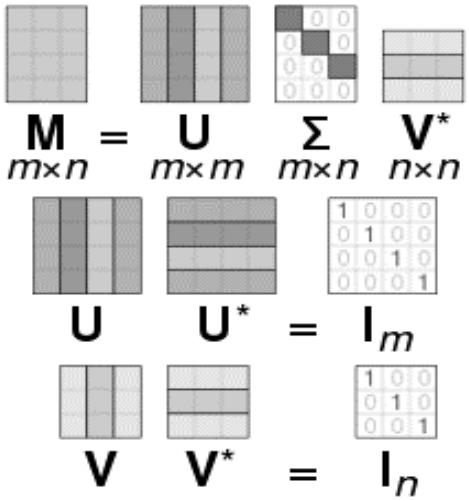

Fig. 1. Representation of matrix factorization in Singular Value Decomposition

Advantages:

- It is very efficient and tends to perform well for most datasets.
- The basis is hierarchical, ordered by relevance

Disadvantages:

- Results are not always best for visualization.
- Strongly focuses on the variance in cases where there is no direct relationship between variance & predictive power can discard useful information.
- Since it operates on fixed matrices and hence it may not be amenable to problems that require adaptive algorithms.
- For problems that can be solved by simpler techniques (Fourier transform, QR decomposition, etc.), SVD can be computationally expensive (Leach).

B. Bayesian Personalized Ranking

Bayesian personalized ranking (BPR) uses the likelihood function and the prior probability to perform analysis (Narapareddy, 2019). Working of BPR in terms of the recommendation system is as follows:

- BPR looks at the user, one item the user interacted with, and one item the user did not (the unknown item). This gives us a triplet (u, i, j) of a user (u), one known item (i), and one unknown item (j) (Narapareddy, 2019).
- It then finds a personalized ranking for a user for all items (i) in the set (Narapareddy, 2019).
- It then tries to maximize the probability of it being true.
- It then performs the final optimization.
- BPR is implemented using matrix factorization which means taking a matrix A of size (all users x all items) and factoring it into a matrix B of size (all users x latent features) and another matrix C of size (latent features x all items) (Narapareddy, 2019).
- After performing BPR we want the approximation where A is approximate to B*C (Narapareddy, 2019).
- We use an optimization method for example Stochastic Gradient Descent (SGD) to arrive at this approximation.
- By using an optimization method, we continuously updating them with new values to maximize the probability for A = B* V (Narapareddy, 2019).

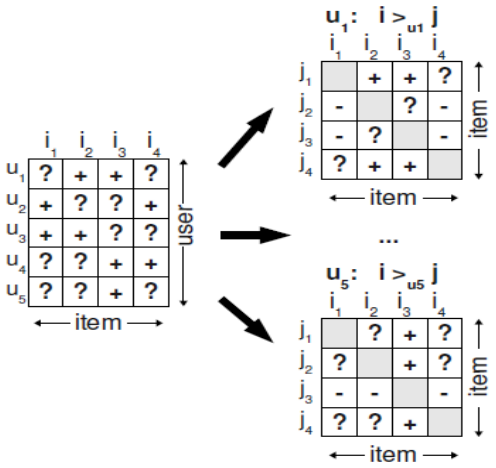

Fig. 2. Item – User representation in Bayesian Personalized Ranking

Advantages:

- Takes into consideration the items a user has not interacted with yet (Victor, 2017) (Narapareddy, 2019).

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 242-248
Published Online October 2020 in IJEAST (http://www.ijeast.com)

- involves pairs of items to come up with more personalized rankings for each user (Victor, 2017) (Narapareddy, 2019).
- Handles the problem of overfitting (Victor, 2017) (Narapareddy, 2019).

Disadvantages:

- Logic is based on Naïve Bayes hence it assumes that all users act independently with each other (Victor, 2017) (Narapareddy, 2019).
- The ordering of each pair of items (i, j) for a specific user is independent (Victor, 2017) (Narapareddy, 2019).

