A Review On Recommendation Systems Using Deep Learning

Rajni Jindal, Kirti Jain

Abstract: As the content on the Internet is increasing day-by-day, recommendation systems play a strong role in showing the best results according to users’ preferences and interests. These systems filter out content not only on the basis of users’ static interests but also consider users’ dynamic choices in predicting the next best content. They are considered as playlist generators for videos and music in Netflix, Youtube; provide hashtags recommendations and help people in connecting with each other with similar interests in social networks; help in e-commerce sites by displaying the next item to be purchased by customer based on ratings of items. Deep learning algorithms have been widely used in so many applications such as detection, classification, segmentation, prediction, recommendation. In this paper, 5 categories of recommendations are taken- entertainment (videos/music), user and item reviews, session-based, learning resources, social networks. A brief review of recent works of deep learning in recommendation systems and how it overcomes the cold-start item and sparsity problems of existing methods of recommendation such as content-based and collaborative filtering is presented in this paper.

Keywords: Recommendation Systems, Deep Learning, Collaborative Filtering

1. Introduction

With the growing population and number of visitors on Internet, it has become necessary to filter out information and present it according to users’ preferences and taste of interests. Recommendation systems help the users’ to get items of their own taste. For this, one must focus on the various attributes such as precision, scalability, accuracy, adaptability, robustness [1] before designing a recommendation system. Also, recommending does not only mean to present products by mapping out visible attributes but it also means to map unobserved product attributes and customer characteristics [2]. Recommendation systems do not help in only e-commerce sites to purchase items [3,4] but also are a critical part in decision making systems like sales and marketing [5]. It helps in picking the best web services as per the requirements [6]. They are also used in the field of Software engineering to help developers in a variety of tasks from reusing code to writing bug reports [7]. Content based and Collaborative Filtering has been extensively used in Recommendation Systems since last 2 decades. It has many applications such as item based recommendation systems [8,9], comparing user profiles [10,11], social networks [12-15], technology enhanced learning [16]. They are also used in various other fields related to recommendation such as predicting users’ behaviour, datasets to be analysed, analysing quality with which prediction will be done, identifying the attributes of both user and item to be analysed [17,18]. There are some drawbacks of above mentioned methods:

a) Cold-Start Problem: This problem occurs when an item is introduced or user is visiting the site for the first time. At that time there are no previous records in the database to make prediction of items for that user.

b) Sparsity Problem: This problem occurs when very few users rate a product. So sometimes it becomes difficult to make a recommendation about that product to a user as it is not possible to find similarities between users.

Efforts are made to resolve these drawbacks by introducing Hybrid methods [19-22] which combine both Content based and Collaborative Filtering (CF). Nowadays, Deep Learning (DL) is heavily applied in recommendation systems as they resolve sparsity and cold-start problems. The rest of the paper is organized as follows. Section 2 describes collection of dataset in this review. Section 3 explains some of the algorithms of deep learning. A summary of research articles is covered in Section 4. Conclusion and future work are represented in Section 5 and 6 respectively.

2. DataSet

In this section, we describe how papers are collected from Google Scholar Search Engine. Keywords used to search research articles are “Recommendation Systems on Deep Learning”, “Convolutional Neural Networks/Recurrent Neural Networks/Autoencoders in Recommendation Systems”. We also narrowed the search by applying the year 2015 onwards. 20 papers are summarized here based on the algorithms of Deep Learning. The algorithms described are Deep Collaborative Filtering Model, Convolutional Neural Network, Recurrent Neural Networks and Autoencoders. The various categories of recommendation systems on the basis of which review is done are: Movie/Video, Item and User Profile, Session-Based Recommendation, Books/Learning Resources and Social Networks. Fig 1 describes the year and Journals matrix. Fig 2 describes the ratio number of papers selected for various categories under each algorithm.
Figure 1: Graph displaying count of papers for various Journals and Year under each Journal

Figure 2: Ratio of papers in various categories of recommendation systems
3. Deep Learning Approaches

We have discussed four approaches of deep learning namely Deep Collaborative Filtering model (DCFM), Convolutional Neural Networks (CNN), Autoencoders and Recurrent Neural Networks (RNN).

