*I.J. Information Engineering and Electronic Business*, **2017, 3, 26-35**
Published Online May 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2017.03.04

# A Survey on the Generation of Recommender Systems

**Rahul Singh**
CSE Department, Chandigarh university, Mohali, 140413, India
Email: rahulcu25@gmail.com

**Kanika chuchra and Akshama Rani**
CSE Department, Chandigarh university, Thapar university, India
Email: kanika.cse@cumail.in, shama1mittal@gmail.com

***Abstract***—In the era of Internet, web is a giant source of information. The constantly growing rate of information in the web makes people confused to decide which product is relevant to them. To find relevant product in today's era is very time consuming and tedious task. Everyday a lot of information is uploaded and retrieved from the web. The web is overloaded with information and it is very essential to cop up with this overloaded and overlooked information. Recommender systems are the solution which can help a user to get relevant information from the bulk of information. Recommender systems provide customized or personalized and non personalized recommendations to interested users. Recommender systems are in its evolution stage. Recommender systems have been evolved from first generation to third generation through second generation. First generation or Web 1.0 recommender systems deal with E-commerce, Second generation or web 2.0 recommender systems use social network and social contextual information for accurate and diverse recommendations, and Third generation recommender systems use location based information or internet of things for generating recommendations. In this paper, three generation of recommender systems and are discussed. Similarity measures and evaluation metrics are used in these generations are also discussed.

***Index Terms*—**Recommender system, Web 1.0, Web 2.0, Web 3.0, similarity measures, evaluation metrics.

## I. INTRODUCTION

The process of making recommendations has been widely used since many years in every aspect of life. Before the dawn of Internet, recommender systems were still there but in personalized form. For example, if a person wanted his daughter to marry, he took recommendations from match makers and family members according to suitability of his daughter. If anyone wanted to visit any tourist place, he normally took recommendations from his friends who were known to that place or who have visited there earlier. Even in the choice of clothing, we generally took recommendations of our companion whether this clothe suits us or not. As internet grows, many users started to take recommendations from the web. Recommender systems give personalized and nonpersonalized suggestions to users. Personalized means give recommendations according to user's preferences such as: if a user wants to go to restaurant, he has to give his preferences about location, food and many more things beforehand for getting good recommendations of restaurants according to his taste and in non personalized recommendations, recommender systems provide recommendations according to the content: For example, news recommender system such as Google news, recommends users news similar to the news that user has been watching. Recommendations are not something which user type in search engine and get the results; it is the search result which comes after matching the user's query. Recommendations are provided according to user's interest. Recommender systems are the way to deal with the overload of information reflected in the increasing volume of information artifacts in the web. Recommender systems analyze existing information on the user activities in order to predict future preferences. Every recommender system implements a different paradigm for generating recommendations in heterogeneous domain. Recommender systems have been recognized as an important tool on web science and e-commerce applications. Many famous websites have been using recommender systems for improving their sales such as Netflix, a web based movie rental service and Amazon, an e-commerce company [15]. Netflix Prize contest was a well known contest in the history of recommender systems, it offered 10 million US dollar for improving the collaborative filtering algorithm by 10.6% accuracy. Many approaches are used in recommender systems and each approach has its advantages and limitations. Currently, the following types of approaches are found in practice:

*Collaborative filtering* [1-5], [7]: In the mid 1990's collaborative filtering algorithms were introduced. This approach is very successful approach among all. Recommender systems based on it recommends items according to the rating behavior of users. Ratings of the recommended items are given by the users similar to active user to whom recommendations are offered or similar to the items previously rated by the active user. But there is a contradiction that two users may have similar preferences in one category but different preferences in another. This approach does not work efficiently in sparse datasets which denotes less number of ratings on the items. It suffers from Cold start problem that means there is no rating for newly added items and newly added users do not rated any item then how recommendations is possible for these kinds of users and items[44].

*Content based filtering* [1-2], [5]: this approach recommends items to active user according to the content information of items.

*Demographic filtering* [1], [5]: this approach considers the ratings of users who have same age, sex and location as active user has for the recommendations.

*Knowledge based filtering* [1]: Recommendations are based on the matching of user's need and set of options available. These recommender systems have knowledge about how a particular item meets the need of a particular user. For example knowledge mined from user's profile will be helpful in recommendations.

*Hybrid filtering methods* [1], [6]: Hybrid denotes combination of two or more than two approaches. Hybrid approach is used to mitigate the limitations of individual approaches. Collaborative filtering approach can be combined with content based, knowledge based and demographic based approaches. The collaborative filtering recommender systems are the most successful approach among all at this time. Collaborative filtering (CF) approach used by recommender systems is a way to predict the usage of items for users based on the similarity among user's preferences and the preferences of other users. Amazon, Epinion adopted social recommender systems for improving the accuracy of their prediction

*Social recommender systems:* Since the introduction of the Tapestry systems by Goldberg, recommender systems have been in existence [7]. Social media has become famous now days. Eminent examples are social resource sharing sites: Delicious (Bookmarks), Flickr (Images), CiteULike (Bibliographic), Youtube (videos), Slashdot (information) and Social networking sites: MySpace (Music), Twitter (micro blogger), Epinion (Product Review), Flixter (movie review), Facebook, Linkedin. A social network models relationship between different users and information is exchanged between them according to their relationship. Data in Social network is usually shown with graphs and matrices. Social network is a set of nodes (actor or user) and edges (relationship between actors). As shown in Fig. 1, Relationships between users are of two types: directional (marriage, cusions) and non directional (friendship, seller–buyer, and employer-employee). In the social recommendations: tag recommendations, people recommendations, and content recommendations take place. In the social network, items can be social entities such as persons or group of persons.

