ISSN No:-2456-2165

Deep-Learing based Recommendation System Survey Paper

Prakash Kumar K, Vanitha V
Department of Information Technology
Kumaraguru collage of technology
Coimbatore, India

Abstract:- With the proliferation of online information, recommender systems have shown to be an effective method of overcoming this information imbalance. The utility of recommendation systems cannot be overstated, as can their ability to ease several concerns associated with excessive choice. Deep learning has had a significant effect in recent years across a variety of research disciplines, including computer vision and natural language processing, contributing not only to astounding results but also to the alluring trait of learning feature representations from scratch. Deep learning's influence is equally ubiquitous, with research demonstrating its usefulness when utilised for recommender systems and information retrieval. The topic of deep learning in recommender systems appears to be growing. The purpose of this study is to provide an in-depth evaluation of recent research endeavours on deep-learning-based recommender systems. Simply put, we explain and categorise deep learning-based recommendation models, as well as provide a consistent appraisal of the research. Finally, we elaborate on current trends and give new perspectives on the industry's game-changing rise.

Keywords:- deeplearning; recommendation system.

I. INTRODUCTION

Recommender frameworks provide as an automatic safeguard against buyer indecision. Given the perilous growth of data available via the online, visitors are regularly greeted by an infinite number of items, films, or restaurants. Personalization is a critical technique for delivering a superior user experience in that capacity. These frameworks have become an integral and critical aspect of many data access frameworks that assist businesses in working with dynamic cycles and are inextricably linked to diverse online spaces like an internet company or media site.

The suggestion lists are generated based on the user's requirements, the product's attributes, past interactions with the customer, and some additional metadata, such as temporal and geographic data. According to the kind of input data, recommender models are often classified as collaborative filtering, content-based recommender systems, or hybrid recommender systems. [1].

Deep learning is seeing a meteoric rise in popularity at the moment. The industry has had tremendous success with previous several years. Academics and industry have been rushing to apply deep learning to a larger range of applications due to its ability to tackle a large number of complex problems while providing state-of-the-art solutions [27]. Additionally, deep learning has altered recommendation structures significantly and created new opportunities for recommender productivity enhancement. attracted significant interest for their ability to overcome the limitations of standard models and achieve high suggestion quality. Deep learning is effective at collecting non-linear data.

It permits the encoding of progressively intricate concepts as data representations at the upper levels through non-trivial user-item interactions. Additionally, it discovers deep data links directly from a variety of available data sources, including contextual, and visual information.

II. OVERVIEW OF RECOMMENDATION SYSTEM AND DEEP LEARNING

A. Recommendation System

Recommender frameworks assess users' preferences and proactively recommend items that consumers may be interested in. Typically, recommendation models are classified as communitarian separating, content-based, or half breed recommender frameworks. Shared separating generates suggestions by obtaining real messages from users, either expressed (for example, user's previous assessments) or understood (for example perusing history). The suggestion of content is mostly based on correlations between objects and the user's assistant data. A diverse range of assistive data, including as words, images, and recordings, might be examined. The term "half and half" refers to a recommender system that integrates at least two distinct types of proposal strategies.

B. Deep Learning

Deep learning is frequently regarded as a subfield of artificial intelligence. The characteristic essence of deep learning is that it acquires deep representations, i.e., acquires various degrees of representations and deliberations from data. For practical reasons, we regard any neural differentiable design to be'deep learning' if it advances a differentiable target task by the use of a version of stochastic angle plunge (SGD). Neural structures have demonstrated enormous capability in both controlled and unsupervised learning tasks [31]. We discuss a distinct demonstration of compositional ideal models that is strongly associated with this work in this paragraph. The following table summarises the many sorts of models.

When it comes to deep learning, each level learns to abstract and combine its incoming data. First, the raw pixels are abstracted and encoded edges; the second layer composes, encodes and encodes edge arrangements; third, a nose, and eyes are mapped; and fourth, an image recognition programme can detect that an image includes an actual face..

A deep learning method, on the other hand, is able to learn on its own which traits should be placed at which level. A certain amount of manual adjustment is still required; for example, the use of different layer counts and layer widths may result in various levels of abstrac.

Deep learning refers to how many levels of processing the data goes through. To be more specific, credit assignment paths (CAPs) in deep learning systems are very long. Transformations from input to output constitute the CAP. Connecting the dots between input and outcome, CAPs indicate possible causal relationships. A feedforward neural network has the same number of hidden layers as a CAP, which is equal to the number of CAPs plus one (as the output layer is also parameterized). When using recurrent neural networks, the CAP depth is theoretically endless since a signal may be sent via several layers. [2] Most studies believe that deep learning requires more than two levels of CAP depth, but there is no commonly agreed-upon threshold for depth. Because of this, CAP of depth 2 has been shown to be a universal approximator. [15] Additional layers have no effect on the network's capacity to approximate functions. It is easier to learn features with more layers in deep models (CAP > 2) than in shallow models

A greedy layer-by-layer strategy may be used to build deep learning architectures.

These abstractions may be disentangled using deep learning, which identifies and prioritises attributes that boost performance.

Because they reduce duplication in the representation of data, deep learning approaches for supervised learning avoid the need for feature engineering by using data compression techniques similar to principal component analysis (PCA).

Unsupervised learning problems may benefit from deep learning methods. Ones that aren't labelled are far more common than data that are. Neural history compressors and deep belief networks are two examples of deep structures that can be taught without supervision.

- A multi-layer perceptron (MLP) is a feed-forward neural network having many (at least one) hidden layers between the information and output layers. In this case, the perceptron may make use of subjective initiation work and is not really concerned with creating a strictly parallel classifier. MLPs may be viewed as stacked layers of nonlinear alterations that are capable of learning multiple levels highlight representations. MLPs are also recognised to be good approximators in general.
- A Convolutional Neural Network (CNN) [45] is a novel type of feed-forward neural network that incorporates convolutional layers and pooling processes. It is capable of capturing global and local highlights and significantly improving the efficacy and precision. It excels in data management when used in conjunction with frameworks such as geography
- The Recurrent Neural Network (RNN) [45] is well-suited for showing sequential data. Unlike feed forward neural networks, RNNs utilise circles and memories to recall earlier calculations. Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are frequently used in practise to overcome the vanishing gradient issue..
- Neural Autoregressive Distribution Estimation (NADE) [81, 152] is an unaided neural network constructed using a top autoregressive model and feedforward neural networks. It is a manageable and efficient assessor for displaying information dissemination and densities.
- Descriptor and producer are the two components of Adversarial Networks (AN). Throughout the entire process, the two neural networks are competing against each other in a "min-max" game framework.
- Attentional Models (AM) are differentiable neural structures that work dependent on delicate substance tending to over an info grouping (or picture). Consideration component is regularly universal and was incepted in Computer Vision and Natural Language Processing areas. Nonetheless, it has additionally been an arising pattern in deep recommender framework research.
- Deep Reinforcement Learning (DRL) [106]. Reinforcement learning works on an experimentation worldview. The entire system chiefly comprises the accompanying segments: specialists, conditions, states, activities and prizes. The mix between a deep neural network and support learning plan DRL which have accomplished human-level execution across different spaces like games and self-driving vehicles. Deep neural networks empower the specialist to get information from crude information and infer productive portrayals without handmade highlights and space heuristics.

C. Why Deep Neural Networks for Recommendation?

Prior to delving into the nuances of recent breakthroughs, it is necessary to understand why deep learning methodologies are being used to recommender systems. Numerous deep recommender frameworks have been developed with a restricted ability to focus for an extended period of time. The Field is unquestionably clamorous for improvement. Now, it is straightforward to examine the necessity for such a diverse array of structures and, maybe, the utility of neural networks in the issue area. In a similar vein, it is good at justifying each suggested design and the scenario in which it would be usually useful. Taken together, this inquiry is particularly pertinent to the issues of assignment, space, and recommender circumstances. Perhaps the most alluring aspects of neural designs are their ability to be (1) differentiable from start to end and (2) to provide suitable inductive predispositions based on the information kind. Assuming the model has an innate design that it can exploit, deep neural networks should be beneficial. For instance, CNNs and RNNs have long used the fundamental structure of vision (as well as human language). Essentially, the consecutive construction of meeting or snap logs is quite rational when compared to the inductive predispositions provided by recurrent/convolutional models [56, 143, 175].

Besides, deep neural networks are additionally composite as in various neural structure squares can be made into a solitary (tremendous) differentiable capacity and prepared from start to finish. The critical benefit here is when managing content- based proposals. It is unavoidable when demonstrating users/things on the web, where multi-modular information is ordinary. For example, when managing

ISSN No:-2456-2165

literary information (audits [202], tweets [44] and so forth), picture information (social posts, item pictures), CNNs/RNNs become crucial neural structure blocks. Here, the customary other option (planning methodology specific highlights and so forth) turns out to be essentially less alluring and thus, the recommender framework can't exploit joint (start to finish) portrayal learning. In some sense, improvements in the field of recommender frameworks are likewise firmly combined with propels research in related modalities (like vision or language networks). For instance, to handle audits, one would need to perform expensive preprocessing (e.g., keyphrase extraction,\ theme displaying and so forth) while fresher deep learning- based methodologies can ingest all text based data start to finish [202]. All things considered, the capacities of deep learning in this viewpoint can be viewed as outlook changing and the capacity to address pictures, text and collaborations in a brought together joint system [197] is unimaginable without these new advances.

With regards to the communication-only situation (i.e., lattice completion or cooperative placement), the critical point here is that deep neural networks are supported when there is a high degree of complexity or a large number of training samples. In [53], the authors used an MLP to predict how connections operate and demonstrated significant performance increases over more conventional approaches such as MF. While these neural models outperform regular AI models, for example, BPR, MF, and CML, it is well established that traditional AI models, for example, BPR, MF, and CML, perform decently well when prepared with force combined with inclination plummet for cooperation-only information [145]. In any case, we may consider these models to be neural designs as well, given they make use of recent breakthroughs in deep learning, such as Adam, Dropout, and Batch Normalization [53, 195].

It is additionally simple to see that, customary recommender calculations (matrix factorization, factorization machines, and so on) can likewise be communicated as neural/differentiable structures [53, 54] and prepared effectively with a system, for example, Tensor stream or Pytorch, empowering productive GPU-enabled preparing and free programmed separation. Thus, in the present exploration environment (and surprisingly modern), there is totally no motivation to not utilize deep learning based devices for advancement of any recommender framework.

Customer experience is improved by recommendation algorithms in a sea of e-commerce that is infinitely wide and whirling. When it comes to internet shopping, recommendation engines are removing the tyranny of choice. It's not only about tackling the issue of irrelevant suggestions with AI; it's also about anticipating the customer's next moves.

Online sales are expected to increase by two fold result effect. In these conditions, firms should provide excellent customer service and provide specific advice in order to distinguish out from their competition. Read on to learn more about how these systems function, the advantages and disadvantages of using them, and the algorithms that power them up.

For deep learning-based recommender systems, non-linear data processing outperforms linear data processing. The main advantages of using DL for recommendations.

To reiterate, we sum up the qualities of deep learning based proposal models that per users may remember when attempting to utilize them for training use.

- **Nonlinear Transformation:** In spite of straight models, deep neural networks is equipped for displaying the non-linearity in information with nonlinear initiations, for example, relu, sigmoid, tanh, and so on This property makes it conceivable to catch the mind boggling and many-sided user thing connection designs. Regular techniques like framework factorization, factorization machine, inadequate straight model are basically direct models.
- For instance, matrix factorization models the user's communication by straightly joining user and thin dormant components [53]; Factorization machine is an individual from multivariate direct family [54]; Obviously, SLIM is a direct relapse model with sparsity requirements. The straight presumption, going about as the premise of numerous conventional recommenders, is distorted and will extraordinarily restrict their demonstrating expressiveness. It is grounded that neural networks can inexact any consistent capacity with a self-assertive exactness by changing the actuation decisions and mixes [58, 59]. This property makes it conceivable to manage complex association designs and definitely mirror user's inclination.
- **Representation Learning**: deep neural network is effective in learning the basic illustrative factors and helpful portrayals from input information. As a general rule, a lot of clear data about things and users is accessible in certifiable applications. Utilizing this data gives an approach to propel our comprehension of things and users, in this manner, coming about in a superior recommender. Accordingly, it is a characteristic decision to apply
- deep neural networks to portrayal learning in proposal models. The benefits of utilizing deep neural networks to help portray learning are in two-folds: (1) it lessens the endeavors close by making a highlight plan. Highlight designing is a work escalated work, deep neural networks empower naturally highlight gaining from crude information in unaided or administered approach; (2) it empowers proposal models to incorporate heterogeneous substance data like content, pictures, sound and even video. Deep learning networks have made forward leaps in interactive media information preparation and shown possibilities in portrayals gained from different sources.
- **Sequence Modelling**: Deep neural networks have shown promising outcomes on various consecutive displaying errands, for example, machine interpretation, regular language understanding, discourse recognition,chatbots, and numerous others. RNN and CNN assume basic parts in these assignments. RNN achieves this with inside memory states while CNN accomplishes this with channels sliding alongside time. The two of them are

broadly appropriate and adaptable in mining. successive designs in information. Displaying consecutive signals is a significant theme for mining the worldly elements of user conduct and thing advancement. For instance, next-thing/basket expectation and meeting based proposal are recommendation applications. Thusly, deep neural network become an ideal fit for this consecutive example mining task.

D. On Potential Limitation

Are there actually any disadvantages and limits with utilizing Deep learning for proposals? In this part, We expect to handle a few regularly referred to contentions against the utilization of Deep learning for recommender frameworks research.

- **Interpretability**: It is noteworthy that Deep learning is acting as a hidden ingredient, and generating reasonable predictions looks to be a very challenging task. The hidden loading and enactments of Deep Neural Networks are often non-interpretable, which is a common criticism levelled against them. It should be noted that this concern has been alleviated by the use of neural consideration models, which has opened the door to deeper neural models that benefit from an increased level of interpretability. Neuronal models, not only in recommender frameworks, have a challenge when it comes to reading individual neurons, but contemporary scenarios with the-workshop models may now at least provide some level of readability, which allows logical reasoning. In the open problems section, we go into further depth on this topic.
- **Data Requirement:** Another possible limitation is that Classifiers is believed to be information hungry, in that it needs sufficient information to adequately support its rich definition. Deep learning is eager for information. However, compared to other fields (such language or vision) where named data is scarce, it is quite straightforward to obtain a large amount of information within the context of recommendation system system research. In both business and academia, million/billion-scale datasets are commonplace.
- **Extensive Hyperparameter Tuning:** The demand for wide hyperparameter adjustment is a third obstacle to learning. However, it is important to stress that hyperparameter tuning is not a specific problem of adapting to AI, but rather an issue of AI in general (e.g., regularisation Even said, Deep learning may sometimes propose additional hyperparameters. For example, a recent study [145] introduces just a single hyperparameter as a conscious addition to the traditional measurement learning computation [60].
- **deeplearning based recommendation**: state-of-art this part, firstly present the classes of learning based models and afterward feature cutting edge research models, intending to distinguish the most eminent and promising progression lately.

