Computers and Electronics in Agriculture 178 (2020) 105779

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A state-of-the-art survey on recommendation system and prospective extensions

Krupa Patel*, Hiren B. Patel

Kadi Sarva Vishwavidhyalaya, Gujarat, India

ARTICLE INFO

Keywords:
Content-base
Collaborative filtering
Crop recommendation system
Recommendation algorithms
Agriculture

ABSTRACT

With the new era of the Internet, we have a large amount of data available in the form of ratings, reviews, graphs, images, etc. However, still, people face difficulty in finding useful information or knowledge from those data. To address these challenges, recommendation systems come into the picture by providing useful content to the user based on users' history and similarity among users. Content-based and collaborative filtering are two major building blocks of recommendation systems. Recommendation systems have been applied into numbers of a domain such as recommending movies, music, course, literature, items, people, links, location, healthcare, agriculture. In the agriculture domain, appropriate crops to cultivate and selecting applicable pesticides based on land quality and types of crops are interesting factors to consider for a country like India. Initially, we review different types of recommendation systems along with its application area. Subsequently, we explore various parameters to evaluate recommendation systems followed by open issues and research challenges. We further study the work carried out by existing researchers in the said domain. As part of our contribution through this research, we have selected the Agriculture domain and proposed our algorithm for recommending crops based on various parameters. As an outcome of our contribution, a crop is recommended to farmers based on his land. Also, the system recommends a list of lands for a given crop. Using statistical analysis, we achieve accuracy from 93% to 97%.

1. Introduction

Recommendation system (RS) is characterized as programs, that display the most appropriate items or products or services to specific users by predicting a user's interest in an item based on the user's history and information regarding the things, the users and also the interactions between things and users (Bobadilla et al., 2013; Adomavicius and Tuzhilin, 2005). Pre-Internet era, people used to share their experiences via word of mouth and they used to make decisions based on others'/own practices. The main challenge in this mechanism was a limited number of opinions/views for taking an appropriate decision under a certain contained environment. However, in this Internet era, one can get opinions/views/suggestions from millions of people across the world with different environments and conditions. Hence, there is a need for a systematic statistical approach to analyze such a huge amount of data and to derive only useful information for the end-user.

The primitive objective of a Recommendation system is to predict or estimate the most relevant data (items/services). For that purpose, a recommendation system predicts a user's interest by analyzing users' past history, the behavior of similar users along with various data processing techniques. Within the mid-1990s once researchers began to focus on issues that expressly rely on the rating structure of the rating matrix (Adomavicius and Tuzhilin, 2005). But, now it becomes more and more powerful by analyzing user's information from different sources to generate personalized recommendations. Nowadays, recommender systems are viewed as independent research space where suggestions for the target user is based on different criteria includes ratings, reviews, comments, tweet, location, etc.

There are a number of application areas in which, recommendation systems playing a major role such as e-commerce wherein recommendation are useful to provide probably interesting items to target users. There is such example includes Amazon's "Amazon item to item collaborative filtering" (Linden et al., 2003). Movie lens and IMDB for recommending movies, social media platforms such as face-book to promote advertising, twitter to recommend people (Rosa et al., 2019). In addition, the recommendation system can be used in the medical domain to suggest diet, treatments, and medicine (Agapito et al., 2016; Wiesner and Pfeifer, 2014). Learners can also get benefited by having a list of prospective interested courses before them (Obeidat et al., xxxx).

* Corresponding author.
E-mail address: krupa.nick24@gmail.com (K. Patel).

https://doi.org/10.1016/j.compag.2020.105779
Received 10 June 2020; Received in revised form 7 August 2020; Accepted 4 September 2020
0168-1699/

Being researchers of the South Asian sub-continent, we believe the most important application area of recommendation system could be the agriculture sector. To forecast a suitable crop for particular land-based on geographical, weather data, individuals estimating amount, and quality of pesticides for a particular crop (Pudumalar et al., 2017; Kumar et al., 2019). Moreover, the recommendation system can also be utilized to find out nearby places; such as gas stations, restaurants, theatres, etc. based on the user's existing location. Hence, the application domain of recommendation system is very wide and subsequently, researchers have developed recommendation algorithms that can broadly be classified among three principal categories viz. (1) content-based(CB), (2) collaborative filtering (CF) and (3) hybrid (Bobadilla et al., 2013; Adomavicius and Tuzhilin, 2005; Ricci et al., 2011). Each method has its own rationale and number of benefits and challenges which we will discuss in the next section.

Recommending crops for land-based on quality and requirement of minerals is called crop recommendation system. Also, recommending pesticides to protect a particular crop from disease is called pesticides recommendation system. Agricultural scientists and researchers actively seek procedures which may increase crop yields, improve productivity, reduce loss due to disease and insects, and increase overall crop quality. To achieve these goals, technology plays a vital role. Importance of a recommendation system in the agricultural domain is to help farmers by increasing harvest and helping them to select the best crops and fertilizers. Many researchers (Pudumalar et al., 2017; Kumar et al., 2019; IEEE et al., 2018; Mallick et al., 2019; Digital Agriculture System for Crop Prediction, 2019) have been working on developing more efficient algorithm by using latest technologies to recommend crop and pesticides to improve quality and profitability. (Pudumalar et al., 2017) use ensemble learning model with multiple classifiers and generate crop recommendation (Kumar et al., 2019) design pest control technique. (IEEE et al., 2018) use multiple classifiers for crop prediction (Mallick et al., 2019) use an apriori model-based approach to generate recommendation base on market requirement (Digital Agriculture System for Crop Prediction, 2019) design a method to identify crop Disease and suggest a way to prevent from those diseases.

Through this paper, we wish to address following research question: How computational technologies such as machine learning can be utilized to make recommendations for the farmers (in form of recommending land, crop, pesticides and fertilizers) so that overall profitability and productivity can be improved? To be specific, (a) Crop should be recommended based on soil properties, water requirement and weather statistics. (b) Land can be recommended from the properties required by a particular crop. (c) Fertilizers can be recommended from the properties of minerals available in the land. (d) Pesticides can be recommended based on particular crop properties and some other external factors.

Through this study, we generate a process of recommendation that takes 13 different soil properties as input and provides the outcome in the form of crop recommendation. Further, we use various machine learning techniques such as Support Vector Machine(SVM) Classifier, Neural Network and K-Nearest neighbor (KNN).

We propose a Crop Recommendation System(CRS) and Land Recommendation System (LRS) for crop and land recommendations. We calculate the error ratio of our recommendation system and compare the same with Pearson similarity error and Cosine similarity error and we get 5% and 7% error respectively for them. Subsequently, we also measure standard deviation for our outcome which turns out to be 7% and 18% respectively for Pearson similarity and Cosine similarity.

