Journal of Physics: Conference Series

A Survey of Deep Learning Approaches for Recommendation Systems

To cite this article: Jun Yi Liu 2018 *J. Phys.: Conf. Ser.* **1087** 062022

A Survey of Deep Learning Approaches for Recommendation Systems

Jun Yi LIU

School of Computing, University of Leeds, Leeds, LS2 9JT, UK

junyiliu0530@qq.com

Abstract. Due to the explosive information, recommendation system has been an important part of people's life. It can suggest or predict information based on the user's preference to help user save time. As deep learning develops, the application of deep neural network in related research is increasingly prevalent. This paper provides a survey of recommendation systems, which focuses on deep learning approaches and the system of applications. The detailed description of each recommendation system is explained, and the related datasets are briefly introduced. In this paper, all references are published ranging from 2014 to 2017, which presents an overview of the current progress in this area.

1. Introduction

The development and popularization of Internet allows people to enjoy the great convenience of the Internet, but people are also nagged by the problem of "information overload", that makes people difficult to find the real results to meet their needs from the massive information. Recommendation system [1], as the effective tool to handle the problem of information overload, has attracted the attention of researchers.

Recommendation system can be divided into three core components: user model, product model and recommendation algorithm. User model is generally used to acquire, represent and store user's data, which can be collected through two ways: the explicit (e.g. rating and thumbs-up) and the implicit (e.g. browsing history and purchase records). Product model refers to the characteristics of the product. Different types of recommendation systems must consider the characteristics of their products, such as image-based recommendation system need to focus on at least colors, texture and shape. Recommendation algorithm mainly discover the user's preferences and interests by mining the history data, and recommend the similar products to the user. More and more researchers concentrate on recommendation algorithms and attempt to the accuracy of personalized recommendation.

The personal recommendation algorithms are normally divided into Collaborative Filtering (CF), Content-Based Filtering (CBF) and hybrid recommendation [2]. Goldberg et al. [3] prosed the definition of collaborative filtering, and apply to Tapestry, which is the first recommender system. Collaborative filtering algorithms recommend the products to a user depending on the user's interests that are able to match other people with similar interests. Content-based filtering algorithms originated from information retrieval [4, 5], and pay more attention on the characteristics of the products. In real applications, more and more new users and new products have increased the workload of CBF, while CF also faces some challenges such as sparsity and cold start. Thus, researchers proposed the more accurate hybrid recommendation, which based on CF and CBF [6].

Since Hinton et al. [7] introduced a fast, greedy unsupervised learning algorithm, called Deep Belief Network (DBN), the deep learning field has attracted more and more researchers [8]. In recent years, many researchers deployed deep learning algorithms into recommendation systems in order to increase the accuracy and solve the problems of recommendation systems. There are several frequent and classical deep learning algorithms such as DBN, Convolutional Neural Network (CNN) [9], Recurrent Neural Network (RNN) [10], Deep Autoencoder [11]. DBN, as a class of deep generative models, is composed of a number of Restricted Boltzmann Machines (RBMs) [12] that is a stochastic neural network consisting of one layer of visual units and one layer of hidden units. CNN is a kind of multi-layer supervised learning neural network, with each module consisting of a convolutional layer and a pooling layer. It uses the Gradient Descent and frequent iterative learning to improve the network accuracy. RNN is used as a class of deep generative architecture for modeling and generating the sequential data. Deep Autoencoder can reduce the dimensionality of data through learning low-dimensional codes.

Depending on the above algorithms, many researchers combined the conventional recommender systems with the deep learning algorithms. For example, Oh et al. [13] exploited a new model, based on DBN, to analyze the user preference for personalized news recommender system. Zahálka et al. [14] presented an interactive and multimodal content-based venue explorer, using CNN to generate the visual features. Lin et al. [15] developed a hierarchical deep CNN framework for clothing retrieval in a recommendation system. Cui et al. [16] proposed a video recommender system, which combines DBN with CF. Lei et al. [17] explored image recommendations according to a dual-net deep network model, which consists two CNNs and a full-connection neural network. Chiliguano et al. [18] presented a music recommender with CNN. Wu et al. [20] developed a real-time customized recommendation service on an e-commerce system with a deep RNN. Zuo et al. [21] proposed a tag-aware recommender system that combined Sparse Autoencoder with user-based CF. Unger et al. [22] introduced a context-aware recommender system by utilizing Autoencoder to analyze latent context data. Deng et al. [23] presented a trust-based recommendation service by using Deep Autoencoder in social networks.