E. Categories of deep learning based recommendation models

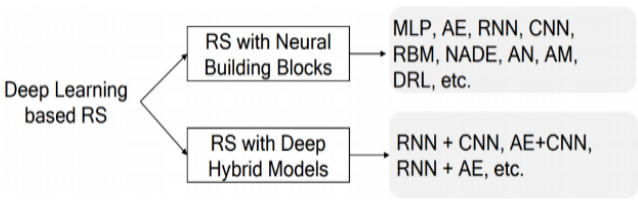


Fig. 1: Categories of deep neural network based recommendation models

Categories	Publications
MLP	[2, 13, 20, 27, 38, 47, 53, 54, 66, 92, 95, 157, 166, 185], [12, 39, 93, 112, 134, 154, 182, 183]
Autoencoder	[34, 88, 89, 114, 116, 125, 136, 137, 140, 159, 177, 187, 207], [4, 10, 32, 94, 150, 151, 158, 170, 171, 188, 196, 208, 209]
CNNs	[25, 49, 50, 75, 76, 98, 105, 127, 130, 153, 165, 172, 202, 206], [6, 44, 51, 83, 110, 126, 143, 148, 169, 190, 191]
RNNs	[5, 28, 35, 56, 57, 73, 78, 90, 117, 132, 139, 142, 174–176], [24, 29, 33, 55, 68, 91, 108, 113, 133, 141, 149, 173, 179]
RBM	[42, 71, 72, 100, 123, 167, 180]
NADE	[36, 203, 204]
Neural Attention	[14, 44, 70, 90, 99, 101, 127, 145, 169, 189, 194, 205], [62, 146, 193]
Adversary Network	[9, 52, 162, 164]
DRL	[16, 21, 107, 168, 198–200]
Hybrid Models	[17, 38, 41, 82, 84, 87, 118, 135, 160, 192, 193]

Table 1: A lookup table for reviewed publications.

- Content - based recommendation structure models are subdivided into eight classes, much as the previous eight Deep Learning models. The relevance of both the recommendation system is influenced by the use of Deep learning. Multilayer Perceptron (MLP) can simulate the ou pas partnerships for both places and things without difficulty; the CNNs are equipped for stripping away nearby and worldwide depictions from non - homogenous information sources like text-based and visual data; RNNs enable the recommender blueprint to display transient elements and successive development of content data.
- It's possible to utilise more than once Deep Learning method in a recommendation model using Deep Hybrid Learning. In order to create a more remarkable half-and-half model, the flexibility of Deep Learning Models allows for the consolidation of a few neural structural hindrances. However, not all of these night Transfer learning algorithms have been applied to their full potential. For example, [31] refers to "half breed" Deep networks, which use both generate and discriminative aspects in their architecture.

Using the previously indicated grouping approach, we have arranged the models in Table 1 in a logical order. Table 2 also includes a summary of the distributions based on the project's perspective. A wide range of projects are addressed in the evaluations. Some projects, such as meeting-based proposal, photo, and video recommendations, have gained attention due of the application of Deep neural networks. It's likely that some of the tasks will not be new to the field of suggestion research; nevertheless, DL offers a bigger opportunity to find improved arrangements (a full audit is included as an aftermath data for service marks in [131]). It would be impossible to keep track of all of your photos and videos without the help of Deep learning techniques. The exhibition of Deep Neural Networks' abilities makes it easy to capture repeated instances of user habits. Deep Neural

Networks The ensuing article will discuss some of the specific tasks..

Data Sources/Tasks	Notes	Publications
Sequential Information	w/t User ID	[16, 29, 33, 35, 73, 91, 117, 133, 143, 160, 173, 175, 189, 194, 198, 205]
	Session based w/o User ID	[55–57, 68, 73, 99, 101, 102, 117, 142, 148, 149]
	Check-In, POI	[150, 151, 165, 185]
Text	Hash Tags	[44, 110, 118, 158, 182, 183, 193, 209]
	News	[10, 12, 113, 135, 169, 200]
	Review texts	[11, 87, 126, 146, 174, 197, 202]
	Quotes	[82, 141]
Images	Visual features	[2, 14, 25, 49, 50, 84, 98, 105, 112, 165, 172, 179, 191, 192, 197, 206]
Audio	Music	[95, 153, 167, 168]
Video	Videos	[14, 17, 27, 83]
Networks	Citation Network	[9, 38, 66]
	Social Network	[32, 116, 166]
	Cross Domain	[39, 92, 166]
Others	Cold-start	[154, 156, 170, 171]
	Multitask	[5, 73, 87, 174, 187]
	Explainability	[87, 126]

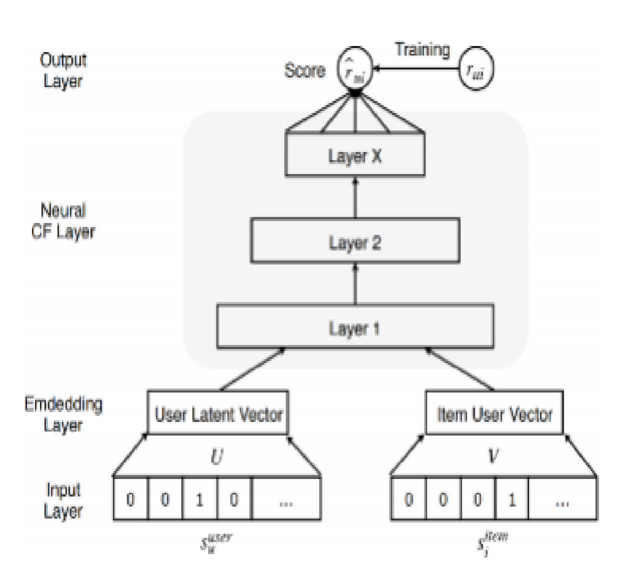
Table 2: Deep neural network based recommendation models in specific application fields.

F. Multilayer Perceptron based Recommendation

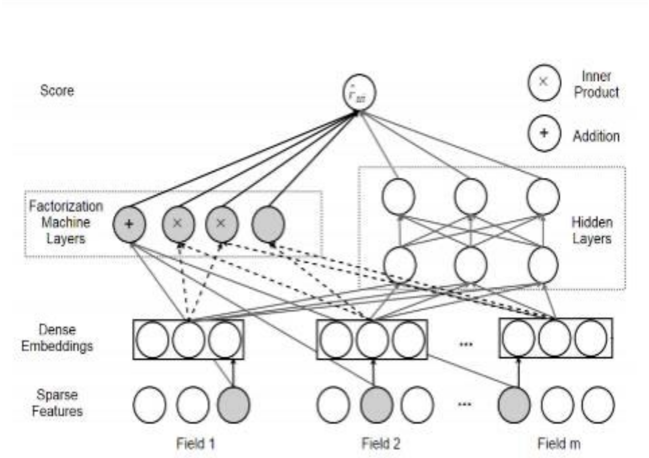
- With the MLP, it has been shown that any measurable capacity is capable of being roughed to any desired degree of accuracy [59]. The foundation of multiple advanced methodologies, it is widely used in a wide range of settings.
- Traditional Recommendation Methods are being extended. There are a lot of preexisting proposal models that are just techniques. Nonlinear changes to current RS drawings may be added using MLP, which can then be decoded into neural expansions.

Filtering with the help of others. There are several times when consumers' preferences and the items emphasised are deemed to have a two-way link. When using lattice factorization, the rating grid is broken down into low-dimensional inert variables for users and things.

.



(a)Neural Collaborative Filtering



(b) Deep Factorization Machine.

Let s^{user}_u and s^{item}_i signify the side data (for example profiles and thing highlights), or only one-hot identifier of user u and thing i. scoring capacity is characterized as follows:

$$\hat{r}_{ui} = f(U^T \cdot s^{user}_u, V^T \cdot s^{item}_i \mid U, V, \theta) \quad (1)$$

where work f (·) addresses the multi-facet perceptron, and θis the boundaries of this network. Customary MF can be seen as an uncommon instance of. Consequently, it isadvantageous to intertwine the neural understanding offramework factorization with MLP to form a more broad model which utilizes both linearity of MF and non-linearity of MLP to upgrade recommendation quality. The entire network can be prepared with weighted square misfortune (for unequivocal criticism) or double cross-entropy misfortune (for certain input). The cross-entropy misfortune is characterized as:

$$\mathcal{L} = -\sum_{(u,i)\in\mathcal{O}\cup\mathcal{O}^-} r_{ui}\log \hat{r}_{ui} + (1 - r_{ui})\log(1 - \hat{r}_{ui}) \quad (2)$$

Negative inspection methods can be used to reduce the number of overlooked samples being prepared. An improvement in the presentation was suggested by subsequent research [112, 134]. Extending the NCF model to include cross-regional ideas, he [92, 166] Sections or columns of the association grid can be substituted for the one-hot identifier to retain the user's achieve better outcomes, according to Xue et al [184] and Zhang et al [195].

Device for In-Depth Factorization. For factorization machines and MLPs, DeepFM [47] is an all-in-one paradigm. Deep learning model and low-order factorization machine interactions are used to highlight correlations that are of interest to the user. You may learn more about FM's use of expansion and external item activities in [119] by looking at Equation (1). MLP's Deep design and non-straight actuations are utilised to highlight in-demand engagements. How to link MLP and FM is illuminated by extensive networks. The wide section has been replaced with a neural knowledge of the factorization machine. A wide and deep model need long-winded highlights, but DeepFM does not. Fig. 2b depicts how

ISSN No:-2456-2165

Deep FM was put together. DeepFM x offers (u,i) sets of m-field data There are yields for FM and MLP, which are referred to as yF M (x) and MLP (x). It was determined that the forecast was correct because of this.

$$\hat{r}_{ui}=\sigma(yFM(x)+yMLP(x)) \quad (3)$$

where σ(·) is the activation function.

- It was proposed by Lian et al.[93] that a "Deep factorization" be used to communicate and implicitly emphasise partnerships in subsequent development. A densely packed network of connections reveals the specific high-demand spotlight companies. Replacement of the third cooperations with Mcp and regularisation using dropout and clump standards were suggested by He et al. [54].
- MLP-based Feature Learning for Feature Recognition. Given the fact that CNNs and RNNs are more expressive, using MLPs is incredibly direct and extremely effective, ignoring the fact that now it generally won't be nearly as expressive. Learning that is both broad and deep. Relapse and classification concerns may be addressed by this overall approach (shown in Figure 3a), however it was first designed for use in Google Play app recommendation [20]. As a single layer perceptron, the general educational segment is a summation of direct models. Perceptrons are used in the Deep learning section. Because these two learning methods are combined, the recommender is able to capture both recollection and hypothesis. The broad learning aspect of remembrance focuses on the capacity to recall the most important details from long-term records. During this period, the Deep Learning component is able to generate additional conjecture by developing broader and more theoretical depictions. This model may be used to improve precision as well as provide a wider range of options.

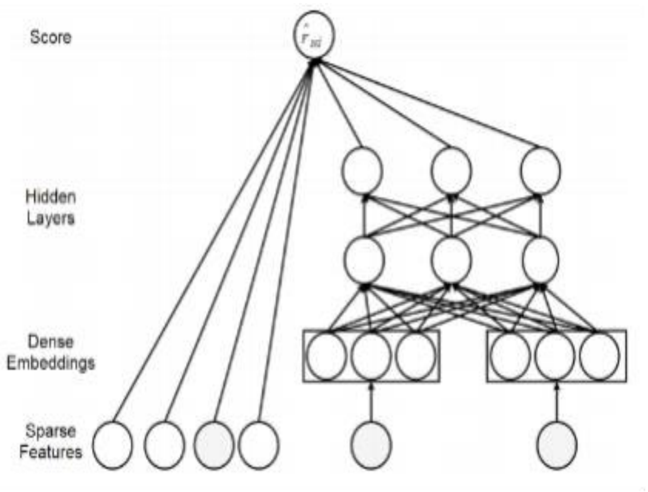
If you want to use WTwide as an official definition of the wide learning process, you may look up the model parameters at this link. The input x,x(x) is the concatenated feature set that includes the raw input feature x and the morphed (e.g. cross product transformation) feature x. (x). By each layer of something like the deep neural component, f () is the activation function, which is defined as W(l) deep a (l) + b (neural). Deep weight and bias are defined by the parameters W (l) and b (l). Combining these two learning models yields the broad and deep learning model. [210]:

$$P(\hat{r}_{ui} = 1|x) = \sigma(W^T_{wide} \{x,\varphi(x)\} + W^T_{deep} a^{(lf)} + bias) \quad (4)$$

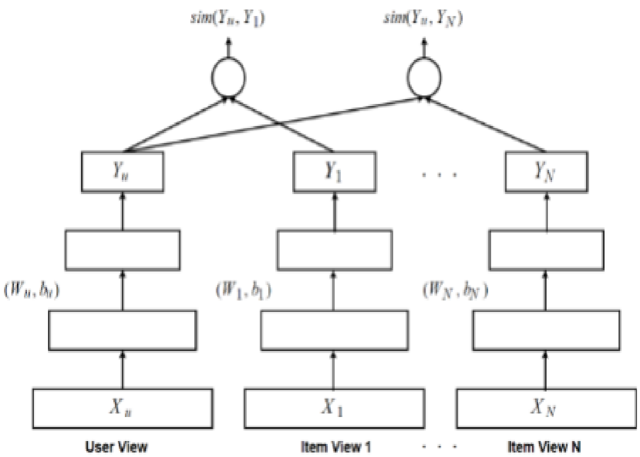
where σ(·) is the sigmoid function, $\hat{r}_{ui}$ is the binary rating label, a(lf) is the final activation. This joint model is optimized with stochastic back-propagation (follow-the-regularized-leader algorithm). Recommending list is generated based on the predicted scores.[210]

- Chen et al. [13] developed a privately associated wide and deep learning model for large-scale mechanical level recommendation problems by widening this model. It employs a productive privately attached network to replace the Deep learning section, which significantly decreases the showing time. The selection of characteristics for broad and deep learning is a major advancement in conveying broad and deep learning. In other words, the framework should be capable of determining whether highlights are maintained or summarised. Additionally, the cross-item adjustment must be physically planned. These preparatory procedures will have a substantial influence on the model's usefulness. The previously discussed Deep factorization-based approach can help alleviate some of the strain associated with inclusion design.
- Researchers at Covington et al. investigated the application of MLP in the production of YouTube propositions. As a result of this strategy, suggestions are broken down into two distinct phases: discovering new talent and positioning yourself for competition. Hundreds of video corpora are extracted by the developing age network. Based on the evaluations of the up-and-comers' closest neighbours, the positioning network provides a top-n list (a few). To be effective in today's environment, recommendation models must be flexible in their construction (e.g., able to be modified, standardised, and combined). MLPs were used by Alashkar et al. [2] to create a model for aesthetics recommendation. Using two identical MLPs, this research shows marked models and mastered runs separately. The bounds of these two or more networks are simultaneously refreshed by lowering the differences between their yields. Master information may be used to regulate the proposal figure's lesson in an MLP system. Despite the fact that mastery requires a huge amount of human work, this statement could not be more true..
- Collaborative Metrics-Based Education (CML). CML [60] substitutes Euclidean distance for the dab result of MF since the speck item does not satisfy the triangular imbalance of distance work. The user and item embeddings are discovered by magnifying the distance between users and their disliked things and reducing the gap between users and their preferred things. MLP is used in CML to extract representations from thin features like as text, images, and labels.

