



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 6      Issue: IV      Month of publication: April 2018**

**DOI: <http://doi.org/10.22214/ijraset.2018.4561>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Survey on Various Methodologies for Recommendation System

Jay Borade<sup>1</sup>, Shraddha Makwana<sup>2</sup>, Puja Gupta<sup>3</sup>, Sayalee Lanjewar<sup>4</sup>

<sup>1</sup>Assistant Professor Information Technology at Fr CRCE, Mumbai University

<sup>2, 3, 4</sup>Final year Information Technology student at Fr CRCE, Mumbai University,

**Abstract:** In recent years the way of exchange of information between any web application and user has changed due to evolution of recommendation system. While considering such application we also have to consider massive data associated with it. Recommendation method aids to make search effortless and grants users with personalized content and services. Considering that intent many algorithms have been designed. All the paper long we have explained various algorithms and methods for recommendation system using model based techniques. The Foremost part of the paper covers traditional techniques and algorithms used in recommendation system. Further, the paper highlights use of clustering and classification in recommendation system. Latter part of the paper illustrates improvement in place recommendation system using KNN algorithm. It ends with providing one of the customized recommendation methods with combination of above mentioned algorithm and data mining.

**Keywords:** Recommendation, Clustering, Classification, K-nearest Neighbors, Rating based Recommendation

## I. INTRODUCTION

Recommender systems are information filtering systems that deals with the problem of information overload [1] by filtering vital information fragment out of large amount of dynamically generated information according to user’s preferences, interest, or observed behavior about item. [2]. Recommender systems are beneficial to both service providers and users [3]. They also reduce transaction costs of finding and selecting items in an online system [4]. The focus of this paper is to provide effectual technique for relevant recommendation system which introduces user with the base of designing recommendation system.

## II. OVERVIEW OF VARIOUS RECOMMENDATION TECHNIQUES

Before taking an insight of various algorithms, it is important to first have better understanding of different recommendation techniques. The use of efficient and accurate recommendation techniques is very important for a system that will provide good and useful recommendations. [5]

The figure [5] below shows assortment of such recommendation techniques –

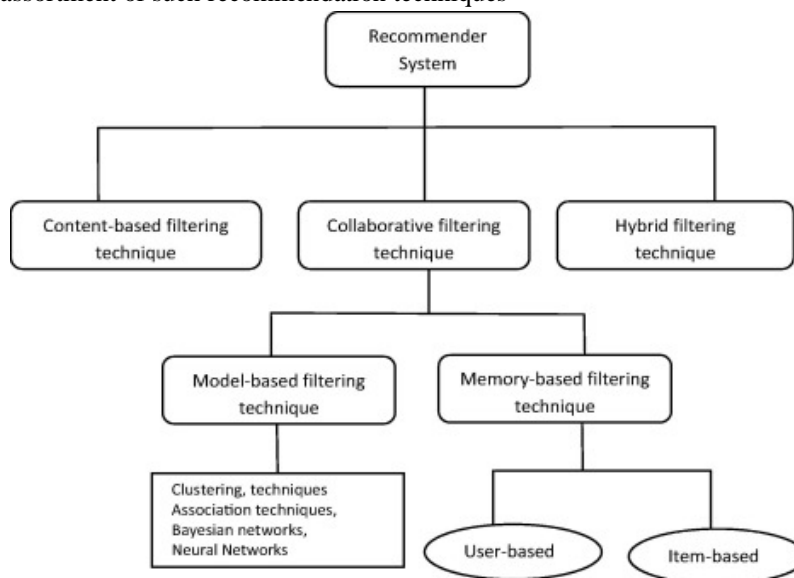


Fig - 1. Different Recommendation Techniques

### A. Content-based filtering

Content-based recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his taste. Taste is based on how the user rated items [6]. Based on such previous taste of user the items are compared and the one which are most similar will be recommended to user.