C. Autoencoders

It uses Collaborative filtering for recommendations. It is used to encode a set of input data, usually to achieve dimensionality reduction (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020). It is a feedforward neural network having an input layer, one hidden layer, and an output layer. The output layer has the same number of neurons as the input layer to reconstruct its own inputs (Oppermann, 2018).
It uses unsupervised learning i.e.; labeled data is not necessary to learn and only a set of input data instead of input-output pairs. A smaller hidden layer than the input layer creates a compressed representation of the data in the hidden layer by learning correlations in the data (Oppermann, 2018).

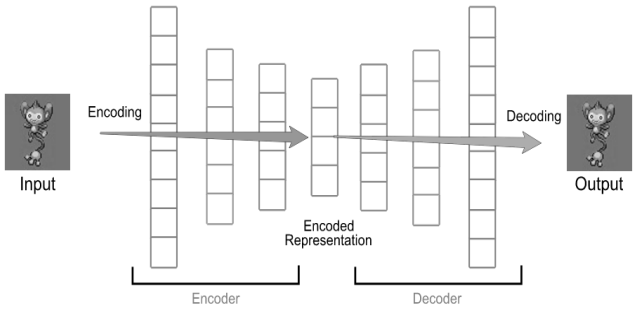

Fig. 3. Encoding and Decoding in Autoencoders

Working of autoencoders in terms of a recommendation system is as follows:

- Consider that there are n users, m items, and a partially filled user-item interaction/rating matrix A with dimension n*m (Oppermann, 2018).
- Each user (u) can be represented by a partially filled vector and each item (i) can be represented by a partially filled vector (Oppermann, 2018).
- Autoencoders directly takes user rating vectors or item rating as input data and obtains the reconstructed rating at the output layer (Oppermann, 2018).
- There are two variants of Autoencoders depending on two types of inputs:
  1) item-based AutoRec (I-AutoRec)
  2) user-based AutoRec (U-AutoRec).
- Both have the same structure.
- I-AutoRec generally performs better than U-AutoRec because the average number of ratings for each item is much more than the average number of ratings given by each user (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020).

Advantages:

- Performs dimensionality reduction (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020) (Bacuet, 2019).
- It can handle heterogeneous data sources and is, therefore, more adaptable in multi-media scenarios (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020) (Bacuet, 2019).
- It has a better understanding of the user demands and item features, thus leading to higher recommendation accuracy (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020) (Bacuet, 2019).
- It can handle input noise better than traditional recommendation models (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020) (Bacuet, 2019).

Disadvantages:

- It does not have interpretability meaning it is impossible to explain the results (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020) (Bacuet, 2019).
- It is very complex to implement (Le, Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering, 2020) (Bacuet, 2019).

D. Restricted Boltzmann Machine

The Restricted Boltzmann Machine (RBM) has an input layer (also referred to as the visible layer) and a hidden layer. In RBM the connections among the neurons are restricted. RBM is considered "restricted" because no two nodes in the same layer share a connection i.e., there are no connections among hidden nodes to hidden nodes or visible nodes to visible nodes. In RBM visible and hidden neuron connections form a bipartite graph. The visible nodes in RBM represent the components that can be observed, and the hidden units represent the dependencies

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 242-248
Published Online October 2020 in IJEAST (http://www.ijeast.com)

present between the components that can be observed (Nayak, 2019).

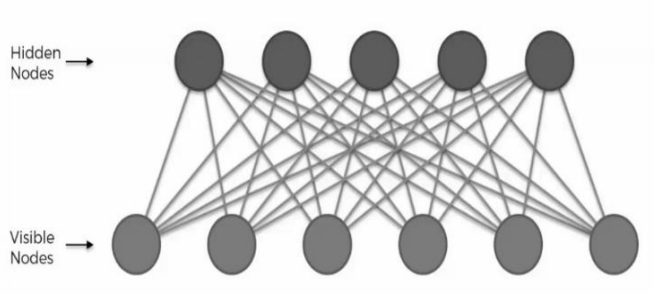

Fig. 4. Architecture of Restricted Boltzmann Machine

It can be trained by using a supervised or unsupervised model depending on the task to be performed.

