A mobile phone recommendation system with user centric voting approach

Mr. Y. Subba Reddy and Prof. P. Govindarajulu

Department of Computer Science
S.V. University, Tirupati, AP, INDIA

Abstract
Mobile phones have become vital in social life. The obvious market competition complemented by advanced functionality and feature is building consumers' mobile phone decision making difficult and challenging. In this study, a user preference /voting based methodology was applied to build a recommendation system for mobile phones selection. A model with user preferences and actual weights of different variants of mobile phones was exercised to group the mobile users according to their interests and to cluster the mobile types based on the weights obtained from the actual attributes linked with user voting. The initial selection of cluster centers is a potential limitation of the popular k-means algorithm. In this paper a variant of clustering process is proposed with better initialization of cluster centers as representatives. The recommendation system was evolved by conducting a controlled experiment that involved 1000 mobile phone users. The information was gathered online through a well organized questionnaire. The analysis results indicate that the use of the proposed system results in higher satisfaction than equal-weight based benchmark systems.
Key words:
RCLUSTER, Theta Similarity, Voting, R-Tree, Top-K query

1. Introduction

Today mobile phones become a basic need of individual, as a communication devise across the globe. The advances in mobile phone technology and the competitive fight among the mobile phone manufacturers created the situation that almost every day a new model of a mobile phone is being introduced in to the market. The endless increase in the options space, presented a tricky challenge in front of the consumers of mobile phones. The major factors that influence consumers in selecting a mobile phone to use include: innovative features, image, price, personal recommendation, durability and portable aspects, influence of media, post-sales service and so on. Among the contemporary mobile communication technologies, the mobile phone became ‘‘the domestic appliance ever invented with the most radiation’’.

Though mobile phones have a number of features in common, manufacturers still try to bring uniqueness to their products by adding some more new features to the existing features. This made the mobile development a challenge and manufacturers welcoming the challenge with a great set of innovative designs. The growing number of brands and models created the fierce market competition. Therefore it is inevitable to run with innovations updates and at the same time it is mostly desirable to know the trending thoughts of potential customers.

This work parted with the introduction of preference based data grouping approach, creation of a datasets through an online questionnaire and online product information search, application of the proposed approach on the datasets, presentation of the findings and the justification of the importance of the proposed methodology through comparative discussions.

The paper is organized into six sections. In this section the existing problems in mobile phone selection along with the need to study the customer behavior is discussed. The second section is dedicated to review the efforts made in the literature to group the datasets based on user preferences in general and the grouping of mobile phone customers based on their preferences in particular. In the next section a preference based data grouping approach with better clustering process is introduced. The idea, methodology along with the proposed algorithm is presented here. In the following sections the way of data collection adopted for this study was presented down with the schema of the datasets collected. The last two sections are assigned for exhibiting the application of the proposed methodology and the discussion of the obtained outcomes.

2. Literature Review

The tremendous growth of the World Wide Web and the materialization of e-commerce have led to the development of systems that can assist the users and manufacturers in a business environment known as recommender systems. Collaborative filtering and content based techniques are used to identify the list of items that will meet the interest to a particular user. The combined growth of Internet and e-commerce, personalized recommendation techniques are getting the results and facing challenges also that result from information

Manuscript received February 5, 2018
Manuscript revised February 20, 2018

overload towards appropriately identifying products /services for customers [6]. The recommendation systems collect the information about user interests and attributes of items. Preferences and profiles are analyzed to advise the user to make right decisions to buy right products [28].Generally people receive recommendations from friends, family, sales people and internet resources. Recommendation systems become part and parcel of everyday life by which people receive recommendation. There are many factors that really influence humans' decisions while buying a product or getting a service. One may want to have a product with the best set of features. E.J Salazar and O. Ortega [12] said that the most of the studies conducted on the evaluation of recommender systems are completely based only on the evaluation of the quality of the recommendations and at the same time operation of the recommendation algorithm used for decision making. J.L. Herlocker et al [18] reviewed the most important decisions in evaluating collaborative filtering recommender systems and the important reviews are: the user tasks being evaluated, various types of analysis techniques and the various types of datasets being used, how to measure the prediction quality, how to evaluate prediction attributes other than quality, and how to perform user-based evaluation of recommendation system a whole.