Fig. 1 shows the structure of this research paper which broadly includes methods of recommendation system, open issues and evaluation parameters. After completion of this introduction section, brief about algorithms of the recommendation system is included in Section 2. Various evaluation matrices to evaluate recommendation systems are covered in Section 3. Open issues and challenges are discussed in Section 4. Sections 5 comprises of literature review and comparison of existing research. Section 6 includes our proposal and experimentation for crop recommendation system. At last, we conclude our research by presenting the future prospective direction in the recommendation system.

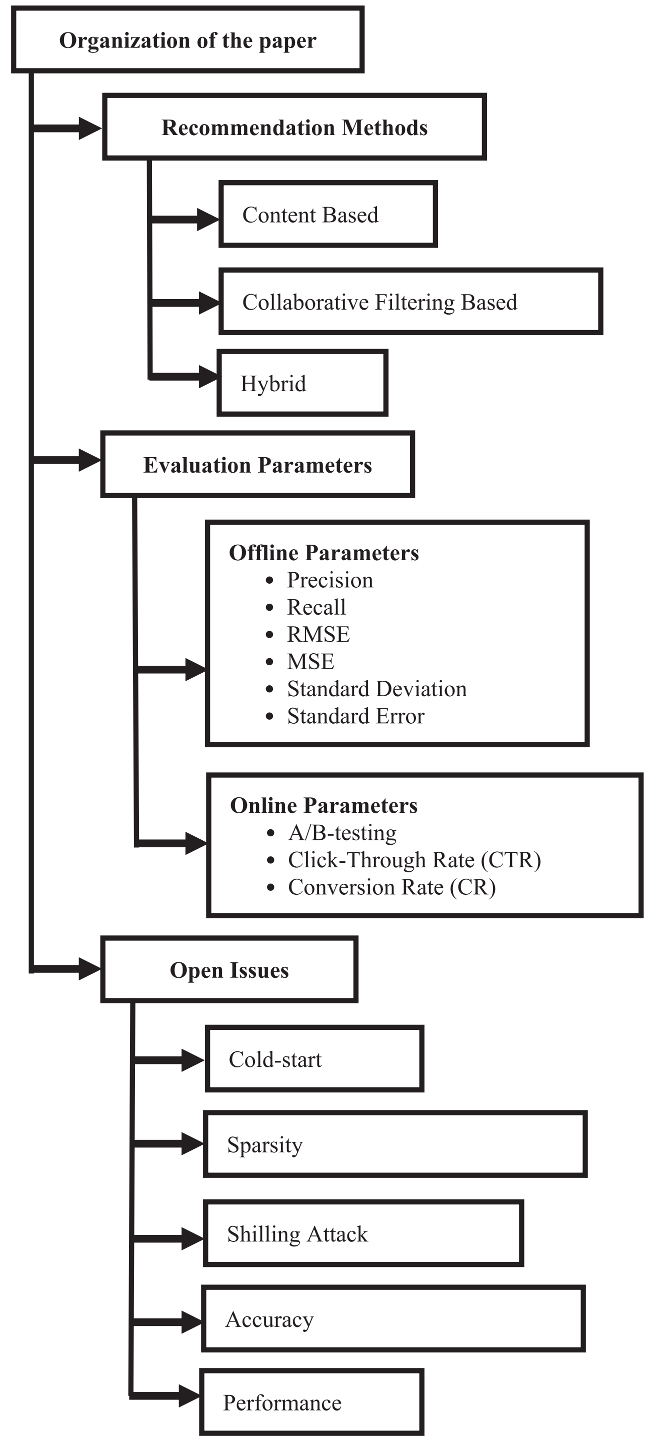

Fig. 1. Organization of research paper.

2. Recommendation algorithms

A recommendation system generates estimation for the target user as a list of item preferences using a correlation between neighbor users and using users' past rating history. Fig. 2 depicts the architecture of a

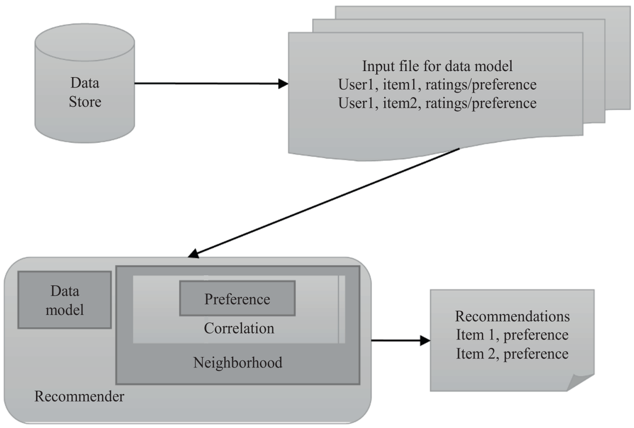

Fig. 2. Architecture of recommendation system.

typical recommendation system. For example, if *f (u, i)* is a function which gives the usefulness of item i for user u; i. e. U X I → R where, R is a totally ordered set (e.g., nonnegative integers within a certain range called rating/preference), U is set of user and I is set of items. So, for each user u ∈ U, we need to choose an item i' ∈ I that maximizes the user's utility.

$$\forall_u \in U,\ i'_u = \mathrm{ArgMax}_{i \in I} f(u, i) \tag{1}$$

2.1. Classification of basic recommendation system

There are multiple diverse algorithms and techniques for creating personalized recommendations (Adomavicius and Tuzhilin, 2005). Moreover, these are usually classified as shown in Fig. 3. In this section, we discuss each recommendation algorithm in detail by taking agriculture as the use case.

2.2. Content-based recommendation systems

The content-based recommendation system generates preferences for the target user based on the user's profile and product features (Bobadilla et al., 2013; Adomavicius and Tuzhilin, 2005). For example, the content-based crop recommendation system is shown in Fig. 4 for recommending crop to a farmer on the basis of land properties and matching those land properties with the required crop property and recommending a list of crops to the farmer.

2.3. Collaborative recommendation systems

Collaborative filtering (CF) recommendation system uses a similarity matrix to generate preferences based on the rating of neighbor users and neighbor items (Bobadilla et al., 2013; Adomavicius and Tuzhilin, 2005). Example of a collaborative-filtering recommendation

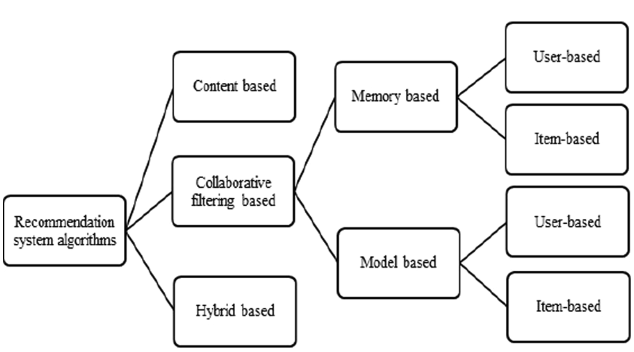

Fig. 3. Classification of recommendation system.