3.1 Deep Collaborative Filtering model (DCFM) -

It is a hybrid model which involves combination of Deep Learning algorithms and Content based/Collaborative Filtering models. This is mainly done to resolve Sparsity and Cold-start problems. One variation is applied in [23] where a hierarchical Bayesian model called Collaborative Deep Learning (CDL) is developed by using Bayesian Stacked Denoising Autoencoders which resolves the sparsity problem of Collaborative Filtering (CF). [24] highlights the challenges of Collaborative Filtering such as matrix factorization and sparsity problems. It bridges the gap between CF and DL by combining matrix factorization of Collaborative Filtering approach with Deep Feature Learning method named marginalized Denoising Auto-Encoder (mDA). It identifies latent factors from both user-item ratings and side information for video recommendations and outperforms other approaches in terms of movies and books recommendations and response prediction. [25] predicts watch time of videos by users based on whether the video was clicked (positive) or not clicked (negative) by combining Neural Networks with CF. This approach considers time-dependent attributes of videos which improves the performance and watch time on recently uploaded videos. The problem of data scarcity of Collaborative Filtering in Point of Interest (POI) recommendation is tackled in [26] by combining CF with Semi-Supervised Learning (SSL). A novel model called PACE (Preference and Context Embedding) is developed which is a combination of CF and SSL based on neural networks for analysing interactions between users and POIs. Deep Collaborative Filtering model called Dual Regularized Matrix factorization with deep neural networks (DRMF) is proposed in [27] consisting of multilayered neural networks by stacking Convolutional Neural Networks and gated Recurrent Neural Networks. Here, matrix factorization is combined with deep neural networks for creating latent representation of users and items after extracting their textual information for item recommendations.

3.2 Convolutional Neural Networks (CNN) –

CNNs are a class of deep neural networks. They are artificial neural networks that have some type of specialization for being able to pick out or detect patterns and make sense of them. This pattern detection makes CNNs so useful for analysis purposes. It has mainly 3 layers: Convolutional Layer, Pooling Layer and Fully Connected Layer. CNNs are widely used in recommendation systems. One variation is used in [28] for recommending learning resources where a content based recommendation algorithm is developed using CNN model to predict rating scores between students and learning resources based on text information in learning resources. In this approach, Language Model is employed for its input to train CNN. Latent Factor Model with L1 norm (LFM-L1R) is employed for its output. CNNs are also widely used in Social Networks. CNN incorporates local attention channel which encodes a few trigger words and represents embeddings of those words; and global channel which encodes all the words and represents embedding of the entire microblog to perform hashtag recommendation task [29] in social networks. Another variation is used in user profiles and item recommendations [30,31]. In [30] word embedding is used to represent questions and user profiles. CNN is then used to recommend experts based on their profiles to best answer a newly posted question. In [31], a variation of CNN - Deep Cooperative Neural Networks (DeepCoNN) is used which consists of 2 parallel neural networks. One of them is responsible for learning user behaviours by analysing reviews written by users and other is responsible for learning items by analysing reviews written for items. Authors compare their results with existing systems like MF, PMF, LDA, CTR, HFT-10, CDL.