L. Chen and H.K. Tsai [19] said that there is a need to study various types of decisions taken by different users in situations where the recommendation systems are influenced by the type of interface used in interacting with recommendation system.

M.E. Alva-Obeso [27] vigorously performed customer-centered evaluation of web systems by using different kinds of tools and questionnaires. B.P.Knijnenburg et al. [3] proposed a well planned framework for customer-centric evaluation of recommender systems and they have used four important steps in order to perform customer centric evaluation of recommender systems. A. Paramythis et al [1] said that formative evaluation is carried out using a set of basic principles involving users as early as possible in the design process and it is very useful in finding what and how to improve in an interactive system. D.N. Chin [9] said that recommendation system design, development and usage are dependent not only on the user experience completely but also dependent on characteristics of the user and the context in which the user is using the recommendation system. J.Nielson [20] said that in order to design the best interface and to find problems in the usage of interface it is required some details such as work experience of user, age, educational qualification, computer knowledge and experience, and so on. J. Nielsen and R.Molich [21] said that heuristic evaluation is one way of analysis of usability and in this evaluation; interface design is submitted to the many evaluators and asked them to evaluate with necessary comments. Here, heuristic evaluation is performed to find good and bad features of an interface. J. Nielsen [22] developed ten effective heuristic rules for usability of an interface design and all these rules and guidelines are listed in a special document of guidelines and this document is named as Usability Aspect Report (UAR), which contains almost all good and bad aspects of user interface design. J. Rieman et al. [25] said that partial functional prototype is needed with partial involvement of users to rectify problems early in the design and development of user interface based recommendation system. This procedure helps to find some of the problems that will arise in the interactions of the recommendation system. N.E.Jacobsen and B.E. John [31] described the cognitive walkthrough consists of two phases: preparation and execution. In the preparation phase the analyst specifies a list of tasks that must be executed from the user and the normal tasks include experience, knowledge, skill, qualification, training and so on. In the execution phase the system analyst asks a series of questions by examining closely the each action of the user.

J.Nielson [23] said that the analyst must first pay attention to know what users do instead of knowing what they say in order to design an effective and easy-to-use interface. F.J. Martin [13] conducted a study on user interface design for effective implementation of recommendation systems. After thorough investigation of conducted study the author said that the most important component in the design and development of recommendation system is the effective user interface. Author said that this study reveals that the user interface represents 50% of the user experience. Srivasthava. J. et al. [34] said that user interest and behavior mining based on the details of web-usage has been using from the long period of time. White R.W. et al. [36] have considered five different contextual information to model based on the user interestingness, and then do recommendation based on it. Nasraoui, O, et al. [30] studied behavior details of a particular website based on tracking user profiles. Xu, J. and Liu, H [37] said that many of the researchers are using selected clustering methods to extract different types of users. Authors said that clustering can be done with respect to users' perspective and cluster the potential users into different types. Clustering can also be done with respect to websites perspective and create URL groups. J. Lang [17] introduced a new technique called combinatorial vote in such a way that a group of agents or voters express their preferences and finally they come to a common decision with respect to a set of non-independent variables to assign.

R. Mukergee, et al. [33] have developed an online movie recommendation system by using rules and principles of voting theory. The developed system allows different types of queries such as instance based, constrained, and unconstrained queries to be executed.

C. Webber, et al. [7] developed a mechanism for modeling, analyzing, and diagnosing conceptions of different types of learners by using a theoretical model of conceptions. Authors applied various techniques from the principles of voting theory for the purpose of group-decision-making. G. Xue, et al. [14] proposed a collaborative system based on traditional simple k-means clustering algorithm for the purpose of smoothening the unrated data details for individual users with respect to the clusters.

