International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I10P214
Volume 70 Issue 10, 124-143, October 2022

Review Article

Text Based and Image Based Recommender Systems: Fundamental Concepts, Comprehensive Review and Future Directions

Anu Mathews[1], N. Sejal[2], K. R. Venugopal[3]

[1]K S Institute of Technology, Department of AI&ML, VTU, Bengaluru, India.
[2] BNM Institute of Technology, Department of AI&ML, VTU, Bengaluru, India.
[3]Bangalore University, Bengaluru, India.

[1]Corresponding Author : anu.mathews77@yahoo.com

Received: 05 July 2022 Revised: 25 September 2022 Accepted: 05 October 2022 Published: 19 October 2022

Abstract - *The exponential growth of data on the Internet leads to the information overload problem, wherein users are presented with a huge mixture of irrelevant and relevant data, making their decision-making process complicated and time-consuming. Recommender Systems are software agents that learn the preferences of individual users and give recommendations accordingly. The availability of exploitable data, including implicit and explicit user feedback, decides these systems' performance. Machine learning algorithms have increased the efficiency of recommender systems by providing recommendations to users based on the users' visual preferences. This paper reviews and classifies recommender systems based on their application domains and provides insights into the underlying concepts, including selecting features and algorithms under each classification. The challenges in developing recommender systems are discussed, considering which e-commerce marketplaces can be transformed to provide better customer satisfaction.*

Keywords - *Feature extraction, Similarity, Precision, recall, Prediction, Recommendation, Convolutional Neural Network (CNN), Deep learning.*

1. Introduction

With the spike in the usage of web services, software agents providing a recommendation of relevant items play a significant role in our lives. Recommender Systems (RS) are powerful tools that help users find useful information regarding the items that might interest them, making their decision-making quick and easy. Considering the revenue generated from e-commerce business, by knowing the individual user's preferences, there is an increase in the chance of purchase of the recommended items and also a decrease in the probability of losing a customer. The underlying filtering algorithm determines the accuracy of recommendations given by RS. The major objective of an RS is to give users exposure to items that can be interesting to them and get the users to purchase those items. Users' interaction with Recommender System consists of input from the user in item ratings and the output to the user in the form of recommendations. The number of ratings the user has to provide, the time taken to register, detailed information about the item to be rated, the rating scale, and user control in setting preferences are some issues related to taking input from the user. Number of good and relevant recommendations, number of recommendations which improve the trust, number of unknown recommendations, the information provided about each recommended item, techniques of generating more recommendations and confidence in prediction are some factors deciding the output to the user [1].

Several research studies have reviewed the different types of recommendation techniques. [2], [3], [4], [5], [6], [7], [8], [9] Studies reveal that the choice of features/attributes and algorithms for information retrieval impact the performance of RS. The other challenge faced in RS is the understanding of dynamic user preferences.

Table 1 gives a summary of the reviewed past surveys. This survey contributes to the existing knowledge of RS by providing a different perspective on the various applications of recommender systems by categorizing them into text- and image-based recommender systems.

The various concepts, methodology/algorithms and various metrics used for the performance evaluation of the different categories of the services provided by recommender systems are reviewed in this paper.

The major points discussed in this paper are as follows:

- Recommender Systems fundamentals: Important techniques with algorithms and the different performance evaluation metrics used in RS are presented.
- Review of the various Text-based and Image-based Recommender systems: RS are categorized into text-based and image-based, depending on the input query

and further classified based on their application domains.
- Algorithms and Performance: The underlying concepts and algorithms used in the different RS are discussed along with their performance.
- Key challenges and research directions in each of the categories of recommender system: Challenges involved in developing recommender systems for each of the different applications are discussed. A combination of techniques has been provided as future research directions to improve user satisfaction and sales margin.

1.1. Research Gaps Identified

With a textual query, there happens to be a semantic gap between the real intention of a user's search and his understanding of the object. It demands a good feature representation as a requirement for image recommendation with high performance, which in turn comes with the requirement for better feature extraction and feature selection techniques. Feature extraction from images using deep learning plays an important role in such a situation because of the difficulty in manually defining a good feature set. Deep Learning techniques can learn a hierarchy of features much more efficiently and can be used for image recommendation tasks with images as input. Transfer learning and fine-tuning can be used to reduce the overhead requirement of large amounts of labelled training data.

The rest of the paper sections are organized as follows. Section 2 discusses the fundamentals of RS and their basic classification. Section 3 reviews the various text-based recommender systems where the input query is in the form of text. Section 4 reviews the various image-based rec- ommender systems, which generate recommendations by analyzing the images. Section 5 concludes the paper by discussing the open research issues in the field of recommender systems and deriving an outlook for further research.

2. Recommender Systems Fundamentals

A recommender system provides relevant suggestions of items to a user by understanding the user's preferences and analyzing the behaviour of this user or other similar users by exploiting user-item interactions. The user preference model incorporates the information necessary for generating recommendations, like the basic user information, browsing patterns and feedback. With the generated user model, the Recommender system provides an item list new to the current user by calculating the relevance scores for each item and presenting the Top N relevant items to the user. Fig. 1 shows the steps involved in RS using machine learning algorithms.

The various methods used in implementing Recommender Systems include probabilistic approaches, Bayesian networks, nearest neighbors algorithms, neural networks, genetic algorithms, fuzzy models etc. Data reduction can be achieved using Singular Value Decomposition, Principal Component Analysis etc. The feedback mechanism in an RS could be explicit feed- back or implicit feedback. In explicit feedback, the user shows positive or negative interest in an item by providing a rating or leaving a review. In contrast, the implicit feedback mechanism acquires the user's preference for searching or buying actions.

2.1. Classification of Recommender Systems

The user model for recommendation varies depending on the implementation technique used to develop the recommender system. The three main categories of recommendation methods to achieve the task of generating recommendations are as follows:

Table 1. Summary of Past Surveys

Author	Year	Approach
Xiwang Yang et al. [3]	2014	Comparison of the various approaches in social recommender systems using collaborative filtering, based on the complexity of model-training and accuracy of recommendations.
J. Borra`s et al. [9]	2014	Survey on the different types of interfaces and the functionalities provided by the various types of recommender systems in tourism.
Denis Kotkov et al. [5]	2016	Classification of serendipity-based recommendation algorithms
Mehdi et al. [6]	2016	Overview of the metrics used to test active learning strategies for collaborative filtering and compare the various approaches utilized in recommender systems.
M. Kunaver et al. [8]	2017	Survey the diversification process's effect on recommendations and the various diversification algorithms.
Norha M et al. [7]	2018	Review about the incorporation of context into recommender systems
Shini et al. [4]	2020	An extensive study on the features of systems providing travel recommendations and their limitations.

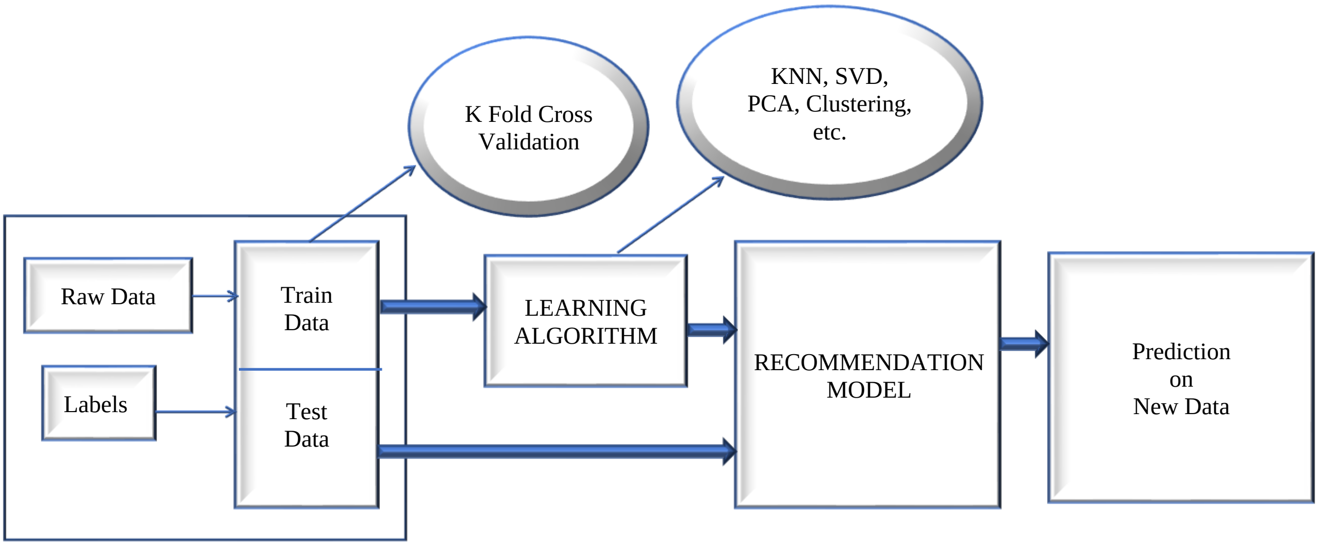

Fig. 1 Basic Steps in Recommender Systems

2.1.1. Collaborative Filtering

Collaborative methods utilize past interactions of users with items stored in the form of a "user-item interactions matrix" to provide new and unique recommendations. The past interactions help to identify similarities between users or items, which can be used for making predictions. The steps involved in the Collaborative filtering algorithm are:

- Create the User Item Utility Matrix from user feedback to represent the degree of preference of each user-item pair.
- Find the distance between the users in the Utility Matrix to find out the users with similar interests.
- Recommend the items with the larger Cosine Similarity. (The larger the cosine, the smaller the angle between two users, implying that they have similar interests.)

"Cold Start" is an issue in Collaborative filtering (CF), wherein generating recommendations for new users or recommending a new item to any user becomes an impossible task. Also, user-item interactions may be too few to be handled efficiently. Algorithms for collaborative filtering are based on two approaches known as model-based and memory-based methods. Memory-based methods work with the recorded user-item interactions without any model and utilize nearest neighbours search, where it finds the nearest users from a target user and suggests the items with maximum popularity among these neighbouring users. An underlying model of user-item interactions is used by model-based methods to generate new predictions.