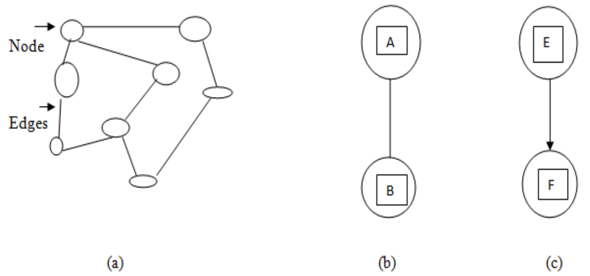


Fig.1. a) Social Network (Node-Actor, Edges-Social-Relationship), b) Nondirectional Relationship (marriage), c) Directional Relationship (Seller- Buyer)

This paper is structured as follows: Section 2 presents the different generation of recommender systems Section 3 describe about similarity measures used in recommender systems, Section 4 presents evaluation metrics used in recommender systems and Section 5 presents the conclusion and future work.

## II. Recommender Sysyetms Generation

In this section, some of the research studies related to the Web 1.0, Web 2.0 and Web 3.0 recommender systems are briefly presented here.

### *A. Web 1.0 Recommender Systems*

Since the Web 1.0 recommender systems developed, item recommendations have become the main area of concern for researchers who investigated many approaches of recommendations. In Table 1, comparison between approaches of first generation recommender systems is shown.

D. Goldberg et al. [7] present a tapestry email system that efficiently uses collaborative filtering approach for filtering the queries of user in mailing list. Users read only those documents which are reviewed by other users before.

J. Herlocker et al. [16] propose neighborhood approach of collaborative filtering for finding similarities between users and items. They use automated collaborative filtering approach for increasing the accuracy of recommendations. They also discuss many evaluation metrics such as coverage, accuracy and ROC.

B. Sarwar et al. [17] propose that item based collaborative filtering is better than user based collaborative filtering. They address two issues of recommender system: scalability of algorithms and quality of recommendations by introducing new item based collaborative filtering approach. As relationship between users is dynamic and relationship between items is static, item based approach takes less online computation time. They basically deal with scalability issue face by user based approach.

G. Adomavicius *et al.* [1] present first generation recommender systems. They also classify recommendation methods into three main approaches: content based, collaborative and Hybrid. Content based recommendations consider the past transaction behavior of users and recommend items that are similar to the previously purchased items by analyzing their content. Collaborative filtering recommend items to users on the basis of rating behavior of like-minded users and hybrid approach combine content based and collaborative approaches to mitigate the limitations of each approach for providing more accurate results. The Web 1.0 recommender systems have two major building blocks: users and items and there is a binary relationship among them. Users rate items on the basis of their preferences and rating can be binary (like or unlike an item) or on the scale of 1 to 5. As shown in Fig. 2, users (U1, U2 U3) and items (I1, I2, I3, I4) are related with each other. Items I1, I2 and I3 are liked by user U1; Items I1 and I2 are liked by user U2; Items I2, I3 and I4 are liked by user U3. On the basis of similarity in user's preferences, Items I3 and I4 liked by U1 and U2 correspondingly are recommended to user U2 according to their rank.

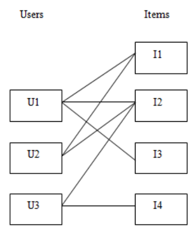


Fig.2. User-Item Relationship (Binary)

Table 1. Web 1.0 Recommender systems

| Approach | Advantages | Limitations | Domain |
|---|---|---|---|
| Collaborative Filtering [7] | Efficient algorithm for filter queries in e-mail system | Security | Electronic Document |
| Collaborative Filtering (Neighborhood) [16] | Good Prediction Accuracy | Scalability | Movie (Movielens) |
| Collaborative Filtering (Item based) [17] | Scalability And High online Performance | Performance decreases as neighborhood size decreases | Movie (Movielens) |
| Content Based Filtering(TF-IDF) [1] | Fast information retrieval in Text based items | Limited content analysis, Overspecialization and New User Problem | News article, Books and Movies |
| Collaborative Filtering (Memory based and Model based) [1] | No overspecialization and limited content analysis problem | New user Problem, New item problem (cold start) and sparsity | News article, Books and Movies |
| Hybrid Filtering (Collaborative and Content based) [1] | Overcome limitations of content based and collaborative based approach | Complex to implement | News article, Books and Movies |
| Collaborative Filtering, Content based and Demographic Filtering [5] | High Precision | Complex to implement | Restaurants |
| Improvement in Collaborative Filtering (kNN) Using pareto dominance [18] | Improvement in Quality Measures | Improvement is applied only on memory based Collaborative Filtering | Movie (Movielens, Netflix) |
| Matrix Factorization Technique(CF) [15] | Accuracy in recommendation as compared to classical Nearest Neighbor approach | Loss of information due to factorization | Movie (Netflix) |
| Collaborative Filtering using Singularity concept [19] | Improvement in Prediction quality and Recommendation quality | Large Formulation Required | Movie ( Movielens, Netflix, Filmaffinity) |
| Collaborative Filtering using Apache Mahout [20] | Provide flexibility using pre-existing algorithms and solves the scalability problem by using hadoop. | Collaborative filtering with Mahout approach is not suitable for time sensitive applications | E-Commerce |

M. J. Pazaani [5] uses a framework of collaborative filtering, content based filtering and demographic filtering for recommending restaurants.F. Ortega *et al.* [18] propose an improvement in traditional collaborative filtering recommender system. In traditional collaborative filtering, a posteriori phase is imprecise means a large number of active user's neighbours are selected. In the neighbours of active user some are less representative and some are more representative. The new approach propose in this report uses pareto dominance[80-20 rule] to perform pre filtering step in k-nearest neighbor selection process for eliminating the less representative users from the neighborhood of active user.