Nowadays the usage of mobile systems in various applications of daily lives of human beings is continuously increasing with rapid speed. Different types of mobile applications are education, banking, hospitals, research, science, engineering, entertainment and controlling of intelligent systems in business. Julia Y. Arana-Llanes et al.[26] said that recommendation systems are mainly designed to facilitate decision making of the users before taking an action. Authors have developed context aware recommender systems with many different features such as heterogeneous object recommendation, explanations of the recommendations, usage of interfaces, and design of multisensory mobile devices. Authors said that there are various stages during the evaluation of a computer system running on mobile systems and these stages are obtained from different types of formative or summative data. Authors found that the purpose of formative evaluation is to find drawbacks or errors in a recommendation system in order to further improvement in the system and to guide the system design and development. Authors also said that summative evaluation purpose is to determine the value or impact of a system. C. Cuadrat Seix [5] said that only few mobile usability studies have been applied till date. Different tools and techniques are available for obtaining data on the evaluation of a context-aware recommender system and each technique is mapped to undergo a specific test. Mobile phone usage is increasing rapidly. The growth percentage of the number of mobile phone users is very high. Ting-peng Liang et al. [35] said that it is very difficult for the customer to select and purchase a mobile because in the market there exists variety of brands and models with different functionalities and they suggested that it is better to develop an intelligent decision making web based system that suggests better alternatives based on the needs of the customer. Authors also said that many factors, features, and criteria must be considered for better decision making with respect to functionality, features, price, screen size, memory capacity, Internet facility, camera features, video conferencing, voice cal, and so on. Recommendation systems for mobile phone purchasing are software systems that can find a limited set of choices from a large set of alternatives. [4] said that mobile phone users are frequently changing their mobile phones and a better decision making system for mobile sales is very useful for mobile sales persons and vendors in order to achieve business benefits. Authors proposed probabilistic based model called phone interest-model based on only mobile web-log data.

Yuan, S.T. and Tsao Y.W [39] presented a personalized and contextualized mobile advertising infrastructure for the purpose of recommending the advertisement. Yang F. and Wang, Z. [38] built a scalable personalized mobile information purchasing platform that can be used for recommending the location based services to the customers. Do, T.M.T. and Gatica-perez, D. [10] have mined user patterns using mobile phone application usage including mobile web usage on mobile phone. Zheng V.W. et al. [40] mined useful knowledge by taking GPS trajectories of many users by considering partial location and activity annotations to provide targeted collaborative location and activity recommendations for each user. Huang, K. et al. [16] used a variety of contextual information such as last used application, time, location, duration, facility, and profiles of users in order to predict whether the user’s application will be open or not.

Pinyapang, S, and Kato, T [31] proposed the relationship between three factors, namely place, purpose, and time. Also they summarized the basic rules to analyze essential data and algorithms for query processing. Harzov, T. et al. [15] executed many surveys of the mobile recommender systems with many illustrations and overviews of the most important techniques, supported functions, and specific computational models. Jyodeep Das, et al. [24] said that recommender systems are actually subclass of information filtering systems that are very useful to the customers in their decision making process by suggesting objects that the customers may prefer.

E. Ephrati and J Rosenschein [11] said that rules and principles of voting theory have been used in many domains successfully for many years in multi-agent systems with respect to group decision making that maximizes benefits of people. Voting theory is very much useful in recommendation systems that maximize the customer preferences. X. Jiang, et al. [37.] implemented successfully a cluster based collaborative filtering system using an iterative clustering method that uses the interrelationships between users and objects. M. O’ Connov and J. Herocker [28] experimented with variety of clustering algorithms in order to partition the item set on the basis of user voting data. Deng-Neng Chen, et al. [8] have presented the design, implementation, and evaluation of an intelligent based recommendation system that is very much useful for the users to select proper mobile phone models based on their individual customer voting or preferences. Developed intelligent web-based recommendation system was empirically tested with respect to its effectiveness and usefulness properties.