Recommendation using a Semantic Deep Structured Model. The Deep Structured Semantic Model (DSSM) [65] is a deep neural network for learning semantic representations of substances and assessing their semantic similarity in a typical constant semantic space. It is often used in data recovery areas and is surprisingly economical for top-n ideas [39, 182]. With cosine work, DSSM projects diverse items into a typical low-dimensional space and detects their similarities. MLP is a component of essential DSSM, which is why we included it in this section. Notably, additional neural layers, such as convolutional and max-pooling layers, may also be efficiently implemented into DSSM..



(a) Wide & Deep Learning;



(b) Multi-View Deep Neural Network.

Deep Semantic Similarity based Personalized Recommendation (DSPR) [182] is a tag-mindful customized recommender where every user xu and thing xi are addressed by label comments and planned into a typical tag space. Cosine likeness sim(u,i) are applied to choose the significance of things and users (or user's inclination over the thing). The misfortune capacity of DSPR is characterized as follows:

$$\mathcal{L} = -\sum_{(u,i*)}\left[\log\left(e^{\text{sim}\,(u,i*)}\right) - \log\left(\sum_{(u,i^-)\in D^-} e^{\text{sim}\,(u,i^-)}\right)\right] \quad (5)$$

where (u,i−) are negative examples which are haphazardly examined from the negative user thing sets. The creators. [183] further developed DSPR utilizing autoencoder to take in low-dimensional portrayals from user/thing profiles.

Multi-View Deep Neural Network (MV-DNN) [39] is intended for cross space recommendation. It regards users as the turn see and every space (assume we have Z areas) as helper see. Evidently, there are Z similitude scores for Z user space sets. Figure 3b shows the construction of MV-DNN. The misfortune capacity of MV-DNN is characterized as:

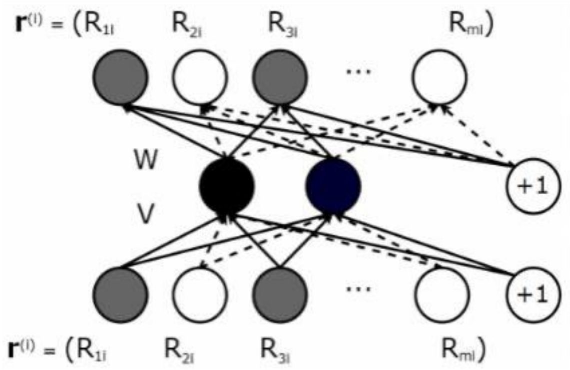
$$\mathcal{L} = \underset{\theta}{\text{argmin}} \sum_{j=1}^{Z} \frac{\exp\left(\gamma \cdot \text{cosine}\left(Y_u, Y_{a,j}\right)\right)}{\sum_{X' \in R^{da}} \exp\left(\gamma \cdot \text{cosine}\left(Y_u, f_a(X')\right)\right)} \quad (6)$$

where θ is the model boundaries, γ is the smoothing factor, Yu is the yield of user see, and is the record of dynamic view.Rda is the information space of view a. MV-DNN is fit for increasing to numerous spaces. In any case, it is in view of the theory that users who have comparative preferences for one space ought to have comparable preferences for other domains.Intuitively, this suspicion may be nonsensical by and large. In this manner, we ought to have some starter information on the connections across various spaces to benefit as much as possible from MV-DNN.

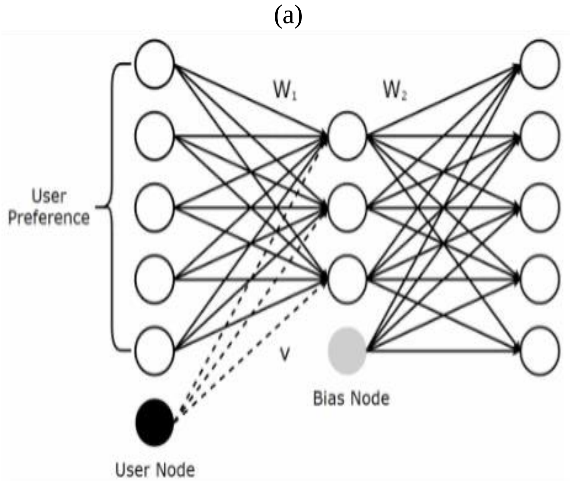
G. Autoencoder based Recommendation

There exist two general methods of applying autoencoder torecommender framework: (1) utilizing autoencoder to learnlower-dimensional element portrayals at the bottleneck layer;or (2) filling the spaces of the communication lattice straight forwardly in the recreation layer. Practically all the autoencoder variations, for example, denoising autoencoder,variational autoencoder, contractive autoencoder andminimized autoencoder can be applied to proposal tasks. Table 3 sums up the recommendation models dependent on the sorts of autoencoder being used.

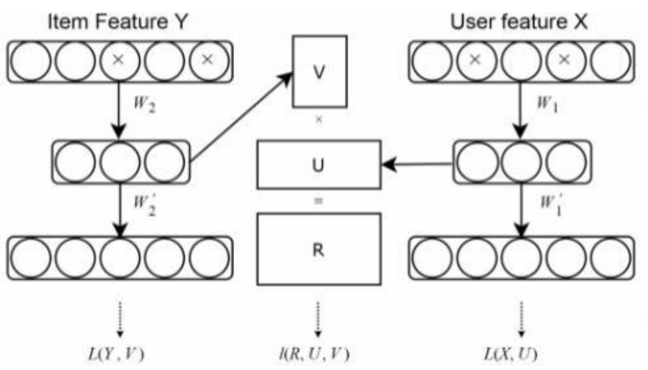
Autoencoder based Collaborative Filtering. One of the fruitful applications is to consider the synergistic separation according to the Autoencoder point of view.



(a)



(b)



(c)

(a) Item based AutoRec; (b) Collaborative denoising autoencoder; (c) Deep collaborative filtering framework.

AutoRec [125] intends to rebuild user halfway vectors r(u) or thin fractional vectors r(i) in the yield layer. Clearly, it comes in two flavours: item-based AutoRec (I-AutoRec) and user-based AutoRec (U-AutoRec), which correspond to the two types of information sources. We just offer I-AutoRec here, as U-AutoRec may be easily deduced. I-design AutoRec's is seen in Figure 4a.. Given information r(i), the remaking is: $h(\mathbf{r}^{(i)}; \theta) = f(W \cdot g(V \cdot \mathbf{r}^{(i)} + \mu) + b)$ where f (·) and g(·) are the actuation capacities, boundary $\theta = \{W, V, \mu, b\}$.

Vanilla/Denoising AE	Variational AE	Contractive AE	Marginalized AE
[114, 125, 136, 137, 159, 177] [70, 116, 170, 171, 188]	[19, 89, 94]	[196]	[88]

Table 3: Summary of four autoencoder based recommendation models

The target capacity of I-AutoRec is planned as follows:

$$\underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| \mathbf{r}^{(i)} - h\left(\mathbf{r}^{(i)}; \theta\right) \right\|_{O}^{2} + \lambda \cdot \text{reg} \quad (7)$$

Here ||·||2O implies that it just considers noticed ratings.The target capacity can be improved by versatile engendering (joins quicker and produces tantamount outcomes) or L- BFGS (Limited-memory Broyden Fletcher Goldfarb Shanno calculation). There are four significant focuses about AutoRec that value seeing before arrangement: (1) I-AutoRec performs better compared to U-AutoRec, which might be because of the greater difference of user incompletely noticed vectors. (2) Different mix of initiation capacities f (·) and g(·) will impact the presentation significantly. (3) Increasing the secret unit size modestly will work on the outcome as extending the secret layer dimensionality gives AutoRec greater ability to display the qualities of the information. (4)

Increasing the number of layers in a Deep network design might result in a modest increase in performance. With CFN [136, 137], AutoRec gains two more advantages: Denoising procedures are shown, which creates CFN more hearty; in addition, side data such as user profiles and things portrayals are included to lessen the frigid starting affect. In addition to fractional noticed vectors, CFN makes a contribution via its two variants: I-CFN and U-CFN, which treat r(i) and r(u) as independent pieces of information. Commotion is used as a solid regularizer to more probable configurations that are lacking parts. To cope with degraded information, the inventors proposed three widely used debasement methods: Gaussian clamour, hiding commotion, and salt-and-pepper disturbance. CFN's expansion also incorporates data from other sources. Despite this, CFN injects side data into each layer rather than merely coordinating it in the main one. Re-creation takes on a new meaning in this way.:

$$h(\{\tilde{\mathbf{r}}^{(i)}, s_i\}) = f(W_2 \cdot \{g(W_1 \cdot \{\mathbf{r}^{(i)}, s_i\} + \mu), s_i\} + b) \quad (8)$$

where si is side information, $\{\tilde{r}(i), s_i\}$ indicates the concatenation of $\tilde{r}(i)$ and si. Incorporating side information improves the prediction accuracy, speeds up the training process and enables the model to be more robust.

Shared Denoising Auto-Encoder (CDAE).The three models audited prior are mostly intended for rating forecast, while CDAE [177] is basically utilized for positioning expectation. The contribution of CDAE is the user's incompletely noticed understood input r(u)pref .The passage esteem is 1 if the user prefers the film, in any case 0. It can likewise be viewed as an inclination vector which mirrors user's inclinations to things. Figure 4b outlines the construction of CDAE.The contribution of CDAE is tainted by Gaussian commotion. The ruined info $\tilde{r}$(u)pref is drawn from a contingent Gaussian conveyance p($\tilde{r}$(u)pref | r(u)pref). The recreation is characterized as:

$$h(\tilde{\mathbf{r}}^{(u)}_{pref}) = f\left(W_2 \cdot g(W_1 \cdot \tilde{\mathbf{r}}^{(u)}_{pref} + V_u + b_1) + b_2\right) \quad (9)$$

where Vu ∈ R K signifies the weight network for the user hub (see figure 4b). This weight network is novel for every user and has a critical effect on the model execution. Boundaries of CDAE are likewise scholarly by limiting the recreation mistake:

$$\underset{W_1, W_2, V, b_1, b_2}{\operatorname{argmin}} \frac{1}{M} \sum_{u=1}^{M} \mathrm{E}_{p\left(\tilde{\mathbf{r}}^{(u)}_{pref} | \mathbf{r}^{(u)}_{pref}\right)} \left[\ell\left(\tilde{\mathbf{r}}^{(u)}_{pref}, h\left(\tilde{\mathbf{r}}^{(u)}_{pref}\right)\right) \right] + \lambda \cdot \text{reg}$$
(10)

Square or logistic loss may be used as loss function l().

SGD is used to update CDAE's parameters initially. A negative sampling strategy was presented in order to sample just a tiny fraction of the negative set (things with which the user has not interacted), greatly reducing the time complexity however, as the authors claimed, without affecting ranking accuracy.

A equations autoencoder for recommendation using understood information was suggested by Muli-VAE and Multi-DAE [94], exhibiting favoured execution over CDAE. Creators used a systematic Bayesian strategy for border evaluation and demonstrated perfect results compared to commonly used probability capabilities.

It's the first model based on an autoencoder that we've found, and it's called Autoencoder-Based Cooperative Filtering (ACF) [114]. Instead of using the initial half-noticed

ISSN No:-2456-2165

vectors, it deteriorates them by whole number assessments. Each r(i) is divided into five half-vectors if the resources (hr falls within the range of [1-5]. The expenditure capacity of ACF, like that of AutoRec and CFN, aims to reduce the mean squared error. However, there are two flaws with ACF: Non-number evaluations are ignored, and incomplete vectors are disintegrated, resulting in a worse prediction accuracy because of the lack of information.

Autoencoder-Based Feature Representation Learning. Using an autoencoder is a great way to learn about the components of an object. The same is true for recommender frameworks, where user/thing content highlights may be included.

"Community-Based Deep Learning" (CDL). CDL [159] is a linear Bayesian model that includes SDAE into probable network factorization as a stacked denoising autoencoder. Bayesian Deep learning [161] suggested an overall structure consisting of two firmly hinged parts: discernment portion (Deep convolutional neural network) and an explicit segment for perfectly joining the Optimization algorithms and recommendation model. A probabilistic comprehension of ordinal SDAE constitutes the insight phase of CDL; the errand explicit section is represented by PMF. CDL is able to adapt the influence of side data with collaboration history because to this tight mix. CDL's generating cycle is outlined here:

- For each layer l of the SDAE: (a) For each column n of weight matrix Wl , draw Wl,∗n ∼ N (0, λ−1wIDl); (b) Draw the bias vector bl ∼ N (0, λ−1w IDl); (c) For each row i of Xl , draw Xl,i∗ ∼ N (σ(Xl−1,i∗Wl +bl),λ−1s IDl).
- For each item i: (a) Draw a clean input Xc,i∗ ∼ N(XL,i∗, λ−1n IIi); (b) Draw a latent offset vector εi ∼N(0, λ−1v ID) and set the latent item vector: Vi= εi + XTL /2 ,i∗ .
- Draw a latent user vector for each user u, Uu ∼ N (0,λ−1uID).
- Draw a rating rui for each user-item pair (u,i), rui ∼N (UTu Vi ,C−1ui).