Y. Koren *et al*. [15] propose the matrix factorization techniques and perform comparison between matrix factorization and classic nearest neighbor technique on Netflix database. In the result matrix factorization is superior to nearest-neighbor techniques for generating item recommendations. It also allows the incorporation of more information such as implicit feedback, confidence levels and temporal effects.

J. Schafer *et al.* [3] explain how a recommender system changes the business of Ecommerce and help Ecommerce sites in increase its sales. They analyze several sites which use recommender systems and more than one recommender system. They create taxonomy of recommender systems and discussed technologies that are used in recommender system and input files that are needed from customers.

J. Bobadilla *et al*. [2] present the recommender systems evolution. As time passed, recommender systems evolve from first generation, second generation to third generation. In first generation recommender systems, e-commerce recommender system comes into scene and various approaches are used to make them efficient and in second generation, social information is also incorporated in recommendation process for improving the prediction results and overcoming the limitations of first generation recommender systems and in third or current generation location aware recommender systems are introduced.

J. Bobadilla *et al*. [19] propose the improvement in traditional similarity measure. In traditional collaborative filtering method, the most similar user are discovered for whom we want to make recommendation. In this report, new approach that takes care of contextual information is considered. Here, Singularity is measured. If there is greater singularity, there is more similarity. The results are tested on Netflix, Movielens databases and shows excellent behavior.

J. Bobadilla *et al.* [14] describe the improvement in K-nearest neighbor (KNN) algorithm, the core of collaborative filtering. The KNN algorithm is non-scalable in nature and has high execution time. This algorithm is based on repeated execution of similarity metric. In this report, new similarity metric HWSimilarity is introduced. This metric has high quality recommendation and employing low- cost hardware circuits.

S. Walunj *et al.* [20] propose a successful implementation of a mahout framework provided flexibility in using pre-existing algorithms. It described challenges of collaborative filtering like Scalability, Synonymous, Grey sheep, Shilling attacks, Diversity. It solved the problem of scalability by using the hadoop platform because it is built on the hadoop framework. Apache mahout offers testimony that a recommendation system provides customizable recommendations enable online companies to perform business more effectively.

F. Gedikli *et al.* [38] describe the importance of explanation in recommender system. Explanations are given with the recommended products and describe why this item is recommended or why system thinks that item is liked by user. This kind of explanation improves effectiveness and quality of recommendation and users are more satisfied. Amazon is using the explanation with recommendation.

D. Melamed et al. [39] present the collaborative information filtering system (Marcol) with pricing mechanism. This is the market based collaborative information filtering system that takes user judgment or feedback as input and good quality recommendation as output. In this report, first time pricing mechanism is employed with collaborative filtering. With this approach user satisfaction is increased.

A. C. M. Fong et al. [40] propose a web recommender system that takes user’s temporal web access behavior as input and provide personalized recommendation on user’s portable or mobile devices. Fuzzy logic is used to represent user’s behavior and to construct knowledge base. By using this approach, user of web enabled mobile device can easily find out the information of his interest.

### *B. Issues in Recommender Systems*

M. A. Ghazanfar *et al.* [32] propose an approach for reducing the error rate in collaborative filtering algorithm caused by gray sheep user problem. They identify gray sheep user by using various clustering approach like K plus means clustering algorithm, various distance measures and improved centroid selection approaches. For generating accurate recommendation, user’s profile is used. This is the first attempt in the direction of solving gray sheep user problem and with this attempt, accuracy, coverage of recommendation is also improved.

Y. Blanco-Fernández et al. [33] present a strategy for solving the problem of overspecialization by using the reasoning techniques taken from semantic web. These techniques are implemented in recommender system for digital Television and these techniques provide accurate recommendation.

B. Lika et al. [34] present a technique for solving the cold start problem. Cold start problem arises in collaborative filtering system. Collaborative filtering approach recommend item to the user on the basis of the rating of the similar user who share the same interest with active user. But if a user new to the system and did not rate any item yet, recommendation can not be possible in this scenario because user’s neighbor can not be find out. If a new item is added to the system and nobody has rated it then this item can also not be recommended. This problem is called cold start problem. Cold Start problem further divided into three parts: a) new user, b) new item, c) new user, new item. In this report B. Lika et al. present a method for removing the new user problem by incorporating demographic data for finding the similar user and then classified the new user in a particular group and employ prediction mechanism.

C. Chung *et al.* [35] propose an approach for shilling attack detection. This approach filters out the malicious rating from the recommender system. Collaborative filtering approach is widely used in the recommendation and due to its open nature, Its suffers from many vulnerabilities. Many attack detection algorithms such as PCA based algorithm, classification based and detector based on SPC (statistical process control) algorithms have been introduced for handling this issue but all of these are restricted by various constraints. Beta-protection (βP) algorithm is introduced for removing this problem.

J. Zhan et al. [36] propose a privacy preserving approach in recommender system. As recommender system is successfully used in e-Commerce, users want relevant and precise recommendation and this can be possible only when two or more than two companies merge their database for overcoming the problem of limited database. Due to privacy disclosure, there can be numerous hazards that can affect the quality of recommendation. To avoid these hazards, cryptology approaches and scalar product protocol is used.

Z. Huang et al. [37] propose a framework for alleviating the problem of sparsity inherent in collaborative filtering, Collaborative filtering approach is one of the most successful approaches which consider user neighbor's data and feedback for recommendation but this data is sparse. The association retrieval and spreading activation algorithms are used to find out transitive association between users from their past behaviors and feedback. These algorithms are compared with traditional collaborative filtering approaches and infer that activation based approaches outperforms.

### *C. Web 2.0 Recommender Systems*

Tag recommendation has become the favoured topic of interest since the growth of social tagging sites. In Table 2, comparison between approaches of second generation recommender system is shown.