### B. Collaborative filtering

The idea of collaborative filtering is to find users in a community that share their tastes [7]. If two users have same or almost same rated items in common, then they have similar tastes. Such users build a group or a so called neighborhood. A user gets recommendations to those items that he hasn't rated before, but that were already positively rated by users in his neighborhood.[8]

- 1) *Memory-based filtering* – Once a neighbor of a user is found, different algorithms can be used to combine the preferences of neighbors to generate recommendations. Memory-based Collaborative Filtering can be achieved in two ways through user-based and item-based techniques. User based collaborative filtering technique calculates similarity between users by comparing their ratings on the same item, and it then computes the predicted rating for an item by the active user as a weighted average of the ratings of the item by users similar to the active user where weights are the similarities of these users with the target item. Item-based filtering techniques compute predictions using the similarity between items and not the similarity between users. It builds a model of item similarities by retrieving all items rated by an active user from the user-item matrix, it determines how similar the retrieved items are to the target item, then it selects the  $k$  most similar items and their corresponding similarities are also determined. Prediction is made by taking a weighted average of the active users rating on the similar items  $k$ [5]. Various similarity measures such as cosine, jaccard, pearson etc. can be used here to find distance between users and items.
- 2) *Model-based filtering* – Model-based approaches, utilize collected user preferences to build models that, provide item relevance predictions. In this way, model-based approaches lead to faster responses at the time of recommendation [9]. Model-based recommendation systems involve building a model based on the dataset of criteria such ratings. In other words, we extract some information from the dataset, and use that as a model to make recommendations without having to use the complete dataset every time. This approach potentially offers the benefits of both speed and scalability.[10] Some important algorithms used in model based recommender system are clustering, association rule, Decision Tree, Artificial Neural Network, Link Analysis, Regression, Bayesian Classifier[5]. In further part of the paper we have discussed more on clustering and classification algorithms for recommendations.

### C. Hybrid filtering

Hybrid filtering technique combines different recommendation techniques in order to gain better system optimization to avoid some limitations and problems of pure recommendation systems [11]. The idea behind hybrid techniques is that a combination of algorithms will provide more accurate and effective recommendations than a single algorithm as the disadvantages of one algorithm can be overcome by another algorithm [12].

## III. CLUSTERING

In many place recommendation system the main objective is to divide items in data set into groups that share some common characteristics. The idea can be to form such clusters for items depending on their characteristics and then recommending based on similarity in each group. The other way could be to recommend those items that have been positively valued by similar users. Various clustering algorithms can be used for recommending items depending on number of clusters that we want to make, size of the dataset, time taken to compute clusters, etc.

### A. Following Is Visualization Of Some Of The Clustering Algorithm When Considering Same Dataset

- 1) *K-means clustering* – In K-means algorithm, one has to provide specific input parameters such as number of clusters and datasets on which clustering has to be performed. The clusters are formed based on the closeness to the center value of the clusters. The initial center value is chosen randomly. K-means clustering is top-down approach.[13]

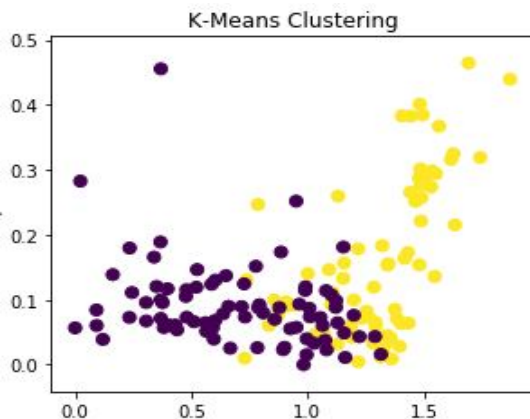


Fig - 2. Visualization of K-means algorithm

- 2) *Agglomerative Clustering* : It is also known as hierarchical clustering, which does not require the user to specify the number of clusters. Initially, each point is considered as a separate cluster, then it recursively clusters the points together depending upon the distance between them. The points are clustered in such a way that the distance between the points within a cluster is minimal and distance between the clusters is maximum[13]. Generally used distance measures are Cosine distance, Euclidean distance, Manhattan distance, etc. Unlike k-means clustering, it is bottom-up approach.