Model-based methods try to predict the users' rating for a particular item by using the following methods:

- *Clustering Algorithms*: DBSCAN, K-means clustering, and K-Nearest Neighbours are RS's most frequently used clustering algorithms.
- *Matrix Factorization-based algorithms*: In these algorithms, the user-item interaction matrix is factorized into two smaller matrices, which can later be used to get the interaction matrix back. Singular Value Decomposition, Probability Matrix Factorization and Non-Negative Matrix Factorization are used for dimensionality reduction. Limitations of Matrix Factorization techniques are:
    - ➢ "Cold start" prevails because of the absence of a feature vector for new items; hence, an item or user not present in the training set cannot be used for querying.
    - ➢ It recommends popular items to all users and does not reflect the specific user's interests.
    - ➢ Complex relations between users and items cannot be captured because these algorithms utilize the simple inner product of feature embeddings.
- *Deep Learning Algorithms*: These are machine learning algorithms based on artificial neural networks. These algorithms address the limitations of the Matrix Factorization methods. The main advantage of these algorithms is their capability to discover low-dimensional features that capture the complex underlying structure of the input with high dimensionality. Deep neural networks, Graph Neural Networks, Deep Belief Networks and Recurrent Neural Networks are some of the commonly used Deep learning architectures in RS.

2.1.2. Content-Based Filtering

Content-based recommendation methods use auxiliary information of users and items, after which a model is built using the available "features" of items previously preferred by the user. Each of the available item's features is then compared to the user's profile, and the "top-N" higher similarity items are suggested to the user as recommendations. The steps involved in the content-based recommendation are:

1. Create the User Item Utility Matrix from user feedback to represent the degree of preference of each user-item pair.
2. Create an Item profile vector for each item using its characteristic features.
3. Create a User profile vector using the Utility matrix to describe user preference.
4. Compute the similarity between the item's feature vectors and the user's preferred items vectors and recommend top items to the user.

2.1.3. Hybrid Recommendations

Hybrid recommendation techniques combine the characteristics of two or more recommendation techniques. It can be used to avoid cold-start and data sparseness problems. Many existing hybrid recommendation models use a combination of CF and other recommendation techniques.

2.2. Similarity Metrics

Similarity metrics are mathematical measures that can be used to determine the similarity between two vectors. Similarity metrics mostly used in RS are:

2.2.1. Cosine Similarity

This metric uses the cosine value of the angle between the vectors.

$$sim(a,b) = \frac{a.b}{||a||.||b||}$$

2.2.2. Euclidean Distance

This is the elementwise squared distance between the two vectors.

2.2.3. Pearson Similarity

It is a coefficient that gives the linear correlation between two variables. For two given users, u and u′, the value of the Pearson Correlation Coefficient is given by:

$$PCC(u,u') = \frac{\sum_{i\in I}(r_{u,i} - \bar{r}_u)(r_{u',i} - \bar{r}_{u'})}{\sqrt{\sum_{i\in I}(r_{u,i} - \bar{r}_u)^2}\sqrt{\sum_{i\in I}(r_{u',i} - \bar{r}_{u'})^2}}$$

where, $r_{u,i}$ and $r_{u',i}$ are the ratings from two users, r_u and $r_{u'}$ are the average of the ratings the users give, and *I* is the items rated by both users.

2.3. Evaluation of Recommender Systems

RS performance can be evaluated offline or online, as discussed below.

2.3.1. Offline Evaluation

These include low-level metrics that can be measured easily. The following are the commonly used offline evaluation metrics:

Precision and Recall

A RS recommends top-ranked N items to the user; hence, precision and recall are computed in the first N items instead of considering all of the items. It leads to the notion of "precision and recall @ k", k being a user-defined integer.

Precision@k is the fraction of recommended items in the top-k set which are relevant.

$$Precision@k = \frac{No: of\ recommended\ items@k\ that\ are\ relevant}{No: of\ recommended\ items\ @k}$$

Recall at k is the proportion of relevant items in the top-k recommendations.

$$Recall@k = \frac{No: of\ recommended\ items@k\ that\ are\ relevant}{Total\ No: of\ relevant\ items}$$

F-score

This is calculated from Precision and Recall by weighting precision and recall equally.

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

Mean Average Precision (MAP)

This is the mean of all the Average Precision. Average precision is calculated as:

$$AP = \sum_{k=1}^{n} p(k)\, rel(k)$$

where $p(k)$ is Precision@k, n is the number of items and $rel(k)$ is 1 if the item with rank k is relevant and 0 otherwise. MAP is given by:

$$MAP = \frac{1}{|U|}\sum_{u=1}^{|U|} AP(u)$$

where |U| gives the total number of users.

Normalized Discounted Cumulative Gain @k:

This measures ranking quality. Cumulative Gain is the sum of the relevance scores given to items and is calculated as:

$$CG(k) = \sum_{i=1}^{k}(rel_i)$$

The "Discounted Cumulative gain" is calculated as:

$$DCG(k) = \sum_{i=1}^{k}\frac{rel_i}{log_2(i+1)}$$

"Ideal Discounted Cumulative Gain" (IDCG) ranking is obtained by arranging the items in descending order by rankings and calculating the DCG. Normalized DCG is calculated as follows:

$$NDCG@k = \frac{DCG_k}{IDCG_k}$$

Mean Reciprocal Rank (MRR)

This is the average reciprocal hit ratio. Reciprocal Rank of a user u, RR(u) is the sum of the relevance score of top L items weighted by Reciprocal Rank and is given by:

$$RR(u) = \sum_i \frac{rel_i}{rank_i}$$

MRR is given by the mean of all users.

$$MRR(u) = \frac{1}{|U|}\sum_{u=1}^{|U|} RR(u)$$

where |U| gives the total number of users.

Mean Absolute Error (MAE)

MAE is the average sum of the absolute value of the difference between the actual and predicted ratings.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

where y_i is the predicted rating and x_i is the actual rating given by the user.

Root Mean Square Error (RMSE)

This evaluates inaccuracies on both positive and negative ratings.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}}$$

Where y_i gives the predicted rating and x_i is the actual rating the user gives.

2.3.2. Online Evaluation

These high-level metrics can be measured only with real-time users after the model is deployed. Click Through Rate (CTR), User Retention Rate, Customer Lifetime Value, and Return on Investment are some of the metrics of this type.

2.4. Applications of Recommender Systems

Recommender systems find their application in various domains of our everyday life, as shown in Fig. 2. Entertainment recommendations for movies, music etc. and content recommendations in the form of personalized newspapers, document recommendations, Web page recommendations, e- learning applications, etc., are quite prevalent today. E-commerce recommendations in the form of recommendations of products such as study materials, personal computers, mobile phones, etc., for consumers to buy and service recommendations in the form of travel recommendations, expert recommendations for consultation, house recommendations for rent, and matchmaking services are some of the most common applications.

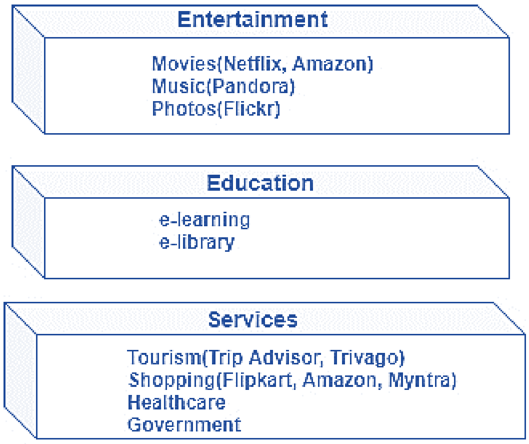

Fig. 2 Applications of Recommender Systems

3. Text-Based Recommender Systems

This section gives a detailed review of various text-based recommender systems based on their application domain.

The various approaches used in text-based recommender systems include text pre-processing, vector space modelling, keyword extraction, topic modelling, text classification, and deep learning.

3.1. Recommender Systems in Social Networks

In social networks, the relationship between users can be utilized for identifying the most appropriate groups of friends to provide improved and much more efficient recommendations.

A recommendation system for the recommendation of friends, which utilizes the life styles of users, has been implemented by Zhibo Wang et al. [99]. A solution for finding users in social media networks based on a uniform network structure together with an algorithm called FRUI (Friend Relationship- Based User Identification) based on friend relationships has been proposed by Xiaoping Zhou et al. [11]. "itop-k" proposed by Panlong Yang et al., allows offloading of task based on social-relationship. In this method, users can offload their tasks to their friends with greater energy and computation efficiency [12].

The discovery of aspect-level influence relationships has been defined by Chuan Hu et al. in [13]. A semisupervised algorithm for aspect extraction utilizing context sentences with labelled aspects is used to model a classifier that can predict the aspects for the unlabelled contexts. A recommender system in the CPSS (Cyber-Physical-Social Systems) domain consists of the discovery of a group based on user behavior similarity, improvement of accuracy by the rating data revision, and group preference modelling with context mining from different sources has been developed by Yin Zhang et al. [14]. A three-layer model for analyzing and identifying users' dynamic social roles for collective decision-making has been proposed by Bo Wu et al. [15]. Recommender System has been developed to provide voting campaigns as recommendations to users using “matrix factorization-based” and “nearest neighbour-based” models. The idea of a meta path connecting different object types through a sequence of relations is used to find the closest neighborhoods for target users. [16]. An approach for predicting link nodes that are close to each other in their latent space are more prone to be linked has been developed by Linhong Zhu et al. [17].

3.1.1. Challenges in Social Network Recommendations

The major challenge in social network recommendations is analyzing social interaction between users. It involves learning the connection between social and behaviour information by properly incorporating sentiment analysis to improve the accuracy of the generated recommendations.

3.2. Information Recommendation

Recommender Systems can provide learning objects in various course management systems.

Silvana V et al. have proposed a recommendation algorithm [18] wherein users' knowledge, reputation and availability to answer questions raised by other users are considered for making recommendations. A system for intent detection, wherein for each question, subjective and objective classes are defined, is designed by Liu et al. [19]. This model used lexical, contextual and syntactical features for classification. Analysis of social interactions in collaborative learning scenarios has been proposed by Claros et al. [20].

The similarity between objects like scientific papers, web pages and social networks can be accurately measured using link similarity by using the computation function proposed by Qin Zhao et al. [21].

3.2.1. Challenges in Information Recommendations

Information recommendation systems need the mining of user preferences for the generation of recommendations. A clear understanding of the knowledge level of users is key to developing such recommender systems.