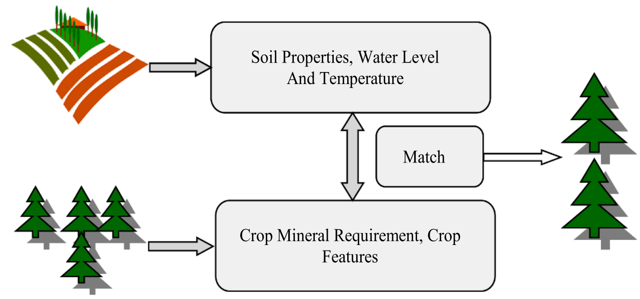

Fig. 4. Content based recommendation system.

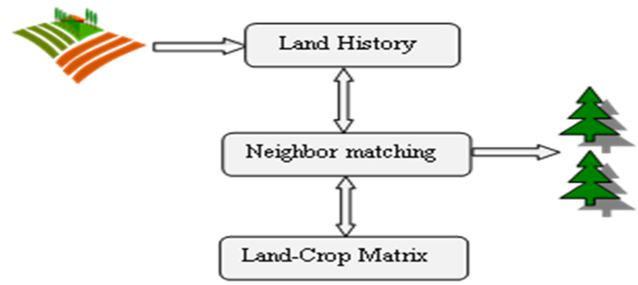

Fig. 5. Collaborative filtering based recommendation system.

system for a crop is shown in Fig. 5. The recommendation for intended land is generated by comparing its properties with other neighbors' land and crop features.

Collaborative filtering recommendation system finds a correlation or similarity between users or items to generate a preference. Most collaborative filtering algorithms use correlation or similarity measure like Pearson and cosine to produce a list of neighbors (Bobadilla et al., 2013). Formulas for Pearson and cosine similarity are presented in equation (2) and (3). After calculating similarity and finding a list of neighbors, the algorithm produces a list of suggestions for target users by using a weighted sum of ratings given by neighbors on target item and a correlation between target users and neighbors by using the formula presented in Eq. (4) (Bobadilla et al., 2013).

$$\mathrm{sim}(u, v) = W_{u,v} = \frac{\sum_{i\in I}(r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i\in I}(r_{u,i} - \bar{r}_u)^2 (r_{v,i} - \bar{r}_v)^2}} \tag{2}$$

where u and v are users and i denotes item.

$$\mathrm{sim}(i, j) = \text{cosÉ}\mu = \frac{\vec{i}\,\text{Â}\cdot\vec{j}}{\|\vec{i}\,\|\text{Â}\cdot\|\vec{j}\,\|} \tag{3}$$

where i and j are items.

$$P_{u,i} = \bar{R}_u + \frac{\sum_{v\in U}(R_{v,i} - \bar{R}_v) \times \mathrm{sim}(u, v)}{\sum_{v\in U}\|\mathrm{sim}(u, v)\|} \tag{4}$$

where u and v are users and i denotes item.

Collaborative filtering has been further categorized into memory-based collaborative filtering and model-based collaborative filtering. In memory-based collaborative filtering the entire rating matrix is used to produce preference and in model-based collaborative filtering rating matrix is split into two parts i.e. train data and test data. Further, train data is used for training the model and the test data is used to assess the model. Once the model has been developed, it could be used to make recommendations. Collaborative filtering can also be categorized into item-based and user-based filtering. User-based techniques use similarities among users and item-based techniques use similarities among items to generate recommendations.

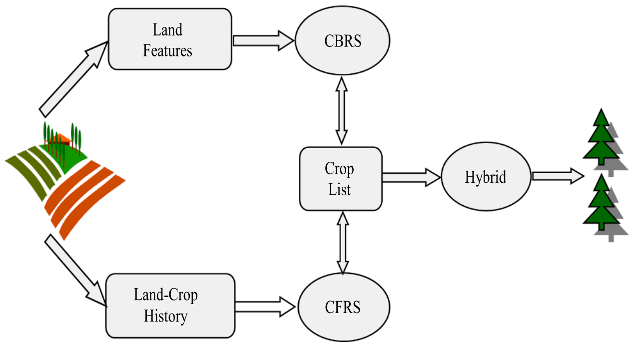

Fig. 6. Hybrid recommendation system.

2.4. Hybrid recommendation systems

A hybrid recommendation the framework uses a combination of content-based recommendation system(CBRS) and collaborative filtering recommendation systems (CFRS) to achieve precisely performance by reducing the drawbacks of the conventional recommendation techniques (Adomavicius and Tuzhilin, 2005). User's side-information, comments, scores, connections, attributes and reviews are used along with the rating matrix in the hybrid recommendation system to enhance performance and precision. Fig. 6 illustrates an example of a hybrid crop recommendation using the soil properties and crop characteristics along with related land characteristics.

Apart from these, researchers also focus on graph base techniques, latent factor model and matrix factorization techniques for recommendation to generate target user preferences (Lu et al., 2015; Almazro et al., 2010).

2.5. Comparison of common techniques

Content-based recommendation and collaborative filtering are two main recommendation algorithms which are widely used by many researchers (Obeidat et al., xxxx; Pudumalar et al., 2017; Kumar et al., 2019; Krisdhamara and BambangPharmasetiawan, 2019; Nassar et al., 2020; Fletcher, 2017; Hong-Xia, 2019; Ifada et al., 2018; IEEE et al., 2018; Mallick et al., 2019; Yu, 2019; Wei et al., 2017; Digital Agriculture System for Crop Prediction, 2019) to generate recommendations. Table 1 shows a comparison between content-based and collaborative filtering recommendation systems. The limited content analysis and over-specialization are the problems of a content-based recommendation system. Researchers use collaborative filtering to address this problem as they provide a wide range of recommendations and generate personalized preference for target users (Obeidat et al., xxxx; Deng et al., 2019; Dong et al., 2016; Fletcher, 2017; IEEE, 2019; Hao et al., 2019; Hao and Zhang, 2018; Chen et al., 2019; Alonso et al., 2019; Hao et al., 2019; Cai and Zhu, 2019; Sang and Vishwakarma, 2017). For example, a recommendation from content-based suggests a crop based on its characteristics and land properties. Collaborative filtering, on the other hand, takes note of the similarity of land and crops and makes recommendations. At last, both methods

Table 2
Primitive issues being addressed by researchers.

Issues → Articles ↓	Cold-start	Sparsity	Shilling attack	Performance/ Accuracy
(Rosa et al., Apr. 2019)				√
(Obeidat et al., xxxx)		√		
(Pudumalar et al., 2017)				√
(Kumar et al., 2019)				√
(Govind et al., 2018)				√
(Deng et al., Jul. 2019)		√		
(Irfan et al., 2019)		√		√
(Krisdhamara and BambangPharmasetiawan, 2019)		√		
(Nassar et al., Jan. 2020)				√
(Dong et al., Apr. 2016)				√
(Fletcher, 2017, 2017)	√	√		
(Hong-Xia, 2019)		√		√
(Ifada et al., 2018)		√		
(An Enhanced Recommendation Algorithm, 2019)				√
(Through et al., 2018)				√
(Mallick et al., 2019)				√
(Hao et al., 2019)			√	
(Hao and Zhang, 2018)			√	
(Chen et al., 2019)			√	
(Alonso et al., 2019)			√	
(Hao et al., 2019)			√	
(Cai and Zhu, 2019)	√			
(Yu, 2019)	√			
(Sang and Vishwakarma, 2017)	√			
(Kumbhar and Belerao, 2018)	√			
(Wei et al., 2017)	√			
(Zhang et al., 2019)		√		

combined and produced the list of crops as a result of the hybrid crop recommendation scheme.

