RECOMMENDATION SYSTEM USING COLLABORATIVE APPROACH

Kamna Singh
Assistant Professor, Ajay Kumar Garg Engineering College, Ghaziabad, UP, India
singhkamna@akgec.ac.in

***Abstract*— Recommendation systems are now popular both commercially and search community, where many approaches have been suggested for providing recommendations. In many cases a system designer that wishes to employ a recommendation system must select between a set of candidate approaches. A first step towards selecting an appropriate algorithm is to decide which properties of the application to focus upon when making this selection. Indeed, recommendation systems have a variety of properties that may affect user experience, such as accuracy, robustness, scalability, and so forth. This paper proposes review three types of experiments, starting with an offline setting, where recommendation approaches are compared without user interaction, then reviewing user studies, where a small group of subjects experiment with the system and report on the experience, and finally describe large scale online experiments, where real user populations interact with the system.**

***Keywords*— K-Mean Algorithm, Algorithm for clothes style specification**

I. INTRODUCTION

Recommender systems nowadays are being deployed in most of the domains where a huge set of items are involved. Such systems typically provision the users with a list of recommended items that might be preferred by them as per their previous likes and dislikes or predict how much they might prefer each item. These systems help users to decide on various choices for desired items, and finding preferred items and selecting in the collection.

There are many algorithms developed in the recent past but none of them are found to be really convincing of producing the desired result with usercustomization choices. This paper used clustering to overcome the drawbacks of past algorithms. As a resultant some unique features for recommending items in a better manner rather follow the previous trends. Some extra feature is also included such as height and weight of users to make system more accurate. It involves customization of recommended products according to the user's specific attributes making it more customers oriented.

This paper proposes an application-based Recommendation System which uses Genetic Algorithm K-mean Clustering Algorithm. It is a subclass of information filtering system that seeks to predict the 'rating' or 'preference' that a user would give to an item, but the problem is here, "Is the rating given is relevant enough to suggest other**?**" The answer is always unpredictable. So, we have come up with the new and fruitful techniques which serve the purpose better.

The main feature of the application-based recommendation system is to produce comparatively better result than previously implemented algorithms and to improve the over-all experience of customer for online shopping of clothes. One more feature can be included that is to help online shopping sites to enhance their recommendation list according to specific user attributes

II. LITERATURE REVIEW

A recommender system using GA K-means clustering in an online shopping market by Kyoung-jae Kim and HyunchulAhn.[2] In this study, they propose a novel clustering algorithm based on genetic algorithms (GAs) to effectively segment the online shopping market. In general, GAs is believed to be effective on NP-complete global optimization problems, and they can provide good near-optimal solutions in reasonable time. Thus, we believe that a clustering technique with GA can provide a way of finding the relevant clusters more effectively. The research in this paper applied K-means clustering whose initial seeds are optimized by GA, which is called GA K-means, to a real-world online shopping market segmentation case. In this study, we compared the results of GA K-means to those of a simple K-means algorithm and self-organizing maps (SOM). The results showed that GA K-means clustering may improve segmentation performance in comparison to other typical clustering algorithms. In addition, our study validated the usefulness of the proposed model as a pre-processing tool for recommendation systems.

This study suggests a new clustering algorithm, GA K-means. We applied it to a real-world case for market segmentation in electronic commerce, and found that GA K-means might result in better segmentation than other traditional clustering algorithms including simple K-means and SOM from the perspective of intra-class inertia. In addition, we empirically examined the usefulness of GA K-means as a pre-processing tool for recommendation model.

However, this study has some limitations. Although we suggest intra-class inertia as a criterion for performance comparison, it is uncertain that this is a complete measure for performance comparison of the clustering algorithms. Consequently, the efforts to develop effective measures to compare clustering algorithms should be done in the future research. The average and standard deviation of SOM and GA K-means Clustering is (3.76,1.342) and (4.51,1.010) respectively.

In this study, we propose a novel clustering algorithm based on genetic algorithms (GAs) to effectively segment the online shopping market. In general, GAs is believed to be effective on NP-complete global optimization problems, and they can provide good near-optimal solutions in reasonable time. Thus, we believe that a clustering technique with GA can provide a way of finding the relevant clusters more effectively. The research in this paper applied K-means clustering whose initial seeds are optimized by GA, which is called GA K-means, to a real-world online shopping market segmentation case. In this study, we compared the results of GA K-means to those of a simple K-means algorithm and self-organizing maps (SOM). The results showed that GA K-means clustering may improve segmentation performance in comparison to other typical clustering algorithms. In addition, our study validated the usefulness of the proposed model as pre-processing tool for recommendation systems.

2.1 Clustering Algorithms

It groups a set of data in d-dimensional feature space to maximize the similarity within the clusters and minimize the similarity between two different clusters. There are various clustering methods and they are currently widely used. Among them, we apply two popular methods, K-means and SOM, [1.] and a novel hybrid method to market segmentation. Before providing a brief description of each method, the following assumptions are necessary. A given population consists of n elements described by m attributes and it is partitioned into K clusters. Xi = (xi1, xi2, . . . ,xim) represents the vector of the m attributes of element i.