3.3. Travel (Location based) Recommendations

Recommender Systems provide suggestions to users in the field of e-tourism. Travel recommendation approaches provide personalized responses based on the travel interest of users. A recommender system built from travelogues and photos usingmetadata like tags, geographic location, etc., linked with these photos has been proposed by Shuhui Jiang et al. [22], which recommends a sequence of travel points instead of individual Points of Interest (POI). The similarity between route and user packages is used to mine and rank famous routes. Then, an optimization of the top-ranked famous routes is done based on the travel records of similar users.

A rating prediction based on location has been proposed by combining interest similarity between users, the user-item geographical connection between user and item, and the user-user geographical connection [23].

A heterogeneous Location Query Browse graph has been proposed to represent people's interactive knowledge about behaviour across cyber and physical spaces. The dependencies between people's browsing, querying, and spatial behaviours have been confirmed by analyzing the influence of context on people’s behaviour [24].

A user's preference for a visited city will be influenced by the specialities of this city or the famous landmarks and the constraints based on time and distance. A recommender system has been proposed, considering all these factors to provide personalized landmark recommendations [25]. A classifier is modelled to detect landmark styles using web photos and landmark images. Style preferences of users are learned based on their past travel records by utilizing the detected landmark styles.

A recommendation framework based on location and preference is proposed by Shanfeng Wang et al. [26] by considering two contradictory objective functions, one of which performs the system for similar item recommendations. In contrast, the other shows the system's ability for diverse item recommendations.

3.3.1. Challenges in Travel Recommendations

Travel recommender systems aim to predict people's movements by utilizing their travel history consisting of their check-in and consuming behaviors, uploaded photos etc. The major challenge in such RS is the sparsity of user data mixed with the massive amount of location data, which might inhibit predicting users' places of interest. Hence, the problem of recommending places to users narrows down to modelling preferences of users from the sparse user data. It should consider the physical range of places travelled by the user, the type of places the user liked and the social influence of a place on the user.

3.4. Online Item (Product) Recommendations

A customer who visits an e-commerce website is presented with a large variety of items to choose from. The customer's decision-making process is eased by means of a product recommender system.

A contextual Multi-Arm Bandit-based clustering approach has been proposed to build recommender systems by considering the recommendation context [27].

A tensor factorization-based recommendation system has been developed, with UPD (user-preference dynamics) values assigning appropriate weights to past user preferences, wherein higher values of UPD down weigh more past user preferences. To identify how users’ preferences change with time, interactions between users and items are captured with a tensor incorporating users, items and time-periods [28].

A hybrid method has been proposed using the semantic information of items for finding user similarity based on liked and disliked items and the users’ previous rating data, which gives the satisfaction level similarity between different users [29].

The inner product of latent vectors used by most recommendation algorithms fails to measure the relationships between same category vectors, like item-item and user-user, which is required for collaborative filtering. The dual-Euclidean distance in latent space has been utilized by "Latent Dual Metric Embedding" for the various types of relationships for collaborative filtering [30].

Music features like speechiness, loudness, and acoustics can provide content-based music recommendations. [31].

A clustering-based e-commerce product recommendation system, which utilizes neighbor factor and time function, has been proposed [32]. Session-based recommendation (SBR) is used to predict a user's next action based on recent series of the users' actions. A framework called VAriational SEssion-based Recommendation using Bayesian inference is proposed [33].

Textual features of a product, like categories, tags, titles and descriptions, play a major role in an e-market place for information retrieval (IR) services and greatly influence consumers' behaviour, increasing sales and revenue. Using the behavior of customers, tag quality attributes have been proposed and are combined utilizing a deep learning-to-rank (L2R) framework, producing results better than those of neural network architectures [34].

A feedforward deep neural network, a combination of different categories of features for the recommendation, called a "deep alliance neural network", has been proposed [35].

Item recommendation using bipartite network embedding technology, which overcomes the data sparsity problem, has been proposed [36].

A meta-path-based heterogeneous recommendation model called HeteroPRS, which uses the meta- information associated with items, has been proposed for personalized recommendations [37].

A sentiment-based rating prediction method has been proposed, which fuses similarity in user sentiment, sentiment influence between users, and similarity in item reputation into a matrix factorization model to predict the ratings [38].

In combination with a collaborative filtering algorithm, deep learning technology can handle the sparsity of data and scalability [39], [40].

Unobserved data are handled equally by algorithms by assigning a uniform weight. These systems have the drawback that the whole data's contributions are not considered, which leads to prediction bias and less efficiency. An all-weighted matrix factorization technique for a recommendation has been proposed [41]. A "frequency- aware weighting scheme" is used for observed data. This system uses a "user-oriented weighting scheme" for unobserved data.

The regression-based MF does not guarantee that the predicted values are in the same order as the user preferences. A framework considering user preference for personalized ranking of Poisson factorization is proposed that uses posteriori based on "learning- to-rank" [42].

3.4.1. Challenges in Online Item Recommendations

The review of online item recommendations reveals that users' preferences for items tend to vary with time, depending on the type of items with which users interact. Different factors of users or other external factors cause variations in user preferences. Incorporating the relationships be- tween item and item, user and user, rather than just the relationships between user and item, improves the recommendation accuracy.

3.5. Drawbacks of Text-based Recommender Systems

Based on the text, RS has the following drawbacks that could result in irrelevant recommendations.

- Human perception subjectivity: Each user has a different perception of items; hence, individual users differ in their interpretation of items. The user himself may not have a clear idea of how to perform a textual search for items of his interest.
- Impreciseness in annotations of items stored in logs. Hence, there is a need for image-based recommender systems, which analyze the visual contents of a given image and generate recommendations based on user preference.

Tables 2 to 5 provide a summary of evaluations of the different Text–Based Recommender Systems reviewed in Section 3.

4. Image-Based Recommender Systems

Image-based recommendation techniques are highly prevalent now. Markov chain models have been used in image recommendations, in which the transition from one state to another is based on probabilistic rules. An algorithm that retrieves images for a given user query has been proposed [43], wherein keyword relevance probability between query keywords and annotated keywords stored in the log are calculated using an absorbing Markov chain to rank the images. The images are re-ranked using their visual features. In the image retrieval method proposed by Konstantinos A et al. [44], images are modelled as vector space points, and Markovian Semantic Indexing is taken as their similarity measure. An Aggregate Markov Chain (AMC) constructed using the users' queries finds the relevance between the keywords.

RS based on relevance feedback has been proposed. An image recommendation that utilizes visual features and user relevance feedback sessions to retrieve relevant images based on user inputs has been proposed [45]. Image features are computed, and cosine similarity gives relevance between images. The images are ranked using clicked frequency and similarity score.

In image-based clothing retrieval, users may be unable to supply a query image of the desired clothing they have in their minds. A solution for this problem has been proposed by Zhuoxiang Chen et al. through relevance feedback [46]. The target image and the feature influencing user responses are considered two different random variables. The retrieved images are refined, and heterogeneous features are re-weighted in a unified Bayesian formulation, learned independently during each search session and helped capture the varying user behavior.

A vertical search engine framework has been proposed for vertical image search, focusing on a specific segment of online content, wherein text and visual features are combined to bridge the semantic gap. ANOVA p-value is used to find the visual synonyms of each term from the visual features of images. Recommendation of images is done by computing pair-wise image cosine similarity [47]. The model proposed by Yin Zheng et al., which performs text document modelling, has been simultaneously used for image classification and annotation [48]. A method that utilizes annotation words, image visual words and class labels to learn a representation has been proposed. Gist and MPEG-7 descriptors of length 1857 are used for global features. First, the feature vectors are extracted using SIMPLE-CEDD (color and edge directivity descriptor). The next search mode combines CBIR and TBIR ("Tag Based Image Retrieval"), wherein the results are linked with a tag, after which the desired images are retrieved [49].

Image retrieval systems based on annotations utilize textual query, a set of images and their annotations to retrieve images based on the matching score of the query and the corresponding annotations [50]. Several works based on annotations of images have been proposed [51], [52], [53], [54], [55].

In image retrieval proposed by Anurag Bhargava et al. [56], the image is categorized into different sets of groups based on the objects present in an image. When a new image arrives, features are extracted from the objects. They are mapped against the various image groups for feature matching, and finally, retrieval of the image is done based on object selection. Features of images have an important role in the image retrieval process; hence, the careful selection of feature extraction techniques is key toachieving accurate image recommendations. Learning the representative features of both users and images in an online environment is challenging because of the extreme diversity in the visual contents of images. The steps involved in developing an image-based recommender system include feature extraction from images, User Preference Modelling and retrieving images based on similarity tothe query image depending on user preference, as shown in Fig. 3. The following parts of this section review the various image-based recommender systems by categorizing them based on the feature extraction techniques used.

Table 2. Comparison of Reviewed Social Network Based Recommendations

Author	Concept	Dataset	Attribute	Algorithm/Method	Performance
Zhibo Wang et al.	Semantic-based Friend recommendation system based on lifestyles	Life styles extracted from the user's smartphone sensors	Features of user's activity From the accelerometer (x,y,z coordinates)	Latent Dirichlet Allocation algorithm. -Friend-matching graph. -K-means clustering	Recall= 95% -Less data and memory usage
Xiaoping Zhou et al.	Probabilistic topic model Bag-of-activity model	Sina Microblog and RenRen(SMNs)	User profile, connections and interaction content	K-means clustering algorithm	Recall= 90%
Panlong Yang et al.	Social-relationship-based task offloading	Trace data from Mobi-Clique application	Mobile and social features	"iTop-K" algorithm	9 times performance the improvement compared to social relationship assignments without priority
Chuan Hu et al.	Detect influence aspects and influence degrees at the aspect level from graphs	-DBLP citation Dataset - Twitter	Profile aspects	Merging-based aspect extraction algorithm -KNN	Precision=90%
Yin Zhang et al.	Multidimensional matrix factorization	-Yelp	-User emotional features	K-Nearest Neighbors (KNN) Clustering	RMSE= 1.245
Bo Wu et al.	Three-layer model for social role analysis	Real user data	User name, age, grade, test scores	NetLogo-based tool	Vote rate=0.5