In the next section, the evaluation of the recommendation method using various matrices and parameters will be explored.

3. Evaluation parameters

In the last section, we addressed fundamental algorithms and their comparison. Different performance parameters are used to assess the accuracy and effectiveness of recommendation algorithms. The assessment criteria are classified into two broad categories: offline and online assessment. Offline parameter utilizes non-real-time data sets to check the efficiency and reliability of the recommendation method. Real-time datasets are being used concerning this online parameter.

3.1. Offline evaluation parameters

The offline evaluation of the recommendation system includes precision, recall, train-test analysis, cross-validation tests, mean square error (MSE), root mean square error (RMSE), standard deviation and standard error (Ricci et al., 2011). The train-testing assessment divides data by a split percentage into the train data and test data. 70–80% of the data are used most frequently for the train and 30–20% is used for

Table 1
Comparison between recommendation systems.

Criteria	Content-Based	Collaborative Filtering
Input	Content or Feature of items/users.	User- Item Rating Matrix.
Privacy	Each users are independent from each other.	Users are not independent from each other.
Diversity	Do not able to generate different categories of Recommendation.	Able to Produce Recommendation from Different Categories.
Problems/ Issues	User Cold-Start, Over-Specialization	User Cold-Start, Item Cold-Start, Sparsity, Shilling Attack

Table 3
Recommendation Approaches Being Undertaken by Researchers.

Content Base	Collaborative Filtering Base		Hybrid	
	Memory	Model	Memory	Model
(Rosa et al., Apr. 2019) (Pudumalar et al., 2017; Kumar et al., 2019) (Through et al., 2018) (Mallick et al., 2019; Kumbhar and Belerao, 2018)	(Obeidat et al., xxxx; Deng et al., Jul. 2019) (Dong et al., Apr. 2016; Fletcher, 2017, 2017) (An Enhanced Recommendation Algorithm, 2019; Hao et al., 2019) (Hao and Zhang, 2018; Chen et al., 2019) (Alonso et al., 2019; Hao et al., 2019) (Cai and Zhu, 2019; Sang and Vishwakarma, 2017)	(Krisdhamara and BambangPharmasetiawan, 2019; Nassar et al., Jan. 2020) (Hong-Xia, 2019; Ifada et al., 2018)	(Yu, 2019)	(Govind et al., 2018; Irfan et al., 2019)

the test results. With this approach, the training data is used to train the model and the test data is for evaluating a model. The same process is performed in n-folds in cross-validation testing and the accuracy at each fold is assessed for each folding model. The accuracy of each fold is ultimately used to produce final accuracy. Other two important parameters are precision and recall which are most widely used to measure accuracy by identifying how many of items are accurately classified as compared with the actual class. Precision at k (nearest neighbors) is the proportion in the top-k set of relevant products. Recall at k is the percentage of relevant items in the recommendations out of total relevant items. For example, If 7 items in 10 of the recommended items are appropriate, then precision is 70% and if 7 out of 10 of the relevant items are recommended by the system then recall is 70%. To evaluate the recommendation system, two other matrices called mean absolute error (MAE) and root mean square error (RMSE) are also used. A common method for calculating the prediction accuracy of the recommendation is MAE. MAE is determined by equation (5). RMSE, which calculates the average error, is given in Eq. (6), where A_t is the actual value and F_t is the forecast value.

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}|A_t - F_t| \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}|A_t - F_t|^2} \tag{6}$$

Standard deviation can be defined as percentage deviation of output value from current or mean value of algorithm. Standard deviation (SD) equation is given in equation (7), where x is the output value, $\bar{x}$ is the mean value and n is number of output.

$$\mathrm{SD} = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \tag{7}$$

3.2. Online evaluation parameters

The parameters included for online evaluation are A/B testing, Click-through Rate (CTR) and Conversion Rate (CR) (Beel et al., 2013; Beel et al., 2013; Garcin et al., 2014; Beel and Langer, 2015). Nowadays, the most useful approach is A/B testing. Let us understand by example: We have two algorithms on crop recommendation to perform A/B testing, dividing your target users (farmer) into two groups and fight two algorithms on crop recommendation. CTR indicates how many of top-n recommended items a user is clicking. For example, out of 10 recommended crops, if the farmer selects 3 crops for consideration then CTR is 30%. The conversion rate determines the number of items purchased or viewed by the user. For instance, when two of the ten recommended plants are cultivated successfully, the conversion rate is 20%.

4. Open issues and research challenges

The evaluation parameters for the analysis of recommendation system algorithms have been addressed in the previous section. To suggest products/services, numerous researchers have developed algorithms. However, the framework of recommendation does have inherent difficulties including cold starts, sparsity, shilling attacks, accuracy and efficiency. In this segment, we discuss each issue in detail.

4.1. Cold-start

Due to a lack of knowledge about attributes and features, it is not possible to generate recommendation which is called *cold-start*. This can further be categorized into two types: (1) New-user cold-start and (2) New-item cold-start. When we are not able to produce preferences for the new user due to lack of search history and rating information is called a new-user cold-start issue. Similarly, when it is not possible to place new-items at the top of the preference list, it is known as new-item cold-start. Collaborative filtering is affected by the new user and new item cold-start problems (Bobadilla et al., 2013; Adomavicius and Tuzhilin, 2005; Ricci et al., 2011). However, the content-based recommendation suffers only from a new user cold-start problem. For example, in crop recommendation system it is not able to recommend such crops in the recommendation list if the crop characteristics are not available.

4.2. Sparsity

Collaborative filtering is subject to the sparsity problem due to inadequate rating information available. In general, there are a large number of items and users in this system so, it is very difficult for every user to rate each item. As a result of this, the rating matrix is very sparse and the recommendation produced using a sparse rating matrix affects the accuracy of the algorithm. Various researchers (Krisdhamara and BambangPharmasetiawan, 2019; Hong-Xia, 2019; Ifada et al., 2018) have proposed a model based on a collaborative approach that reduces problems of sparsity in recommendation systems. This is achieved using matrix factorization techniques, prediction methods, and external knowledge base from other social media platforms.