The remaining of the paper is organized as follows. Section 2 introduces several datasets in experiments. Section 3 reviews the deep learning models in some recommendation systems. Section 4 describes the applications. Section 5 presents the conclusions and some thoughts of future.

2. Datasets

2.1. The News Recommender [13] Dataset

The News Recommender dataset is collected on two implemented platforms: Google Chrome Extension “Daum News Tracker,” and Android application “KECI News.” The collection consists of timestamps, IP address, twitter ID, and URL of articles. The training data includes 1500 articles from 70 users, and the test data includes 2000 articles of 50 users.

2.2. The Venue Recommender [14] Dataset

The Venue Recommender dataset consists of the visual content, the associated text content, user information, venue list and the user-venue mapping matrix, which are split into two very different metropolises: New York and Amsterdam. The part of New York contains 7,246 venues, and the corresponding 1,072,181 multimedia collections. The part of Amsterdam contains 693 venues and the corresponding 55,990 multimedia collections. The main challenge is to process such large-scale data in an interactive, responsive manner, and to make recommendation depending on the comparatively smaller-scale and sparse data.

2.3. The Clothing Retrieval Recommender [15] Dataset

This dataset contains 161,234 clothes images collected by crawling the images from the Yahoo shopping sites. The dataset is categorized by a hierarchical tree which is defined according to the product database from the Yahoo shopping sites.

2.4. The Video Recommender [16] Dataset

In this system, the video data derived from Netflix movie score data published in 2005. It contains 480,189 users in about 17,770 movies and the corresponding about 10 hundred million score records. The dataset is split into two parts: 20% for training and the rest for testing.

2.5. The Image Recommender [17] Dataset

The image and user's information in dataset are from Flickr through its API. It contains 101,496 images, 54,173 users, 6,439 groups and 35,844 tags. In consideration of the inactivity and diver interests of user, users that have less than 40 or more than 200 favorite images are filtered out from testing. To improve the accuracy of training, users that have interests in less than 80 or more than 280 clusters are filtered out from training. Finally, there are 8,616 users for training and 15,023 users for testing. For each user, 20 images are randomly selected from favorite images and the rest for testing.

2.6. The Music Recommender [18] Dataset

The dataset derived from Million Song Dataset [19] which consists of audio features and metadata of a million popular music tracks. Considering the mismatches between song ID and metadata, and the size of the dataset, only the users with more than 1000 played songs and the song identifiers of 1,500 most played songs are selected. The filtered dataset has 65,327 triplets.

2.7. The E-Commerce Recommender [20] Dataset

This dataset comes from the web log of June 1st, 2015 of the Kaola e-commerce system. It contains 232,326 records, 37,667 unique users, and 1584 different items. These records can be grouped into 27,985 sessions: 60% sessions for training; 20% sessions for validation; and the rest for testing.

2.8. The Tag-Aware Recommender [21] Dataset

In the experiments, two real web datasets are evaluated: Last.fm and Del.icio.us [24]. The first dataset comes from Last.fm online music system, which allows users to tag music tracks and artists. The second dataset come from a popular social bookmarking system Del.icio.us, which allows users to tag personal web bookmarks. To reduce the computation, tags are used more than 5 times in Last.Fm and selected in Del.icio.us 15 times. Moreover, 80% of the dataset is training data, and the other 20% is testing data.

2.9. The context-aware Recommender [22] Dataset

The dataset collected from a variety of sensors of an Android application, such as location, time, ringer mode, speed, battery, activity recognition, microphone, light, accelerometer, rotation, gyroscope, and Wi-Fi. The experiment lasts over a period of 4 weeks, and collects 60 undergraduate students' records. The data of 6 students that rated less than 20 points of interest (POIs) was filtered out. Finally, the system collected 7,416 events (like, dislike, check-in) and 669,760 raw sensor records.