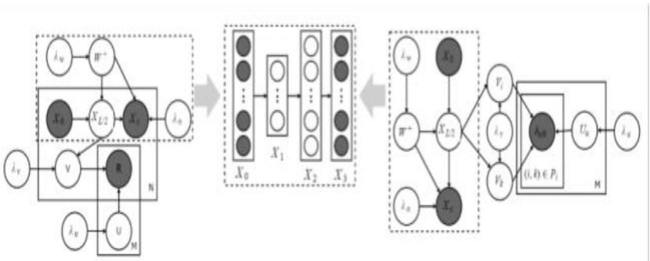


Fig. 5: Graphical model of collaborative deep learning (left) and collaborative deep ranking (right).

Weight array and dispositions matrix Wl and bl for layer l are addressed by the layer-specific Xl. There are six hyper-parameters (w, n, v, and u) and a certainty boundary (Cui) that may be used to determine the confidence of perceptions. CDL's visual model may be seen in Figure 5(left) of this document. EM-style calculations were employed by the designers to become adept at working with the boundaries. In every loop, it fixes U and V immediately, and then W and b are refreshed as a result of this repair. To avoid straying too close to the ideal, the authors devised an inspection-based computation [161].

Deep Ranking in a Team Environment (CDR). CDR [188] is designed specifically for the top-n proposal in a paired system. There have been a few studies that show that a pair paradigm is better at determining a record's age. Additionally, CDR outperforms CDL when it comes to predicting position. Construction of CDR may be shown in Figure 5. CDR's first and second generating interaction phases are identical to CDL's. As a result, the 3rd and 4th stages are no longer necessary:

Framework for Deeply Collaborative Filtering. With it, deep learning and synergistic separation models may be linked together as a single system. Deep element learning algorithms may be used to build half-and-half communitarian models using this approach. This general framework may be observed in the previously stated works, such as [153, 159, 167]. These are the official characteristics of the Deep cooperative filtering system:

$$\arg\min_{U,V} \ell(R,U,V) + \beta(\| U \|_F^2 + \| V \|_F^2) + \gamma\mathcal{L}(X,U) + \delta\mathcal{L}(Y,V) \quad (11)$$

This is the model's shortcoming, where l() is a compromise boundary to modify the affects of these three segments, and X and Y are side data. Connecting side data with idle variables is facilitated by using L(X,U) and L(Y,V) as pivots to link Deep Learning with community-oriented modelling frameworks. A denoising autoencoder based collective sifting model was presented on top of this system (mDA-CF). When compared to CDL, the mDA-CF researches an even more computationally efficient form of autoencoder: the underestimated denoising autoencoder. By underestimating out the undermined input, mDA-CF is more flexible than CDL, which reduces calculation costs for searching through acceptable debased adaption of contribution. Items and users are also included into mDA-CF, but CDL just considers the effects of things highlighted.

For the adaptation of object highlight representations, AutoSVD++ [196] makes use of contractive autoencoder [122] and then coordinates this data into SVD++ [79], an illustrative recommendation model. The model under consideration imposes the following advantages: Contractive autoencoder, in contrast to other autoencoder variants, captures minute information fluctuations; (2) it reveals the recognised criticism to further increase precision; (3) a productive preparation calculation is designed to save preparation time. contractive.HRCD [170, 171] is a mixture shared model dependent on autoencoder and timeSVD++ [80]. It is a period mindful model which utilizes SDAE to take in thing portrayals from crude highlights and targets tackling the chilly thing issue.

H. Convolutional Neural Networks based Recommendationome

Convolution Convolution and pooling procedures in neural networks enable them to analyse large amounts of unstructured multimedia input. It is common to use CNNs for feature extraction in CNN-based recommendation models

Learning about Feature Representation by Using CNNs. CNNs may be used to emphasise a wide range of information, including images, text, sound, video, and so on, from a variety of sources.

Image Feature Extraction Using CNNs. Research by Wang et al. [165] examined the effects of visual highlights on POI recommendation and developed a visual substance-enhanced POI recommendation framework.(VPOI).

VPOI gets CNNs and removes all of the image highlights from them. Based on PMF, the proposed model investigates the cooperation between (1) visual substance and inactive user factors; and (2) visual substance and inert area factors. Using visual data (for example, photos of food and commodities from the café) in a proposal for an eatery was overused by Chu et al. [25]. MF, BPRMF and FM are testing their exhibitions with CNN's removal of visual highlights and content depiction. Visual data seems to have a limited impact on the presentation, according to the results. A visual Bayesian customised positioning (VBPR) computation was designed by combining visual highlights (acquired using CNNs) with network factorization. [50] By examining the user's design consciousness and the progression of visual components that users consider while making decisions, he et al. [49] expanded VBPR. When it comes to clothing recommendations, Yu et al. [191] used CNNs to get a better understanding of what makes a piece fashionable and what doesn't. A CNN-based personalised label recommendation algorithm was suggested by Nguyen et al. [110]. Convolution and max-pooling layers are used to extract visual highlights from a collection of images. Personalized suggestions are generated using user data. The goal of BPR is to increase the contrasts between relevant and irrelevant labels in this network. CNN-based image suggestion was suggested by Lei et al [84]. This network consists of two CNNs that are used to learn how to depict an image, and an MLP that is used to demonstrate user preferences. It compares two images (one that the user clearly likes and one that the user abhors) against one another. T (User Ut, positive picture I+t, negative picture It) is the triad of data needed to prepare the experiment.

Expecting that the distance among user and positive picture D(π(Ut),φ(I+t)) ought to be nearer than the distance among user and negative pictures D(π(Ut),φ(I−t)), where D(·) is the distance metric (for example Euclidean distance). ConTagNet [118] is a setting mindful tag recommender framework. The picture highlights are learned by CNNs. The setting portrayals are handled by a two layers completely associated feedforward neural network. The yields of two neural networks are connected and taken care of into a softmax function to foresee the likelihood of applicant labels.

CNNs for Text Feature Extraction. Deep CoNN [202] receives two equal CNNs to demonstrate user practices and properties from audit messages. This model mitigates the sparsity issue and upgrades the model interpretability by misusing rich semantic portrayals of survey messages with CNNs. It uses a word installing method to plan the survey messages into a lower-dimensional semantic space just as keep the words groupings data

The removed audit portrayals then, at that point go through a convolutional layer with various bits, a max pooling layer, and a full-associated layer successively. The yield of the user network xu and thin network xi are at long last connected as the contribution of the forecast layer where the factorization machine is applied to catch their communications for rating expectation. Catherine et al. [11] referenced that DeepCoNN possibly functions admirably when the survey text composed by the objective user for the objective thing is accessible at test time, which is outlandish. Accordingly, they broadened Graph CNNs for Recommendation.Graph convolutional Networks is an integral asset for non-Euclidean information for example, informal communities, information charts, protein-communication networks, and so forth [77]. Corporations in the proposal region can likewise be seen as such an organized dataset (bipartite chart). Subsequently, it can likewise be applied to recommendation undertakings. For instance, Berg et al. [6] proposed considering the recommendation issue as a connection forecast task with diagram CNNs. This system makes it simple to coordinate user/thing side data, for example, interpersonal networks and thin connections into recommendation models. Ying et al. [190] proposed utilizing diagram CNNs for proposals in Pinterest10. This model creates thin embeddings from both diagram structures and highlight data with irregular walk and chart CNNs, and is reasonable for exceptionally enormous scope web recommenders. The proposed model has been conveyed in Pinterest to address an assortment of true recommendation assignments.

it by acquainting an idle layer with the objective user target- thing pair. This model doesn't get to the surveys during approval/test can in any case stay great precision. Shen et al. [130] fabricated an e-learning assets recommendation model. It utilizes CNNs to separate thin highlights from text data of learning assets like presentation and content of learning material, and follows a similar technique of [153] to perform recommendations. ConvMF [75] consolidates CNNs with PMF likewise as CDL. CDL utilizes autoencoder to get familiar with the thing highlighting portrayals, while ConvMF utilizes CNNs to learn significant level thingportrayals. The principle benefit of ConvMF over CDL is that CNNs can catch more precise context oriented data of things by means of word installing and convolutional pieces. Tuan et al. [148] proposed utilizing CNNs to learn highlight portrayals structure thing content data (e.g., name, depictions, identifier and classification) to improve the precision of meeting based recommendation

CNNs based Collaborative filtering. Straightforwardly applying CNNs to vanilla collective sifting is likewise feasible. For the model, He et al. [51] proposed utilizing CNNs to further develop NCF and introduced the ConvNCF. It utilizes external items rather than speck items to display the user's collaboration designs. CNNs are applied over the consequence of external items and could catch the high-request relationships among embeddings measurements. Tang et al. [143] introduced successive recommendation (with user identifier) with CNNs, where two CNNs (progressive and vertical) are utilized to demonstrate the association level consecutive examples and skip practices for arrangement mindful proposal

Graph CNNs for Recommendation. Graph convolutional Networks is an integral asset for non-Euclidean information for example, informal communities, information charts, protein-communication networks, and so forth [77]. Corporations in the proposal region can likewise be seen as such an organized dataset (bipartite chart). Subsequently, it can likewise be applied to recommendation undertakings. For instance, Berg et al. [6] proposed considering the recommendation issue as a connection forecast task with diagram CNNs. This system makes it simple to coordinate user/thing side data, for example, interpersonal networks and thin connections into recommendation models. Ying et al. [190] proposed utilizing diagram CNNs for proposals in Pinterest10. This model creates thin embeddings from both diagram structures and highlight data with irregular walk and chart CNNs, and is reasonable for exceptionally enormous scope web recommenders. The proposed model has been conveyed in Pinterest to address an assortment of true recommendation assignments.

Recurrent Neural Networks based Recommendation

RNNs are amazingly appropriate for successive information preparation. Thusly, it turns into a characteristic decision for managing the worldly elements of corporations and successive examples of user practices, just as side data with consecutive signals, like writings, sound, and so on

Session-based Recommendation without User Identifier. In numerous certifiable applications or sites, the framework typically doesn't trouble users to sign in so it has no admittance to user's identifier and her extensive stretch utilization propensities or long haul interests. Nonetheless,the meeting or treat instruments empowers those frameworks to get user's momentary inclinations. This is a moderately neglected undertaking in recommender frameworks because of the outrageous sparsity of preparing information. Ongoing headways have shown the viability of RNNs in addressing this issue [56, 142, 176].

GRU4Rec. Hidasi et al. [56] proposed a meeting based recommendation model, GRU4Rec, based GRU (displayed in Figure 6a). The info is the real condition of meeting with 1-of-N encoding, where N is the quantity of things. The network will be 1 if the comparing thing is dynamic in this meeting, in any case 0. The yield is the probability of being the following in the meeting for everything. To productively prepare the proposed structure, the creators proposed a meeting equal scaled down clusters calculation and an examining strategy for yield. The positioning misfortune which is additionally begat TOP1 and has the accompanying structure:

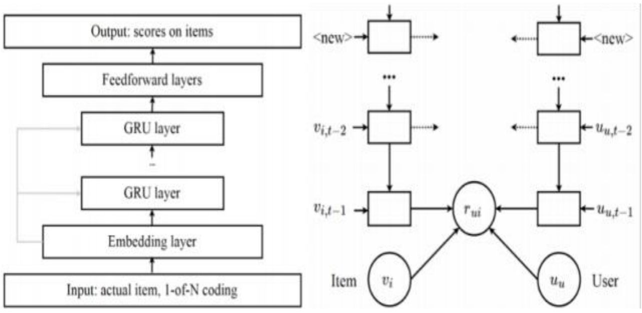
$$\mathcal{L}_s = \frac{1}{S}\sum_{j=1}^{S}\sigma\left(\hat{r}_{sj} - \hat{r}_{si}\right) + \sigma\left(\hat{r}_{sj}^2\right) \quad (12)$$

where S is the example size, r̂si and r̂sj are the scores on bad thing I and positive thing j at meeting s, σ is the calculated sigmoid capacity. The last term is utilized as a regularization. Note that, BPR misfortune is likewise practical. A new work [55] tracked down that the first TOP1 misfortune and BPR misfortune characterized in [56] super from the slope evaporating issue, thus, two novel misfortune capacities: TOP1-max and BPR-max are proposed.

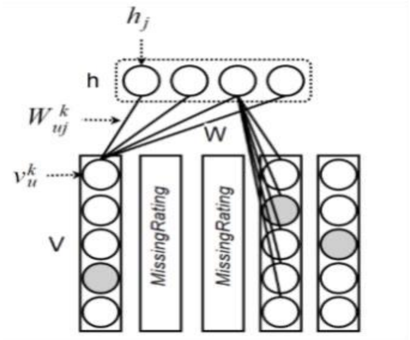
The later paper [142] suggested many new strategies for working with this model: (1) expand snap groupings through succession preprocessing and dropout regularisation; (2) adapt to transient changes by pre-preparing with complete preparation data and tweaking the model with subsequent snap arrangements; (3) refine the model with favoured data via an educator model; and (4) use thing inserting to reduce the number of boundaries for faster calculation. Wu et al. [176] developed a recommendation model for a real online business site based on meetings. It makes use of basic RNNs to forecast what the user will purchase next based on their snap history. To keep computation costs down, it retains just a subset of the most recent states while exploding the more experienced states into a single historical state. This technique enables you to fine-tune the trade-off between computation costs and expectation precision. Adrana et al. [117] developed a multileveled recurrent neural network for recommendation based on meetings. When user identities are provided, this approach can manage both meeting and thoughtful recommendation. The three meeting-based models stated before do not consider any additional data. Two augmentations [57,132] demonstrate that side data has an effect on the quality of meeting suggestion.

Hidasi et al. [57] proposed an equal design for meeting-based recommendation that incorporates representations from character one-hot vectors, image inclusion vectors, and text highlight vectors. The yields of these three GRUs are weighted and included into a non-direct initiation to forecast the meeting's subsequent events. Smirnova et al. [132] developed a methodology for putting together recommender systems based on dependent RNNs. It incorporates configuration data into the info and yield layers. The trial results for these two models suggest that models that incorporate additional data outperform those that rely simply on authentic conversations. Despite the success of RNNs in meeting-based recommendation, Jannach et al. [68] demonstrated that a simple area technique may attain the same precision as GRU4Rec. Typically, consolidating the neighbourhood using RNNs tactics results in the best performance. This paper argues that a few baselines used in current research are not always legitimated or appropriately appraised. A more in-depth discussion is available in [103].

Recommendation in Sequence with User Identifier.Not at all like meeting-based recommenders, which often do not include user identities. The preceding investigations handle the subsequent proposal task through the use of user-recognizable pieces of evidence..



(a)



(b)

(a) recommendation based on serssion with RN network; Recurrent recommender network; (b) Restricted Boltzmann Machine based C F.