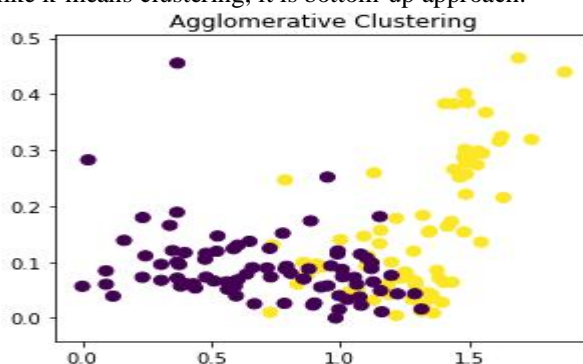


Fig - 3. Visualization of Agglomerative algorithm

- 3) *Affinity Propagation*: It does not require the number of cluster to be estimated and provided before starting the algorithm. It makes no assumption regarding the internal structure of the data points. [13]

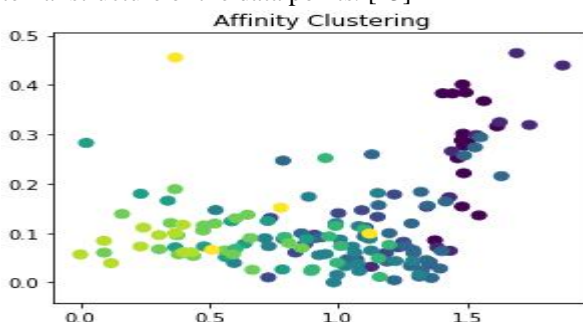


Fig - 4. Visualization of Affinity algorithm

Depending on proximate study of various above clustering algorithm we can determine the performance. So, before performing recommendation it is better to visualize our dataset for each clustering algorithm. After making clusters, the interests of other users in a cluster can be averaged and used to make recommendations for individual users. A good clustering method will produce high quality clusters in which the intra-cluster similarity is high, while the inter-cluster similarity is low[5].

#### IV. RECOMMENDATION USING CLASSIFICATION

Classification is one of the Data Mining techniques that is mainly used to analyze a given dataset and takes each item of it and assigns this item to a particular class[15]. Classification has two phases. The first phase is learning phase in which the classification algorithm is applied on training data set in order to make the model learn and second phase is testing phase where model is tested against a test dataset to determine accuracy of model.

##### A. Decision tree

Decision Trees are a type of Supervised Machine Learning where the data is continuously split according to a certain criteria. The Decision tree consists of decision nodes and leaves. The leaves are the decisions or the final outcomes and the decision nodes are where the data is split.[18]

- 1) *ID3 Algorithm* - On every cycle of the algorithm it emphasizes through every unused attribute of the set and finds the entropy of that attribute. At that point chooses the attribute which has the smallest entropy (or biggest data gain) value. [15]
- 2) *C4.5 Algorithm* - C4.5 is an algorithm that generates decision tree and it uses information for it. It uses a single-pass pruning process to mitigate over-fitting. Pruning results in many improvements. C4.5 can work with both continuous and discrete data by specifying ranges or thresholds for continuous data thus turning continuous data into discrete data.

##### B. M5P Algorithm

This algorithm is based on decision tree for numeric data prediction and at each leaf it stores a linear regression model that predicts the class value of items that reach the leaf and it combines a predictable decision tree with the possibility of linear regression functions at the nodes [16].

##### C. Artificial Neural network

Artificial neural networks are stimulated by biological neural networks that correspond to brain image for information processing. Similar to human brains, neural networks are also consisting of processing units (artificial neurons) and connections (weights) between them. The processing units convey received information on their outgoing connections to other units. The most important feature of these networks is their adaptive nature where "learning by example" replaces "programming" in solving problems [5].

##### D. K-Nearest Neighbor Classification

In pattern recognition, the KNN algorithm is a method for classifying objects based on closest training examples in the feature space.[17] Predictions are made for a new item say  $x$ , by searching through the entire training set for the  $K$  most similar items and summarizing the output variable for those  $K$  instances. For regression this might be the mean output variable, in classification this might be the mode class value and to determine which of the  $K$  items in the training dataset are most similar to a new input a distance measure is used[19].