Table 3. Comparison of Reviewed Information Recommendations

Author	Dataset	Concept	Attribute	Algorithm/Method	Performance
Silvana V. Aciar et al.	Moodle platform with students	Recommendations based on reputation, knowledge anduser availability	Knowledge, reputation and availability of user	Text Mining -TFIDF	Accuracy=80%
Zhe Liu et al.	Replyz (Twitter-based QAsite)	Intent detection modelled as a binary classification problem: subjective and objective. -Question subjectivity modelled using syntactic, lexical and contextual features	-Lexical (N-gram), -Syntactical features, -Contextual features (hashtags, emoticons. mentions)	Text classifier using lexical features, meta-features and part-of-speech (POS) tagging -SVM classifier	Precision=85.4%
Ivan Claros et al.	Case study At Universidad del Cauca (Colombia)	Application of Social network analysis metrics for a Collaborative Learning Experience	Degree (individual performance, team outcomes, satisfaction), Betweenness and Closeness (efficiency, autonomy) centrality	Cliques - Pearson's correlation between the contributions of an individual	SD=0.7
Qin Zhao et al.	CiteSeer dataset	Semantic Similarity, Link similarity	Keywords	TF-IDF for calculating the top k keywords	Precision>69.6% (CiteSeer dataset) Precision> 81.5% (Cora dataset)

Table 4. Comparison of Reviewed Travel Recommendations

Author	Dataset	Concept	Attribute	Algorithm/Method	Performance
Shuhui Jiang et al.	Flickr images	Learning travel attributes of users and routes by Topical Package Model	Users' and routes' multi attributes (time, topical interest, season preference, cost) - Users and routes attributes (consumption capability, preferred season)	-Cosine distance similarity -TF-IDF for tag score	Mean Average Precision is 0.33% higher than CF
Guoshuai Zhao et al.	Yelp	-Relevance between ratings and user item geographicallocation distances	User and item latent features	-Machine Learning (Gradient Descent)	RMSE =1.036 MAE = 0.791
Junge Shen et al.	Flickr, TripAdvisor	Style-Oriented Landmark Recommendation	Deep CNN features for landmark style	-Laplacian -SVM classification -Cosine distance to evaluate the landmark similarity -CNN for Feature Extraction	Accuracy=0.891

4.1. SIFT, LBP, HSV

Social image sharing websites allow the uploading and tagging of images, but the same image may be tagged differently by different users due to the diversities of users' interests. A recommendation algorithm using the "user-image-tag model" is proposed by Jing Zhang et al. [57], wherein a re-ranking of tags of social images is done based on the image content to take advantage of tag information efficiently.

Sharath Chandra et al. developed a machine learning model for user personality based on the image features to generate recommendations by considering the factors influencing different personality users to like an image [58].

A recommendation algorithm called MM-VBPR is proposed by G. Li et al. [59]. First, an algorithm that utilizes the cross-modal semantic correlations between image features generates a list of recommendations. Lei Zhu et al. propose ALM (Augmented Lagrangian Multiplier) to compute the optimal hash codes [60].

Table 5. Comparison of Reviewed Product Recommendations

Author	Dataset	Concept	Attribute	Algorithm/Method	Performance
Linqi Song et al.	Yahoo! Front Page	-Click behaviour of users utilized to obtain their item preferences - Contextual MAB (Multi-Arm Bandit) approach	Demographic features geographic features; behavior categories	Adaptive Clustering Recommendation (ACR)	CTR(Click through Rate) higher than those of UCB1 and ϵ -greedy by 58-69 percent
Dimitrios Rafailidis et al.	Last. fm, MovieLens	Tensor factorization technique	User, location, activity	First-order optimization algorithm of the nonlinear conjugate gradient -Combination of User Preference Dynamics and Coupled tensor factorization	Accuracy=45%
Bushra Al-hijawi et al.	HetRec 2011	Semantic-based Similarity, Satisfaction-based Similarity	Genre, actors, director and origin	Jaccard Similarity	MAE=0.6
Ting Zhong et al.	YOOCHOOSE 1/64 dataset - DIGINETICA	Session-based Recommen-dation	Item category, click time and latent factors	-Bayesian inference: for parameter estimation -VASER with deterministic attention (VASER-DA); -VASER with variational attention (VASER-VA)	VASER-DA Recall@ 20=71.85% VASER-VA Recall@ 20 =72.12%
Fabiano M. Belém et al.	Elo7 (largest Brazilian e-marketplace)	The queries and clicks issued by consumers are used for high-quality tags to describe the products	Tag quality attributes and search-oriented attributes	Deep Multilayer Perceptron Architecture	Precision@10= 0.653 Re-call@ 10= 0.589 NDCG@ 10= 0.733
Libo Zhang et al.	MovieLens-100K, MovieLens-1 M,Epinions	Collaborative filtering with deep learning technology	User's rating, User/Item ID	Quadric polynomial regression to obtain latent features -Forward propagation algo-rithm - ReLU as the activation Function of DNN -Neural Collaborative Fil-tering (NCF)	MAE=0.6586
Hongmei Li et al.	Lastfm, Foursquare	All-weighted matrix factorization model by the combination of weighting schemes of observed and unobserved data	User's ID rating, of user or item a	WRMF fast eALS (Weight Regularized Matrix Factorization Elementwise Alternating Least Squares)	MRR(Mean Reciprocal Rank) =0.3
Wang Zhou et al.	Amazon, Movie Douban	Deep learning for model top-N recommendation	Item attributes, user's profile, time, venue, demographic features	-CNN to extract latent features - LDA -Three-layer denoising autoencoder	MAE=0.621 RMSE=0.841
Ming-Syan Chen et al.	Last. fm MovieLens100 K	Personalized Ranking on Poisson Factorization	Rating matrix	Hierarchical Poisson factorization -Point-wise Learning to Rank	Precision@10 =45.8%

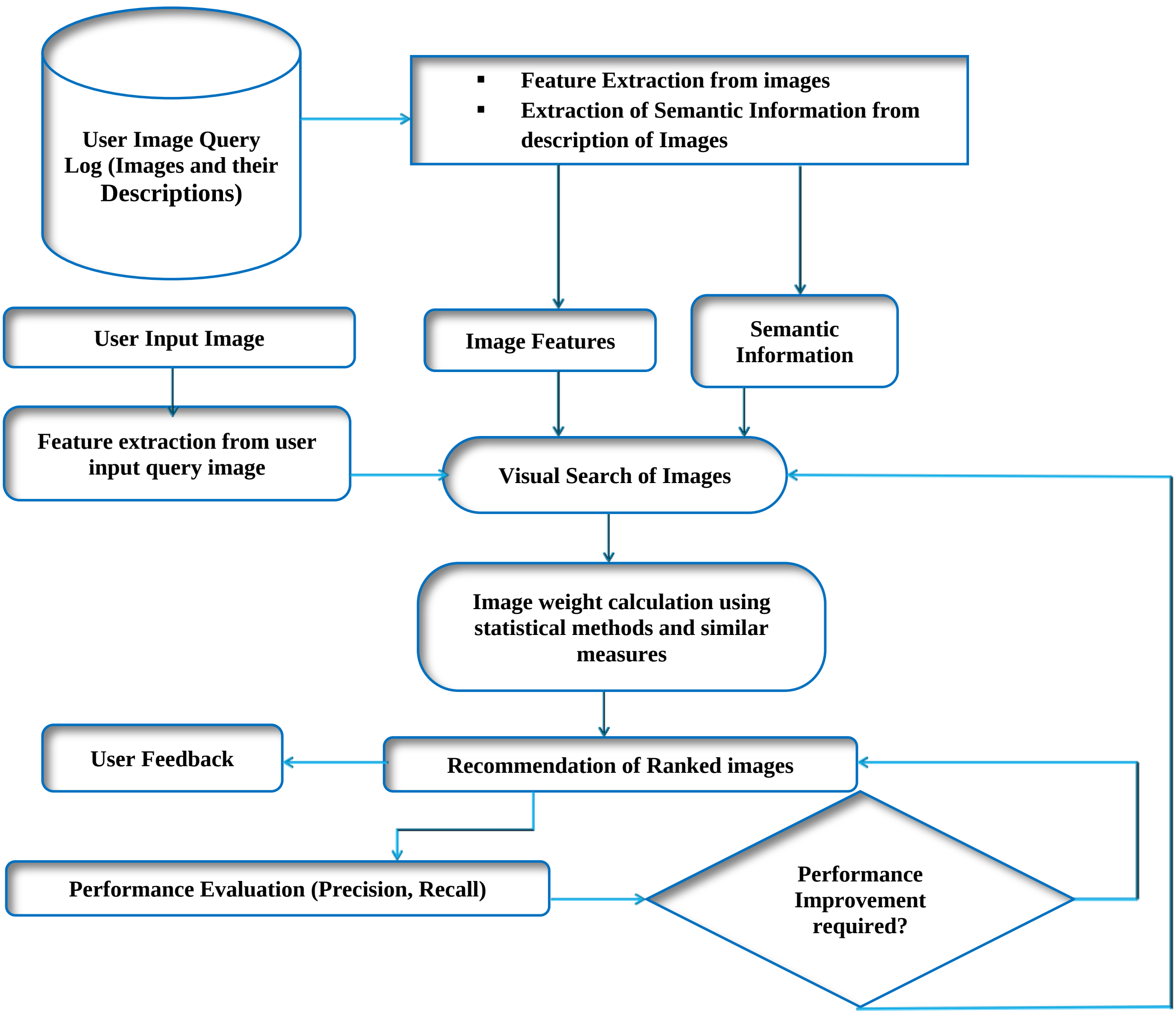

Fig. 3 Steps in Image Recommendation

Recommend-Me proposed by Duc et al. does the task of automatically recommending good query regions to users [61]. SIFT descriptor represents each region. First, an inverted index technique eliminates images with insufficient similarities compared to the input image. The BoVW model is used to represent images. A branch-and-bound algorithm identifies top region pairs having the highest similarity scores. The advantage of this method is that users will not have to try all possible query regions.

4.2. CNN and Deep Learning

The popularity of deep learning in recommender systems is of high relevance today. Nonlinear Transformation, Representation Learning, Sequence Modelling, and Flexibility are the strengths of deep learning-based recommendation models [62]. With deep learning, the input image is allowed to propagate forward through a pre-trained network and the features of the image are obtained by taking the output from the layer before the fully connected layers.

An approach for location-oriented clothing recommendations from social photos in which a multilabel convolutional neural network probes the uneven distribution of clothing attributes and a Support Vector Machine has been proposed to analyze the correlation between location and clothing [63].

A technique motivated by Bayesian Personalized Ranking (BPR) and neural networks have been proposed by Wei Niu et al. for image recommendations in social sharing communities [64]. In this model, the basic Neural Personalized Ranking model is enhanced with contextual factors like geographic features, user tags and visual factors.