The mobile phone selection process relies on several features possessed by the manufacturers. The recent trends in advance technologies, the competition strategies in product development, identified as a key factor for the growing number of varieties and models. Cross comparisons, as a result, become a difficult task and the need of computer-aided decision systems to assist consumers in exploration for information on mobile products that can best meet the preferences of customers. In [2] authors proposed a classic personalized recommender system that can uncover information about the features of mobile phones and provides specialized services to potential customers. Triangular Fuzzy Numbers with Fuzzy Near Compactness is employed in their work by which user preferences and product attributes were technically expressed. Accordingly similarities were used to recommend optimal products that best satisfy the needs.

3. The scope and need to improve the existing processes

Though the literature has evidence for numerous recommendation system techniques, the efforts for mobile device recommendations appeared less. Therefore there exists a large gap and scope for recommendation systems for volatile product types like mobile phones. Most of the clustering based recommendation systems made use of traditional clustering techniques which have many restrictions/limitations.

In the literature various proposals are made to improve the effectiveness of the recommendation systems in general and to improve the mobile device recommender system in particular. However the basic problem of k-means clustering in terms of initial cluster centers information is tranquil persist. Furthermore the problems linked with primary memory, search time and space are quite challenging. With the aim to address the issues the present work proposed methodologies for improved object clustering with weighted preferences assisted by new means for fast searching.

4. Proposed methodology and Algorithms

4.1Prelimanaries

4.1.1 R-Tree

The proposed algorithms make use of R-Tree data structure for quick search processing. R-tree indexing is an extension of the B^+-tree indexing technique. B^+-tree supports indexing only for one dimensional data values whereas R-tree supports multidimensional indexing facilities for very large datasets particularly for spatial datasets. Each node of the R-tree is represented as a minimum bounding rectangle (MBR) and this rectangle is represented with lower left corner and upper right corner. Each node of the R-tree stores a set of pointers and each pointer points to a child node in that path. All the child nodes of a particular parent node are represented in the form of a set of rectangles based on ranges of values. Actual objects are stored only in the leaf nodes of the R-tree. Therefore, during searching all paths must be traversed either in depth first traversal order or breadth first traversal order depending on the application where R-tree is employed for effective implementation and management of the data. In the literature many variants of the R-tree are available for index implementation in various real world applications. Some of the R-tree variants are – R^+-tree, R^*-tree, bR-tree, KRR-tree and so on. In general, R-tree is efficient and effective multidimensional indexing tree but when scalability of the R-tree is considered still it is not up to the mark when very large datasets are selected for use. Hence, in such cases special techniques and efficient pruning techniques are required to make the desired application more scalable in the case of real time applications.

4.1.2 Queries

Top-k queries are frequently used in information retrieval systems. Top-k query returns a set of k objects in their ranking order from the specified dataset. Top-k query returns a set of objects with respect to the preferences of customer whereas reverse top-k query returns a set of customers based on the results obtained from the many top-k queries.

4.2 Algorithms

4.2.1 Minimum similarity:

This sub process is used to check the threshold value while comparing two objects.

Algorithm Minimum_Similarity (p, q)

Input

p: is the object (mobile) presents in the leaf node of the R-Tree

q: is the queried object (mobile)

Output

A numeric value representing the similarity measure between two objects a = total list of customers referenced the mobile object p

b = total list of customers referenced the mobile object q

similarity = $\frac{a \cap b}{a \cup b}$
return similarity

4.2.2 Reverse Top-k computation: This sub process picks up the list of customers who preferred a product object.

Algorithm Reverse_Top-k_Full ( )

Reverse_Topk [][] = create two dimensional array with rows as cell phone tuples and columns as

customers

```
for  i =1 to number of cell_phones do
{
   Col=0
  for j=0 to number of elements in each row in top-k result_set
   {
     for k=0 to number of elements in row
     {
       if  (topkresultset == i  ) then
         reverseTopk[i-1][col++]= j+1
     }
  }
}
```

4.2.3 Construct cluster centers: This sub process decides the initial set of cluster centers as representatives.

Algorithm Construct_Cluster_Centers (M, Limit)

Input

M: mobile versus customer matrix

Limit: minimum support value of customers

Output

Cluster representative set C of size r

```
C = null
for each row r of M do
  if(number of customers in the r >= limit)
    C = C U r
   end_if
end_for
return C
```

4.2.4 Algorithm reverseTopklist(obj)

Input

Obj: mobileObject

Output

List of customers

```
for i =1 to number of mobiles do
{
    if (mobileList[i][1] = Obj ) then
      return  ith row list in reverseTopk[i]
   End_if
}
End_for
```

4.2.5 Algorithm RCLUSTER (Threshold, Root, D)

Input

Threshold: user specified similarity limit

Root: indexed tree

D: the dataset

C: A subset of D representing cluster centers of size m.