2.10. The Trust-Aware Recommender [23] Dataset

In this system, there are two datasets: the Epinions dataset and the Flixster dataset. Epinions is a product review website that allows users to rate products, submit personal reviews, and specify whom to trust. Flixster is a social networking service that allows users to rate movies and add friends. Table 1 shows the statistics of the two datasets.

Table 1 The statistics of two datasets.

	Epinions	Flixster
Users	49,290	104,9445
Items	139,738	492,359
Ratings	664,824	8,238,597
Trust Relations	487,181	26,771,123

3. The Deep Learning Models

To achieve higher precision, Oh et al. proposed a deep neural network model which is modified from DBN. The aim of the model is to build user profiles. The input layer has 5 different features: term frequency (TF), inverted term frequency (ITF), title (T), first sentence (FP), and cumulated preference weight (CP). Values of TF and ITF are greater than 0, value of CP is between 0 and 1, and values of T and FP are Boolean values. Through 3-layer perceptron calculation, the model can output the users' profile. Zahálka et al. utilized a deep CNN which conceived by Krizhevsky et al. [25], trained on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 dataset of 1000 semantic concepts [26]. It contains 8 layers: 5 convolutional layers and 3 fully connected layers, the output of which is the set of the total feature representation of all individual images. Lin et al. also used a deep CNN for learning visual features. The deep CNN proposed by Krizhevsky et al. [25] is used for pre-training, but it is non-optimal for the clothing domain. To this end, a new fully connected layer is added between the 7th and 8th layer of the original network to bridges mid-level features and semantics, simultaneously learning clothing image representations and a set of binary coded functions.

Cui et al. deployed a DBN model to figure out the user profile and find neighborhood relationships. DBN consists of multi-layer RBM, and the process of training DBN consists of two stages: initialization and tuning. In this DBN model, Markov Chain Monte Carlo process (MCMC) is deployed to simulate the Maximum Likelihood law method with Gibbs sampling as the transfer function. Lei et al. proposed a comparative deep learning (CDL) architecture to learn the relative distances between every two images and one user. The CDL model consists of three sub-networks, one is for user, and the others are for positive and negative images. Chiliguano et al. recreated a similar CNN model to classify music genre [30]. The CNN model consists of 2 convolutional layers, 2 max-pooling layers and one full-connection layer. Finally, a logistic regression layer with 10 softmax units classifies the music genre.

Wu et al. used a deep RNN (DRNN) architecture to model user's history and make real-time recommendations in e-commerce system. This DRNN is inspired by [32], but it has several differences. First, the role of this DRNN is to track the user's browsing patterns. Users often view a number of web pages to find the desired product. The intuition of this DRNN model is to make a real-time recommendation to reduce the number of web pages to help users find product quickly. Zuo et al. adopted sparse autoencoder neural network to process tag information. The network consists of three layers: one input layer, one hidden layer and one output layer. The input layer and hidden layer form an encode, and the identical hidden layer and output layer form a decoder. Unger et al. used an autoencoder to extract latent context features from a set of mobile sensors with sensor data (e.g. location, time, light, etc.) in low dimensionality. Yu et al. presented a deep network to learn the deep CNNs and tree classifier for large-scale detection of privacy-sensitive object. The first phase of network is to learn the common representations for all the privacy-sensitive object classes via the 3 commonly-shared convolutional layers, the second phase is to learn the class-specific representations for the visually-similar privacy-sensitive object classes in the same group, and the third is to replace the flat softmax-lay for the tree classifier via the last layer. Deng et al. utilized a deep autoencoder and decoder model to learn the initial values of latent features of users and items. The encoder network transfers the high-dimensional input data to the low-dimensional code, on the contrary, the decoder network reconstructs the original data from the low-dimensional code. In this paper, continuous RBM (CRBM) is used to pretrain weights with a visible layer and a hidden layer that correspond to the input data and the output data respectively.

4. Applications

Considering the frequent changes of hot news topics, Oh et al. introduced a personalized news recommender system. The system consists of two phases: preference analysis and news recommendation. In the phase of preference analysis, a modified 3-layer DBN is employed to capture user preference. The news recommendation utilizes the profiles to rank the upcoming news that refers to the Term Frequency - Inverted Document Frequency (TF-IDF) scores. In the evaluation, this news recommender system achieves average accuracy of 54% from the top-10 ranked news in 5 days, which is higher 46% than recommendation of news randomly 37%. Zahálka et al. introduced an interactive multimedia content-based venue recommender system called City Melange. This system consists of three steps: data collection, user and venue topical analysis, and interactive city exploration. The dataset is used to generate the semantic topics for each venue and social media user by clustering on visual features that produced by CNN and text features that produced by Latent Dirichlet Allocation (LDA). During the evaluation, two datasets of New York and Amsterdam demonstrate that City Melange is capable to recommend suitable venues.