A probabilistic model, which utilizes the shape feature of images along with the item's contextual information for the recommendation of fashionable goods, has been proposed by Yufeng Duan et al. [65]. A method proposed by Sagar Verma et al. learns part-based similarity for image recommendation [66].

A method for using the style features in the visual recommendation to understand user preferences has been proposed by Ming He et al. based on the fact that users decide on a particular product depending on whether the product suits their style [67].

A model has been proposed by Guibing Guo et al., which understands the user preference towards a particular image by identifying the semantic information of the sub-objects that appear in the image for improved recommendation [68]. Yuan Meng et al. have proposed a deep neural network using reciprocal social influence for image recommendation [69].

Jing Zhang et al. [70] suggested a method for recommending social images, wherein re-ranked tags are used to construct a tag tree of social images. An image-based product search in an online shopping system has been proposed by Farhan Ullah et al. [71]. The product type is learned using Random Forests (RF) classifier, and closely matched similar products are retrieved for a recommendation. The JPEG coefficients are used for feature extraction.

Sabahi et al. propose a method based on Hopfield neural networks (HNN) for efficient image retrieval and reduction of the semantic gap [72].

Zechao Li et al. [73] propose a weakly supervised deep embedding model, which can simultaneously address image-to-tag retrieval, CBIR, TBIR and tag-to-tag retrieval. For image similarity, the dual sparse reconstruction approach uses the CNN-extracted features and the tag information provided by users.

A fine-tuned CNN for fully automated retrieval of images has been proposed by Filip Radenovic et al. [74]. A Generalized-Mean pooling layer that generalizes both max and average pooling is proposed. The clustering of images is done to construct a 3D model for each cluster.

A heterogeneous Social-aware movie recommendation (SMR) network, which utilizes the textual description from a deep RNN and a deep CNN-based visual representation of movie posters, along with social relationships and user ratings, has been developed by Zhou Zhao et al. [75].

A model has been proposed by Junmei Lv et al., which consists of a module for sharing knowledge among different modalities, a module for learning the interest relevance between target items and different historical items and the item similarity recommendation module, using the ResNet50 model for visual feature extraction and BERT (Bidirectional Encoder Representation from Transformers) for textual feature extraction [76].

A method that effectively deals with similarity bias has been proposed by Chen et al. [77]. A Faster R-CNN extracts region-level features from the image regions, and a bi-directional gated recurrent unit extracts textual features to convert them into word-level features.

A method has been proposed by Khawaja Tehseen et al. that uses features like colour, texture and shape to find the relevant images. This method finds the interest points of the objects using Fast Retina Keypoints and identifies the feature sets for the classification of images from multiple categories [78].

Ahmad Alzu'bi et al. have proposed an architecture for feature extraction using two parallel CNNs [79]. The features of the image are extracted using the convolutional layers. An efficient bilinear root pooling performs dimensionality reduction of image features at the pooling layer.

Srinidhi Hiriyannaiah et al. propose an image recommendation system using Autoencoder neural network for classification [80]. The CNN Inception v3 extracts image classes. The Inception v3 output images are used in feature extraction for a recommendation.

Explainable recommendation, give explanations on the reason behind an item being recommended to a user. But these systems have the drawback that they cannot provide explanations with visual and textual modalities. Also, these systems fail to provide explanations for the user's changing preferences, which could ultimately reduce customers' satisfaction. "Attentive Recurrent Neural Network" has been proposed by P. Liu et al., with textual and visual fusion, providing "explainable" recommendations [81].

Wei Zhou et al. propose a recommendation method for fashion products, which gives similar product and mix-and-match recommendations by utilizing textual attributes of products and image features [82].

Several image recommender systems have been developed for travel recommendations [83], [84], [85], [86]. A hierarchical model for the recommendation of the social contextual image has been developed by Le Wu et al., wherein social influence, upload history and owner admiration are considered along with basic latent user interest [87].

Jinhui Tang et al. propose a solution by using adversarial learning to create a multimedia recommender model [88]. The model is trained to oppose an adversary who tries to decrease the model's accuracy.

Yujie Lin et al. have developed an outfit recommendation method, which uses a CNN with mutual attention to extracting visual features for outfit matching and a gated RNN with a "Cross-Modality Attention" mechanism that transforms visual features into a sentence for the abstractive comment generation [89]. Hierarchical attention based on Food Rec- ommendation(HAFR), developed by Xiaoyan Gao et al., determines the food that similar users are usually inclined to, infers the preference that a user has for a particular food at the ingredient level and uses the recipe's visual images to learn user preference [90]. A model is proposed by Qiang Cui et al. [91], which overcomes the problem of a cold start by including visual and textual information.

Most personalized fashion recommendations based on images ignore hidden features, such as the texture and quality of the clothes and also fail to give textual explanations. A model has been proposed by Q. Wu et al. [92], which uses images of products and historical reviews to provide visual and textual explanations. The different knowledge transfer techniques for vision recognition tasks are classified into six categories based on the origin and destination of knowledge being transferred [93].

A Visual Search and Recommendation system for online product recommendation has been proposed by Devashish Shankar et al., wherein, VisNet learns embeddings to capture the visual similarity [94]. The Euclidean distance between embeddings of two images gives the measure of similarity between two given images, and the k-Nearest-Neighbor searches for similar images.

An approach that discovers visual relationships between objects depending on their appearance is proposed by Julian John McAuley et al. [95]. It is based on graphs of related images and recommends which clothes and accessories match each other.

Eye gaze data of users can be used to understand their interest in recommending products, as proposed by Zaman et al. [96], [97]. The work proposed in [98] retrieves image features from convolutional layers, which are then transferred to the decoder as local image features at each step.

4.3. Comparison of Results of CNN Feature extraction with traditional methods

As seen in Figure 4, feature extraction with CNN improves the accuracy of image retrieval compared to handcrafted features extracted using SIFT and JPEG coefficients.

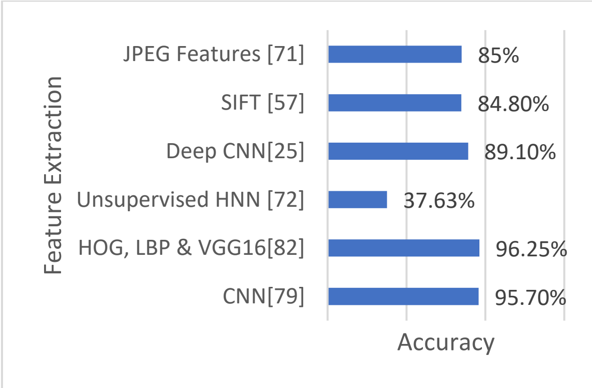

Fig. 4 Accuracy comparison of feature extraction techniques

Table 6 summarises the evaluation of the different Image – Based Recommender Systems reviewed in Section 4.

Table 6. Comparison of Reviewed Image Based Recommendations

Author	Concept	Dataset	Attribute	Method	Performance
D.Sejal et.al	Vertical image search using ANOVA Cosine Similarity	myntra.com	Visual features	ANOVA p Pair-wise image cosine similarity	Relevance score accuracy higher by 15.26% for top-10images compared to iLike.
Jing Zhang et.al	User-image-tag model for personalizedsocial image recommendation	NUS-WIDE	Descriptive visual words and HSV features	- Scale-Invariant Feature Transform (SIFT) descriptorsby Difference of Gaussian. - K-means clustering	Average NDCG (Normalized Discounted Cumulative Gain) = 0.848
Zhuoxiang Chen et al.	Image retrieval using Relevance feedback	TMALL.com TAOBAO.com	Garment categories, colors, length, button shape	-Bayesian classification -Feature re-weighting	Success rate= 80.45%
Yufeng Duan et al.	Combination of image shape featureof items with their contextual information	Amazon	Shape features of image	-Image Shape Feature Matrix Factorization: integratesCNN into the PMF	RMSE=1.09 (PMF =1.46 Con-vMF= 1.19) -Improves over ConvMF consistently from 4.7% to 8.6%.
Sagar Verma et al.	Image Recommen-dation by using part-based similar-ity	Fashion144K Fashion550k DeepFashion dataset	Texture-based features	-Six-layer CNN -Visual Attention Module - LSTM, -spatial transformer and a texture	Retrieval accuracy= 0.784 as compared to FashionNet accu-racy =0.764

				encoding layer(dictionary learning, feature pooling and classifier learning)	
Ming He et al.	Incorporating style features into collaborative learning	Amazon, Tradesy	Style features	- Hierarchical gram matrix - Style-aware Bayesian Personalized Ranking - CNN	Area Under the ROC Curve =0.7767 as compared to: BPR-MF=0.6058 VBPR=0.7608 DEEPSTYLE=0.7653
Yuan Meng et al.	Reciprocal social influence	Flickr	Deep visual features	CNN - deep neural network for image recommendation	- Precision@5=0.4032 as compared to: (CDL=0.2576, SBPR=0.3192, SCDL=0.3346) Improvement of 20.5% -Recall@5= 0.1008as compared to: (CDL=0.0644, SBPR=0.7898, SCDL=0.0836) Improvement of 20.57%
Jing Zhang et al.	-User interest tree using deep features and tag trees	NUS-WIDE	Deep features	CNN	Precision@5=0.82 Recall@5=0.11
Farhan Ullah et al.	Content-based image retrieval	Amazon	JPEG features	- JPEG coefficients - Random Forests (RF) classifier	Precision= 85%
F. Sabahi et al.	CBIR using Hopfield Neural Network	Corel Dataset	-color -texture (Gabor wavelet Wavelet moments)	Unsupervised HNN	Accuracy = 0.332-0.3763 as compared to: (FFBP=0.304 – 0.3522) MAP=0.35– 0.67 as compared to: (FFBP= 0.34 – 0.56)
Zechao L i et al.	Deep Collaborative Embedding (DCE) model	MIR Flickr and NUS-WIDE	CNN features	CNN (AlexNet)	Mean F1=0.718 Normalized Discounted Cu- mulative Gain NDCG @1000on MIRFlickr= 0.512 (AS com-pared to Multi correlation Probabilistic Matrix Factorization =0.437)
Filip Radenovic´ et al.	CNN Image Retrieval withno human annotation	-Flickr - Oxford5K buildings - Paris6K -Holidays	Hessian affine local features	-CNN Generalized mean pooling(GeM) - BoW -Structure-from-Motion(SfM) -Lw (learned discriminative Whitening)	mAP=91%
Junmei Lv et al.	Multimodal interest-related	-MovieLens 1M -	768-dimensional Textual feature	-Multimodal visual Bayesian	-Hit Ratio HR@10=0.8293