Output

Set of clusters

```
For each object oi in C do
   Cluster set ci = Theta-Similarity-Query (Root, Threshold, oi )
```

End-For.

4.2.6 Algorithm Theta-Grouping (Root, theta, q)

INPUT

Root: root node of the R-tree

theta: is the similarity measure threshold value

q: is the query object

OUTPUT

result-set: is the set of similar objects

1. node = create a new tree node
2. node = Root
3. if (minimum-similarity(node ,q) ≥ theta) then
4. result-set = result-set UNION p for every sub-tree (node)
5. end-if
6. if (node.type = leaf-node) then
7. for every pi in the node do
8. reverse pi vector = execute reverse top-k (pi)
9. if (minimum-similarity(pi , q) ≥ theta) then
10. result-set = result-set UNION pi
11. en-dif
12. end-for
13. else
14. for every sub-tree of node do
15. if (maximum–similarity(sub-tree , q)) ≥ theta then
16. node = sub-tree(node)
17. end-if
18. end-for
19. end-if
20. if (node is not empty) then
21. Theta-Grouping (node, theta, q)
22. end-if
23. return (result-set)

4.3 Process Description

The RCLUSTER algorithm makes use of the similarity search algorithm described above. RCLUSTER provides the set of clusters by grouping the given dataset D. In each step of the iterative process a cluster is formed around a representative object as centre. The process ends when all the representative elements of the representative dataset C have been exhaustive.

The process, Theta-Grouping, returns all the similar objects of the given representative object oi. The object may represent any real world entity such as tuple, product, service, patient, medicine, profile, mobile, wine and so on. R-tree index structure used in this process facilitates fast searching. In each iteration a node is examined to test the value of the maximum-similarity which should be greater than or equal to the theta value. Then all the nodes within the sub-tree of the node or recursively searched and all the tuples of each node are processed based on the minimum similarity condition some tuples or objects are added to the result set. Whenever a leaf node is referenced Jaccard similarity measure is applied to all the objects of the leaf node by executing reverse top-k query for each object and at the sometimes similarity measure, similar (p, q) greater than or equal is also tested and the corresponding object is added to the result set during the computation of the similarity measure different types of pruning techniques are applied.

5. Data collection

Two types of datasets are the candidates for the proposed work. The first dataset is the set of tuples that constitutes the attributes like model name, price, battery life, memory size, and other common features of a mobile device. This data was collected by searching around e-business websites. The second dataset is the collection of tuples that represent the preferences/votes given by customers online/offline for each feature attribute of the first dataset. The required information for this data set is aided by a well organized questionnaire.

5.1 The schema of the mobile feature dataset

The feature set of the first dataset include the attributes: price level. screen size level, processor Speed level, screen quality level, RAM & ROM , camera ,resolution level, internet speed range, phone type, Weight range, Blue Tooth, and security level. For the purpose of unbiased survey for brand/company name it was given as a hypothetical name.

Table 1: Example data entry for product attributes

Example data entry

price level	screen size level	processor Speed level	screen quality level	RAM & ROM	camera resolution level	internet speed range	phone type	Weight range	Blue Tooth	security level.
1	1	3	1	1	2	3	2	5	1	2
1	1	1	1	1	1	1	1	1	1	1
2	4	3	5	3	3	2	4	1	3	5
3	1	5	4	3	1	2	5	3	4	1
3	3	2	4	2	4	1	4	3	4	3

.