Cui et al. introduced a video recommender system based on DBN and CF. This system transfers the User-Item Matrix to 0-1 matrixes and use them as the input of DBM model, the parameters of the DBN structure are settled by training the dataset. During the evaluation, the proposed system achieves a higher accuracy than the content-based recommendation and CF recommendation. Lei et al. introduced an image recommender system based on a CDL method to achieve the hybrid representations. In this paper, word2vector [27] is used to generate user vectors by random sampling through the social data to form a set of triplets for training of CDL. The personalized image recommendation task with the CDL method achieves much better performance than two disconnected sub-networks as well as SIDL [28], Bag of Words (BoW), ImageNet [25], LMNN [29], and Social+LMNN [28]. Chiliguano et al. introduced a hybrid recommender system that combines real-world users' information and high-level audio data. A CNN architecture is used to represent an audio segment in a multi-dimensional vector, which represents the probability of a song to belong to a specific music genre. Then Estimation of Distribution Algorithms (EDAs) [31] are used to model user profiles. Finally, the contend-based filtering module computes the similarity between user profiles and song vectors, and recommend top-N songs based on the ranked similarities. Although this hybrid recommender system makes recommendations successfully, the accuracy is similar to the content-based recommender system, partly because a limited number of genres for song representation.

Wu et al. introduced a real-time recommendation service for e-commerce system through exploit current viewing history of the user. The real-time recommendation service is implemented by the DRNN model. First, a DRNN model extracts how users browse the web pages. During the evaluation, the DRNN model is truly outperforms CF model. This real-time recommendation service has been successfully deployed on the NetEase's e-commerce system, Kaola. Zuo et al. explored tag-aware recommender systems based on the sparse autoencoder model. In social tagging systems, representations of users' profiles contain three elements: users, items and tags [33], the sparse autoencoder model used in tag-aware recommender systems can achieve better performance than traditional user-based CF and clustering-based CF. Unger et al. introduced a novel context-aware recommendation model based on the latent context. This model consists of latent context collection using autoencoder and Principal Component Analysis (PCA), user-item data collection, explicit context extraction, rating model generation, latent context matrix factorization recommendation (LCMF). Yu et al. introduced an image privacy protection system by identifying sensitive objects via deep multi-task learning. Deng et al. introduced trust-aware recommendations in social networks. The recommendation process consists of deep autoencoder and matrix factorization (MF) to synthesize the users' interests and their trusted friends' interests with the impact of community effect for recommendations. During the evaluation, the proposed approach outperforms the other state-of-the-art methods on accuracy.

5. Conclusion

As the application of recommender systems becomes more and more widespread, such as music, video, social network, etc., recommender system has not only been an active area but also a challenging task. In recent years, a trend of the research is to address the issue by establishing deep neural networks and learning the latent information features in order to model the relationship of information. Most of conventional method construct handcrafted features for recommendation based on domain knowledge, which is time-consuming, inefficient, the sparsity problem and cold-start problem. While the methods with deep learning model are driven by data, which are capable to extract features from raw data automatically.

In this paper, we present several recommender systems based on deep learning techniques. The deep learning neural network for recommender system is devoted to extract users and items feature or latent and explicit features. For example, DBN is usually used to build users' profile, CNN is usually used to extract image or visual features, and autoencoder is usually used to find latent or implicit features. More and more research concentrates on applying to the real-world conditions and providing real-time prediction. As the accuracy is improved, the cost-efficiency and low resource consumption in practical application would be more important gradually.

References

[1] Resnick P, Varian H R. Recommender systems [M]. ACM, 1997.

[2] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. IEEE Transactions on Knowledge & Data Engineering, 2005, 17(6):734-749.

[3] Goldberg D, Oki B M, Oki B M, et al. Using collaborative filtering to weave an information tapestry [J]. Communications of the Acm, 1992, 35(12):61-70.