	item similarity model	Amazon Clothing, Shoes, Jewelry	vector - Visual pre-trained features	personalized ranking algorithm -Hierarchical sampling statistics model -BERT (Bidirectional Encoder Representation from Transformers) to pre-fetch textual features -ResNet50 model	-NDCG@10=0.5898
Yin Zheng et al.	-Topic Modeling of Multimodal Data	-LabelMe -UIUC-Sports -MIR Flickr	Gist and MPEG-7 descriptors (Edge Histogram Descriptor, Homogeneous Texture Descriptor, Color Structure De-scriptor, Color Layout De-scriptor, Scalable Color De-scriptor)	-SupDeepDocNADE (Supervised Deep Document Neural Autoregressive Distribution Estimator) -SIFT -RBF kernel -SVM classifier	MAP=0.69
Anurag Bhargava et.al	Object-based image retrieval framework	IAPR TC12	-HSV -RGB	-Feature extraction by Speeded Up Robust Fea-tures - SVM classifier	Precision=0.38 Recall=0.35 F1=0.36
Lei Zhu et al.	CBIR using Un-supervised Visual Hashing with Semantic Assistant	NUS-WIDE Wiki, Flickr,	Visual features	Hash Fuction-linear regression model - Similarity-Hamming distance	mAP=52.81%
Budikova et.al	Search-based Image Annotation with Relevance Feedback	Profiset data	VGG-16 descriptors for visual similarity	- Cosine distance - ConceptRank -CBIR-PPP-Codes technique	CBIR+TextRank Precision=83.5% Text+CBIRRank Precision= 79.1% Text2Vec Precision=78.2%
Zhigang Ma et.al	Shared Feature Sub-space Uncovering	- MSRA-MM 2.0 database (Microsoft Research Asia) - NUS-WIDE	Colour and texture	l2,l1 norm regularization for sparse feature selection	MAP=0.061
G. Li et al	Hierarchical Sampling Statistics And Multimodal Visual Bayesian Personalized Ranking combined for hybrid recommendation	"Wisdom Tourist" : Combination of a well-designed questionnaire survey and the automatic crawling (Ctrip.com) of multimodal data.	Color, texture, Shape and VGG features -gender, district , age, job wage of user	Multimodal Visual Bayesian Personalized Ranking	RMSE=0.917 MAE= 0.826
Khawaja Tehseen et al.	BoW in combination with local image	ImageNet Corel-1000, Caltech-101	Color (RGB), Shape	- Fast Retina Keypoints (Keypoint Descriptor)	Average Recall= 80% Average Precision= 0.95

	features & spatial information	ALOT		-Bag-of-Words (BoW)	
Ahmad Alu'bi etal.	CBIR with Com-pact Deep Convolu-tional Features	Holidays, Oxford, Ukbench	Deep convolutional features	Very deep architecture (Compact Root Bilinear CNN) and medium architecture	Retrieval accuracy =95.7% (Oxford5K)
Srinidhi Hiriyannaiah et al.	Visual recommendation using Deep Visual Ensemble similarity metric	Amazon 2014, 2015 and Street2Shop	Deep visual features	-Convolutional Autoencoder (CAE) for feature extraction -CNN for classification	AUC= 0.8693 Recall= 74.47%
Wei Zhou et al.	Fashion recommendation by combining textual mining and CBIR techniques	Zalora, Uniqlo, H and M, ASOS	Colour(descriptor of Pantone colours), texture(HOG and LBP)	-VGG-16 net classifier on Caffe library	Accuracy= 96.25%
Ren Xingyi, et al.	POI recommendation using SM-Twitter LDA	Twitter API	Text, image, timestamp, location and hashtag of tweets	Collapsed Gibbs Sampling to obtain latent Variables - ImageNet network (CNN) for feature extraction - SVM classifier with Gaussian kernel	MAP=0.7
Jinhui Tang et al.	Adversarial Multimedia Recommendation (AMR)	Pinterest Amazon	Deep image features	-SGD learning algorithm -DNN (for extraction of image deep features) -LFM for prediction of user preference	HR@10= 0.2360 NDCG@ 20=0.1296
Yujie Lin et al.	Neural outfit recommendation (NOR), which provides outfit recommendations and generates abstractive comments	Polyvore	Deep Visual features	-CNN with mutual attention mechanism to extract visual features of outfits -Gated RNN with cross-modality attention mechanism for abstractive comment generation	Precision= 9.40 Recall= 10.29

4.4. Challenges in Image Recommendations

The exhaustive review of image-based recommendations concludes that the semantic gap between the representation of features of images and the visual understanding of humans can lead to irrelevant recommendations. Hence, selecting visual features and having a discriminative feature representation is crucial for effective image recommendations. This demands the need for an efficient feature extraction algorithm. Training deep neural networks for large-scale image datasets is a complex task. The effect of the feature dimension on the recommendation performance has to be considered. Also, model training time and average recommendation time must be kept to a minimum.

5. Conclusion & Future Directions

RS came up with a solution to the "information overload" problem and provided users with more pragmatic and personalized information services. The capability to model the complex interaction patterns between users and items, including multimodal information such as text and images, and understanding the changing behaviour of the user and dynamic evolution of items is crucial for developing an efficient recommender system.

The various algorithms, evaluation metrics, issues and challenges involved in the development of recommender systems, especially in the case of image recommendations, wherein, given a query image, the recommender system has to retrieve a ranked list of images semantically similar to this query image, while also considering the user preference, have been discussed in this paper.

Future research directions include :

- Utilization of Deep Learning models, which combine multiple deep neural networks for modelling complex interactions between users and items and the dynamically changing user preferences for a more satisfying user experience.
- Developing a Context-aware recommendation model by incorporating the user's context like physical contexts (location, date, time), health, mood, job status, weather conditions etc.
- Ensuring the utilization of the most efficient feature extraction algorithm from images/text to have the best performance concerning the generation of recommendations.
- Reduction of the semantic gap by the fusion of different local features. Fusing local features with global features is also a challenge for future research.

References

[1] Kirsten Swearingen and Rashmi Sinha, "Beyond Algorithms: An Hci Perspective on Recommender Systems," *In ACM SIGIR 2001 Workshop on Recommender Systems*, vol. 13, pp. 1–11, 2001.

[2] Jesu´ S Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutie´Rrez, "Recommender Systems Survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

[3] Xiwang Yang, Yang Guo, Yong Liu, and Harald Steck, "A Survey of Collaborative Filtering-Based Social Recommender Systems," *Computer Communications*, vol. 41, pp. 1–10, 2014.

[4] Shini Renjith, A Sreekumar, and M Jathavedan, "An Extensive Study on the Evolution of Context-Aware Personalized Travel Recommender Systems," *Information Processing & Management*, vol. 57, no. 1, pp. 102078, 2020.

[5] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen, "A Survey of Serendipity In Recommender Systems," *Knowledge-Based Systems*, vol. 111, pp. 180–192, 2016.

[6] Mehdi Elahi, Francesco Ricci, and Neil Rubens, "A Survey of Active Learning in Collaborative Filtering Recommender Systems," *Computer Science Review*, vol. 20, pp. 29–50, 2016.

[7] Norha M Villegas, Cristian Sa´Nchez, Javier D´Iaz-Cely, and Gabriel Tamura, "Characterizing Context-Aware Rec-ommender Systems: A Systematic Literature Review," *Knowledge-Based Systems*, vol. 140, pp. 173–200, 2018.

[8] Matevzˇ Kunaver and Tomazˇ Pozˇrl, "Diversity in Recommender Systems–A Survey," *Knowledge-Based Systems*, vol. 123, pp. 154–162, 2017.

[9] Joan Borra`S, Antonio Moreno, and Aida Valls, *"Intelligent Tourism Recommender Systems: A Survey," Expert Systems with Applications*, vol. 41, no.16, pp. 7370–7389, 2014.

[10] Dr. R.Surendiran, "Development of Multi-Criteria Recommender System," *SSRG International Journal of Economics and Management Studies*, vol. 4, no. 1, pp. 31-35, 2017. *Crossref*, https://doi.org/10.14445/23939125/IJEMS-V4I1P106.

[11] Xiaoping Zhou, Xun Liang, Haiyan Zhang, and Yuefeng Ma, "Cross-Platform Identification of Anonymous Identical *Users* in Multiple Social Media Networks," *IEEE Transactionson Knowledge and Data Engineering*, vol. 28, no. 2, pp. 411–424, 2015.

[12] Panlong Yang, Qingyu Li, Yubo Yan, Xiang-Yang Li, Yan Xiong, Baowei Wang, and Xingming Sun, " Friend is a Treasure: Exploring and Exploiting Mobile Social Contacts for Efficient Task Offloading," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5485–5496, 2015.

[13] Chuan Hu and Huiping Cao, "Aspect-Level Influence Discovery from Graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 7, pp. 1635–1649, 2016.

[14] Yin Zhang, "Grorec: A Group-Centric Intelligent Recommender System Integrating Social, Mobile and Big Data Technologies," *IEEE Transactions on Services Computing*, vol. 9, no.5, pp. 786–795, 2016.

[15] Bo Wu, Xiaokang Zhou, Qun Jin, Fuhua Lin, and Henry Leung, "Analyzing Social Roles Based on A Hierarchical Model and Data Mining for Collective Decision-MakingSupport," *IEEE Systems Journal*, vol. 11, no.1, pp. 356–365, 2015.

[16] Xiwang Yang, Chao Liang, Miao Zhao, Hongwei Wang, Hao Ding, Yong Liu, Yang Li, and Junlin Zhang, "Collaborative Filtering-Based Recommendation of Online SocialVoting," *IEEE Transactions on Computational Social Systems*, vol. 4, no. 1, pp. 1–13, 2017.

[17] Linhong Zhu, Dong Guo, Junming Yin, Greg Ver Steeg, and Aram Galstyan, "Scalable Temporal Latent Space Inference for Link Prediction In Dynamic Social Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2765–2777, 2016.

[18] Silvana V Aciar, Gabriela I Aciar, Cesar Alberto Collazos, and Carina Soledad Gonza´Lez, "User Recommender System Based on Knowledge, Availability, and Reputation From in-Teractions In Forums," *IEEE IBERO-American Journal of Learning Technologies*, vol. 11, no. 1, pp. 18–22, 2016.