The Recurrent Recommender Network (RRN) [175] is a non-parametric recommendation model based on Recurrent Neural Networks (RNNs) (displayed in Figure 6b). It is capable of illustrating the cyclical advancement of objects and changes in user preferences through time. RRN illustrates dynamic user state uut and thing state vit using two LSTM networks as the structural square. Meanwhile, taking into account fixed qualities such as user long-term interests and thin static highlights, the model also combines the user's and thing's fixed idle credits: uu and vi. The projected rating of item j by user I at time t is as follows:

$$r\hat{}ui|t=f(uut,vit,uu,vi) \quad (13)$$

where uut and vit are learned from LSTM, uu and vi are learned by the standard matrix factorization. The optimization is to minimize the square error between predicted and actual rating values.

Wu et al. [174] further developed the RRNs model by displaying text surveys and appraisals all the while. Dissimilar to most content survey upgraded recommendation models [127, 202], this model plans to create audits with a person level LSTM network with user and thin idle states. The survey age errand can be seen as a helper assignment to work with rating expectation. This model can further develop the rating forecast exactness, yet can't produce intelligent and coherent audit messages. NRT [87] which will be presented in the accompanying content can produce clear audit tips. Jing et al. [73] proposed a performance learning structure to at the same time foresee the returning season of users and suggest things. The returning time forecast is roused by an endurance examination model intended for assessing the likelihood of endurance of patients. The creators changed this model by utilizing LSTM to appraise the returning season of customers. The thing proposal is additionally performed through LSTM from user's past meeting activities. Dissimilar to previously mentioned meetings put together recommendations which center with respect to suggesting in a similar meeting, this model means to give between meeting proposals. Li et al. [91] introduced a conduct serious model for successive proposals. This model comprises two segments: neural thing implanting and discriminatory practices learning.The later part comprises two LSTMs for meeting and inclination practices adapting individually. Christakopoulou et al. [24] planned an intelligent recommender with RNNs. The proposed structure expects to address two basic assignments in intuitive recommender: ask and react. RNNs are utilized to handle the two assignments: anticipate questions that the user may ask dependent on her new behaviors(e.g, watch occasion) and foresee the reactions. Donkers et al. [35] planned a novel kind of Gated Recurrent Unit to address singular users for next thing recommendation.

Feature Representation Learning with RNNs. For side data with consecutive examples, utilizing RNNs as the portrayal learning instrument is a prudent decision. Dai et al. [29] introduced a co-transformative idle model to catch the co-advancement nature of users' and things' dormant highlights. The collaborations among users and things assume a significant part in driving the progressions of user inclinations and thing status. To display the authentic corporations, the creator proposed utilizing RNNs to consequently take in portrayals of the impacts from float, advancement and co-development of user and thin highlights

For side data with consecutive examples, utilizing RNNs as the portrayal learning device is a prudent decision. Dai et al. [29] introduced a co-developmental inert model to catch the co-advancement nature of users' and things' dormant highlights. The associations among users and things assume a significant part in driving the progressions of user inclinations and thing status. To demonstrate the recorded associations, the creator proposed utilizing RNNs to consequently take in portrayals of the impacts from float, advancement and co-development of user and thin highlights

Bansal et al. [5] proposed utilizing GRUs to encode the content successions into an idle factor model. This mixture model tackles both warm-start and cold-start issues. Moreover, the creators embraced a performance regularizer to forestall overfitting and mitigate the sparsity of preparing information. The principal task is evaluating expectation while the assistant assignment is a meta-information (for example labels, kinds) forecast. Okura et al. [113] proposed utilizing GRUs to learn more expressive conglomeration for user perusing history (perused news), and suggest news stories with idle factor model.

The outcomes show a huge improvement contrasted and the customary word-based approach.The framework has been completely conveyed to online creation administrations and serving more than ten million novel users everyday.Li et al. [87] introduced a perform multiple tasks learning system, NRT, for anticipating evaluations just as producing literary tips for users all the while. The produced tips give compact ideas and expect user's experience also, sentiments on specific items. The rating expectation task is demonstrated by

ISSN No:-2456-2165

non-direct layers over thing and user idle elements U ∈ R ku ×M , V ∈ R kv ×M , where ku and kv (not really equivalent) are dormant factor measurements for users and things. The anticipated rating rui and two inactive factor lattices are taken care of into a GRU for tips age. Here, rui is utilized as setting data to choose the assumption of the produced tips. The perform multiple tasks learning system empowers the entire model to be prepared eciently in a start to finish worldview. Tune et al. [135] planned a transient DSSM model which coordinates RNNs into DSSM for proposal.

In light of customary DSSM, TDSSM supplant the left network with thin static highlights, and the right network with two sub-networks to display user static highlights (with MLP) and user transient highlights (with RNNs).

Restricted Boltzmann Machine based Recommendation Salakhutdinov et al. [123] proposed a confined Boltzmann machine based recommender (displayed in Figure 6c). Supposedly, it is the primary proposal model that is based on neural networks.

The apparent unit of RBM is restricted to twofold qualities, subsequently, the rating score is addressed in a one-hot vector to adjust to this limitation. For instance, [0,0,0,1,0] addresses that the user gives a rating score 4 to this thing. Let hj, j = 1, ..., F mean the secret units with fixed size F . Every user has a novel RBM with shared boundaries. Assume a user evaluated m films, the quantity of noticeable units is m, Let X be a K × m network where xiy = 1 if user u appraised film I as y and xyI = 0 in any case. Then, at that point:

$$p(v_i^y = 1 \mid h) = \frac{\exp\left(b_i^y + \sum_{j=1}^{F} h_j W_{ij}^y\right)}{\sum_{l=1}^{K} \exp\left(b_i^l + \sum_{j=1}^{F} h_j W_{ij}^l\right)}, \; p(h_j = 1 \mid X) = \sigma\left(b_j + \sum_{i=1}^{m}\sum_{y=1}^{K} x_i^y W_{ij}^y\right)$$
(14)

where Wyij addresses the load on the association between the rating y of film I and the secret unit j, byi is the predisposition of rating y for film I, bj is the predisposition of covered up unit j. RBM isn't manageable, however the boundaries can be learned by means of the Contrastive Divergence (CD) calculation [45]. The authors further proposed utilizing a restrictive RBM to join the implied criticism. The embodiment here is that users certainly tell their inclinations by giving evaluations, paying little mind to how they rate things.

The above RBM-CF is user based where a given user's appraising is cinched on the apparent layer. Comparably, we can without much of a stretch plan a thing based RBM-CF in the event that we clip a given thing's appraisal on the apparent layer. Georgiev et al. [42] proposed to consolidate the user based and thing based RBM-CF in a brought together system. For the situation, the apparent units are resolved both by user and thing covered up units. Liu et al. [100] planned a cross breed RBM-CF which fuses thing highlights (thing classifications). This model is likewise founded on restrictive RBM. There are two differences between this half breed model with the contingent RBM-CF with verifiable criticism: (1) the restrictive layer here is displayed with the parallel thing sorts; (2) the contingent layer influences both the secret layer and the noticeable layer with diverse associated loads.

Vanilla Attention	Co-Attention
[14, 44, 70, 90, 99, 101, 127, 145, 169, 189]	[62, 146, 193, 194, 205]

Table 4: Categories of neural attention based recommendation models.

Neural Attention based Recommendation

Consideration system is spurred by human visual attention. For instance, individuals just need to zero in on explicit pieces of the visual contributions to comprehend or remember them. Consideration component is equipped for sifting through the uninformative highlights from crude information sources and lessen the side effects of loud information. It is an instinctive, however viable strategy and has collected significant consideration over the new years across regions, for example, PC vision [3], normal language preparation [104, 155] and discourse acknowledgment [22, 23]. Neural consideration can not just be utilized in combination with MLP, CNNs and RNNs, yet in addition address a few errands autonomously [155]. Incorporating consideration instruments into RNNs empowers the RNNs to deal with long and loud data sources [23]. In spite of the fact that LSTM can take care of the long memory issue hypothetically, it is as yet risky when managing long-range conditions. Consideration system gives a superior arrangement and assists the network with better remembered inputs. Consideration based CNNs are fit for catching the most instructive components of the sources of info [127]. By applying a consideration component to the recommender framework, one could use a consideration system to sift through uninformative substance and select the most delegate things [14] while giving great interpretability. Although the neural consideration system isn't by and large an independent Deep neural strategy, it is as yet beneficial to talk about it independently because of its inescapable use.

Consideration model figures out how to take care of the contribution with consideration scores. Figuring the consideration scores lives at the core of neural consideration models.

In view of the way for figuring the consideration scores, we group the neural consideration models into (1) standard vanilla consideration and (2) co-consideration. Vanilla consideration uses a defined setting vector to figure out how to join in while co-consideration is worried about taking in consideration loads from two-successions. Self-consideration is an extraordinary instance of co- consideration. Late works [14, 44, 127] show the ability of consideration components in improving proposal execution. Table 4 sums up the consideration based proposal models

Recommendation with Vanilla Attention Recommendation with Co-Attention Chen et al. [14] proposed a mindful synergistic filtering model by presenting a two-level consideration component to the inert factor model. It comprises thing level and segment level consideration. The thing level consideration is utilized to choose the most delegate things to portray users. The part level consideration means to catch the most instructive highlights from interactive media assistant data for every user. Tay et al. [145] proposed a memory-based consideration for community oriented measurement learning. It presents an inert connection vector learned through thoughtfulness regarding CML. Jhamb et al. [70] proposed utilizing a consideration instrument to work on the exhibition of autoencoder based CF. Liu et al. [99] proposed a momentary consideration and memory need based model, in which both long and transient user interests are integrated for meeting based recommendation. Ying et al. [189] proposed a progressive consideration model for consecutive recommendation. Two consideration networks are utilized to display user long haul and transient interests. Acquainting a consideration system with RNNs could altogether work on their exhibition. Li et al. [90] proposed such a consideration based LSTM model for hashtag recommendation. This work takes the upsides of both RNNs and consideration components to catch the consecutive property and perceive the useful words from microblog posts. Loyala et al. [101] proposed an encoder-decoder engineering with consideration for user meetings and plans demonstrating. This model comprises two RNNs and could catch the progress consistencies in a more expressive manner.

For recommender projects, vanilla attention may also be used in conjunction with CNNs. According to Gong et al. [44], a consideration-based CNNs framework for hashtag suggestion in microblogs was presented by the researchers. It views hashtag suggestions as a problem of characterisation with several names. The suggested model includes a global and a local consideration channel. Convolution channels and max-pooling layers make up the global channel. The global channel's contribution encodes every word. Nearby consideration channels have consideration layers with a set window size and limit for selecting suitable phrases (known as trigger words in this work). Because of this, the subsequent layers only use trigger words. In a recent study [127], Seo et al. used two neural networks similar to those described in [44] (minus the last two layers) to gather highlights from user and thin audit messages and predict rating scores with dab items in the final layer of the neural network architecture. Using CNNs and consideration, Wang et al. [169] developed a unified model for article suggestion that takes into account the various ways in which editors make decisions and how those differences manifest themselves in their work.

According to Zhang et al. [194], an integrated approach called AttRec enhances the execution of successive proposals by leveraging the power of both soul and metric learning simultaneously. As a result, it uses self-regard to take in the user's short-term goals from her new collaborations and reaps the advantages of metric-based learning. Self-consideration was suggested by Zhou et al. [205] for the presentation of user diverse behaviour. As far as recommendation jobs go, self-consideration is a simple but practical component that has proven superior performance than CNNs and RNNs. We believe that everyone has the potential to replace a variety of complicated neuronal models, and more study is warranted. Audit-based recommendations with multi-pointer co-consideration were presented by Tay et al. (146). Using both user and item audits, the model may employ co-consideration to choose data surveys. Using both visual and written data, Zhang et al. [193] suggested a hashtags model - based approach that leverages co-attention. A neural co-consideration approach for tailored placement errands with meta-way was suggested by Shi et al. [62].

Neural AutoRegressive based Recommendation

In order to estimate the document inclination on the boundary using RBM-CF, we often apply the Contrastive Divergence calculation [81], since RBM is not adjustable. It is claimed that NADE, as opposed to RBM, is a controlled dissemination assessor that offers an advantageous option. Zheng et al. [204] proposed a cooperative differentiating system based on NADE after being encouraged by RBM-CF (CF-NADE). Dispersion of user feedback may be modelled using CF-NADE. This section demonstrates the CF-NADE in action via a step-by-step guide. For example, let's say we had 4 films: M1 (rating 4); M2 (rating 2); M3, and M4 (rating 3). (rating 4). (The rating is 5) Rules of the CF-NADE chain is used to estimate the joint probability of the rating vector r: p(r) = ÎD i=1 p(rmoi |rmoi), where D is the number of items the user has assessed, o denotes the D-tuple in the stages (1, 2,...,D),

Preferably, the request for films should coincide with the time stamps associated with assessments. Nonetheless, careful examination demonstrates that uneven sketching also produces excellent displays. Additionally, this model may be connected to a Deep model. Zheng et al. [203] therefore advocated combining several criticisms in order to overcome the sparsity issue using rating networks. Du et al. [36] enhanced this model further by including a user-thinking co-autoregressive technique that results in increased performance in both rating evaluation and tailored placement assignments.

Recommendation through Deep Reinforcement Learning The majority of proposal models treat the suggestion interaction as a static cycle, which makes it impossible to anticipate and respond optimally to users' worldly expectations. DRL has just begun to garner consideration [21, 107, 168, 198–200] when submitting tailored proposals. Zhao et al. [199] suggested a DRL structure, DEERS, for a proposal that received both positive and negative feedback in a subsequent collaboration. Zhao et al. [198] explored the page-wise suggestion scenario using DRL; they proposed that the proposed system DeepPage may adaptively improve a page of items based on the user's ongoing behaviours. Zheng et al. [200] presented a news recommendation framework, DRN, using DRL to address the following three challenges: (1) rapid changes in information content and user preference; (2) fusing users' return paths (to assistance); and (3) increasing the diversity of suggestions. Chen et al. [16] developed a robust Deep Q-learning calculation to handle the unstable prize evaluation issue through the use of two systems: specified testing replay and hypothesised redeemed reward. Choi et al. [21] advocated using RL and bi-grouping to address the cool starting issue. Munemasa et al. [107] advocated that shop suggestions be made using DRL.

Reinforcement For instance, logical suggestion execution in real-world applications is an example of a learning strategy. Deep neural networks broaden the scope of

ISSN No:-2456-2165

reinforcement learning and enable the demonstration of alternative or extra data for planning constant recommendation strategies.