[4] Belkin N J, Croft W B. Information filtering and information retrieval: two sides of the same coin?[J]. Communications of the Acm, 1992, 35(12):29-38.

[5] Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval: addison wesley [J]. Computer Science & Information Technology (CS & IT, 1999.

[6] Burke R. Hybrid recommender systems: survey and experiments [J]. User Modeling and User-Adapted Interaction, 2002, 12(4):331-370.

[7] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[M]. MIT Press, 2006.

[8] Li D. A tutorial survey of architectures, algorithms, and applications for deep learning [J]. Apsipa Transactions on Signal & Information Processing, 2014, 3:-.

[9] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: a convolutional neural-network approach.[J]. IEEE Transactions on Neural Networks, 1997, 8(1):98-113.

[10] Martens J, Sutskever I. Learning recurrent neural networks with hessian-free optimization[C]// International Conference on Machine Learning, ICML 2011, Bellevue, Washington, Usa, June 28 - July. DBLP, 2012:1033-1040.

[11] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786):504.

[12] Fischer A, Igel C. An Introduction to restricted boltzmann machines[C]// Iberoamerican Congress on Pattern Recognition. Springer, Berlin, Heidelberg, 2012:14-36.

[13] Oh K J, Lee W J, Lim C G, et al. Personalized news recommendation using classified keywords to capture user preference[C]// International Conference on Advanced Communication Technology. IEEE, 2014:1283-1287.

[14] Zahálka J, Rudinac S, Worring M. Interactive multimodal learning for venue recommendation [J]. IEEE Transactions on Multimedia, 2015, 17(12):2235-2244.

[15] Lin K, Yang H F, Liu K H, et al. Rapid clothing retrieval via deep learning of binary codes and hierarchical search[J]. 2015:499-502.

[16] Cui H, Qin X. The Video Recommendation system based on DBN[C]// IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. IEEE, 2015:1016-1021.

[17] Lei C, Liu D, Li W, et al. Comparative deep learning of hybrid representations for image recommendations[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016:2545-2553.

[18] Chiliguano P, Fazekas G. Hybrid music recommender using content-based and social information[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:2618-2622.

[19] Mcfee B, Bertin-Mahieux T, Ellis D P W, et al. The million song dataset challenge[C]// International Society for Music Information Retrieval Conference, Ismir 2011, Miami, Florida, Usa, October. DBLP, 2012:909-916.

[20] Wu S, Ren W, Yu C, et al. Personal recommendation using deep recurrent neural networks in NetEase[C]// IEEE, International Conference on Data Engineering. IEEE, 2016:1218-1229.

[21] Zuo Y, Zeng J, Gong M, et al. Tag-aware recommender systems based on deep neural networks [J]. Neurocomputing, 2016, 204:51-60.

[22] Unger M, Bar A, Shapira B, et al. Towards latent context-aware recommendation systems [J]. Knowledge-Based Systems, 2016, 104:165-178.

[23] Deng S, Huang L, Xu G, et al. On deep learning for trust-aware recommendations in social networks [J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28(5):1164-1177.

[24] Cantador I, Brusilovsky P, Kuflik T. Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011)[C]// ACM Conference on Recommender Systems, Recsys 2011, Chicago, Il, Usa, October. DBLP, 2011:387-388.

[25] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[26] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[27] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [J]. Computer Science, 2013.

[28] Liu S, Cui P, Zhu W, et al. Social embedding image distance learning[C]// ACM International Conference on Multimedia. ACM, 2014:617-626.

[29] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification [M]. JMLR.org, 2009.

[30] Kereliuk C, Sturm B L, Larsen J. Deep learning and music adversaries [J]. IEEE Transactions on Multimedia, 2015, 17(11):2059-2071.

[31] Liang T, Liang Y, Fan J, et al. A hybrid recommendation model based on estimation of distribution algorithms [J]. Journal of Computational Information Systems, 2014, 10(2):781-788.

[32] Hermans M, Schrauwen B. Training and analysing deep recurrent neural networks [J]. Advances in Neural Information Processing Systems, 2013:190-198.

[33] Marinho L B, Schmidtthieme L. Collaborative tag recommendations [J]. 2007:533-540.