Xu *et al.* [21] propose a tag recommendation algorithm that recommends only high quality tags. High quality tags denote appropriate tags that do not include spam and noise because user annotates tags to resources in free form and some of these tags are noise and spam and these tags are not play major role in the tag recommendation. With the help of tag co-occurrence frequency, relationship between two tags is derived. Tag co-occurrence frequency denotes how many times two tags are attached to the same item. If two tags are attached to the items frequently, tag co-occurrence frequency will be high and tags are closely related otherwise co-occurrence frequency will be low.

D. Lee [22] proves item based similarity shows better result than user based similarity in unilateral relation. Unilaterally relationship denotes one sided relationship (e. g. in micro blogging sites, users are connected with each other without mutual agreement or if user find other's content relevant or worth, he starts following him).

Singurbjőrnsson *et al.* [23] analyze how and what kind of tags are annotated to items by users and global co–occurrence frequency metric, a metric is used to measure the relationship among tags.

A. Ji *et al.* [42] present a novel approach for recommender system by incorporating social tagging information in collaborative filtering. They use Naïve bayes classifier on candidate tag set that is collection of tags of each user for finding similarity between two users.

A. A. Barragáns-Martínez *et al.* [24] propose a tag based recommender system that improves coverage and diversity of recommendations. They use tags to build user and item tag clouds and compare these clouds for recommendations. User tag cloud is made up of tags that user yet not annotated to item and item tag cloud is made up of tags that are annotated to an item.

Golder *et al.* [41] analyze the structural and dynamic aspect of collaborative tagging system and discuss about the tag frequency, popularity and stability in the proportion of tags in the resources. They also explains the problem incur in tagging information such as synonym, homonym, polysemy.

N. Zheng *et al.* [25] explore tag and time is two important factors in social tagging systems. They integrate these factors into collaborative filtering to show better performance in recommendation results. They propose three strategies: tag weight, time weight, fusion of tag and time weight to generate ratings in user-item matrix.

K. Tso-sutter *et al.* [26] propose tag aware recommender system that incorporates tagging information into traditional recommender systems (user based and item based CF). Social tagging Systems are based on three building blocks: users, items, and tags for generating recommendations. These building blocks have three dimensional correlations with each other. To incorporate tagging information into Collaborative filtering, three dimensional correlations <users, items, tags> is reduced to three two dimensional correlations such as < users, items >, < users, tags >, < items, tags>. In three dimensional correlation users (U1, U2, U3) annotates tags (T1, T2, T3, T4, T5, T6, T7, T8) to items (I1, I2, I3). As shown in Fig. 3, Users, items and tags have ternary relationship and this relationship splits into three binary relations.

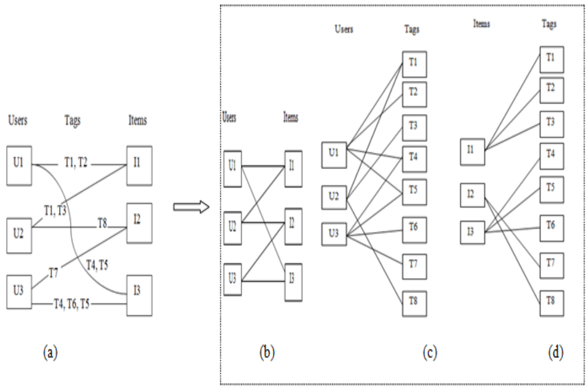


Fig.3. (a) Users-Tags-Items relation (ternary), (b) Users-Items relation, (c) Users-Tags relation, (d) Items-Tags relation.

Zhao *et al.* [27] present a novel tag based collaborative filtering approach which uses semantic distance between user generated tags as a measure for finding similarity between users. This approach selects neighbors of user effectively. WorldNet is used to find the distance between tags.

H. Kim *et al.*[28] propose a hybrid approach for tag aware recommender system that mitigate the limitations of existing approaches in social tagging systems. They explore association rule, bigram approach and trust relationship for tag and item recommendations in their proposed framework.

O. Arazy *et al.* [29] present a framework that improves the accuracy of recommender systems and they explain four constructs such as homophily, trust, tie strength and social capital and these construct will impact the advice taking capability of recipients.

Table 2. Web 2.0 Recommender systems

| Approach | Advantages | Limitations | Domain |
|---|---|---|---|
| Collaborative Tagging [21] | High quality tags | Loss of information | My web 2.0 |
| Traditional Collaborative Filtering with social network [22] | Overcome the shilling attacks | Semantically rich information is important | Social bookmarking sites (Delicious) |
| Collective knowledge in tagging [23] | Gracefully handle the expansion of vocabulary | Not implemented online | Online photo sharing site (Flickr) |
| Collaborative Filterinng, Content Based and Tagging [24] | Tag remains present whether item is present or not | Does not consider time in the weight of tag | TV program (quevo.tv) |
| Collaborative Tagging with Temporal Information [25] | Information drift with time is considered | Approach considers only one dataset | Social bookmarking site (CiteULike) |
| Incorporation of Tag with standard CF algorithms (Fusion approach) [26] | Compare performance with or without incorporation of tags in CF algorithms and proves better quality recommendation results with tagging information | Only item recommendation is considered (tag recommendation is not considered) | Music community website (Last.Fm) |
| Collaborative Filtering with Tagging information [27] | Semantic information of tag is added in recommendation process and generate accurate recommendations | Community wisdom is not considered | Bookmarking system (Dogear with IBM Lotus connections) |
| Social Tagging System with Hybrid Framework [28] | Implicit Trust information is incorporated | Scalability Problem | Movie (Movielens) |
| Recommender System Incorporating social Network information [29] | Improvement in Recommendation | Privacy | Theoretical Framework (Social Network Information) |

*D. Web 3.0 Recommender Systems*

Web 3.0 recommender systems come into existence after the increasing use of mobile devices. Recommender system using internet of things and location based recommender systems are becoming more widespread. In Table 3, comparison between approaches of third generation recommender system is shown.