Adversarial Inference Network IRGAN [162] is the primary model for applying GAN to data recovery areas. To demonstrate its abilities in 3 data recovery operations, the developers demonstrated its search, suggestion, and question-answering capabilities in detail. In this research, we're primarily concerned with finding ways to make use of IRGAN as a suggestion engine. Right off the bat, we present the overall structure of IRGAN. Customary GAN comprises a discriminator and a generator.

Likely, there are two schools of deduction in data recovery, that is, generative recovery and discriminative recovery. Generative recovery accepts that there is a fundamental generative cycle among reports and questions, and recovery undertakings can be accomplished by producing pertinent archive d given an inquiry q. Discriminative recovery figures out how to foresee the pertinence score 'r' given marked important inquiry record sets. The point of IRGAN is to join these two musings into a bound together model, and make them play a minimax game like generator and discriminator in GAN. Generative recovery means to create applicable records like ground truth to trick the discriminative recovery model.

Officially, let p true(d | qn,r) referred to the user's pertinence (inclination) dispersion. The generative recovery model pθ(d|qn,r) attempts to estimate the genuine pertinence dispersion. Discriminative recovery fφ(q,d) attempts to recognize important reports and non-significant archives Like the target capacity of GAN, the general goal is defined as follows:

$$J^{G^*,D^*} = \min_\theta \max_\phi \sum_{n=1}^{N} \left(\mathbb{E}_{d \sim p_{true}(d|q_n,r)} [\log D(d \mid q_n)] + \mathbb{E}_{d \sim p_\theta(d|q_n,r)} [\log (1 - D(d \mid q_n))] \right) \quad (15)$$

Where D(d|qn) = (f(q,d)), the sigmoid capacity is addressed by, and the bounds for generative and racially biased recovery are defined by and, respectively. Angle plunge may be used to further adjust the boundary and. The above target circumstance is designed for pointwise significance testing. The best positioning records may be achieved in certain cases if you examine the world from a paired perspective. Assume that a softmax work provides p (d|qn,r) in this case:

$$p_\theta(d_i \mid q, r) = \frac{\exp\left(g_\theta(q, d_i)\right)}{\sum_{d_j} \exp\left(g_\theta(q, d_j)\right)} \quad (16)$$

gθ (q,d) is the opportunity of record d being produced from inquiry q. In genuine word recovery framework, both gθ (q,d) and fφ(q,d) are task-explicit. They can either have something very similar or various details. The creators displayed them with a similar capacity for comfort, and characterize them as: gθ (q,d) = sθ (q,d) and fφ(q,d) = sφ(q,d). In the thing recommendation situation, the creators embraced the lattice factorization to plan s(·). It may very well be subbed with other progressed models, for example, factorization machines or neural networks.

He and his colleagues [52] suggested a Bayesian customised positioning strategy that incorporates suffering preparing for ill-disposed custom placement. In this game, the initial BPR is evenhanded, and the opponent adds turmoil or modifications to the BPR's misery. GAN-based depiction learning in heterogeneous bibliographic networks has been presented by Cai et al. [9], and it has the potential to successfully handle the problem of customized reference recommendation. We suggested using GANs to generate negative instances for the memory network-based streaming recommender [164]. Observations have shown that the suggested GAN-based sampler might fundamentally work for the presentation.

Deep Hybrid Models for Recommendation: Many neural structural tiles can be merged because to Deep Convolutional neural Networks' tremendous versatility to formulate even more astonishing and expressive models. Our recommendation is that the half-breed model be thoughtfully and methodically designed for the specific duties, notwithstanding the many possible techniques of mix. Models that are shown to work in various application areas are summarised below.

CNNs and Autoencoder: Collective Knowledge Based Embedding (CKE) [192] joins CNNs with autoencoder for pictures including extraction. CKE can be seen as a further advance of CDL. CDL just considers thin text data (for example edited compositions of articles and plots of films), while CKE uses primary substance, text based substance and visual substance with various installing procedures. Underlying data incorporates the properties of things also, the connections among things and users. CKE embraces the TransR [96], a heterogeneous network inserting strategy, for deciphering primary data. Likewise, CKE utilizes SDAE to take in highlight portrayals from literary data. Concerning visual data, CKE receives stacked convolutional auto-encoders (SCAE). SCAE utilizes convolution by supplanting the completely associated layers of SDAE with convolutional layers. The proposal cycle is done in a probabilistic structure like CDL.

CNN's and RNN's: Deep mixture models incorporating RNNs and CNNs were suggested by Lee et al. [82]. An effort to compile a list of statements based on the written or spoken responses to an enquiry is known as a "statement proposal" CNNs are used in the process of analysing tweets in order to extract the most important neighbourhood semantics and direct them to a distributed vector. These differential vectors are also created using LSTM in order to record the relevance of target phrases to the specific tweet being spoken about. Figure 12 depicts the overall design (a). It was hypothesised by Zhang et al. (193) that a blend of RNN and CNN models may be used for recommending hashtags. The developers employed CNNs and LSTM to extract text summaries from tweets and images associated with the tweets. During this moment, the designers have offered a founder instrument to demonstrate the link affects and balance between the devotion of writings and images. In an encoder-decoder architecture for connection proposal, Ebsesu et al. [38] developed a neural referencing network that coordinates CNNs with RNNs. As an encoder, CNNs are used to capture the elongated circumstances in the reference environment. There is a decoder in the RNNs, which can learn the probability of a word with in alluded to paper's title based on all previous words and CNNs' depictions..

RNN's and Autoencoder: The synergy Learning algorithm previously referred to is lacking in strength and is unable to demonstrate the groups of text data. RNNs and de - noising auto - encoders were also abused by Wang et al. [160] to circumvent these restrictions. Powerful Intermittent Network (PIN) was the first moniker for the RNN theory. The inventors suggested the CRAE continuous Bayesian model - based approach in light of the robust Recurrent network. When RNNs replace feedforward neural layers in CRAE's encoding and deciphering components, it allows CRAE to capture the consecutive data in thin content data. To prevent overfitting, the designers devised an ace card noise removal and a beta-pooling technique.

RNNs with DRL : Wang et al. [163] treatment recommendation. The structure can take in the remedy strategy from the marker sign and assessment signal. Trials exhibit that his framework could deduce and find the ideal medicines naturally. We accept that this is a significant subject and benefits society greatly.

III. USING THE FUTURE RESEARCH DIRECTIONS

Making precise recommendations involves a thorough having a thorough knowledge of the product's features as well as the genuine demands and preferences of the end user [1, 85]. An extensive supply of auxiliary data is often used in this process. Setting data, for example, customises administrations and goods depending on the conditions and environmental aspects of the user [151] and reduces the effect of a hard freeze; implicit input shows users' understood objective advantages of the accessible information. In addition, there are few research evaluating how users' views (e.g., Twitter or Facebook posts) are influenced by online media (e.g., the Internet of Everything) and the actual world. One may learn about a user's global advantages or ambitions from these new data assets, which can be included using the deep learning technique. Deep learning's ability to analyse a wide range of input sources, such as video highlights, expands its range of possibilities.

It is critical and widely used in mechanical applications [20, 27]. In any case, the majority of contemporary models involve physical creation and selection of highlights, which is laborious and tiresome. By reducing manual mediation, deep neural networks are a viable method for programmed highlight generation [129]. Additionally, there is an advantage to depiction gathering from unstructured messages, images, or data that exists in the 'wild' without creating complicated component design workflows. Additional research on explicitly specifying deep elements for recommender systems is essential to minimise human effort while also improving proposal quality. A fascinating forward-looking research question is how to construct neural networks in such a way that they maximise the accessibility of various types of information. The Joint Representational Learning system [197] is one current project that may pave the way for models of this type. Joint (perhaps multi-modular) depictions of users and objects are likely to emerge as a new trend in recommender frameworks research. To this purpose, a deep picking up perspective from this point of view would be the best way to develop more effective inductive predispositions (cross breed brain structures) in a start to finish design. For instance, considering information in several modalities (text, images, collaboration) in order to improve suggestion execution.

REFERENCES

[1.] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering 17, 6 (2005), 734–749.

[2.] Taleb Alashkar, Songyao Jiang, Shuyang Wang, and Yun Fu. 2017. Examples-Rules Guided Deep Neural Network for Makeup Recommendation. In AAAI. 941–947.

[3.] Jimmy Ba, Volodymyr Mnih, and Koray Kavukcuoglu. 2014. Multiple object recognition with visual attention. arXiv preprint arXiv:1412.7755 (2014).

[4.] Bing Bai, Yushun Fan, Wei Tan, and Jia Zhang. 2017. DLTSR: A Deep Learning Framework for Recommendation of Long-tail Web Services. IEEE Transactions on Services Computing (2017).

[5.] Trapit Bansal, David Belanger, and Andrew McCallum.2016. Ask the gru: Multi-task learning for deep text recommendations. In Proceedings of the 10th ACM Conference on Recommender Systems. 107–114.

[6.] Rianne van den Berg, Thomas N Kipf, and Max Welling.2017. Graph convolutional matrix completion. arXiv preprint arXiv:1706.02263 (2017).

[7.] Basiliyos Tilahun Betru, Charles Awono Onana, and Bernabe Batchakui. 2017. Deep Learning Methods on Recommender System: A Survey of State-of-the-art. International Journal of Computer Applications 162, 10

[8.] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction 12, 4 (2002), 331–370.

[9.] Xiaoyan Cai, Junwei Han, and Libin Yang. 2018. Generative Adversarial Network Based Heterogeneous Bibliographic Network Representation for Personalized Citation Recommendation. In AAAI.

[10.] S. Cao, N. Yang, and Z. Liu. 2017. Online news recommender based on stacked auto-encoder. In ICIS. 721–726

[11.] Rose Catherine and William Cohen. 2017. Transnets:Learning to transform for recommendation. In Recsys. 288–296

[12.] Cheng Chen, Xiangwu Meng, Zhenghua Xu, and thomas Lukasiewicz. 2017. Location-Aware

Personalized News Recommendation With Deep Semantic Analysis. IEEEAccess 5 (2017), 1624–1638.

[13.] Cen Chen, Peilin Zhao, Longfei Li, Jun Zhou, Xiaolong Li, and Minghui Qiu. 2017. Locally Connected Deep Learning Framework for Industrial-scale Recommender Systems. In WWW.

[14.] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. 2017. Attentive Collaborative Filtering: Multimedia Recommendation with Item- and Component-Level Aention. (2017).

[15.] Minmin Chen, Zhixiang Xu, Kilian Weinberger, and Fei Sha. 2012. Marginalized denoising autoencoders for domain adaptation. arXiv preprint arXiv:1206.4683 (2012).

[16.] Shi-Yong Chen, Yang Yu, Qing Da, Jun Tan, Hai-Kuan Huang, and Hai-Hong Tang. 2018. Stabilizing reinforcement learning in a dynamic environment with application to online recommendation. In SIGKDD. 1187–1196.

[17.] Xu Chen, Yongfeng Zhang, Qingyao Ai, Hongteng Xu, Junchi Yan, and Zheng Qin. 2017. Personalized Key Frame Recommendation. In SIGIR.

[18.] Xu Chen, Yongfeng Zhang, Hongteng Xu, Yixin Cao,Zheng Qin, and Hongyuan Zha. 2018. Visually Explainable Recommendation. arXiv preprint arXiv:1801.10288 (2018).

[19.] Yifan Chen and Maarten de Rijke. 2018. A Collective Variational Autoencoder for Top-N Recommendation with Side Information. arXiv preprint arXiv:1807.05730 (2018).

[20.] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, and others. 2016. Wide & deep learning for recommender systems. In Recsys.7–10.

[21.] Sungwoon Choi, Heon Seok Ha, Uiwon Hwang, Chanju Kim, Jung-Woo Ha, and Sungroh Yoon. 2018. Reinforcement Learning based Recommender System using Biclustering Technique. arXiv preprint arXiv:1801.05532

[22.] Jan Chorowski, Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. End-to-end continuous speech recognition using attention-based recurrent NN: first results. arXiv preprint arXiv:1412.1602 (2014).

[23.] Jan K Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. 2015. Attention-based models for speech recognition. In Advances in Neural Information Processing Systems. 577–585.

[24.] Konstantina Christakopoulou, Alex Beutel, Rui Li, Sagar Jain, and Ed H Chi. 2018. Q&R: A Two-Stage Approach toward Interactive Recommendation. In SIGKDD. 139–148.

[25.] Wei-Ta Chu and Ya-Lun Tsai. 2017. A hybrid recommendation system considering visual information for predicting favorite restaurants. WWWJ (2017), 1–19

[26.] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. Proceedings of the 25th international conference on Machine learning. 160–167.

[27.] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In Recsys. 191–198.

[28.] Hanjun Dai, Yichen Wang, Rakshit Trivedi, and Le Song. 2016. Deep coevolutionary network: Embedding user and item features for recommendation. arXiv preprint. arXiv preprint arXiv:1609.03675 (2016)

[29.] Hanjun Dai, Yichen Wang, Rakshit Trivedi, and Le Song. 2016. Recurrent coevolutionary latent feature processes for continuous-time recommendation. In Recsys.29–34.

[30.] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath.2010. e YouTube Video Recommendation System. In Recsys.

[31.] Li Deng, Dong Yu, and others. 2014. Deep learning: methods and applications. Foundations and Trends® in Signal Processing 7, 3–4 (2014), 197–387.

[32.] Shuiguang Deng, Longtao Huang, Guandong Xu, Xindong Wu, and Zhaohui Wu. 2017. On deep learning for trust-aware recommendations in social networks. IEEE transactions on neural networks and learning systems 28, 5 (2017), 1164–1177.

[33.] Robin Devooght and Hugues Bersini. 2016.Collaborative filtering with recurrent neural networks. arXiv preprint arXiv:1608.07400 (2016).

[34.] Xin Dong, Lei Yu, Zhonghuo Wu, Yuxia Sun, Lingfeng Filtering Model with Deep Structure for Recommender

[35.] Tim Donkers, Benedikt Loepp, and Jurgen Ziegler. 2017. Sequential user-based recurrent neural network recommendations. In ¨Recsys. 152–160.

[36.] Chao Du, Chongxuan Li, Yin Zheng, Jun Zhu, and Bo Zhang. 2016. Collaborative Filtering with User-Item Co- Autoregressive Models. arXiv preprint arXiv:1612.07146

[37.] Gintare Karolina Dziugaite and Daniel M Roy. 2015. Neural network matrix factorization. arXiv preprint arXiv:1511.06443 (2015).

[38.] Travis Ebesu and Yi Fang. 2017. Neural Citation Network for Context-Aware Citation Recommendation.(2017).