The cell values of tuples were filled based on the information obtained for attributes. The values for innovative features were given as 0/1/2 and so on based on the existence of the feature in a particular mobile devise. The rating attributes are given values following a range from 1 to 5 with respect to the level of availability of features the device have. The other attribute values are filled based on the similar information. This table represents the estimated strength of the mobile variants listed and acts as the product table (first dataset) for the further computations to be followed. For the purpose of this study the data about120 mobile devises were recorded.

5.2 The schema of the preference dataset

Table 2: Example data entry for user preferences

price level	screen size level	processor Speed level	screen quality level	RAM & ROM	camera resolution level	internet speed range	phone type	Weight range	Blue Tooth	security level.
3	5	3	3	4	4	4	3	5	5	5
2	2	2	2	1	1	2	5	3	4	1
2	2	2	2	3	2	3	4	2	4	2
3	4	4	5	5	4	4	4	5	5	3
1	5	1	5	4	1	1	1	1	1	1

The attributes set of the second dataset include the voting/preferences/weights given by users or customers. The information furnished in the questionnaire is converted into preference dataset (Second dataset) and it looks like the following.

In the above table the cell values are ranged from 1 to 6.The values are according to the responses (weights/votes) given through the questionnaire. For the purpose of this study 745 responses were recorded.

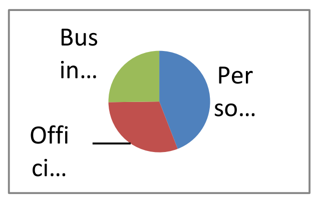

Fig. 1 Purpose of mobile phone

6. Descriptive data analysis

The product dataset having 120 records representing the features of mobile devises was used for the experiment. To incorporate the user votes into the clustering process a dataset consisting of 745 preference records was used.

57 percent of the respondents are in the age group "20-24".65 percent of the respondents are male.44 percent of the respondents are buying the mobile phones for business purpose. Major portion (30%) of the respondents prefers large screen sizes. The same type of responses is observed towards high quality features without considering the price limits. More than half of the respondents are keen about the security features of the mobile devises. Some of the facts are presented in the following figures.

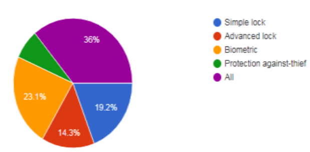

Fig. 2 Preference towards security

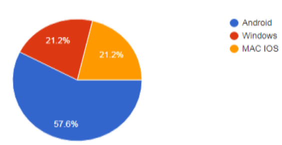

Fig. 3 Preferences towards OS

7. Results and Discussion

The execution is performed for various cases with varying parameters. The following is the description of case wise results. Table 3 shows the fixed things and Table 4 is the result of the varying cases.

Table 3: Fixed thing (Theta similarity =0.2)

Total tuples =	120
Maximum weights =	745
Theta similarity =	0.2
Maximum attributes =	11
vectorCommonThreshold	100