Table 3. Web 3.0 Recommender systems

| Approach | Advantages | Limitations | Domain |
|---|---|---|---|
| Collaborative recommender with space and time similarity [30] | Location and time based item similarity is better than Rating based similarity in IoT environment | Expensive Implementation | Customized NFC Tagged Item |
| Location aware recommender system(LARS) [31] | Better Quality Recommendation | Complex Implementation | Spatial Rating Dataset (FourSquare) |

M. Muñoz-Organero et al. [30] they discuss collaborative filtering algorithms using Internet of Things. This recommender is relying on user- object interaction and space-time interaction patterns. They used time and location for finding the similar user in this IoT based recommender. For finding similarities, they used NFC (near field communities) based social networking information from users who belongs to the same region. In IoT recommender, Sensed user-object pattern and user-location pattern is far better than user rating approach for finding similarities.NFC is one of the major components of RFID tags. It is used for face to face communication. It is used to identify the content embedded in physical object. NFC technology in mobile devices is used to communicate with other NFC

Device as shown in Fig. 4 and allows peer to peer communication between two mobile device and read or writes RFID tags. By touching the RFID enabled device, user can read description of that object for example Bar code.

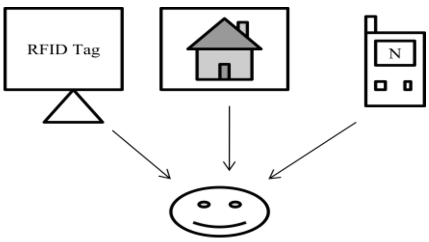


Fig.4. RFID Tagged Objects [30].

J. Levandoski et al. [31] propose a recommender system named Location aware recommender system (LARS). This Recommender system considers location based rating and spatial properties of user. This system enhances scalability and recommendation quality of system. LARS knows item location by using a technique that is called travel penality.

## III. Similarity Measure

Similarity measure is used to calculate the similarity between users. In Memory based collaborative filtering, user to user similarity and item to item similarity is find out by using these similarity measure. In these measures, user's rating is considered as shown in Table 4 and on the basis of rating behavior, similarity is calculated. Memory based CF such as kNN algorithm is dependent on these similarity measures. Pearson correlation coefficient [1], [18], Cosine similarity [1], [18], Adjusted cosine similarity [2], Mean squared difference [1], [18], Euclidean Distance [2], Jaccard index [43]. Pearson correlation coefficient measures the relationship between two variables. Cosine similarity is used to measure the similarity between two vectors and cosine angle between them. Adjusted cosine is used to measure the similarity between two items rated by the entire user to find out the item to item similarity. It overcomes the drawback of cosine similarity by considering the rating scale of different user. Mean squared difference also calculate the similarity between two user by considering difference in two user's rating. Euclidean distance calculates the distance between two points and on the basis of distance, similarity can be found. Jaccard index is new similarity measure between two set.

Table 4. User- Item Rating Matrix

| Users | Item | | | |
|---|---|---|---|---|
| | ***I*** | ***J*** | ***K*** | ***L*** |
| A | 1 | 3 | $\varepsilon$ | 4 |
| B | 3 | $\varepsilon$ | 2 | 5 |
| C | ε | 4 | 3 | $\varepsilon$ |

**Pearson Correlation coefficient**

$$\text{Corr(a,b)} = \frac{\sum_{i\in S_{a,b}}(r_{a,i}-\bar{r}_a).(r_{b,i}-\bar{r}_b)}{\sqrt{\sum_{i\in S_{a,b}}(r_{a,i}-\bar{r}_a)^2 . \sum_{i\in S_{a,b}}(r_{b,i}-\bar{r}_b)^2}} \quad (1)$$

Where $S_{a,b}$ be the set of items rated by both user a and b, $r_{a,i}$ be the rating given by user a to item i , $r_{b,i}$ be the rating given by user b to item i and $\bar{r}_a$, $\bar{r}_b$ be the average rating of user a and user b. Corr(a,b)= [-1,0,1] ; Here,-1 shows negative correlation, 0 shows independent and 1 shows positive correlation.

**Cosine Similarity**

$$\text{Cos}(a,b) = \frac{\sum_{i\in S_{a,b}} r_{a,i}.r_{b,i}}{\sqrt{\sum_{i\in S_{a,b}} r_{a,i}^2 . \sum_{i\in S_{a,b}} r_{b,i}^2}} \quad (2)$$

Where $S_{a,b}$ be the set of items rated by both user a and b, $r_{a,i}$ be the rating given by user a to item i, $r_{b,i}$ be the rating given by user b to item i.

$Cos(\theta) = \frac{a.b}{\|a\|\|b\|}$, where a.b is a dot product of two vectors $\vec{a}$ ,$\vec{b} \Longrightarrow \| a \|\| b \| \cos\theta$.

**Adjusted Cosine Similarity**

$$\text{Acos(i,j)} = \frac{\sum_{x\in X}(r_{x,i}-\bar{r}_x).(r_{x,j}-\bar{r}_x)}{\sqrt{\sum_{x\in X}(r_{x,i}-\bar{r}_x)^2} . \sqrt{\sum_{x\in X}(r_{x,j}-\bar{r}_x)^2}} \quad (3)$$

Where X is a set of users whose rating is considered for finding out similarity between two items. $\bar{r}_x$ is the average rating of user.