[19] Zhe Liu and Bernard J Jansen, "Understanding and Pre-Dictating Question Subjectivity in Social Question and an- Swering," *IEEE Transactions on Computational Social Systems*, vol. 3, no. 1, pp. 32–41, 2016.

[20] Iva´N Claros, Ruth Cobos, and Ce´Sar A Collazos, "An Ap- Proach Based on Social Network Analysis Applied to a Col- Laborative Learning Experience," *IEEE Transactions on Learn-Ing Technologies*, vol. 9, no. 2, pp. 190–195, 2015.

[21] Qin Zhao, Cheng Wang, Pengwei Wang, Mengchu Zhou,and Changjun Jiang, "A Novel Method on Information Recommendation Via Hybrid Similarity," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 32, pp. 448–459, 2016.
[22] Shuhui Jiang, Xueming Qian, Tao Mei, and Yun Fu, "Personalized Travel Sequence Recommendation on Multi- Source Big Social Media," *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 43–56, 2016.
[23] Guoshuai Zhao, Xueming Qian, and Chen Kang, "Service Rating Prediction by Exploring Social Mobile Users' Geo- Graphical Locations," *IEEE Transactions on Big Data*, vol. 3, no. 1, pp. 67–78, 2016.
[24] Yongli Ren, Martin Tomko, Flora Dilys Salim, Jeffrey Chan, Charles LA Clarke, and Mark Sanderson, "A Location-Query-Browse Graph for Contextual Recommenda-Tion," *IEEE Transactions on Knowledge and Data Engineering, vol. 30*, no. 2, pp. 204–218, 2017.
[25] Junge Shen, Zhiyong Cheng, Meihong Yang, Bing Han, and Shuying Li, "Style-Oriented Personalized Landmark Recommendation," *IEEE Transactions on Industrial Electron- ICS*, vol.66, no. 12, pp. 9768–9776, 2019.
[26] Shanfeng Wang, Maoguo Gong, Yue Wu, and Mingyang Zhang, "Multi-Objective Optimization for Location-Based and Preferences-Aware Recommendation," *Information Sci- Ences*, vol. 513, pp. 614–626, 2020.
[27] Linqi Song, Cem Tekin, and Mihaela Van Der Schaar, "On-Line Learning in Large-Scale Contextual Recommender Systems," *IEEE Transactions on Services Computing*, vol. 9, no. 3, pp. 433– 445, 2014.
[28] Dimitrios Rafailidis and Alexandros Nanopoulos, "Mod- Eling Users Preference Dynamics and Side Information in Recommender Systems," *IEEE Transactions on Systems, Man,and Cybernetics: Systems*, vol. 46, no. 6, pp. 782–792, 2015.
[29] Bushra Alhijawi, Nadim Obeid, Arafat Awajan, and Sara Tedmori, "Improving Collaborative Filtering Recommender Systems Using Semantic Information," In *2018 9th Interna- Tional Conference on Information and Communication Systems(ICICS)*, IEEE pp.127–132. 2018.
[30] Yuanbo Xu, Yongjian Yang, Jiayu Han, Xiang Li, andEn Wang, "Improving Recommendations by Embedding Multi-Entity Relationships With Latent Dual-Metric Learning," *IEEE Access*, vol. 7, pp. 9817–9826, 2019.
[31] Parmar Darshna, "Music Recommendation Based on Con Tent and Collaborative Approach & Reducing Cold StartProblem," In *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, IEEE, pp. 1033–1037, 2018.
[32] Kai Wang, Tiantian Zhang, Tianqiao Xue, Yu Lu, and Sang-Gyun Na, "E-Commerce Personalized Recommendation Analysis by Deeply-Learned Clustering," *Journal of VisualCommunication and Image Representation*, vol. 71, pp. 102735, 2020.
[33] Ting Zhong, Zijing Wen, Fan Zhou, Goce Trajcevski, and Kunpeng Zhang, "Session-Based Recommendation Via Flow-Based Deep Generative Networks and Bayesian Infer-Ence," *Neurocomputing*, 2020.
[34] Fabiano M BeléM, Rodrigo M Silva, Claudio MV De Andrade, Gabriel Person, Felipe Mingote, Raphael Ballet, Helton Alponti, Henrique P De Oliveira, Jussara M Almeida, and Marcos a GonçAlves. "Fixing The Curse of the Bad Product Descriptions"–Search-Boosted Tag Recom-Mendation for E-Commerce Products," *Information Process- Ing & Management*, vol. 57, no. 5, pp. 102289, 2020.
[35] Xuna Wang, Qingmei Tan, and Lifan Zhang, "A Deep Neural Network of Multi-Form Alliances for Personalized Rec-Ommendations," *Information Sciences*, 2020.
[36] Yiyeon Yoon, Juneseok Hong, and Wooju Kim, "Item Rec- Ommendation by Predicting Bipartite Network Embeddingof User Preference," *Expert Systems With Applications*, pp. 113339, 2020.
[37] Mukul Gupta and Pradeep Kumar, "Recommendation Gen-Eration Using Personalized Weight of Meta-Paths In Het- Erogeneous Information Networks," *European Journal of Op-Erational Research*, vol. 284, no. 2, pp.660–674, 2020.
[38] Xiaojiang Lei, Xueming Qian, and Guoshuai Zhao, "RatingPrediction Based on Social Sentiment From Textual Reviews," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1910–1921, 2016.
[39] Libo Zhang, Tiejian Luo, Fei Zhang, and Yanjun Wu, "A Recommendation Model Based on Deep Neural Network," *IEEE Access*, vol. 6, pp. 9454–9463, 2018.
[40] Wang Zhou, Jianping Li, Malu Zhang, Yazhou Wang, andFadia Shah, "Deep Learning Modeling for Top-N Recommendation With Interests Exploring,"*IEEE Access*, vol. 6, pp. 51440–51455, 2018.
[41] Hongmei Li, Xingchun Diao, Jianjun Cao, and Qibin Zheng, "Collaborative Filtering Recommendation Based on All-Weighted Matrix Factorization and Fast Optimization," *IEEE Access*, vol. 6, pp. 25248–25260, 2018.
[42] Li-Yen Kuo, Chung-Kuang Chou, and Ming-Syan Chen, "The Framework of Personalized Ranking on Poisson Factorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 1, pp. 287–301, 2019.
[43] D Sejal, V Rashmi, KR Venugopal, SS Iyengar, and LM Patnaik, "Image Recommendation Based on Keyword Relevance Using Absorbing Markov Chain and Image Features," *International Journal of Multimedia Information Retrieval*, vol. 5, no. 3, pp. 185–199, 2016.
[44] Konstantinos A Raftopoulos, Klimis S Ntalianis, Dionyssios D Sourlas, and Stefanos D Kollias, "Mining User Queries With Markov Chains: Application to Online Image Retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 2, pp. 433–447, 2011.
[45] D Sejal, D Abhishek, KR Venugopal, SS Iyengar, and LM Patnaik. Ir Urfs Vf: Image Recommendation With User Relevance Feedback Session and Visual Features In Vertical Image Search. *International Journal of Multimedia Information Retrieval*, vol. 5, no. 4, pp. 255–264, 2016.

[46] Zhuoxiang Chen, Zhe Xu, Ya Zhang, and Xiao Gu, "Query- Free Clothing Retrieval Via Implicit Relevance Feedback," *IEEE Transactions on Multimedia,* vol. 20, no. 8, pp. 2126–2137, 2017.

[47] D Sejal, T Ganeshsingh, KR Venugopal, SS Iyengar, and LM Patnaik, "Acsir: Anova Cosine Similarity Image Recommendation In Vertical Search," *International Journal of Multimedia Information Retrieval*, vol. 6, no. 2, pp. 143–154, 2017.

[48] Yin Zheng, Yu-Jin Zhang, and Hugo Larochelle, "A Deep and Autoregressive Approach for Topic Modeling of Multi-Modal Data," *IEEE Transactions on Pattern Analysis and Ma-Chine Intelligence*, vol. 38, no. 6, pp. 1056–1069, 2015.

[49] Meenaakshi N Munjal and Shaveta Bhatia, "A Novel Tech- Nique for Effective Image Gallery Search Using JP System," In *2019 International Con- Ference on Machine Learning, Big Data, Cloud and Parallel Computing (Comitcon)*, IEEE, pp. 25–29. 2019.

[50] Xin-Jing Wang and Lei Zhang. Annotation-Based Image Retrieval, 2009.

[51] Gustavo Carneiro, "Artistic Image Analysis Using Graph-Based Learning Approaches," *IEEE Transactions on Image Pro-Cessing*, vol. 22, no. 8, pp. 3168–3178, 2013.

[52] Haiyu Song, Pengjie Wang, Jian Yun, Wei Li, Bo Xue, andGang Wu, "A Weighted Topic Model Learned From Local Se- Mantic Space for Automatic Image Annotation," *IEEE Access*, vol. 8, pp. 76411–76422, 2020.

[53] Petra Budikova, Michal Batko, and Pavel Zezula, "Multimodal Image Retrieval for Search-Based Image Annotation With Rf," In *2018 IEEE International Symposium on Multime- Dia (ISM)*, IEEE, pp. 52–60, 2018.

[54] Ran Li, Yafei Zhang, Zining Lu, Jianjiang Lu, and YulongTian, "Technique of Image Retrieval Based on Multilabel Image Annotation," In *2010 Second International Conference on Multimedia and Information Technology*, IEEE, vol. 2, pp. 10–13, 2010.

[55] Zhigang Ma, Feiping Nie, Yi Yang, Jasper RR Uijlings, andNicu Sebe, "Web Image Annotation Via Subspace-Sparsity Collaborated Feature Selection," *IEEE Transactions on Multi-Media*, vol. 14, no. 4, pp. 1021–1030, 2012.

[56] Anurag Bhargava, Shashi Shekhar, and KV Arya, "An ObJect Based Image Retrieval Framework Based on Automatic Image Annotation," In *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, IEEE, pp. 1–6.,2014.

[57] Jing Zhang, Ying Yang, Qi Tian, Li Zhuo, and Xin Liu, "Personalized Social Image Recommendation Method Basedon User-Image-Tag Model," *IEEE Transactions on Multimedia,* vol. 19, no. 2, pp. 2439–2449, 2017.

[58] Sharath Chandra Guntuku, Joey Tianyi Zhou, Sujoy Roy, Weisi Lin, and Ivor W Tsang, "Who Likes What and Why?' Insights Into Modeling Users' Personality Based on Image 'Likes," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 130–143, 2016.