[39.] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In WWW. 278–288.

[40.] Ignacio Fernandez-Tob ´ ´ıas, Ivan Cantador, Marius Kaminskas, and Francesco Ricci. 2012. Cross-domain recommender systems: A survey ´ of the state of the art. In Spanish Conference on Information Retrieval. 24.

[41.] Jianfeng Gao, Li Deng, Michael Gamon, Xiaodong He, and Patrick Pantel. 2014. Modeling interestingness with deep neural networks. (June 13 2014). US Patent App. 14/304,863.

[42.] Kostadin Georgiev and Preslav Nakov. 2013. A non-iid framework for collaborative filtering with restricted boltzmann machines. In ICML. 1148–1156.

[43.] Carlos A Gomez-Uribe and Neil Hunt. 2016. e netix recommender system: Algorithms, business value, and innovation. TMIS 6, 4 (2016), 13.

[44.] Yuyun Gong and Qi Zhang. 2016. Hashtag Recommendation Using Aention-Based Convolutional Neural Network.. In IJCAI. 2782–2788.

[45.] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. MIT Press.http://www.deeplearningbook.org.

[46.] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial Yuan, and Fangxi Zhang. 2017. A Hybrid Collaborativenets. In NIPS. 2672–2680.

[47.] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. In IJCAI. 2782– Systems. In AAAI. 1309–1315.

[48.] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778.

[49.] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one- class collaborative filtering. In WWW. 507–517.

[50.] Ruining He and Julian McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In AAAI. 144–150.

[51.] Xiangnan He, Xiaoyu Du, Xiang Wang, Feng Tian,Jinhui Tang, and Tat-Seng Chua. 2018. Outer Product-based Neural Collaborative Filtering. (2018).

[52.] Xiangnan He, Zhankui He, Xiaoyu Du, and Tat-Seng Chua. 2018. Adversarial Personalized Ranking for Recommendation. In SIGIR. 355–364.

[53.] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In WWW. 173–182.

[54.] Xiangnan He and Chua Tat-Seng. 2017. Neural Factorization Machines for Sparse Predictive Analytics. (2017).

[55.] Balazs Hidasi and Alexandros Karatzoglou. 2017. Recurrent neural networks with top-k gains for session-based recommendations. ´ arXiv preprint arXiv:1706.03847 (2017).

[56.] Balazs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent ´ neural networks.

[57.] Balazs Hidasi, Massimo adrana, Alexandros Karatzoglou, and Domonkos Tikk. 2016. Parallel recurrent neural network architectures ´ for feature-rich session-based recommendations. In Recsys. 241–248

[58.]]Kurt Hornik. 1991. Approximation capabilities of multilayer feedforward networks. Neural networks 4, 2 (1991), 251–257.

[59.] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. 1989. Multilayer feedforward networks are universal Wenyao Xu, and Aidong Zhang. 2015. Multi-modal learning approximators. Neural networks 2, 5 (1989), 359–366.

[60.] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. 2017. Collaborative metric learning. In WWW. 193–201.

[61.] Cheng-Kang Hsieh, Longqi Yang, Honghao Wei, Mor Naaman, and Deborah Estrin. 2016. Immersive recommendation: News and event recommendations using personal digital traces. In WWW. 51–62.

[62.] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. 2018. Leveraging Meta-path based Context for Top-N Recommendation with A Neural Co-Aention Model. In SIGKDD. 1531–1540.

[63.] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In ICDM.

[64.] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely Connected Convolutional Networks.. In CVPR, Vol. 1. 3.

[65.] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In CIKM. 2333–2338.

[66.] Wenyi Huang, Zhaohui Wu, Liang Chen, Prasenjit Mitra, and C Lee Giles. 2015. A Neural Probabilistic Model for Context Based Citation Recommendation. In AAAI.2404–2410.

[67.] Drew A Hudson and Christopher D Manning. 2018. Compositional aention networks for machine reasoning. arXiv preprint arXiv:1803.03067 (2018).

[68.] Dietmar Jannach and Malte Ludewig. 2017. When Recurrent Neural Networks Meet the Neighborhood for Session-Based Recommendation. In Recsys.

[69.] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. Recommender systems: an introduction.

[70.] Yogesh Jhamb, Travis Ebesu, and Yi Fang. 2018. Aentive Contextual Denoising Autoencoder for Recommendation. (2018).

[71.] X. Jia, X. Li, K. Li, V. Gopalakrishnan, G. Xun, and A.Zhang. 2016. Collaborative restricted Boltzmann machine for social event recommendation. In ASONAM. 402–405.

[72.] Xiaowei Jia, Aosen Wang, Xiaoyi Li, Guangxu Xun, for video recommendation based on mobile application usage. In 2015 IEEE International Conference on Big Data(Big Data). 837–842.

[73.] How Jing and Alexander J Smola. 2017. Neural survival recommender. In WSDM. 515–524.

[74.] Muhammad Murad Khan, Roliana Ibrahim, and Imran Ghani. 2017. Cross Domain Recommender Systems: A Systematic Literature Review. ACM Comput. Surv. 50, 3 (June 2017).

[75.] Donghyun Kim, Chanyoung Park, Jinoh Oh, Sungyoung Lee, and Hwanjo Yu. 2016. Convolutional matrix factorization for document context-aware recommendation.In Recsys. 233–240.

[76.] Donghyun Kim, Chanyoung Park, Jinoh Oh, and Hwanjo Yu. 2017. Deep Hybrid Recommender Systems via Exploiting Document Context and Statistics of Items.Information Sciences (2017).

[77.] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks.arXiv preprint arXiv:1609.02907 (2016).

[78.] Young-Jun Ko, Lucas Maystre, and Matthias Grossglauser. 2016. Collaborative recurrent neural networks for dynamic recommender systems. In Asian Conference on Machine Learning. 366–381.

[79.] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In SIGKDD. 426–434.

[80.] Yehuda Koren. 2010. Collaborative filtering with temporal dynamics. Commun. ACM 53, 4 (2010), 89–97.

[81.] Hugo Larochelle and Iain Murray. 2011. The neural autoregressive distribution estimator. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 29–37.

[82.] Hanbit Lee, Yeonchan Ahn, Haejun Lee, Seungdo Ha, and Sang-goo Lee. 2016. ote Recommendation in Dialogue using Deep Neural Network. In SIGIR. 957–960.

[83.] Joonseok Lee, Sami Abu-El-Haija, Balakrishnan Varadarajan, and Apostol Paul Natsev. 2018. Collaborative Deep Metric Learning for Video Understanding. (2018).

[84.] Chenyi Lei, Dong Liu, Weiping Li, Zheng-Jun Zha, and Houqiang Li. 2016. Comparative Deep Learning of Hybrid Representations for Image Recommendations. In Proceedings of the IEEE Conference on Computer Vision and Paern Recognition. 2545–2553.

[85.] Jure Leskovec. 2015. New Directions in Recommender Systems. In WSDM.

[86.] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th international conference on World wide web. 661–670.

[87.] Piji Li, Zihao Wang, Zhaochun Ren, Lidong Bing, and Wai Lam. 2017. Neural Rating Regression with Abstractive Tips Generation for Recommendation. (2017).

[88.] Sheng Li, Jaya Kawale, and Yun Fu. 2015. Deep collaborative filtering via marginalized denoising auto-encoder. In CIKM. 811–820.

[89.] Xiaopeng Li and James She. 2017. Collaborative Variational Autoencoder for Recommender Systems. In SIGKDD.

[90.] Yang Li, Ting Liu, Jing Jiang, and Liang Zhang. 2016. Hashtag recommendation with topical attention-based LSTM. In COLING.

[91.] Zhi Li, Hongke Zhao, Qi Liu, Zhenya Huang, Tao Mei,and Enhong Chen. 2018. Learning from History and Present:User Behaviors. In SIGKDD. 1734–1743.

[92.] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2017. CCCFNet: A Content-Boosted Collaborative Filtering Neural Network for Cross Domain Recommender Systems. In WWW. 817–818.

[93.] Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. 2018. xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems. arXiv preprint arXiv:1803.05170 (2018).

[94.] Dawen Liang, Rahul G Krishnan, Mahew D Homan, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. arXiv preprint arXiv:1802.05814 (2018).

[95.] Dawen Liang, Minshu Zhan, and Daniel PW Ellis. 2015. Content-Aware Collaborative Music Recommendation Using Pre-trained Neural Networks.. In ISMIR. 295–301.

[96.] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In AAAI. 2181–2187.

[97.] Juntao Liu and Caihua Wu. 2017. Deep Learning Based Recommendation: A Survey.

[98.] Qiang Liu, Shu Wu, and Liang Wang. 2017. DeepStyle: Learning User Preferences for Visual Recommendation.(2017)

[99.] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: Short-Term Attention/Memory Priority Model for Session-based Recommendation. In SIGKDD. 1831–1839.

[100.] Xiaomeng Liu, Yuanxin Ouyang, Wenge Rong, and Zhang Xiong. 2015. Item Category Aware Conditional Restricted Boltzmann Machine Based Recommendation. In the International Conference on Neural Information Processing. 609–616.

[101.] Pablo Loyola, Chen Liu, and Yu Hirate. 2017. Modeling User Session and Intent with an Aention-based Encoder-Decoder Architecture. In Recsys. 147–151.

[102.] Pablo Loyola, Chen Liu, and Yu Hirate. 2017. Modeling User Session and Intent with an Aention-based Encoder-Decoder Architecture. In Recsys (RecSys ’17).

[103.] Malte Ludewig and Dietmar Jannach. 2018. Evaluation of Session-based Recommendation Algorithms. CoRR abs/1803.09587 (2018).

[104.] Minh-hang Luong, Hieu Pham, and Christopher D Manning. 2015. Ective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025

[105.] Next-item Recommendation via Discriminatively Exploiting Julian McAuley, Christopher Targe, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In SIGIR. 43–52.

[106.] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, and others. 2015. Human-level control through deep reinforcement learning. Nature 518, 7540 (2015), 529.

[107.] Isshu Munemasa, Yuta Tomomatsu, Kunioki Hayashi, and Tomohiro Takagi. 2018. Deep Reinforcement Learning for Recommender Systems. (2018).

[108.] Cataldo Musto, Claudio Greco, Alessandro Suglia, and Giovanni Semeraro. 2016. Ask Me Any Rating: A Content-based Recommender System based on Recurrent Neural Networks. In IIR.

[109.] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. 2015. Deep learning applications and challenges in big data analytics. Journal of Big Data 2, 1 (2015), 1.

[110.] Hanh T. H. Nguyen, Martin Wistuba, Josif Grabocka, Lucas Rego Drumond, and Lars Schmidt-ieme. 2017. Personalized Deep Learning for Tag Recommendation. ICML. 833–840.

[111.] Xia Ning and George Karypis. 2010. Multi-task learning for recommender systems. In Proceedings of 2nd Asian Conference on Machine Learning. 269–284.

[112.] Wei Niu, James Caverlee, and Haokai Lu. 2018. Neural Personalized Ranking for Image Recommendation. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 423–431.

[113.] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based News Recommendation for Millions of Users. In SIGKDD.

[114.] Yuanxin Ouyang, Wenqi Liu, Wenge Rong, and Zhang Xiong. 2014. Autoencoder-based collaborative filtering. In the International Conference on Neural Information Processing. 284–291.

[115.] Weike Pan, Evan Wei Xiang, Nathan Nan Liu, and Qiang Yang. 2010. Transfer Learning in Collaborative Filtering for Sparsity Reduction. In AAAI, Vol. 10. 230–235.

[116.] Yiteng Pana, Fazhi Hea, and Haiping Yuan. 2017. Trust-aware Collaborative Denoising Auto-Encoder for Top- N Recommendation. arXiv preprint arXiv:1703.01760

[117.] Massimo adrana, Alexandros Karatzoglou, Balazs Hidasi, and Paolo Cremonesi. 2017. Personalizing session-based recommendations ´ with hierarchical recurrent neural networks. In Recsys. 130–137.

[118.] Yogesh Singh Rawat and Mohan S Kankanhalli. 2016.ConTagNet: exploiting user context for image tag recommendation. In Proceedings of the 2016 ACM on Multimedia Conference. 1102–1106.

[119.] S. Rendle. 2010. Factorization Machines. In 2010 IEEE International Conference on Data Mining.

[120.] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-ieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In Proceedings of the twenty-fih conference on uncertainty in artificial intelligence. 452–461.

[121.] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In Recommender systems handbook. 1–34.

[122.] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. 2011. Contractive auto-encoders: Explicit invariance during feature extraction. In collaborative filtering. In ICML. 791–798.

[123.] Ruslan Salakhutdinov, Andriy Mnih, and Georey Hinton. 2007. Restricted Boltzmann machines for CF

[124.] Adam Santoro, David Raposo, David G Barrely, Mateusz Malinowski, Razvan Pascanu, Peter Baaglia, and Tim Lillicrap. 2017. A simple neural network module for relational reasoning. In NIPS. 4967–4976.

[125.] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. 2015. Autorec: Autoencoders meet collaborative filtering. In WWW. 111–112.

[126.] Sungyong Seo, Jing Huang, Hao Yang, and Yan Liu. 2017. Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In Recsys. 297–305.

[127.] Sungyong Seo, Jing Huang, Hao Yang, and Yan Liu.2017. Representation Learning of Users and Items for Review Rating Prediction Using Attention-based Convolutional Neural Network. In MLRec.

[128.] Joan Serra and Alexandros Karatzoglou. 2017. Getting deep recommenders fit: Bloom embeddings for sparse binary input/output networks. In Recsys. 279–287.

[129.] Ying Shan, T Ryan Hoens, Jian Jiao, Haijing Wang, Dong Yu, and JC Mao. 2016. Deep Crossing: Web-scale modeling without manually crafted combinatorial features. In SIGKDD. 255–262.

[130.] Xiaoxuan Shen, Baolin Yi, Zhaoli Zhang, Jiangbo Shu, and Hai Liu. 2016. Automatic Recommendation Technology for Learning Resources with Convolutional Neural Network. In the International Symposium on Educational Technology.30-34

[131.] Yue Shi, Martha Larson, and Alan Hanjalic. 2014.Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. ACM Computing Surveys (CSUR) 47, 1 (2014), 3.

[132.] Elena Smirnova and Flavian Vasile. 2017. Contextual Sequence Modeling for Recommendation with Recurrent Neural Networks.(2017).

[133.] Harold Soh, Sco Sanner, Madeleine White, and Greg Jamieson. 2017. Deep Sequential Recommendation for Personalized Adaptive

[134.] Bo Song, Xin Yang, Yi Cao, and Cong fu Xu. 2018. Neural Collaborative Ranking. arXiv preprint arXiv:1808.04957 (2018).