4.3. Shilling attack

A shilling attack or profile injection attack is produced by entering fraudulent ratings, comments and reviews to make items more popular or less popular (Collaborative filtering, 2000). Push and nuke are two forms of shilling attacks. In the push attack type, an attacker enters false profiles for target objects and in the nuke attack; an attacker adds a profile to reduce the popularity of target objects. Several researchers (Hao et al., 2019; Hao and Zhang, 2018; Chen et al., 2019; Alonso et al., 2019) have developed an algorithm that detects fake ratings, reviews, comments using text analysis, sentiment analysis and different

Table 4
Techniques used by researchers for specific outcome.

Methodology → Outcome ↓	Classification	Clustering	Association rule mining	Neural network Deep learning	Opinion mining/sentiment analysis
Accuracy/ Performance	(Pudumalar et al., 2017; Kumar et al., 2019) (Govind et al., 2018; Through et al., 2018)	(Pudumalar et al., 2017)	(Mallick et al., 2019)	(Nassar et al., Jan. 2020; Rosa et al., Apr. 2019)	(Dong et al., Apr. 2016)
Sparsity	(Irfan et al., 2019; Hong-Xia, 2019)	(Obeidat et al., xxxx; Deng et al., Jul. 2019) (Krisdhamara and BambangPharmasetiawan, 2019; Hong-Xia, 2019) (Ifada et al., 2018)	(Obeidat et al., xxxx)	(Sang and Vishwakarma, 2017)	
Cold-start				(Sang and Vishwakarma, 2017)	(Kumbhar and Belerao, 2018)
Shilling attack	(Hao et al., 2019; Hao and Zhang, 2018) (Chen et al., 2019; Alonso et al., 2019) (Hao et al., 2019)				

statistical parameters.

4.4. Accuracy and performance

Inherent issues such as cold-start, sparsity and shilling-attacks have a significant effect on the accuracy and performance of the recommendation systems. The accuracy of recommendations is computed by different parameters described in the preceding section. A trade-off between them is always possible when we think about the quality and effectiveness of the approach proposed. More processing time and system efficiency are needed to increase the accuracy of the recommendation by integrating content data from different sources to generate the most precise recommendation. When creating a recommendation process, selecting parameters for optimization is often a serious issue. We can now boost both parameters to a certain degree with the availability of a fast processing device and computer system, but they continue to compete with each other.

In the next section, we will discuss various researchers' works to resolve the problems inherent in the recommendation systems.

5. Related work

In this section, we are addressing work undertaken to resolve issues such as cold start, sparsity, shilling, accuracy and performance. We present our survey using three dimensions and reviewing them in detail. Table 2 represents the topics addressed by current researchers. Table 3 discusses the algorithm of recommendation used in the represent research, and Table 4 represents an overview of the different approaches taken by other researchers.

We find that a majority of scholars are implementing a recommendation framework that enhances efficiency and precision by solving problems such as sparsity and cold start. Through using clustering and association rules for recommending course, the Sparsity issue has been resolved by (Obeidat et al., xxxx). Some researchers use Support Vector Machine (SVM) classification and decision tree classification methods to generate their recommendations (Pudumalar et al., 2017; Kumar et al., 2019). Recommendation approaches using the Naive Bayes Classification, Support Vector Machine, Singular Value Decomposition, Random Forest to enhance Recommendation accuracy are addressed in (Govind et al., 2018; IEEE et al., 2018; Irfan et al., 2019). Collaborative filtering based on memory with a clustering approach to solve sparsity is given in (Deng et al., 2019; Hong-Xia, 2019; Ifada et al., 2018). Some researchers use neural network and opinion mining methods to solve the problem of cold-start (Fletcher, 2017; Sang and Vishwakarma, 2017; Kumbhar and Belerao, 2018; Wei et al., 2017).

Table 3 compares existing algorithms of creating recommendations for applications such as e-commerce, agriculture, social media, smart city, medical, etc. Most researchers use memory-based collaborative filtering algorithms. The creation of a model-based recommendation is discussed by (Krisdhamara and BambangPharmasetiawan, 2019; Nassar et al., Jan. 2020; Hong-Xia, 2019; Ifada et al., 2018). The mixture of collaborative filtering and content-based recommendations can also be used by adopting a hybrid approach to address sparsity problems (Yu, 2019; Govind et al., 2018; Irfan et al., 2019). Table 4 displays the different approaches used to achieve the desired result by researchers.

We have seen that content-based and collaborative filtering is widely used recommendation algorithms. However, both have some drawbacks such as; content-based recommendation suffers from over-specialization, which cannot recommend more varieties (Lu et al., 2015) and Collaborative filtering algorithm suffers from cold-start, sparsity and shilling attack (Adomavicius and Tuzhilin, 2005; Hao et al., 2019). The algorithm must learn user preferences from the ratings given by users to produce reliable recommendations. Despite all of this, collaborative filtering gives more improved and diverse recommendations.

Table 5
Description of different land properties.

Properties	Description
pH value	Most soils have pH values in the range of 3.5 and 10for higher rainfall areas it ranges from 5 to 7, and drier areas range from 6.5 to 9.
EC	Electrical conductivity is properties of soil which give the ability of a material to transmit electrical charges and expressed on mS/m and dS/m.
Organic Carbon (OC)	Soil organic carbon denotes carbon component of soil which has idea range between 0.51% and 0.75%.
Nitrogen (N)	N includes all form of nitrogen like organic and inorganic. Measured in Kg/H.
Phosphorus (P)	Phosphorus plays a major role in a crop to store and transfer energy for growth of the crop. Ideal value range is 23 to 57 kg/ha.
Potassium (K)	It is required for reproduction of crop. Its ideal value range is 145–337 kg/ha.
Sulphur (S)	It is essential for chlorophyll formation and building blocks of amino acids. Its ideal value is greater than 10 ppm.
Zinc (Zn)	It is an important component of various enzymes, which are responsible for driving many reactions in crops. Its ideal value is greater than 0.6 ppm.
Boron (B)	It is present in the soil in many forms like water solvable and form of acid and it is necessary for the growth of large crops. Its ideal value is greater than 0.5 ppm.
Iron (Fe)	Iron is needed to produce chlorophyll its ideal value is greater than 4.5 ppm.
Manganese (Mn)	It plays a key role in physiological processes, particularly photosynthesis. Its value is greater than 2.0 ppm.
Copper (Cu)	Copper (Cu) is one of the necessary eight essential plant micronutrients. Required for many activities in plants and for chlorophyll and seed production.

Department of Agriculture, Cooperation & Farmers Welfare Ministry of Agriculture and Farmers Welfare Government of India	
Soil Health Card	
Soil Health Card Number - GJ/2016-17/39946330/1	
Validity - From: To:	
Farmer's Details	
Farmer Name	-
Father's/Husband Name	
Address	XXXXX
Mobile No.	-XXXX-XX
Gender	XXXXX
Category	XXXXX
Soil Sample Details	
Date of Sample Collection	15-08-2016
Survey No., Khasra No./ Dag No.	696,377
Farm Size	0.33 Irrigated
Geo Position (GPS)	#VALUE!