2.1.1 K-means Clustering Algorithm

The K-means method is a widely used clustering procedure that searches for a nearly optimal partition with a fixed number of clusters. It uses an iterative hill-climbing algorithm. The process of K-means clustering is as follows:

(1) The initial seeds with the chosen number of clusters, K, are selected and an initial partition is built by using the seeds as the centroids of the initial clusters.
(2) Each record is assigned to the centroid that is nearest, thus forming a cluster.
(3) Keeping the same number of clusters, the new centroid of each cluster is calculated.
(4) Iterate Step (2) and (3) until the clusters stop changing or stop conditions are satisfied.

The K-means algorithm has been popular because of its easiness and simplicity for application. However, it also has some drawbacks. First, it does not deal well with overlapping clusters and the clusters can be pulled out of center by outliers.[2]

2.1.1 K-means Clustering Algorithm

The K-means method is a widely used clustering procedure that searches for a nearly optimal partition with a fixed number of clusters. It uses an iterative hill-climbing algorithm. The process of K-means clustering is as follows:

(1) The initial seeds with the chosen number of clusters, K, are selected and an initial partition is built by using the seeds as the centroids of the initial clusters.
(2) Each record is assigned to the centroid that is nearest, thus forming a cluster.
(3) Keeping the same number of clusters, the new centroid of each cluster is calculated.
(4) Iterate Step (2) and (3) until the clusters stop changing or stop conditions are satisfied.

The K-means algorithm has been popular because of its easiness and simplicity for application. However, it also has some drawbacks. First, it does not deal well with overlapping clusters and the clusters can be pulled out of centre by outliers.[2]

2.1.3 GA K-means Clustering Algorithm

Genetic algorithms are stochastic search techniques that can search large and complicated spaces. [2] It is based on biology including natural genetics and evolutionary principle. In particular, GAs are suitable for parameter optimization problems with an objective function subject to various hard and soft constraints. The GA basically explores a complex space in an adaptive way, guided by the biological evolution of selection, crossover, and mutation. This algorithm uses natural selection survival of the fittest – to solve optimization problems.

3.1 Algorithm for specification matching

1. Input parameters:size, height, weight, colour, location, price, neck style, pattern.
2. According to these parameters find the exact matching in the database as per the specifications, i.e. for height and weight –it is used to restrict the different size that is to be displayed.
3. Colour, neck style, pattern have predefined values which can be used in clustering algorithm.

Color	Color Value
Blue	.05
White	.1
Grey	.15
Black	.2
Green	.25
Orange	.3
Red	.35
Yellow	.4
Brown	.45
Maroon	.5
Pink	.55

Figure 2: colour value

Neck Style	Neck Style Value
Polo	.01
Round	.02
V-Neck	.03
Collar	.04

Figure 3: neck value

Pattern	Pattern Value
Strip	.03
Printed	.06
Plain	.09

Fig4. Pattern value

3.2 Algorithm for function specification

1. Compute the function value=
(color value + neck style value + pattern value) / 3
2.This function value computesthe set of related cluster which can be displayed to the user asrecommendations.
3. These recommended product function value is increased by a factor of 0.01 for each successfulrecommendation.)
4.Repeat the above steps for eachuser.

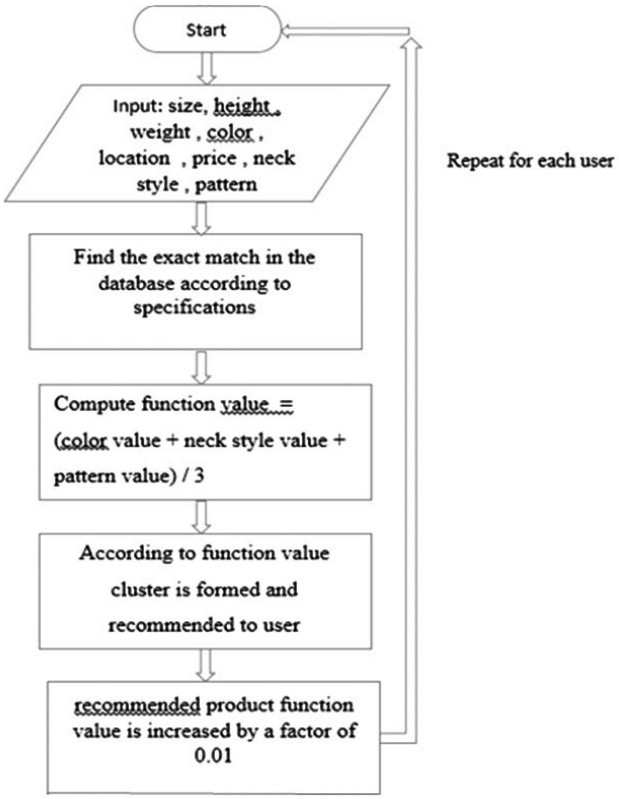

Figure 5: Flow Chart for recommendation System

In this study, we have used realworld data from online shopping sites for assessing the performance of our proposed model. The research data was collected from an online shopping site in India which contains all kinds of clothes for online shopping such as T- SHIRTS, SHIRTS, JEANS etc. In the case of a shopping site, customers generally have a clear objective, and they have a strong demand for buying clothes according to their desire and needs. Thus, they are usually open-minded about providing their personal informationto get appropriate service. As a result, the target company of our research possesses detailed and accurate customer information that may be used as a strong attribute to be used as a source of for recommending them products. Consequently, we have tried to build a recommendation model for the users of this website.the proposed system is able to implement the recommendation system using the function value as an average of attributes and cluster the products around function value if user's attributes function value lie in which cluster that clusters product are recommended to user. Results have found to be very impressing and accurate in recommending products to users.