By considering list of features and criteria on which it has to be split we can design classification algorithm for our recommendation system. One of the use case can be to take users preference for any item as our testing instance and previously stored characteristics for that item as training data and then applying classification algorithm on them to provide user-personalized recommendation.

#### V. IMPROVING RECOMMENDATION USING KNN ALGORITHM

Whenever we apply recommendation filter for finding say places like food place or hotels then before applying above model based recommendation technique it is better to apply KNN algorithm first. For example consider that we want to recommend places based on their rating to user then we can apply following steps –

- A. Start
- B. Find nearby places using KNN algorithm.
- C. After that make clusters of above remaining restaurants using any clustering algorithm.
- D. Input parameter for clustering would be latitude and longitude of places in considered case.
- E. Sort clusters based on their distance from the user.
- F. For each cluster Sort that cluster based on rating End for
- G. Stop

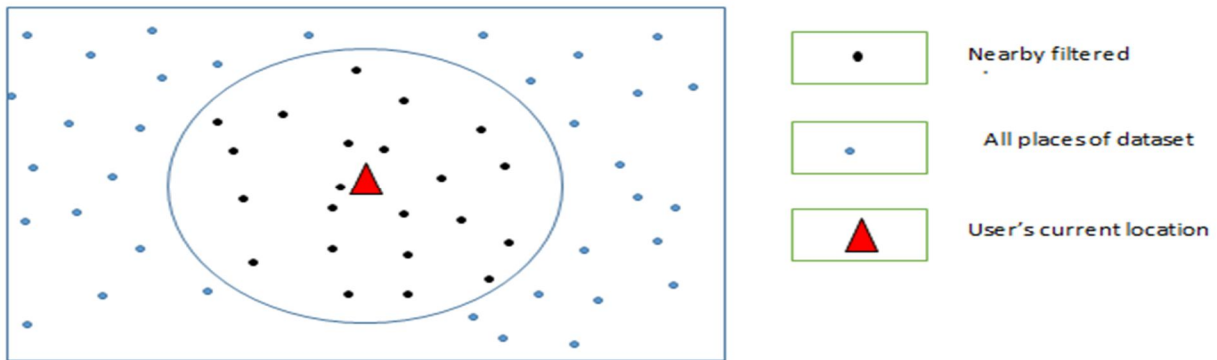


Fig - 5. Applying KNN algorithm on data set

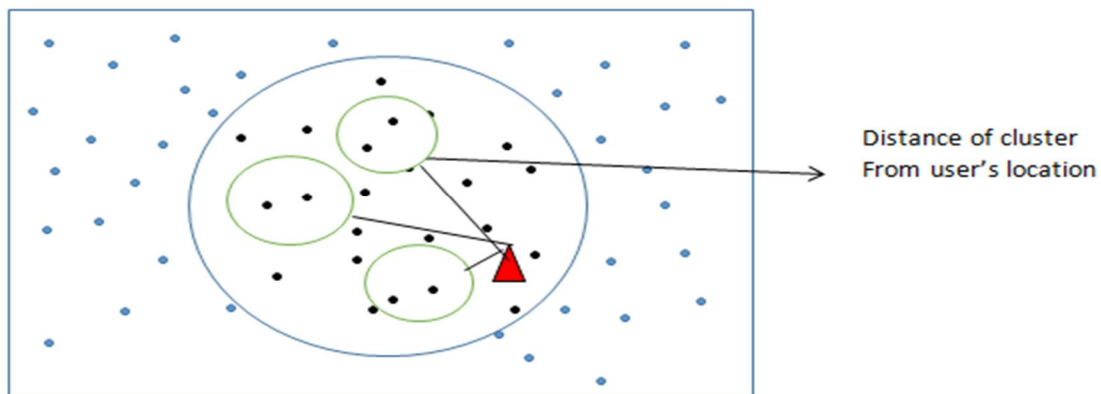


Fig - 6. Calculating distance of clusters from user's location

It can be stated that applying multiple clustering and classification algorithms for a specific system provide better output results for the user. The recommender system developed by it helps to gain better performance with fewer drawbacks.