Fig. 7a. Soil card farmer detail.

Soil Test Results

STL GANDHINAGAR

Soil Type: Sandy soil

Sr.No.	Parameter	Test Value	Unit	Rating	Normal Level
1	pH	7.8		Moderately alkaline	7, Neutral
2	EC	0.48	dS/m	Normal	0 - 1 dS/m
3	Organic Carbon (OC)	0.66	%	Medium	0.51 - 0.75%
4	Available Nitrogen (N)	--	kg/ha	--	
5	Available Phosphorus (P)	24	kg/ha	Medium	23 - 57 kg/ha
6	Available Potassium (K)	308	kg/ha	Medium	145 - 337 kg/ha
7	Available Sulphur (S)	40.69	ppm	Sufficient	> 10 ppm
8	Available Zinc (Zn)	0.22	ppm	Deficient	> 0.6 ppm
9	Available Boron (B)	5.66	ppm	Sufficient	> 0.5 ppm
10	Available Iron (Fe)	3.04	ppm	Deficient	> 4.5 ppm
11	Available Manganese (Mn)	10	ppm	Sufficient	> 2.0 ppm
12	Available Copper (Cu)	1.6	ppm	Sufficient	> 0.2 ppm

Fig. 7b. Soil card parameters.

Memory-based collaborative filtering uses full matrix information to generate recommendations, therefore more time and memory will be needed; to solve this problem, researchers focus on model-based approach with reduced time and memory requirements to generate recommendations. From our literature survey, we found that so many researchers use matrix factorization, Singular Value Decomposition, machine learning algorithms and data mining techniques along with collaborative filtering. The recommendation system uses data such as ratings, reviews, tags, opinions posted on different websites and analyzed using natural language processing technology and text mining algorithms. This information is often used for the purpose of assessing issues like sparsity, cold-start (Obeidat et al., xxxx). cross-domain recommender system based on kernel-induced knowledge transfer, called KerKT developed by (Zhang et al., 2019) to solve sparsity issue.

Our survey states that, while the recommendation framework has a wide variety of applications, it focuses primarily on e-commerce and the area of social media. Very few studies have been conducted on the use of recommendation system technology to solve real-life problems in the world. The development of a recommendation system for agriculture to help farmers in developing countries like India by providing them with state-of-the-art technology to guide what crops and pesticides are required to produce good quality crops is a great help. Therefore, in our study, we intend to develop a system of recommendations for agriculture to produce the list of best crops to cultivate. Some part of our work has been covered in the below details.

6. Agriculture recommendation system

6.1. Objective

As we know, agriculture is an important sector for improving the economy of developing countries such as the subcontinent countries of South Asia and even though there is growing evidence that technological innovation has a key role to play in increasing agricultural production and enhancing food security; the research and development (R&D) sector in agriculture has a virtuous scope of improvement, specifically for the developing countries like India. As a result of rising demand for food, price volatility, climate change, decreasing natural resources for agricultural production, and rising input costs; agricultural research has again begun to pick up momentum since the turn of the century. The analysis shows that, for a number of different reasons, the overall area of land used for agriculture is constantly being reduced. It is good for society to cultivate high-quality crops through the adaptation of good and modern technology, as a result of this recommendation technology is designed to meet all of these criteria. A system of recommendations of crops and pesticides based on soil quality, the types of soil, crop characteristics, environmental condition and water demands have been developed by very few researchers. Some of them use classification techniques such as support vectors machine classification

Table 6
List of top five recommended crop.

Crop recommendations	
Land	Top Five Crops Recommended for a Given Land
GJ/2018–19/https://doi.org/109025250/5	Groundnut, Wheat, Cotton, Raya, Orange
GJ/2017–18/https://doi.org/1083400/1	Lady finger, Caster Seed, Mango, Bengal Grams

Table 7
List of lands for crop.

Land recommendation	
Crop	Land
Tomato	1. GJ507347-2016–17-40052579 2. GJ520522-2018–19-88565026 3. GJ6092423-2016–17-36632708 4. GJ6096930-2017–18-87821242

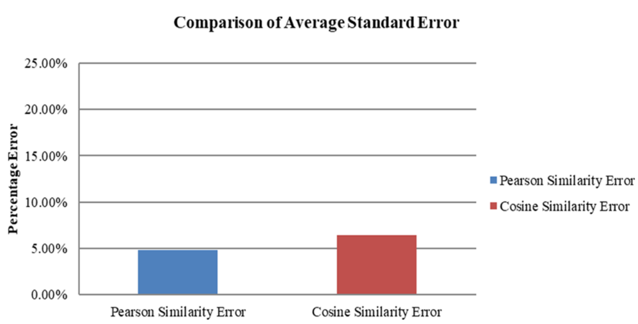

Fig. 8. Standard error.

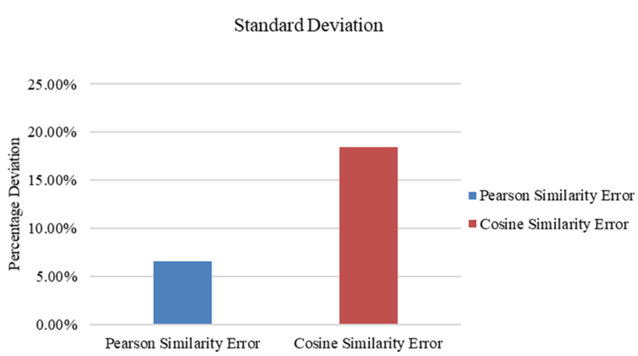

Fig. 9. Standard deviation.

based on land properties (Pudumalar et al., 2017). One of them uses ensemble learning to suggest crops (Adomavicius and Tuzhilin, 2005). Pesticides recommendations based on the identification of plant conditions established by (Kumar et al., 2019). Further research has also been carried out to recommend crop using apriori association rule techniques.

The objective of our research is, therefore, to develop an agricultural recommendation algorithm to suggest crops for farmers using different parameters such as the soil types, soil quality, crop characteristics, climate, water demands, etc. In this section, we will review an algorithm for recommending crops according to their content. The experiment we are discussing in this section is just part of our research to show the importance of the agriculture recommendation system for recommending crops and how it will be beneficial to improve society. We are currently working on a real-time agriculture recommendation system for suggesting crop and pesticides by considering various parameters and machine learning technology.

6.2. Crop recommendation system

As part of the contribution through this research, we propose an algorithm for suggesting crop depending on a content-based recommendation that generates a list of suggested crops for a particular land. This algorithm takes into consideration various land properties such as soil types, pH, Electric Conductivity (EC), Organic Carbon (OC), Nitrogen (N), Phosphorus (P), Sulfur (S), Zinc (Zn), Boron (B), Iron (Fe), Manganese (Mn) and Copper (Cu). A description of each property is presented in Table 5 based on soil test data from the government of India (https://soilhealth.dac.gov.in/). Soil health card data sample is also displayed in Figs. 7a and 7b, which represent farm details including location and owner detail and soil test results respectively.