**Mean Squared Difference**

$$\text{Msd(a,b)} = 1 - \frac{1}{\#S_{a,b}} \sum_{i\in S_{a,b}} \left(\frac{r_{a,i}-r_{b,i}}{max-min}\right)^2 \quad (4)$$

Where $S_{a,b}$ be the set of items rated by both user a and b,$\#S_{a,b}$ is the cardinality of set $S_{a,b}$ . $r_{a,i}$ be the rating given by user a to item i , $r_{b,i}$ be the rating given by user b to item i. max and min is the maximum and minimumrating value of the system.

**Euclidean Distance**

$$\text{d(a,b)} = \sqrt{(b_1-a_1)^2(b_2-a_2)^2..(b_{n-}a_n)^2}$$
$$\text{d(a,b)=d (b,a)} = \sqrt{\sum_{i=1}^{n}(b_i-a_i)^2} \quad (5)$$

Where a $(a_1, a_2 \ldots. a_n)$ and b $(b_1, b_2 \ldots\ldots b_n)$ are two points in Euclidean n- space.The distance from a to b and b to a is given above.The position of point in Euclidean n-space is Euclidean vector.

**Jaccard index**

$$\text{J(U,V)} = \frac{U\cap V}{U\cup V} \quad (6)$$

This coefficient measure similarity between two sets U and V.

Here Size of intersection is divided by the size of union of two sets U and V. $0 \le J(U,V) \le 1$

## IV. Evaluation Metric

Evaluation is an important part of any recommender system because without evaluation we can not infer whether results of the system are right or not. Comparison between two systems can be possible only after performing evaluation. With the help of evaluation, improvement in the system can be done. Quality of techniques, algorithm, and procedure is evaluated for good prediction and recommendation. Evaluation metrics can be classified into different metrics as shown in Fig. 5.

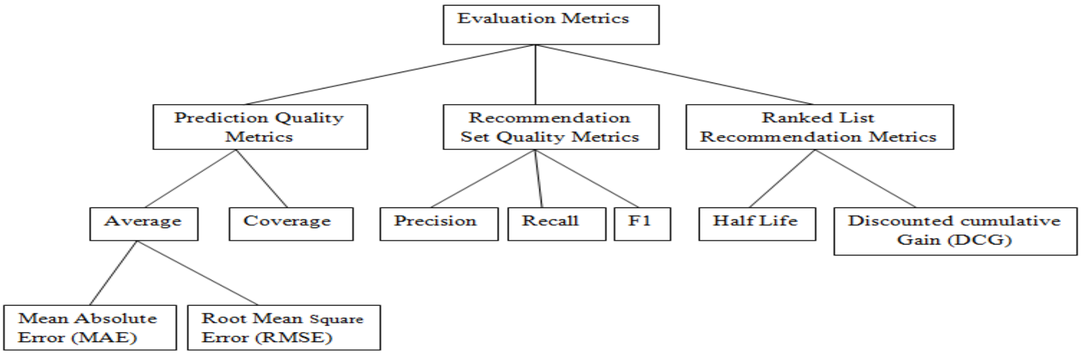


Fig.5. Evaluation Metrics

### A. *Prediction Quality Metric*

This metric is used to evaluate the prediction of the system. Prediction is to know what will user like in future. Accuracy and coverage are important metric for evaluating the prediction. Accuracy is further divided into *MAE*, *RMSE*, and *NMAE* [1], [2], [16], [17], [18].

Suppose $D_a$ ={ i ∈ I : $p_{a,i} \neq \varepsilon \wedge r_{u,i} \neq \varepsilon$ }, set of items rated by user a having prediction and rating value is not equal to null.Error in the prediction is calculated by taking the difference between actual rating and predicting rating, $|p_{a,i} - r_{a,i}|$ informs the error in the system. Here MAE and RMSE is discussed.

$$\boldsymbol{MAE} = \frac{1}{\#A}\sum_{a\in A}\left(\frac{1}{\#D_a}\sum_{i\in D_a}|p_{a,i} - r_{a,i}|\right) \tag{7}$$

$$\boldsymbol{RMSE} = \frac{1}{\#A}\sum_{a\in A}\sqrt{\frac{1}{\#D_a}\sum_{i\in D_a}(p_{a,i} - r_{a,i})^2} \tag{8}$$

**Coverage (*C*)**

It is used to measure the capacity of user's neighbours to predict new items. User's coverage is shown in equation [2], [18]

$$C = \frac{1}{\#A}\sum_{a\in A} 100 \times \frac{\#\{i\in I\ |r_{a,i}=\varepsilon \wedge p_{a,i}\neq\varepsilon\}}{\#\{i\in I|r_{u,i}=\varepsilon\}} \tag{9}$$

### B. *Recommendation Set Quality Measure*

These measures evaluate the quality of recommendation. User confidence is based on the accuracy and quality of the recommendations offered by the system. Recommendation quality measure informs whether system is recommended relevant content or not.

Suppose $X_a$ is the set of recoomendation to user a and $Z_a$ is the set of N recommendations Offered to user a. Here, $\theta$ is the threshold value of relevancy and It is assumed that all user accept N recommendations [1], [2], [8], [16], [17], [18].

**Precision (*P*)**

It is the ratio between the relevant recommended items to the total number of item recommended.

$$P = \frac{1}{\#A}\sum_{a\in A}\frac{\#\{i\in Z_a|r_{a,i}\geq\theta\}}{N} \tag{10}$$

**Recall(*R*)**

It is the ratio between the relevant recommended items to the total number of relevant items.

$$R = \frac{1}{\#A}\sum_{a\in A}\frac{\#\{i\in Z_a|r_{a,i}\geq\theta\}}{\#\{i\in Z_a|r_{a,i}\geq\theta\} + \#\{i\in Z_a^c|r_{a,i}\geq\theta\}} \tag{11}$$

**F1 Metric**

The Combination of precision and recall is called F1 metric.

$$\text{F1} = \frac{2\times P\times R}{P+R} \tag{12}$$

### C. *Ranked list Recommendation Metric*

When the amount of recommended items is enormously large then user considers only first fiew items in the list of recommendation. The error occurs in these recommendation are more serious than last recommendation in the list. The ranking Metric considers this scenario. Half Life and Discounted cumulative gain (DCG) are the two most important ranked metrics [2].