Three steps are repeated over and over throughout the training process:

- Forward pass: every input is combined with individual weight and one overall bias, and the result is passed to the hidden layer which may or may not activate (Nayak, 2019).
- Backward pass: each activation is combined with individual weight and an overall bias, and the result is passed to the visible layer for reconstruction [9].
- Visible layer: the reconstruction is compared against the original input to determine the quality of the result (Nayak, 2019).

It decides which input features are important and how they should be combined to form patterns i.e., which input feature will be part of what family of neural nets that are all designed to recognize inherent patterns in data (Nayak, 2019).

Working of RBM is as follows:

- Consider a (user-item interaction) matrix A with the dimension n*m, where n is the number of users, and m is the number of items (Le, Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering, 2020).
- An entry Aij (row i, column j) corresponds to user (i's) rating for item (j) (Le, Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering, 2020).
- The rows of A encode each user's preference over all movies, and the columns of A encode each item's ratings received by all users (Le, Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering, 2020).
- The whole user-item interaction matrix is a collection of each user's ratings and since each user interacts with different items, each will have a unique RBM graph (Le, Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering, 2020).
- The variable is initialized as a set of edge potentials that are tied across all such RBM graphs (Le, Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering, 2020).
- In the training phase, RBM characterizes the relationship between the ratings and hidden features using conditional probabilities (Le, Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering, 2020).
- In the optimization phase, the variable initialized as a set of edge potentials is optimized (Le, Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering, 2020).

Advantages:

- The learning algorithm is intuitive i.e., with a big enough dataset, this algorithm can effectively learn arbitrary mappings between input and output (Educba.com, 2019) (Pole, 2015).
- It can solve an imbalanced data problem (Educba.com, 2019) (Pole, 2015).
- It can overcome the problem of noisy labels by correcting incorrect label data during reconstruction (Educba.com, 2019) (Pole, 2015).
- The problem of unstructured data is rectified by a feature extractor that transforms the raw data into hidden units (Educba.com, 2019) (Pole, 2015).
- It can find missing values (Educba.com, 2019) (Pole, 2015).

Disadvantages:

- It is difficult to interpret learned features required in recommendation systems when we use RBM (Educba.com, 2019) (Pole, 2015).
- It suffers from inaccuracy
- Training time because training is intractable
- Estimating probability accurately in RBM is difficult (Educba.com, 2019) (Pole, 2015).

E. Deep Neural Network

Deep Neural Network (DNN) is a neural network that consists of layers other than the input layer and the output layer called the hidden layers. It uses sophisticated mathematical modeling to process data in complex ways. Each layer performs specific types of sorting and ordering in a process that some refer to as

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 242-248
Published Online October 2020 in IJEAST (http://www.ijeast.com)

"feature hierarchy". It represents a specific form of machine learning where technologies using aspects of artificial intelligence seek to classify and order information in ways that go beyond simple input/output protocols (Techopedia, 2018).

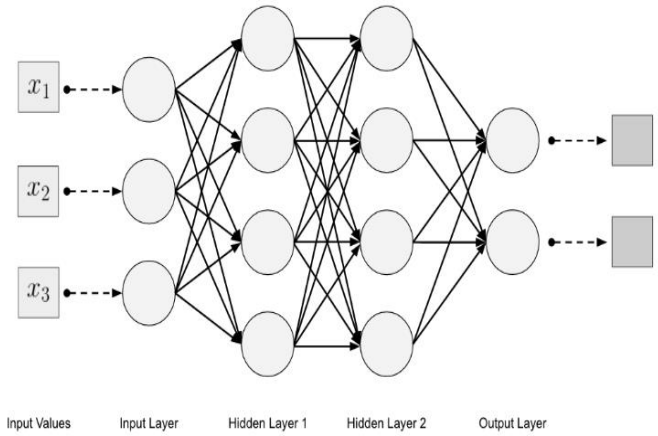

Fig. 5. Architecture of Deep Neural Network

General working of DNN is as follows:

- The first step is to pre-process the data and convert it into a format that can be fed to the neural network.
- To do this we can use techniques like tokenization, word2vec transformers, etc.
- We split the data into training, test, and validation. The splitting could be in any format i.e., 80-10-10 or 75-15-10, etc. depending on the type of data we are using.
- We create two columns of the data which are the feature and label columns.
- We feed the features and the labels to the neural network to perform training and validation.
- Training data – the model learns from this data.
- Validation data – the model checks its performance on this data.
- We then use previously unseen data i.e., the test data to measure model accuracy and loss.
- We use different hyperparameters to improve the performance of the model based on its performance on the validation data before testing the model.
- Hyperparameters are the parameters that can be manually changed by us to change the model performance.
- It consists of parameters like number of layers, number of neurons in each layer, activation function, optimization function, etc.
- Once we get the desired loss and accuracy from the validation data, we then test the model on testing data to get the test accuracy and loss.

Advantages:

- Robustness to natural variations in data is automatically learned (RF Wireless World, 2018).
- When the amount of data is huge, it performs better as compared to other machine learning models (RF Wireless World, 2018).
- Flexible to be adapted to new problems in the future (RF Wireless World, 2018).
- Gives high-cardinality outcomes (Arya, 2018).

Disadvantages:

- Large amounts of data are required to train deep learning models. Though big companies can gather and work with abundant data, small companies may not be able to do so. Also, data availability for some sectors can be sparse.
- Requires large amounts of processing power (Arya, 2018).
- Retraining of models is required in case we are dealing with different domains (Arya, 2018).
- More time consuming as compared to other techniques.
- Prone to overfitting (MC.AI, 2019).

III. Implementation and Result

We implemented the five algorithms mentioned in the literature review on the lastfm dataset. First, we performed pre-processing where we removed nan values in the dataset and took only the required columns in the dataset. The columns in the dataset we used were user id, artist id, and plays (number of times the user played the songs of that artist). The column plays is an implicit type of data as the user does exactly specify how many times they have played the songs of that particular artist. The user may not listen to every artist in the dataset, this is the latent feature in the music dataset. We then give this data to the model to perform training and testing. After implementing these five algorithms on the lastfm dataset we got the following result shown in Table 1. We have used parameters - loss, accuracy, and root mean square error (RMSE) as a means of comparison for the recommendation. RSME gives the error rate with which the algorithm provides recommendations. RMSE and loss should be as low as possible. As we can see in Table 1 given below, DNN gives the best accuracy and lowest RMSE proving to be the ideal algorithm among these to perform music recommendation.

Table -1 Performance Comparison of Algorithms

Parameter	Algorithms				
	SVD	**BPR**	**Autoencoders**	**RBM**	**DNN**
RMSE	**0.67**	**0.66**	**0.62**	**0.59**	**0.55**
Loss	**0.63**	**0.58**	**0.56**	**0.53**	**0.5**
Accuracy	**0.75**	**0.83**	**0.86**	**0.87**	**0.92**

IV. CONCLUSION

After implementing the five algorithms SVD, BPR, autoencoders, RBM, and DNN on the lastfm dataset we have concluded that DNN performs better of all with minimum loss and RMSE and highest accuracy among the five algorithms we tested.

V. ACKNOWLEDGEMENT

The author and all the co-authors involved, sincerely acknowledge the efforts of all the people involved in helping them write this paper. They also appreciate the reviewers for their time and effort in reviewing this paper.

VI. REFERENCES

1 Arya, T. (2018, July 19th). *Drawbacks of Deep Learning*. (Stanford University) Retrieved September 11, 2020, from Stanford Management Science and Engineering:https://mse238blog.stanford.edu/2018/07/tanuarya/drawbacks-of-deep-learning/

2 Bacuet, Q. (2019, April 2). *How Variational Autoencoders make classical recommender systems obsolete*. (Medium) Retrieved August 12, 2020, from https://medium.com/snipfeed/how-variational-autoencoders-make-classical-recommender-systems-obsolete-4df8bae51546

3 Educba.com. (2019, August 5). *Introduction to Restricted Boltzmann machine*. (educba.com) Retrieved September 20, 2020, from https://www.educba.com/restricted-boltzmann-machine/