## VI.CONCLUSION

Increasing need of recommendation engine in every e-commerce business demands a detailed study of various recommendation algorithms. Taking that into account, this paper is intended for those users who want to generate customized recommender system using the combination of various algorithms and data mining technique. In order to fulfil that purpose it is necessary to finalize the algorithm to be used for the system and hence the initial part of the paper introduces such techniques and algorithms. It also elaborates some of the well –known model based recommendation methods like clustering and classification with comparative overview of various algorithms for such methods.

## REFERENCES

- [1] Joseph A. Konstan · John Riedl, "Recommender systems: from algorithms to user experience, User Model User-Adapt Interact, Vol-22 , pp. 101-123, March 201
- [2] C. Pan, W. Li, "Research paper recommendation with topic analysis" ,Computer Design and Applications IEEE, Vol-4,2010
- [3] Pu P, Chen L, Hu R, "A user-centric evaluation framework for recommender systems", Proceedings of the fifth ACM conference on Recommender Systems(RecSys'11), ACM, Vol-612, pp. 57-164, Sep. 2010.
- [4] Hu R, Pu P., "Potential acceptance issues of personality-ASED recommender systems", Proceedings of the fifth ACM conference on Recommender Systems(RecSys'09), New York City, NY, USA, pp. 22-50, October 2009
- [5] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, "Recommendation systems: Principles, methods and Evaluation", Vol-16, pp.261-273, June 2015
- [6] L. Candillier, F. Meyer, F. Fessant, and K. Jack, "State-of-the-art recommender systems", April 2009
- [7] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," IEEE Internet Computing, Vol-7, pp. 76–80, January 2003.



- [8] Daniar Asanov , Berlin Institute of Technology, Germany, “Algorithms and Methods in Recommender Systems”, 2011
- [9] Ignacio Fernández-Tobías, Pedro G. Campos, Iván Cantador, Fernando Díez, “A Contextual Modeling Approach for Model-based Recommender Systems” , Superior Universidad Autónoma de Madrid 28049 Madrid, Spain
- [10] A Computer Science Comprehensive Exercise, Carleton College, Northfield, MN, “Model based recommender system
- [11] G. Adomavicius, J. Zhang, “Impact of data characteristics on recommender systems performance ACM Trans Manage Inform System1”, Vol-3 , Issue-1, 2012
- [12] J.B. Schafer, D. Frankowski, J. Herlocker, S.Sen, “Collaborative filtering recommender systems”, The Adaptive Web, LNCS, Vol- 4321, Springer, Berlin Heidelberg, Germany, pp. 291-324, 2007
- [13] <https://www.kaggle.com/dhanvajothimani/basic-visualization-and-clustering-in-python>
- [14] G. Linden, B. Smith, J. York, “Amazon.com recommendation: item-to-item collaborative filtering”, IEEE Internet Computer, Vol-7, Issues-1, pp. 76-80, 200
- [15] Prasad A. Jagdale, Prasad A. Jagdale, Dr. K. Rajeswari, “A Comparative Study of Classification Techniques in Data Mining Algorithms”, International Journal on Recent and Innovation Trends in Computing and Communication, Vol-6, Issue-1, pp. 142-14
- [16] Ritu Sharma, Mr. Shiv Kumar, Mr. Rohit Maheshwari,” Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets”, International Journal of Computer Science and Mobile Computing, Vol-4, Issue-12, pp. 125-134, Dec. 2015
- [17] S B Imandoustet, Mohammad Bolandraftar, Int. Journal of Engineering Research and Applications [www.ijera.com](http://www.ijera.com) Vol. 3, Issue 5,pp- 605-610 Sep-Oct 201
- [18] Mayur Kulkarni, “Decision Trees for Classification: A Machine Learning Algorithm”, Sep. 201
- [19] Brwonlee, “K-Nearest Neighbors for Machine Learning”, April 2016.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)