Crop Recommendation System (CRS) mentioned beneath is based on the contents that use the properties mentioned in Table 5 and suggest the list of five high priority crops (as mentioned in Table 6) on the corresponding properties between the crop and the land for matching soil properties. The algorithm consists of two parts: (1) To compute the rank of each crop and (2) display top-N crop. The algorithm takes two input; one is land soil card detail and other is required property value for each crop. Primitively, the algorithm compares the land with the crop based on their properties mentioned above. If the comparison falls in a predefined range, then we generate a rank for the combination of crop and land. Later, based on the value of the(higher)rank generated, the crop is recommended for that particular land. Further, we sorted each crop for land and generate a top recommendation for given land.

Algorithm CRS: Crop Recommendation System
Computer Rank:
```
For All L in Land Matrix do
  For All C in Crop Property Matrix do
    For All K in Land Properties do
      If L(i,k) is in range of C(j,2k-1) and C(j,2k)
        Rank(i,j) = rank(i,j) + 1
      End if
    End for
  End for
Sort Rank of each Land in Descending Order
End for
```
Display Top-N Crop:
```
For all L in Land Matrix do
  For all C in Crop do
    Display (L_i, C_j, Rank_j)
  End for Crop.
End for Land Matrix.
```

To provide flexibility to the farmers for identifying lands appropriate for a particular crop, we have developed a separate algorithm called land recommendation system (LRS) which is shown as under. Our LRS algorithm will generate a list of top-N suitable Land for each Crop based on Crop Property (as mentioned in Table 7). The algorithm LRS takes two input; one is land soil card detail and other is required property value for each crop. Compute matching property between land and crop for each crop and generate ordered ranking of land for the given crop and suggest which land is most suitable to the given type of crop.

Algorithm LRS: Land Recommendation System
Computer Rank:
```
For All L in Land Matrix do
  For All C in Crop Property Matrix do
    For All K in Land Properties do
      If L(i,k) is in range of C(j,2k-1) and C(j,2k)
        Rank(i,j) = rank(i,j) + 1
      End if
    End for
  End for
Sort Rank of each Crop in Descending Order
End for
```
Display Top-N Land:
```
For all C in Crop do
  For all L in Land Matrix do
    Display (C_i, L_j, Rank_j)
  End for Land Matrix.
End for Crop.
```

6.3. Discussion and experimental study

We experimented with R programming on our sample data obtained from the Indian Government Soil testing card website (https://soilhealth9.dac.gov.in). For the evaluation of our algorithm; a sample database was developed from the given website. The data set consists of 86 lands of Gujarat state and 24 crops. Lands and plants are classified according to 12 properties that we have already described in Table 5. Table 6 show the result of our experiment as a list of recommended crops for given land identified by survey number of that land. We have kept N (number of top N recommended crops) a variable which can be modified to suggest more/less number of crops based on the priority (higher to lower).

The outcome of Algorithm LRS has been demonstrated in Table 7.

To evaluate our recommendation systems, we calculate the cosine and Pearson similarity between the lands using crop ranking generated by our algorithms and compare them with the actual land similarity. Subsequently, we assess our crop recommendation algorithms using evaluation parameters such as standard error and standard deviation. Details of these parameters have been discussed in Section 3.

Fig. 8 depicts standard error for both the methods wherein one can see 4.80% and 6.45% similarity in Pearson and cosine respectively. Hence, one may conclude that the accuracy of our algorithms is therefore between 93% and 96% relative to the actual recommendation. Also, the standard deviation between expected and real data was determined. As per the diagram of standard deviation shown in Fig. 9 we can see that there is a deviation of 18% and 7% in cosine and Pearson similarity techniques respectively. Through this relation, we infer that the standard deviation for a Pearson's measure of similarity is less than the cosine similarity measure. Therefore, we conclude that the Pearson similarity fits better with the recommendation than that of cosine similarity. This shows that our crop recommendation systems produce more accurate results as compared to other conventional mechanisms. This shows that content-based recommendation system ensures compliance with the recommendation of crops.

7. Conclusion and future work

We have addressed the system and methods of recommendation in this research paper. We have also discussed challenges, parameters of evaluation and the application of the agricultural recommendation system. We have also addressed a recommendation framework for suggesting crop, which includes part of our research work. We may conclude from our research that applying the recommendation system technology in agriculture to improve quality crop production is beneficial for farmers and the ecosystem itself. We are currently working on a machine learning framework for recommending crops and pesticides for agricultural recommendations. Further experiments with our crop recommendation algorithm will be conducted in future by including more parameters such as weather data, demographic conditions and market conditions. We were also intended to work on a recommendation for pesticides and pest control techniques to increase the production of quality crops and to maximize the profits to farmers. Once, a farmer has selected a particular crop based on the land from our recommendation system, in the next part of our research, we wish to recommend (a) pesticides for the selected crop and (b) fertilizers for the land by incorporating machine learning techniques. We acknowledge the Government of India for making the data publically available.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2020.105779.