4 Kumar, D. V. (2020, March 23rd). *Singular Value Decomposition (SVD) & Its Application In Recommender System*. (Analytics India Magazine) Retrieved August 25, 2020, from Analytics India Magazine: https://analyticsindiamag.com/singular-value-decomposition-svd-application-recommender-system/

5 Le, J. (2020, June 27). *Recommendation System Series Part 6: The 6 Variants of Autoencoders for Collaborative Filtering*. (Towards Data Science) Retrieved August 12, 2020, from https://towardsdatascience.com/recommendation-system-series-part-6-the-6-variants-of-autoencoders-for-collaborative-filtering-bd7b9eae2ec7

6 Le, J. (2020, August 17). *Recommendation System Series Part 7: The 3 Variants of Boltzmann Machines for Collaborative Filtering*. (Towards Data Science) Retrieved September 20, 2020, from https://towardsdatascience.com/recsys-series-part-7-the-3-variants-of-boltzmann-machines-for-collaborative-filtering-4c002af258f9

7 Leach, S. (n.d.). Singular Value Decomposition - A Primer. Providence. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Singular_value_decomposition

8 MC.AI. (2019, December 9th). *Eight Deep Learning Pros and Cons*. (MC.AI) Retrieved September 11, 2020, from MC.AI: https://mc.ai/eight-deep-learning-pros-and-cons/

9 Nait-Meziane, M. (2017, October 19th). *What are the advantages and disadvantages of using the singular value decomposition algorithm?* (Research Gate) Retrieved August 25, 2020, from Research Gate: https://www.researchgate.net/post/What_are_the_advantages_and_disadvantages_of_using_the_Singular_Value_Decomposition_SVD_algorithm

10 Narapareddy, A. (2019, February 4). *Recommender system using Bayesian personalized ranking*. (Towards Data Science) Retrieved June 15, 2020, from https://towardsdatascience.com/recommender-system-using-bayesian-personalized-ranking-d30e98bba0b9

11 Nayak, M. (2019, April 17). *An Intuitive Introduction Of Restricted Boltzmann Machine (RBM)*. (Medium) Retrieved September 20, 2020, from https://medium.com/datadriveninvestor/an-intuitive-introduction-of-restricted-boltzmann-machine-rbm-14f4382a0dbb

12 Oppermann, A. (2018, April 15). *Deep Autoencoders For Collaborative Filtering*. (Towards Data Science) Retrieved August 12, 2020, from https://towardsdatascience.com/deep-autoencoders-for-collaborative-filtering-6cf8d25bbf1d

13 Pole, I. (2015, April 30). *Restricted Boltzmann Machine - A comprehensive study with a focus on Deep Belief Networks*. (Slideshare.net) Retrieved September 20, 2020, from https://www.slideshare.net/poleindraneel/413688151-restricted-boltzmann-machine-a-comprehensive-study-with-a-focus-on-deep-belief-networks

14 RF Wireless World. (2018, July 18). *Advantages of Deep Learning | disadvantages of Deep Learning*. (RF Wireless World) Retrieved September 11, 2020, from RF Wireless World: https://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-Deep

15 Learning.html#:~:text=Drawbacks%20or%20disadvantages%20of%20Deep%20Learning&text=%E2%9E%A8It%20requires%20very%20large,increases%20cost%20to%20the%20users.

16 Seo, J. D. (2018, November 22). *[Paper Summary] Matrix Factorization Techniques for Recommender Systems*. (Towards Data Science) Retrieved June 15, 2020, from Towards Data Science: https://towardsdatascience.com/paper-summary-matrix-factorization-techniques-for-recommender-systems-82d1a7ace74

17 Techopedia. (2018, April 13th). *Deep Neural Network*. (Techopedia) Retrieved September 11, 2020, from Techopedia: https://www.techopedia.com/definition/32902/deep-neural-network

18 Victor. (2017, August 24). *ALS Implicit Collaborative Filtering*. (Medium.com) Retrieved June 15, 2020, from https://medium.com/radon-dev/als-implicit-collaborative-filtering-5ed653ba39fe

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED
All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS
Free and unrestricted access to research for all.

GLOBAL REACH
Connecting researchers and professionals worldwide.

TIMELY PUBLICATION
We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143