3.3 Autoencoders –

An autoencoder is a type of artificial neural network, which is not meant for producing classes at the output but to reproduce its input at the output. One variation of autoencoders is implemented in [32] where a model based on autoencoders (A-COFILS) is proposed to reduce the sparsity problem of Collaborative Filtering models. Here, Collaborative Filtering is converted to Supervised Learning (COFILS). Its 5 main steps are: matrices creation (ratings of users for items are specified), data normalization (missing ratings of the matrix are filled with zeros), Autoencoder Architecture selection (adjusting the user item matrix so that it gets fit within the range of Autoencoders' activation function), generating supervised dataset (features of user and items are mapped) and applying regressor (regressor model is trained using the generated supervised learning data, for prediction purpose). Extraction, mapping and prediction are 3 main parts of this approach. Singular Value Decomposition (SVD) is used for extraction part but it is not able to extract non-linear features. So, to resolve this issue, Stacked Denoising Autoencoder (A-COFILS) is used. In mapping phase, ratings of users for items are specified. In prediction phase, supervised Learning is applied for making a model to learn. This approach has a limitation, as the number of new users increase after a threshold, the quality of the system degrades. So, to determine this threshold is still a task of future work. A novel model AutoRec is developed in [33] which is based on autoencoder framework for video recommendation and outperforms existing neural network model for Collaborative Filtering. In [34], 3 problems of recommendation systems are highlighted via Collaborative Filtering - Sparsity Problem, Cold Start Problem, Trustworthiness Problem. Trust Information of users is helpful for new users (cold start users) on social networks. To overcome these issues, a model called Deep Learning based Matrix Factorization (DLMF) is proposed in [34] which is used for trust-aware recommendations in social networks. Another model of autoencoders in the field of social networks is implemented in [35], which is used to tag users' profile to extract deep features from them. In this way, user recommendation performance is increased by updating users' profile by this proposed method. One more limitation of Collaborative Topic Regression (CTR) is highlighted in [36] as the learned representation of items may be not effective. So, to overcome this issue, Stacked Denoising Autoencoder
Deep Neural Networks designed for capturing information from sequences and they really work well with time series data. They make use of temporal dynamic behavior of users in recommendation systems. Existing video recommendation methods consider that users’ interests are static. [37] resolves this problem by considering dynamic users’ interests for videos. It uses RNN and considers three factors named: (1) Video semantic embedding which represents videos according to their content information, (2) User Interest modelling which represents users’ choice of playing the video, (3) User Relevance Mining - which provides additional attributes for improving the performance of recommendation. RNN is widely used in session-based recommendations. In [38] kNN is used to measure the performance of recommendation of next item in an anonymous session. Combination of kNN with RNN is effective in improving recommendation. Also, methods are done on short session based data instead of long term based sessions. [39] has used RNN combined with GRU for item-to-item recommendation in long term based sessions. GRU is modified by including session-parallel mini-batches, mini batch based output sampling and ranking loss function. The proposed method outperforms all the baselines used earlier. Another application of RNN can be seen in social networks. In [40] the product purchasing rate of users is determined after mentioning those products in tweets of Twitter. The proposed algorithm outperforms the best baseline by up to 46.7%. Another variation of Recurrent Neural Networks - Long Short Term memory - LSTM-RNN for Hashtag Recommendation on Twitter data is implemented in [41]. This approach consists of four components - word embeddings generation (skip-gram model), sentence composition (CNN), tweet composition (RNN), Hashtag Classification. Previously implemented techniques such as SVM or Collaborative Filtering make use of TF-IDF which represents tweets but ignores meaningful information hidden in those tweets. But proposed algorithm outperforms other methods. A novel model called Multi-View Recurrent Neural Network (MV-RNN) is implemented in [42] to have a good impact on the problem of item cold start. It extracts latent features from images and textual information of items. Three features are taken into consideration: concatenation, fusion by addition, fusion by reconstruction, thus giving rise to 3mDAE (Marginalized Denoising AutoEncoder).

### 4. Discussion

A summary of all research articles is given below in Table 1-5. Results are given in form of Mean Average Precision (MAP), Precision, Recall, F-Score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Square Error (MSE), Area Under Curve (AUC), Mean Reciprocal Rank (MRR).

<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>DATASET</th>
<th>RESULTS / CONCLUSION</th>
</tr>
</thead>
</table>
| Deep Collaborative Filtering model by combining Bayesian Stacked Denoising Autoencoders to remove sparsity problem of CF [23] | 3 datasets: 1) CiteULike-a 2) CiteULike-t 3) Netflix | MAP - 0.0514  
Recall – 70.42  
Conclusion: Performance is boosted when DL for content information is combined with CF for ratings matrix. |
| Deep Collaborative Filtering model by combining marginalized Denoising Auto-Encoder (mDA) for movies recommendation and response prediction [24] | 3 datasets: 1) MovieLens-100K 2) MovieLens-1M 3) Advertising | RMSE: MovieLens-100K - 0.8849±0.0167, MovieLens-1M - 0.8304±0.0057,  
AUC on Advertising Dataset: 0.8115  
Conclusion: This approach outperforms other approaches in terms of movies and response prediction. |
| A-COFILS: a model based on Stacked Denoising Autoencoder. 3 main parts are: Extraction, mapping and prediction. A-COFILS is used for extracting non-linear features. In mapping phase, ratings of users for items are specified. In prediction phase, supervised Learning is applied for making a model to learn [32]. | 3 datasets: 1) MovieLens 2) R3 Yahoo! Music 3) MovieTweetings | MAE: MovieLens 100K - 0.697, MovieLens - 1M - 0.661, Yahoo Music - 0.895, MovieTweetings - 1.304;  
RMSE: MovieLens 100K - 0.885, MovieLens - 1M - 0.838, Yahoo Music - 1.152, MovieTweetings - 1.747  
Conclusion: A-COFILS (Autoencoder COFILS) reduces MAE compared to COFILS (Collaborative Filtering to Supervised Learning) on all datasets. |
| Deep collaborative filtering model including Neural Networks for predicting watch time of videos by considering time-dependent attributes of videos [25] | YouTube | Weighted, per-user loss - 34.6%  
Conclusion: Deep Collaborative Filtering model outperforms previous approaches for matrix factorization on YouTube. |
Autorec - Autoencoders for video recommendation [33]

2 datasets:
1) MovieLens
2) Netflix

RMSE:
MovieLens - 0.782,
Netflix - 0.823

Conclusion: The proposed model Autorec is based on autoencoder framework and outperforms existing neural network model for Collaborative Filtering.