[59] Guangli Li, Tao Zhu, Jin Hua, Tian Yuan, Zhengyu Niu, Tao Li, and Hongbin Zhang, "Asking Images: Hybrid Recommendation System for Tourist Spots by Hierarchical Sampling Statistics and Multimodal Visual Bayesian Personalized Ranking," *IEEE Access*, vol. 7, no. 126539–126560, 2019.

[60] Lei Zhu, Jialie Shen, Liang Xie, and Zhiyong Cheng, "Unsupervised Visual Hashing With Semantic Assistant Forncontent-Based Image Retrieval," *IEEE Transactions on Knowl-Edge and Data Engineering,* vol. 29, no. 2, pp. 472–486, 2016.

[61] Thanh Duc Ngo, Sang Phan, Duy-Dinh Le, and Shin'ichi Satoh. Recommend-Me: Recommending Query Regions forImage Search. In *Proceedings of The 29th Annual ACM Sym- Posium on Applied Computing*, pp. 913–918, 2014.

[62] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Computing Surveys (CSUR)*, vol. 52, no. 1, pp. 1–38,2019.

[63] Xishan Zhang, Jia Jia, Ke Gao, Yongdong Zhang, Dongming Zhang, Jintao Li, and Qi Tian, "Trip Outfits Advisor: Location-Oriented Clothing Recommendation," *IEEE Trans- Actions on Multimedia,* vol. 19, no. 11, pp. 2533–2544, 2017.

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

[65] Ryosuke Saga and Yufeng Duan, "Apparel Goods Recom- Mender System Based on Image Shape Features Extracted by a Cnn," In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE pp. 2365–2369, 2018.

[66] Sagar Verma, Sukhad Anand, Chetan Arora, and Atul Rai, "Diversity In Fashion Recommendation Using SemanticParsing," In *2018 25th IEEE International Conference on ImageProcessing (ICIP)*, IEEE, pp. 500–504, 2018.

[67] Ming He, Shaozong Zhang, and Qian Meng, "Learning to Style-Aware Bayesian Personalized Ranking for Visual Recommendation," *IEEE Access*, vol. 7, pp. 14198–14205, 2019.

[68] Guibing Guo, Yuan Meng, Yongfeng Zhang, Chunyan Han, and Yanjie Li, "Visual Semantic Image Recommendation," *IEEE Access*, vol. 7, pp. 33424–33433, 2019.

[69] Yuan Meng, Chunyan Han, Yongfeng Zhang, Yanjie Li, and Guibing Guo, "Image Recommendation With Reciprocal Social Influence," *IEEE Access*, vol. 7, pp. 132279–132285, 2019.

[70] Jing Zhang, Ying Yang, Li Zhuo, Qi Tian, and Xi Liang, "Personalized Recommendation of Social Images by Constructing a User Interest Tree With Deep Features and Tag Trees," *IEEE Transactions on Multimedia*, vol. 21, no. 11, pp. 2762–2775, 2019.

[71] Farhan Ullah, Bofeng Zhang, and Rehan Ullah Khan, "Image-Based Service Recommendation System: A Jpeg-Coefficient Rfs Approach," *IEEE Access*, vol. 8, pp. 3308–3318, 2019.

[72] F Sabahi, M Omair Ahmad, and MNS Swamy, "An Un-Supervised Learning-Based Method for Content-Based Image Retrieval Using Hopfield Neural Network," In *2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS)*, pp. 1–5. IEEE, 2016.

[73] Zechao Li, Jinhui Tang, and Tao Mei, "Deep Collaborative Embedding for Social Image Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2070–2083, 2018.

[74] Filip Radenović, Giorgos Tolias, and Ondřej Chum, "Fine-Tuning Cnn Image Retrieval With No Human Annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1655–1668, 2018.

[75] Zhou Zhao, Qifan Yang, Hanqing Lu, Tim Weninger, Deng Cai, Xiaofei He, and Yueting Zhuang, "Social-Aware Movie Recommendation Via Multimodal Network Learning," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 430–440, 2017.

[76] Junmei Lv, Bin Song, Jie Guo, Xiaojiang Du, and Mohsen Guizani, "Interest-Related Item Similarity Model Based on Multimodal Data for Top-N Recommendation," *IEEE Access*, vol. 7, pp. 12809–12821, 2019.

[77] Hui Chen, Guiguang Ding, Zijia Lin, Sicheng Zhao, Xi- Aopeng Gu, Wenyuan Xu, and Jungong Han, "Acmnet: Adaptive Confidence Matching Network for Human Behav-Ior Analysis Via Cross-Modal Retrieval," *ACM Transactions on Multimedia Computing, Communications, and Applications(TOMM)*, vol. 16, no. 1s, pp. 1–21, 2020.

[78] Khawaja Tehseen Ahmed, Shahida Ummesafi, and AmjadIqbal, "Content-Based Image Retrieval Using Image FeaturesInformation Fusion. *Information Fusion*," vol. 51, pp. 76–99, 2019.

[79] Ahmad Alzu'bi, Abbes Amira, and Naeem Ramzan. Content-Based Image Retrieval With Compact Deep Convolutional Features. *Neurocomputing*, vol. 249, pp. 95–105, 2017.

[80] Srinidhi Hiriyannaiah, GM Siddesh, and KG Srinivasa, "Deep Visual Ensemble Similarity (Dvesm) Approach for Visually Aware Recommendation and Search In Smart Community," *Journal of King Saud University-Computer and Information Sciences*, 2020.

[81] Peng Liu, Lemei Zhang, and Jon Atle Gulla, "Dynamic Attention-Based Explainable Recommendation With Textualand Visual Fusion," *Information Processing & Management*, vol. 57, no. 6, pp. 102099, 2020.

[82] Wei Zhou, PY Mok, Yanghong Zhou, Yangping Zhou, Jialie Shen, Qiang Qu, and KP Chau "Fashion Recommendations Through Cross-Media Information Retrieval," *Journalof Visual Communication and Image Representation*, vol. 61, pp.112– 120, 2019.

[83] Yeran Sun, Hongchao Fan, Mohamed Bakillah, and Alexander Zipf, "Road-Based Travel Recommendation Using Geo-Tagged Images," *Computers, Environment and UrbanSystems*, vol. 53, pp. 110–122, 2015.

[84] Ren Xingyi, Song Meina, E Haihong, and Song Junde, "Social Media Mining and Visualization for Point-of-Interest Recommendation," *The Journal of China Universities of Posts and Telecommunications*, vol. 24, no. 1, pp. 67–86, 2017.

[85] Zhibin Zhang, Cong Zou, Ruifeng Ding, and Zhenzhong Chen, "VCG: Exploiting Visual Contents and Geographical Influence for Point-of-Interest Recommendation," *Neuro-Computing*, vol. 357, pp. 53–65, 2019.

[86] Pengpeng Zhao, Chengfeng Xu, Yanchi Liu, Victor S Sheng, Kai Zheng, Hui Xiong, and Xiaofang Zhou, "Photo2trip: Exploiting Visual Contents in Geo-Tagged Photos for Personalized Tour Recommendation," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[87] Le Wu, Lei Chen, Richang Hong, Yanjie Fu, Xing Xie, andMeng Wang, "A Hierarchical Attention Model for Social Contextual Image Recommendation," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[88] Jinhui Tang, Xiaoyu Du, Xiangnan He, Fajie Yuan, Qi Tian,and Tat-Seng Chua, "Adversarial Training Towards Robust Multimedia Recommender System," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 5, pp. 855–867, 2019.

[89] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, JunMa, and Maarten De Rijke, "Explainable Outfit Recommen- Dation With Joint Outfit Matching and Comment Generation," *IEEE Transactions on Knowledge and Data Engineering*,2019.

[90] Xiaoyan Gao, Fuli Feng, Xiangnan He, Heyan Huang, Xinyu Guan, Chong Feng, Zhaoyan Ming, and Tat-Seng Chua, "Hierarchical Attention Network for Visually-Aware Food Recommendation," *IEEE Transactions on Multimedia*, vol. 22, no. 6, pp. 1647–1659, 2019.

[91] Qiang Cui, Shu Wu, Qiang Liu, Wen Zhong, and Liang Wang, "MV-RNN: A Multi-View Recurrent Neural Network for the Sequential Recommendation," *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[92] Qianqian Wu, Pengpeng Zhao, and Zhiming Cui, "Visual and Textual Jointly Enhanced Interpretable Fashion Recommendation," *IEEE Access*, vol. 8, pp. 68736–68746, 2020.

[93] Ying Lu, Lingkun Luo, Di Huang, Yunhong Wang, and Liming Chen, "Knowledge Transfer In Vision Recognition: A Survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 2, pp. 1–35, 2020.

[94] Devashish Shankar, Sujay Narumanchi, HA Ananya, Pramod Kompalli, and Krishnendu Chaudhury, "Deep Learning Based Large Scale Visual Recommendation andSearch for E-Commerce," *Arxiv Preprint Arxiv:1703.02344*, 2017.

[95] Julian Mcauley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel, "Image-Based Recommendations on Styles and Substitutes," *In Proceedings of the 38th International ACM SIGIR Conference on Research and Development In Information Retrieval*, pp. 43–52, 2015.

[96] ABM Fahim Shahriar, Mahedee Zaman Moon, Hasan Mahmud, and Kamrul Hasan, "Online Product Recommendation System by Using Eye Gaze Data," In *Proceedings of TheInternational Conference on Computing Advancements*, pp.1–7, 2020.

[97] Nelson Silva, Tobias Schreck, Eduardo Veas, Vedran Sabol, Eva Eggeling, and Dieter W Fellner, "Leveraging Eye-Gaze and Time-Series Features to Predict User Interestsand Build a Recommendation Model for Visual Analysis," *In Proceedings of The 2018 ACM Symposium on Eye Tracking Research & Applications*, pp. 1–9, 201

[98] Gaurav & Mathur, Pratistha, "An Attention Mechanism and GRU Based Deep Learning Model for Automatic Image Captioning," *International Journal of Engineering Trends and Technology*, vol. 70, pp. 302-309, 2022.
Crossref, https://doi.org/10.14445/22315381/IJETT-V70I3P234.

[99] Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi, and Zhi Wang, "Friendbook: A Semantic-Based Friend Recommendation System for Social Networks," *IEEE Transactions on Mobile Computing*, vol.14, no. 3, pp. 538–551, 2014.