**Half Life (*HL*)**

IT specifies the interest of user decreases exponentially when user move down in the list of recommendation.

$$\text{HL} = \frac{1}{\#A}\sum_{a\in A}\sum_{i=1}^{n}\frac{max\left(r_{a,p_i} - d_r, 0\right)}{2^{(i-1)/(\alpha-1)}} \tag{13}$$

**Discounted Cumulative Gain (*DCG*)**

It specifies user's interest decreases logarithmically.

$$\text{DCG} = \frac{1}{\#A}\sum_{a\in A}\left(r_{a,p_i} + \sum_{i=2}^{K}\frac{r_{a,p_i}}{\log_2 i}\right) \tag{14}$$

Here, recommendation list is represented by $p_1,\ p_2 ......... p_n$ , $r_{a,p_i}$ denotes user a's rating for item $p_i$,$\alpha$ denotes number of items in the list and there is half chance of user will review it.

## V. Conclusion and Future Work

Recommender systems proved to be an important tool of coping up with the information overload problem. E-commerce sites successfully used this for improving their sales. Then, recommender systems are used in Social commerce and social networking sites to remove the sparsity and many mores problems of traditional recommender systems. Nowdays, location based and RFID tag based recommender systems are used. In this paper, a brief survey on three generation of recommender system is presented. Similarity measures and evaluation metrics are used in these generations are also discussed.

## REFERENCES


[1] G. Adomavicius, and A. Tuzhilin, "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, pp. 734-749, 2005.

[2] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109-132, 2013.

[3] J. Schafer, J. Konstan, and J. Riedl, "Recommender Systems in E-Commerce," *In proceeding of ACM conference on Electronic commerce in Computing Systems*, pp. 158-166, 1999.

[4] N. Antonopoulus, and J. Safer "Cinema screen recommender agent: combining collaborative and content based filtering," *IEEE Intelligent Systems*, vol. 21, pp. 35-41, 2006.

[5] M. J. Pazaani, "A Framework for collaborative, content based and Demographic Filtering," *Artificial Intelligent Review*, vol. 13, pp. 393-408, 1999.

[6] R. Burke, "Hybrid recommender systems: survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, pp. 331-370, 2002

[7] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using Collaborative Filtering to weave an information Tapestry," *Communications of the ACM*, pp.61-70, 1992.

[8] H. Ma, T. CH. Zhou, M. R. Lyu, and I. King, "Improving recommender systems by incorporating social contextual information," ACM Transactions on Information Systems, vol. 29, pp. 1-23, 2011.

[9] S. Tan, J. Bu, CH. Chen, and X. He, "Using rich social media information for music recommendations via hypergraph model," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 7, pp. 1-22, 2011.

[10] C .Chen, J. Zeng, X. Zheng, D. Chen, "Recommender systems Based on Social Trust Relationships", *IEEE 10th International Conference on e-Business Engineering*, pp. 32-37, 2013.

[11] W. Carrer-Neto, M. L. Hernandez-Alcaraz, R. Valencia-Garcia, and, F. Garcia-Sanchez, "Social knowledge based recommender systems. Application to the movie domain," *Expert Systems with Applications*, vol. 39, pp. 10990-11000, 2012.

[12] A. A. Kardan, and M. Hooman, "Targeted Advertisement in Social Networks using Recommender Systems," *7th International Conference on e-Commerce in Developing Countries: With Focus on e-security (ECDC)*, pp. 1-13, 2013.

[13] Y. Li, C. Wu, and C. Lai, "A social recommender mechanism for e-commerce: combining similarity, trust, and relationship," *Decision Support Systems*, Vol. 55, pp. 740-752, 2013.

[14] J. Bobadilla, A. Hernando, F. Ortega and A. Gutiérrez, "Collaborative filtering based on significances," Information *Sciences* vol.185, pp.1–17, 2012.

[15] Y. Koren, R. Bell, and C. Volinsky, " Matrix Factorization Techniques For Recommender Systems," *IEEE Computer Society*, vol. 42, pp. 30-37, 2009.

[16] J. Herlocker, J. Konstan, A. Brochers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 230-237, 1999.

[17] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Proceedings of the 10th international conference on World Wide Web*, pp. 285-295, 2001.

[18] F. Ortega, J. Sánchez, J. Bobadilla and A. Gutiérrez, "Improving collaborative filtering-based recommender systems results using Pareto dominance," *Information Sciences*, vol.239, pp.50-61, 2013.

[19] J. Bobadilla, F. Ortega and A. Hernando, "A collaborative filtering similarity measure based on singularities," *Information Processing and Management*, vol.48, pp.204-217, 2012.

[20] S. Walunj and K. Sadafale, "An Online Recommendation System for E-commerce Based on Apache Mahout Framework,'' *ACM*, 2013.

[21] Z. Xu, Y. Fu, J. Mao and D. Su, "Towards the semantic web: collaborative tag suggestions," *In Collaborative web tagging workshop Scotland*, 2006.

[22] D. lee, "How to measure the information similarity in unilateral relations: the case study of *Delicious*," *In 10 Proceedings of the International Workshop on Modeling Social Media, ACM*, 2010.