References

Adomavicius, G., Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. 17 (6), 734–749. https://doi.org/10.1109/TKDE.2005.99.
Agapito, G., et al., 2016. DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion. Int. Conf. Wirel. Mob. Comput. Netw. Commun., no. August 2017, doi: 10.1109/WiMOB.2016.7763190.
Almazro, D., Shahatah, G., Albdulkarim, L., Kherees, M., Martinez, R., Nzoukou, W., 2010. A survey paper on recommender systems.
Alonso, S., Bobadilla, J., Ortega, F., Moya, R., 2019. Robust model-based reliability approach to tackle shilling attacks in collaborative filtering recommender systems. IEEE Access 7(c), pp. 41782–41798, doi: 10.1109/ACCESS.2019.2905862.
Beel, J., Langer, S., 2015. A comparison of offline evaluations, online evaluations, and user studies in the context of research-paper recommender systems. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9316, 153–168. https://doi.org/10.1007/978-3-319-24592-8_1.
Beel, J., Genzmehr, M., Langer, S., Nürnberger, A., Gipp, B., 2013. A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation. ACM International Conference Proceeding Series, 7–14. https://doi.org/10.1145/2532508.2532511.
Beel, J., Langer, S., Genzmehr, M., Gipp, B., Breitinger, C., Nürnberger, A., 2013. Research paper recommender system evaluation: A quantitative literature survey. ACM International Conference Proceeding Series, 20(April 2013), 15–22. https://doi.org/10.1145/2532508.2532512.
Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A., 2013. Recommender systems survey. Knowledge-Based Syst. 46, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012.
Cai, Y., Zhu, D., 2019. Trustworthy and profit: A new value-based neighbor selection method in recommender systems under shilling attacks," Decis. Support Syst., vol. 124, no. March, p. 113112, 2019, doi: 10.1016/j.dss.2019.113112.
Chen, K., Chan, P.P.K., Zhang, F., Li, Q., 2019. Shilling attack based on item popularity and rated item correlation against collaborative filtering. Int. J. Mach. Learn. Cybern. 10 (7), 1833–1845. https://doi.org/10.1007/s13042-018-0861-2.
"Collaborative filtering," A State Art Surv. Shilling Attack Collab. Filter. Based Recomm. Syst., vol. 12, no. 6, p. 783, 2000, doi: 10.3156/jfuzzy.12.6_65_1.
Deng, J., Guo, J., Wang, Y., Jul. 2019. A Novel K-medoids clustering recommendation algorithm based on probability distribution for collaborative filtering. Knowledge-Based Syst. 175, 96–106. https://doi.org/10.1016/j.knosys.2019.03.009.
Digital Agriculture System for Crop Prediction & Disease Analysis Based on Machine Learning, 2019. 21(X), 1065–1070.
Dong, R., O'Mahony, M.P., Schaal, M., McCarthy, K., Smyth, B., Apr. 2016. Combining similarity and sentiment in opinion mining for product recommendation. J. Intell. Inf. Syst. 46 (2), 285–312. https://doi.org/10.1007/s10844-015-0379-y.
Fletcher, K.K., 2017. A method for dealing with data sparsity and cold-start limitations in service recommendation using personalized preferences. In: Proceedings - 2017 IEEE 1st International Conference on Cognitive Computing, ICCC 2017, 2017, pp. 72–79, doi: 10.1109/IEEE.ICCC.2017.17.
Garcin, F., Faltings, B., Donatsch, O., Alazzawi, A., Bruttin, C., Huber, A., 2014. Offline and online evaluation of news recommender systems at swissinfo.ch. RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems, 169–176. https://doi.org/10.1145/2645710.2645745.
Govind, B.S.S., Tene, D.R., Saideep, K.L., 2018. Novel recommender systems using personalized sentiment mining. In: 2018 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCT 2018, 2018, doi: 10.1109/

CONECCT.2018.8482394.

Hao, Y., Zhang, F., 2018. Detecting shilling profiles in collaborative recommender systems via multidimensional profile temporal features. IET Inf. Secur. 12 (4), 362–374. https://doi.org/10.1049/iet-ifs.2017.0012.

Hao, Y., Zhang, F., Chao, J., 2019. An ensemble detection method for shilling attacks based on features of automatic extraction. China Commun. 16 (8), 130–146. https://doi.org/10.23919/JCC.2019.08.012.

Hao, Y., Zhang, F., Wang, J., Zhao, Q., Cao, J., 2019. Detecting Shilling Attacks with Automatic Features from Multiple Views. Secur. Commun. Networks 2019. https://doi.org/10.1155/2019/6523183.

Hong-Xia, W., 2019. An improved collaborative filtering recommendation algorithm.

Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique. IEEE, 2018.

An Enhanced Recommendation Algorithm Based on Modified User-Based Collaborative Filtering. IEEE, 2019.

Ifada, N., Prasetyo, E.H., Mularab, 2018. Employing sparsity removal approach and fuzzy C-means clustering technique on a movie recommendation system.

Irfan, R., Khalid, O., Khan, M.U.S., Rehman, F., Khan, A.U.R., Nawaz, R., 2019. SocialRec: A context-aware recommendation framework with explicit sentiment analysis. IEEE Access 7, 116295–116308. https://doi.org/10.1109/ACCESS.2019.2932500.

Krisdhamara, K.M., BambangPharmasetiawan, W.P., 2019. Improvement of collaborative filtering recommendation system to resolve sparsity problem using combination of clustering and opinion mining methods Krisdhamara.

Kumar, A., Sarkar, S., Pradhan, C., 2019. Recommendation system for crop identification and pest control technique in agriculture.

Kumbhar, N., Belerao, K., 2017. Microblogging reviews based cross-lingual sentimental classification for cold-start product recommendation. In: 2017 Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2017, pp. 1–4, 2018, doi: 10.1109/ICCUBEA.2017.8463697.

Linden, G., Smith, B., York, J., 2003. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Comput. 7 (1), 76–80. https://doi.org/10.1109/MIC.2003.1167344.

Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G., 2015. Recommender system application developments: A survey. Decis. Support Syst. 74, 12–32. https://doi.org/10.1016/j.dss.2015.03.008.

Mallick, P.K., Balas, V.E., Bhoi, A.K., Zobaa, A.F., 2019. Development of a Model Recommender System for Agriculture Using Apriori Algorithm, vol. 768. Springer Singapore.

Nassar, N., Jafar, A., Rahhal, Y., Jan. 2020. A novel deep multi-criteria collaborative filtering model for recommendation system. Knowledge-Based Syst. 187. https://doi.org/10.1016/j.knosys.2019.06.019.

R. Obeidat, R. Duwairi, and A. Al-Aiad, A Collaborative recommendation system for online courses recommendations.

Pudumalar, S., Ramanujam, E., Rajashree, R.H., Kavya, C., Kiruthika, T., Nisha, J., 2017. Crop recommendation system for precision agriculture.

Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., Ricci, F., 2011. Recommender systems handbook. In: Recommender Systems Handbook. https://doi.org/10.1007/978-0-387-85820-3.

Rosa, R.L., Schwartz, G.M., Ruggiero, W.V., Rodriguez, D.Z., Apr. 2019. A knowledge-based recommendation system that includes sentiment analysis and deep learning. IEEE Trans. Ind. Inform. 15 (4), 2124–2135. https://doi.org/10.1109/TII.2018.2867174.

Sang, A., Vishwakarma, S.K., 2018. A ranking based recommender system for cold start & data sparsity problem. In: 2017 10th Int. Conf. Contemp. Comput. IC3 2017, vol. 2018-Janua, no. August, pp. 1–3, 2018, doi: 10.1109/IC3.2017.8284347.

Wei, J., He, J., Chen, K., Zhou, Y., Tang, Z., 2017. Collaborative filtering and deep learning based recommendation system for cold start items. Expert Syst. Appl. 69, 1339–1351. https://doi.org/10.1016/j.eswa.2016.09.040.

Wiesner, M., Pfeifer, D., 2014. Health recommender systems: Concepts, requirements, technical basics and challenges. Int. J. Environ. Res. Public Health 11 (3), 2580–2607. https://doi.org/10.3390/ijerph110302580.

Yu, P., 2019. Merging attribute characteristics in collaborative filtering to alleviate data sparsity and cold start. In: Proc. 2019 IEEE 3rd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2019, no. Itnec, pp. 569–573, 2019, doi: 10.1109/ITNEC.2019.8729461.

Zhang, Q., Lu, J., Wu, D., Zhang, G., 2019. A cross-domain recommender system with kernel-induced knowledge transfer for overlapping entities. IEEE Trans. Neural Networks Learning Syst. 30 (7), 1998–2012. https://doi.org/10.1109/TNNLS.2018.2875144.