[135.] Yang Song, Ali Mamdouh Elkahky, and Xiaodong He. 2016. Multi-rate deep learning for temporal recommendation. In SIGIR. 909–912.

[136.] Florian Strub, Romaric Gaudel, and Jeremie Mary. 2016. Hybrid Recommender System based on Autoencoders. In ´ Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. 11–16.

[137.] Florian Strub and Jeremie Mary. 2015. Collaborative Filtering with Stacked Denoising AutoEncoders and Sparse Inputs. In NIPS Workshop.

[138.] Xiaoyuan Su and Taghi M Khoshgoaar. 2009. A survey of collaborative filtering techniques. Advances in artificial intelligence 2009 (2009), 4.

[139.] Alessandro Suglia, Claudio Greco, Cataldo Musto, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2017. A Deep Architecture for Content-based Recommendations Exploiting Recurrent Neural Networks. In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. 202–211.

[140.] Yosuke Suzuki and Tomonobu Ozaki. 2017. Stacked Denoising Autoencoder-Based Deep Collaborative Filtering Using the Change of Similarity. In WAINA. 498–502.

[141.] Jiwei Tan, Xiaojun Wan, and Jianguo Xiao. 2016. A Neural Network Approach to vote Recommendation in Writings. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management.65–74.

[142.] Yong Kiam Tan, Xinxing Xu, and Yong Liu. 2016.Improved recurrent neural networks for session-based recommendations. In Recsys.17–22.

[143.] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In WSDM. 565–573.

[144.] Jiaxi Tang and Ke Wang. 2018. Ranking Distillation: Learning Compact Ranking Models With High Performance for Recommender System. In SIGKDD.

[145.] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Latent Relational Metric Learning via Memory-based Attention for Collaborative Ranking. In WWW.

[146.] Yi Tay, Anh Tuan Luu, and Siu Cheung Hui. 2018. Multi-Pointer Co-Aention Networks for Recommendation. In SIGKDD.

[147.] Trieu H Trinh, Andrew M Dai, ang Luong, and V Le. 2018. Learning longer-term dependencies in rnns with auxiliary losses. arXiv preprint arXiv:1803.00144 (2018).

[148.] Trinh Xuan Tuan and Tu Minh Phuong. 2017. 3D Convolutional Networks for Session-based Recommendation with Content Features. In Recsys. 138–146.

[149.] Bartlomiej Twardowski. 2016. Modelling Contextual Information in Session-Aware Recommender Systems with Neural Networks. In Recsys.

[150.] Moshe Unger. 2015. Latent Context-Aware Recommender Systems. In Recsys. 383–386.

[151.] Moshe Unger, Ariel Bar, Bracha Shapira, and Liorn Rokach. 2016. Towards latent context-aware recommendation systems. Knowledge Based Systems 104 (2016), 165–178.

[152.] Benigno Uria, Marc-Alexandre Cotˆ e, Karol Gregor, Iain Murray, and Hugo Larochelle. 2016. Neural autoregressive distribution estimation. Journal of Machine Learning Research 17, 205 (2016), 1–37.

[153.] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In NIPS. 2643–2651.

[154.] Manasi Vartak, Arvind iagarajan, Conrado Miranda, Jeshua Bratman, and Hugo Larochelle. 2017. A Meta-Learning Perspective on Cold-Start Recommendations for Items. Advances in Neural Information Processing Systems. 6904–6914.

[155.] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017.Attention is all you need. In Advances in Neural Information Processing Systems. 5998–6008.

[156.] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. 2017. DropoutNet: Addressing Cold Start in Recommender Systems. In Advances in Neural Information Processing Systems. 4957–4966.

[157.] Jeroen B. P. Vuurens, Martha Larson, and Arjen P. de Vries. 2016. Exploring Deep Space: Learning Personalized Ranking in a Semantic Space. In Recsys.

[158.] Hao Wang, Xingjian Shi, and Dit-Yan Yeung. 2015.Relational Stacked Denoising Autoencoder for Tag Recommendation.In AAAI.3052–3058.

[159.] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In SIGKDD. 1235–1244.

[160.] Hao Wang, SHI Xingjian, and Dit-Yan Yeung. 2016. Collaborative recurrent autoencoder: Recommend while 2017. IRGAN: A Minimax Game for Unifying Generative learning to fill in the blanks. In NIPS. 415–423.

[161.] Hao Wang and Dit-Yan Yeung. 2016. Towards Bayesian deep learning: A framework and some existing methods. TKDE 28, 12 (2016), 3395–3408.

[162.] Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong,Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang.and Discriminative Information Retrieval Models. (2017).

[163.] Lu Wang, Wei Zhang, Xiaofeng He, and Hongyuan Zha. 2018. Supervised Reinforcement Learning with Recurrent Neural Network for Dynamic Treatment Recommendation. In SIGKDD. 2447–2456.

[164.] Qinyong Wang, Hongzhi Yin, Zhiting Hu, Defu Lian, Hao Wang, and Zi Huang. 2018. Neural Memory

Streaming Recommender Networks with Adversarial Training. In SIGKDD.

[165.] Suhang Wang, Yilin Wang, Jiliang Tang, Kai Shu, Suhas Ranganath, and Huan Liu. 2017. What Your Images Reveal: Exploring visuals contents for point of intrest recommendation INn WWW.

[166.] Xiang Wang, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. 2017. Item Silk Road: Recommending Items from Information Domains to Social Users. (2017).

[167.] Xinxi Wang and Ye Wang. 2014. Improving content-based and hybrid music recommendation using deep learning. In MM. 627–636.

[168.] Xinxi Wang, Yi Wang, David Hsu, and Ye Wang 2014. Exploration in interactive personalized music recommendation: a reinforcement learning approach. TOMM 11, 1 (2014), 7.

[169.] Xuejian Wang, Lantao Yu, Kan Ren, Guangyu Tao, Weinan Zhang, Yong Yu, and Jun Wang. 2017. Dynamic Aention Deep Model for Article Recommendation by Learning Human Editors Demonstration. In SIGKDD.

[170.] Jian Wei, Jianhua He, Kai Chen, Yi Zhou, and Zuoyin Tang. 2016. Collaborative filtering and deep learning based hybrid recommendation for cold start problem. IEEE, 874–877.

[171.] Jian Wei, Jianhua He, Kai Chen, Yi Zhou, and Zuoyin Tang. 2017. Collaborative filtering and deep learning based recommendation system for cold start items. Expert Systems with Applications 69 (2017), 29–39.

[172.] Jiqing Wen, Xiaopeng Li, James She, Soochang Park,and Ming Cheung. 2016. Visual background recommendation for dance performances using dancer-shared images. 521–527.

[173.] Caihua Wu, Junwei Wang, Juntao Liu, and Wenyu Liu. 2016. Recurrent neural network based recommendation for time heterogeneous feedback. Knowledge-Based Systems 109 (2016), 90–103.

[174.] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, and Alexander J Smola. 2016. Joint Training of Ratings and Reviews with Recurrent Recommender Networks. (2016).

[175.] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J Smola, and How Jing. 2017. Recurrent recommender networks. In WSDM. 495–503.SIGKDD.

[176.] Sai Wu, Weichao Ren, Chengchao Yu, Gang Chen, Dongxiang Zhang, and Jingbo Zhu. 2016. Personal recommendation using deep recurrent neural networks in NetEase. In ICDE. 1218–1229.

[177.] Yao Wu, Christopher DuBois, Alice X Zheng, and Martin Ester. 2016. Collaborative denoising auto-encoders for top-n recommender systems. In WSDM. 153–162.

[178.] Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, and Tat-Seng Chua. 2017. Aentional factorization machine: Learning the weight of feature interactions via attention networks. arXiv preprint arXiv:1708.04617 (2017).

[179.] Ruobing Xie, Zhiyuan Liu, Rui Yan, and Maosong Sun. 2016. Neural Emoji Recommendation in Dialogue Systems. arXiv preprint arXiv:1612.04609 (2016).

[180.] Weizhu Xie, Yuanxin Ouyang, Jingshuai Ouyang, Wenge Rong, and Zhang Xiong. 2016. User Occupation recommendation using a deep-semantic similarity model with negative sampling. In CIKM. 1921–1924.(2017).

[181.] Caiming Xiong, Victor Zhong, and Richard Socher. 2016. Dynamic coaention networks for question answering. arXiv preprint arXiv:1611.01604 (2016).

[182.] Zhenghua Xu, Cheng Chen, omas Lukasiewicz, Yishu Miao, and Xiangwu Meng. 2016. Tag-aware personalized

[183.] Zhenghua Xu, Thomas Lukasiewicz, Cheng Chen, Yishu Miao, and Xiangwu Meng. 2017. Tag-aware personalized recommendation using a hybrid deep model.(2017)

[184.] Hong-Jian Xue, Xinyu Dai, Jianbing Zhang, Shujian Huang, and Jiajun Chen. 2017. Deep Matrix Factorization Models for Recommender Systems.. In IJCAI. 3203–3209.

[185.] Carl Yang, Lanxiao Bai, Chao Zhang, an Yuan, and Jiawei Han. 2017. Bridging Collaborative Filtering and Semi- Supervised Learning: A Neural Approach for POI Recommendation. In SIGKDD.

[186.] Lina Yao, an Z Sheng, Anne HH Ngu, and Xue Li. 2016. ings of interest recommendation by leveraging heterogeneous relations in the internet of things. ACM Transactions on Internet Technology (TOIT) 16, 2 (2016), 9.

[187.] Baolin Yi, Xiaoxuan Shen, Zhaoli Zhang, Jiangbo Shu, and Hai Liu. 2016. Expanded autoencoder recommendation framework and its application in movie recommendation. In SKIMA. 298–303.

[188.] Haochao Ying, Fuzhen Zhuang, Fuzheng Zhang,Yanchi Liu, Guandong Xu, Xing Xie, Hui Xiong, and Jian Wu. 2018. Sequential Recommender System based on Hierarchical Attention Networks. In IJCAI.

[189.] Haochao Ying, Liang Chen, Yuwen Xiong, and Jian Wu. 2016. Collaborative deep ranking: a hybrid pair-wise recommendation algorithm with implicit feedback. In PAKDD. 555–567.

[190.] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. arXiv preprint arXiv:1806.01973 (2018).

[191.] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, and Zheng Qin. 2018. Aesthetic-based clothing recommendation. In WWW. 649–658.

[192.] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. 2016. Collaborative knowledge base embedding for recommender systems. In SIGKDD. 353–362.

[193.] Qi Zhang, Jiawen Wang, Haoran Huang, Xuanjing Huang, and Yeyun Gong. Hashtag Recommendation for Multimodal Microblog Using Co-Attention Network. In IJCAI.

[194.] Shuai Zhang, Yi Tay, Lina Yao, and Aixin Sun. 2018. Next Item Recommendation with Self-Attention. arXiv preprint arXiv:1808.06414 (2018).

[195.] Shuai Zhang, Lina Yao, Aixin Sun, Sen Wang,Guodong Long, and Manqing Dong. 2018. NeuRec: On Nonlinear Transformation for Personalized Ranking. arXiv preprint arXiv:1805.03002 (2018).

[196.] Shuai Zhang, Lina Yao, and Xiwei Xu. 2017. AutoSVD++: An Ecient Hybrid Collaborative Filtering Model via Contractive Auto encoders. (2017).

[197.] Yongfeng Zhang, Qingyao Ai, Xu Chen, and W Bruce Cro. 2017. Joint representation learning for top-n recommendation with heterogeneous information sources. In CIKM. 1449–1458.

[198.] Xiangyu Zhao, Long Xia, Liang Zhang, Zhuoye Ding, Dawei Yin, and Jiliang Tang. 2018. Deep Reinforcement Learning for Page-wise Recommendations. arXiv preprint arXiv:1805.02343 (2018).

[199.] Xiangyu Zhao, Liang Zhang, Zhuoye Ding, Long Xia, Jiliang Tang, and Dawei Yin. 2018. Recommendations with Negative Feedback via Pairwise Deep Reinforcement Learning. arXiv preprint arXiv:1802.06501 (2018).

[200.] Guanjie Zheng, Fuzheng Zhang, Zihan Zheng, Yang Xiang, Nicholas Jing Yuan, Xing Xie, and Zhenhui Li. 2018. DRN: A Deep Reinforcement Learning Framework for News Recommendation. In WWW. 167–176.

[201.] Lei Zheng, Chun-Ta Lu, Lifang He, Sihong Xie, Vahid Noroozi, He Huang, and Philip S Yu. 2018. MARS: Memory Attention-Aware Recommender System. arXiv preprint arXiv:1805.07037 (2018). 575.

[202.] Lei Zheng, Vahid Noroozi, and Philip S. Yu. 2017. Joint Deep Modeling of Users and Items Using Reviews for Recommendation. In WSDM.

[203.] Yin Zheng, Cailiang Liu, Bangsheng Tang, and Hanning Zhou. 2016. Neural Autoregressive Collaborative Filtering for Implicit Feedback. In Recsys.

[204.] Yin Zheng, Bangsheng Tang, Wenkui Ding, and Hanning Zhou. 2016. A Neural Autoregressive Approach to Collaborative Filtering. In ICML.

[205.] Chang Zhou, Jinze Bai, Junshuai Song, Xiaofei Liu, Zhengchao Zhao, Xiusi Chen, and Jun Gao. 2017. ATRank: An Attention-Based User Behavior Modeling Framework for Recommendation. arXiv preprint arXiv:1711.06632 (2017).

[206.] Jiang Zhou, Cathal Gurrin, and Rami Albatal. 2016. Applying visual user interest profiles for recommendation & personalisation. (2016).

[207.] Fuzhen Zhuang, Dan Luo, Nicholas Jing Yuan, Xing Xie, and Qing He. 2017. Representation Learning with Pair- wise Constraints for Collaborative Ranking. In WSDM. 567–575.

[208.] Fuzhen Zhuang, Zhiqiang Zhang, Mingda Qian, Chuan Shi, Xing Xie, and Qing He. 2017. Representation learning via Dual-Autoencoder for recommendation. Neural Networks 90 (2017), 83–89.

[209.] Yi Zuo, Jiulin Zeng, Maoguo Gong, and Licheng Jiao. 2016. Tag-aware recommender systems based on deep neural networks. Neurocomputing 204 (2016), 51–60.

[210.] Shuai Zhang, Lina Yao, Aixin Sun, Yi Tay. "Deep Learning Based Recommender System" , ACM Computing Surveys, 2019 111.