[23] B. Sigurbjörnsson and R. V. Zwol, "Flickr tag recommendation based on collective knowledge," *In Proceedings of the 17th international conference on worldwide web*, pp. 327–336, ACM, 2008.

[24] A. B. Barragáns-Martínez, M. Rey-López, E. Costa-Montenegro and F. A. Mikic-Fonte et al., "Exploiting Social Tagging in a Web 2.0 Recommender System," *IEEE Internet Computing*, vol. 14, pp. 23-30, 2010.

[25] N. Zheng and Q. Li, "A recommender system based on tag and time information for socialtagging systems," *Expert Systems with applications*, Vol. 38, pp. 4575-4587, 2011.

[26] K. Tso-sutter, L. Marinho and L. Schmidt-Thieme, "Tag-aware recommender systems by fusion of collaborative filtering algorithms," *In Proceedings of the 2008 ACM symposium on Applied computing*, pp. 1995-1999, 2008.

[27] S. Zhao, N. Du, A. Nauerz, X. Zhang, Q. Yuan and R. Fu, "Improved recommendation based on collaborative tagging behaviors," *In Proceedings of the 13th international conference on Intelligent user interfaces*, pp. 413-416, 2008.

[28] H. Kim and H. Kim, "A framework for tag-aware recommender systems," *Expert Systems with Applications*, vol. 41, pp. 4000-4009, 2014.

[29] O. Arazy, N. Kumar, and B. Shapira, "Improving Social Recommender Systems," *IT Professional*, vol. 11, pp. 38-44, 2009.

[30] M. Muñoz-Organero, G. A. Ramírez-González, P. J. Muñoz-Merino and C. D. Kloos, "A Collaborative Recommender System Based on Space-Time Similarities," IEEE Pervasive Computing, Vol. 9, pp. 81-87, 2010.

[31] J. Levandoski, M. Sarwat, A. Eldawy, and M. Mokbel, "LARS: A Location-Aware Recommender System," 2012 IEEE 28th International Conference on Data Engineering (ICDE), pp. 450-461, 2012.

[32] M. A. Ghazanfara and A. Prügel-Bennett, "Leveraging clustering approaches to solve the gray-sheep users

problem in recommender systems,"*Expert Systems with Applications*, Vol. 41, pp. 3261-3275, 2014.
[33] Y. Blanco-Fernández, J. J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer and M. López-Nores, "Providing Entertainment by Content-based Filtering and Semantic Reasoning in Intelligent Recommender Systems," *IEEE Transactions on Consumer Electronics*, Vol. 54, pp. 727-735, 2008.
[34] B. Lika, K. Kolomvatsos and S. Hadjiefthymiades, "Facing the cold start problem in recommender system," *Expert Systems with Applications*, Vol. 41, pp. 2065-2073, 2014.
[35] C. Chung, P. Hsu and S. Huang, "βP: A novel approach to filter out malicious rating profiles from recommender systems," *Decision Support Systems*, Vol. 55, pp. 314-325, 2013.
[36] J. Zhan, C. Hsieh, I. Wang, T. Hsu, C. Liau and D. Wang, "Privacy-Preserving Collaborative Recommender Systems," *IEEE Transactions on Systems, Man, and Cybernetcis, Part C : Applications and Reviews*, Vol. 40, pp. 472-476, 2010.
[37] Z. Hung, H. Chen and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," *ACM Transactions on Information Systems (TOIS)*, Vol. 22, pp. 116-142, 2004.
[38] F. Gedikli, D. Jannach and M. Ge, "How should I explain? A comparison of different explanation types for recommender systems,"*International Journal Human-Computer Studies*, Vol. 72, pp. 367-382, 2014.
[39] D. Melamed, B. Shapira and Y. Elovici, "Marcol : A Market-Based Recommender System,"*IEEE Intelligent Systems*, Vol. 22, pp. 74-78,2007.
[40] A. C. M. Fong, B. Zhou, S. C. Hui, G. Y. Hong and A. Do, "Web Content Recommender System based on Consumer Behavior Modeling," *IEEE Transactions on Consumer Electronics*, Vol. 57, pp. 962-969. 2011.
[41] S. A. Golder and B. A. Huberman, "Usage patterns of collaborative tagging systems," *Journal of Information Science*, vol. 32 pp. 198–208, 2006.
[42] A. Ji, C. Yeon, H. Kim and G. Jo, "Collaborative tagging in recommender systems," *In Advances in artificial intelligence (AI2007)*, pp. 377–386, 2007.
[43] M.Y.H.AI-Shamir, "Power coefficient as a similarity measure for memory-based collaborative recommender systems," *Expert Systems with Applications*, 2014.
[44] Sharma, R., Vinayak. S., and Singh, R.K., "Guide Me: A Research Work Area Recommender System", Intelligent Systems and Applications 9 30-37 2016.

## Authors' Profiles

**Rahul Singh:** Belongs to Kanpur, Uttar Pradesh, India and born on April 15, 1989. He is an Assistant Professor in CSE Department of Chandigarh University. He has received his Masters degree from Thapar University. His areas of interests are Theory of Computation, Software Engineering and sentimental analysis.

**Kanika Chuchra:** Belongs to Jalalabad, Punjab, India and born on Jan 18, 1991. She is an Assistant Professor in CSE Department of Chandigarh University. She has received his Masters degree from GNDU Amritsar. Her areas of interests are Data Mining and text mining.

**Akshama:** Belong to Jind Haryana, India and born on July 2, 1990. She has received his Masters degree from Thapar University Patiala. Her areas of interests are Big data and Recommender System.

**How to cite this paper:** Rahul Singh, Kanika chuchra, Akshama Rani,"A Survey on the Generation of Recommender Systems", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.9, No.3, pp.26-35, 2017. DOI: 10.5815/ijieeb.2017.03.04