Taking real time and dynamic interests of users’ using RNN [37].

Google+ website - cross network dataset

Precision - 0.0350,
Recall - 0.0399,
F-Score - 0.0340

Conclusion: The proposed method considers users’ dynamic and real time interests.

**Table 2: Summary of research articles for Session-Based Recommendations**

<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>DATASET</th>
<th>RESULTS / CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrent Neural Networks with Gated Recurrent Units (GRU4REC) and session-based kNN for recommending next item in an anonymous session [38].</td>
<td>2 datasets: 1) Public e-commerce dataset used in the TMall competition (TMALL), 2) Music playlists from the platforms last.fm (LFM), artofthemix.org (AOTM), and 8tracks.com (8T).</td>
<td>Conclusion: The kNN method outperformed GRU4REC both in terms of the Hit Rate and MRR in majority cases except for last.fm and AOTM.</td>
</tr>
<tr>
<td>Recurrent Neural Networks with Gated Recurrent Unit (GRU) for item-to-item recommendation in long term based sessions [39].</td>
<td>2 datasets: 1) RecSys Challenge 2015, 2) YouTube like OTT video service platform.</td>
<td>Recall for RCS15 - 0.5065, Recall for Video - 0.5508, MRR for RCS15 - 0.2048, MRR for Video - 0.3381 Conclusion: The proposed method outperforms all the baselines used earlier.</td>
</tr>
</tbody>
</table>

**Table 3: Summary of research articles for Books / Learning Resources Recommendations**

<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>DATASET</th>
<th>RESULTS / CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Collaborative Filtering model by combining marginalized Denoising Auto-Encoder (mDA) to remove sparsity problem of CF [24]</td>
<td>Book-Crossing</td>
<td>RMSE: 3.6513 Conclusion: This approach outperforms other approaches in terms of books recommendations.</td>
</tr>
<tr>
<td>Convolutional Neural Networks for recommending new and unpopular learning resources [28]</td>
<td>Book-Crossing</td>
<td>MSE - 0.28981, Precision - 0.76158 Conclusion: The proposed algorithm is feasible in recommending new and unpopular learning resources.</td>
</tr>
</tbody>
</table>

**Table 4: Summary of research articles for Recommendations in Social Networks**

<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>DATASET</th>
<th>RESULTS / CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Collaborative Filtering model by combining CF and Neural Networks for Point of Interest (POI) recommendation [26]</td>
<td>Gowalla and Yelp check-in dataset</td>
<td>Conclusion: Neural networks can be used to model non-linear complex interactions between users and POIs but at the same time, smoothing is applied on the data to resolve the problem of data scarcity of Collaborative Filtering.</td>
</tr>
<tr>
<td>CNN incorporates local attention channel and global channel to perform hashtag recommendation task [29].</td>
<td>Microblogging Data.</td>
<td>Precision - 0.443, Recall - 0.362, F1 Score - 0.398 Conclusion: The proposed method outperforms the other methods which consider only local or global information.</td>
</tr>
<tr>
<td>Autoencoder for overcoming the problem of trustworthiness for first time users on social networks [34].</td>
<td>Epinions dataset and Flixster dataset</td>
<td>RMSE: Epinions - 1.0736, Flixster - 0.7853, F-Measure: Epinions - 0.7975, Flixster - 0.8742</td>
</tr>
</tbody>
</table>
Conclusion: The proposed trust-aware algorithm outperforms the other state-of-the-art methods on social networks recommendations.

Autoencoders for tagging users' profile to extract deep features from them [35].

2 website datasets: Last.fm Del.icio.us

Conclusion: The proposed algorithm outperforms CF in terms of precision, recall and rank score. It also reveals that deep learning algorithms work better if there are a different number of hidden layers.

RNN for determining the product purchasing rate of users after mentioning those products in tweets of Twitter [40].

Twitter - words used as "cameras" and "mobile"

Accuracy: Mobile - 80.1, Camera - 79.2, Both - 77.7

Conclusion: The proposed algorithm outperforms the best baseline by up to 46.7%.