Table 2: Input pattern for algorithm

Product_no	Function_value
67	0.09000000000000000000
68	0.06000000000000000000

Table 3. Output pattern for algorithm

Product_id	Function_val
24	.12
56	.12
103	.123
66	.13
20	.133
25	.14
46	.14
104	.143
2	.146
1	.15
82	.152
13	.153
28	.153
33	.16
40	.16
63	.16
64	.16
73	.16
74	.16
16	.17
19	.176
54	.19
57	.19
65	.19
59	.20

5.1 Performance Evaluation

Let say we have enter specification:

COLOR: Blue (Co=0.05)

NECK STYLE: Round (N=0.02)

PATTERN: Printed (P=0.06)

Function Value= (Co+N+P)/3=(0.05+0.02+0.06)/3=0.13/3=0.0433

CLUSTERING RESULT:

Max. Value of center1: 0.11000000000000000000

Min. Value of center1: 0.03000000000000000000

Max. Value of center2: 0.31000000000000000000

Min. Value of center2: 0.12000000000000000000

Cluster1 Center(C1)	Cluster2 Center(C2)	Interclass Separability(C1~C2)	Intraclass Separability		Cluster Separability(Max of C1~Min of C2)
			C1(Max-Min)	C2(Max-Min)	
0.06672027972028	0.16776923076923	0.10104895104895	0.08000	0.19000	0.010000000000

Comparison with existing State-of-the-Art Technologies

IV: CONCLUSION AND FUTURE SCOPE

This paper include two extra parameters height and weight through which we calculate the size of T-SHIRT more accurately. Every brand has its own size parameter, i.e. each has different height and width of T-SHIRT at same specification (small, large, medium, etc.). By using height and weight of user we recommend the T-SHIRT more appropriately to users. This makes our recommendation System more concise from other recommendation system.

This paper provide a recommendation system only for T-SHIRT and small database. In future, trying to incorporate more variety of clothes with large dataset and also trying to incorporate Genetic Algorithm with K-mean algorithm so that make use of machine learning to produce better result with large knowledgebase.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Trans. on Knowl. and Data Eng., vol. 17, pp. 734–749, June 2005. [Online]. Available :http://dx.doi.org/10.1109/TKDE.2005.99

[2] Kyoung-jae Kim, HyunchulAhn "A recommender system using GA K-means clustering in an online shopping market" –Science Direct Expert Systems with Applications 34 (2008)1200–1209

[3] Linden, G., Smith, B., and York, J. (2003). Amazon. com recommendations: Item- to-item collaborative filtering. Internet Computing, IEEE, 7(1),76-80.

[4] Ricci, F., Rokach, L., Shapira, B., and Kantor, P.B. (2011). Introduction to Recommender Systems Handbook.Springer.

[5] Yu Liu ,JingwenNie, Lexi Xu, Yue Chen , Bingyu Xu, "Clothing Recommendation System Based on Advanced User-Based Collaborative Filtering Algorithm" Springer Nature Singapore Pte Ltd. 2018 S. Sun et al. (eds.), Signal and Information

6] AtharvPandit , Kunal Goel , Manav Jain , Neha Katre,"A Review on Clothes Matching and Recommendation Systems based on user Attributes", Volume 09, Issue 08 (August 2020) IJERT 2020

[7] Qingqing Tu, Le DongAn Intelligent Personalized Fashion Recommendation System, 2010 International Conference on Communications, Circuits and Systems (ICCCAS)

[8] Pradeep Kumar Singh, PijushKanti Dutta Pramanik, Avick Kumar Dey, Prasenjit Choudhury,"Recommender Systems: An Overview, Research Trends, and Future Directions", January 2021,International Journal of Business and Systems Research 15(1):14–52

[9] JoeranBeel BelaGipp Stefan Langer Corinna Breitinger, "Research-paper recommender systems: a literature survey, Erschienen in: International Journal on Digital Libraries ; 17 (2016), 4. - S. 305-338

ABOUT THE AUTHOR

Kamna Singh is an Assistant Professor in Ajay Kumar Garg Engineering college Ghaziabd. She is B.Tech. from UCER, Allahabd and MTech from JIIT, Noida. Her research area includes Internet of Things, cyber security, information security, Computer Network and Data Mining. She has published many papers on information security in reputed Journals. She has published various research paper in Scopus Indexed International Journal.