Recurrent Neural Networks - Long Short Term memory - LSTM-RNN for Hashtag Recommendation on Twitter data [41].

Public Twitter Streaming API was used to collect tweets from 1st June - 31st August 2015.

Accuracy - 28.6, HitRate - 86.4

Conclusion: Previously implemented techniques such as SVM or Collaborative Filtering make use of TF-IDF which represents tweets but ignores meaningful information hidden in those tweets. But proposed algorithm outperforms other methods.

Table 5: Summary of research articles for User Profile and Item Reviews

<table>
<thead>
<tr>
<th>ALGORITHM</th>
<th>DATASET</th>
<th>RESULTS / CONCLUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep Collaborative Filtering model by stacking Convolutional Neural Networks and gated Recurrent Neural Networks for Ratings prediction [27]</td>
<td>4 Datasets: 1) Amazon Instant Video (AIV) 2) Apps for Android (AA) 3) Kindle Store (KS) 4) Yelp</td>
<td>MAE: AIV - 0.6982, AA - 0.9000, KS - 0.5507, YELP - 0.7615, RMSE: AIV - 0.9426, AA - 1.1789, KS - 0.7736, YELP - 0.9843 Conclusion: The model proposed as DRMF achieves gains in rating prediction.</td>
</tr>
<tr>
<td>CNN is used to recommend experts based on their profiles to answer a newly posted question in Community Question Answering (CQA) [30].</td>
<td>Real world dataset is collected from StackOverflow</td>
<td>Conclusion: CNN outperforms this experiment in comparison to other methods such as TF-IDF, Language Model, SSRM, LR, LDA, STM.</td>
</tr>
<tr>
<td>A variation of CNN named Deep Cooperative Neural Networks (DeepCoNN) for making use of valuable information written in reviews of users and reviews for items for recommendation systems [31].</td>
<td>3 Datasets 1) Yelp: Restaurant Reviews 2) Amazon: Product Reviews 3) Beer: Beer Reviews</td>
<td>MSE: Yelp - 1.441, Amazon - 1.268, Beer - 0.273 Average of all these - 0.994. Conclusion: The proposed algorithm DeepCoNN achieved 8.5% and 7.6% improvements on datasets of Yelp and Beer, respectively. On Amazon, it outperformed all the baselines and gained 8.7% improvement on average. Overall, 8.3% improvement is attained by the proposed model on all three datasets.</td>
</tr>
<tr>
<td>Stacked Denoising Autoencoder (SDAE) combined with Probabilistic Matrix Factorization (PMF) to further extend it to Relational Stacked Denoising Autoencoder (RSDA) for improving the performance of tag recommendation [36].</td>
<td>3 datasets: 1) CiteULike-a 2) CiteULike-t 3) MovieLens</td>
<td>Conclusion: The proposed model outperforms state-of-the-art methods.</td>
</tr>
<tr>
<td>Multi-View Recurrent Neural Network (MV-RNN) and Marginalized Denoising AutoEncoder (3mDAE) are combined to have a good impact on the problem of item Taobao and Amazon</td>
<td>Recall: Taobao - 1.174, Amazon - 2.995. MAP:</td>
<td></td>
</tr>
</tbody>
</table>
5. Conclusion
We presented deep learning in various recommendation systems such as entertainment, user and item reviews, social networks, session-based and learning resources. Deep learning algorithms resolve the cold-start item and sparsity problems of existing methods of recommendation such as content-based and collaborative filtering. Deep Collaborative Filtering model (DCF) is heavily used in the category of entertainment. Mainly, Autoencoder is used combinedly with Collaborative Filtering. Performance is boosted when DL for content information is combined with CF for ratings matrix. In session-based recommendations, RNN with Gated Recurrent Units is used majorly used and outperforms in terms of Hit Rate, MRR and Recall. CNN is feasible to recommend new learning resources with a good precision level. CNNs, RNNs and Autoencoders are used for Hashtag and users’ profile recommendations in social networks and give a better result in their respective objectives than the existing algorithms.

6. Future Work
Hybrid models by combining Deep learning approaches and Collaborative Filtering can be applied on distributed optimization algorithms to reduce computational costs [24]. Heterogeneous context graphs, timestamp attributes, location and dynamic users’ interests can be explored in future for analyzing user preference in various recommendation systems [26,32,37].

References
Proceedings of the sixth ACM conference on Recommender systems (pp. 35-42). ACM.