Table 4: Result set 1

Serial No	K in top-k	Execution time Seconds	Clusters	Clusters data	Representatives
1	20	8	12	[106, 108, 17, 24, 26, 3, 30, 32, 34, 43, 49, 53, 56, 6, 62, 68, 79, 81, 93, 96] [106, 108, 17, 26, 3, 30, 32, 33, 34, 43, 49, 50, 54, 6, 62, 66, 68, 81, 90] [13, 37, 62],[17, 26, 29, 30, 34, 41, 49, 50, 53, 54, 62, 79, 90, 93] [27, 41, 88],[17, 34],[17, 50, 90],[17, 33, 41, 57],[17, 34, 50, 52, 90],[27, 34, 41] [56],[108, 13, 37, 62],	3,6,13,17,24,26,27,29,30,32,33,34,41,43,49,50,52,53,54,56,62,66,68,79,81,88,89,90,93,96,106,108, repcount is = 32
2	40	10	20	[1, 105, 106, 108, 109, 111, 114, 13, 16, 17, 19, 20, 21, 24, 26, 27, 29, 3, 30, 32, 33, 34, 37, 4, 41, 43, 48, 49, 50, 52, 53, 54, 56, 57, 58, 6, 62, 64, 66, 68, 74, 75, 79, 81, 88, 89, 9, 90, 93, 94, 96],[48, 6, 78, 88],[46, 58],[114, 19, 20, 33, 41, 57],[114, 19, 34, 52, 58] [105],[23, 92],[105, 16, 19, 32, 43, 48, 52, 6, 68, 81, 88, 9],[25] [108, 13, 21, 30, 37, 54, 62, 64, 79, 94],[86],[32, 43, 50, 66, 68, 77, 81, 90] [43, 68, 9],[108, 13, 15, 30, 37, 39, 54, 62, 64, 79, 94],[21, 96],[6, 65],[78],[32, 77, 81] [3, 4, 78],[106, 108, 26, 43, 49, 68, 74, 89, 9, 94]	1,3,4,6,9,13,16,17,19,20,21,23,24,25,26,27,29,30,32,33,34,37,41,43,46,48,49,50,52,53,54,56,57,62,64,66,68,74,75,78,79,81,88,89,90,93,94,96,105,106,108,109,111,114, repcount is = 54
3	60	7	15	[1, 105, 106, 108, 109, 111, 114, 13, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 29, 3, 30, 32, 33, 34, 37, 38, 39, 4, 41, 43, 46, 48, 49, 50, 52, 53, 54, 56, 57, 58, 6, 62, 64, 65, 66, 68, 7, 72, 73, 74, 75, 77, 78, 79, 8, 81, 82, 86, 88, 89, 9, 90, 92, 93, 94, 95, 96],[2, 4, 5],[109, 114, 13, 16, 17, 20, 23, 26, 28, 3, 30, 34, 37, 39, 4, 40, 45, 49, 52, 54, 62, 64, 70, 74, 79, 89, 94, 95, 96],[111, 26, 3, 4, 49, 5, 74, 89, 94, 95, 96] [38, 6, 78, 83, 88, 92],[32, 43, 48, 68, 73, 77, 81, 88, 9],[42, 58, 7, 8, 97] [17, 28, 33, 46, 50, 57, 90],[2, 56],[10, 65, 71, 86],[109, 114, 16, 17, 23, 28, 32, 34, 43, 50, 52, 68, 72, 77, 81, 88, 9, 90],[109, 114, 16, 17, 18, 23, 27, 28, 41, 52] [20, 40],[42, 8],[16, 17, 28, 33, 46, 52, 57, 7] BUILD SUCCESSFUL (total time: 7 seconds)	1,2,3,4,6,7,8,9,13,15,16,17,18,19,20,21,23,24,25,26,27,28,29,30,32,33,34,37,38,39,40,41,43,45,46,48,49,50,52,53,54,56,57,58,62,64,65,66,68,70,72,73,74,75,77,78,79,81,82,86,88,89,90,92,93,94,95,96,105,106,108,109,111,114,119,120, repcount is = 76
4	80	7	10	[1, 10, 105, 106, 108, 109, 11, 111, 112, 114, 116, 117, 119, 120, 13, 15, 16, 17, 18, 19, 2, 20, 21, 23, 24, 25, 26, 27, 28, 29, 3, 30, 31, 32, 33, 34, 35, 37, 38, 39, 4, 40, 41, 42, 43, 44, 45, 46, 48, 49, 5, 50, 52, 53, 54, 55, 56, 57, 58, 6, 62, 63, 64, 65, 66, 68, 69, 7, 70, 72, 73, 74, 75, 77, 78, 79, 8, 80, 81, 82, 83, 84, 86, 88, 89, 9, 90, 91, 92, 93, 94, 95, 96, 97],[103, 106, 108, 117, 120, 35, 48, 60, 73, 78, 88],[43, 50, 65, 66, 68, 69, 71, 86, 88, 9, 90, 91],[107, 109, 114],[100, 106, 108, 117, 120] [11, 111, 113, 117, 120, 13, 16, 17, 19, 23, 27, 28, 30, 37, 41, 45, 51, 52, 54, 62, 70, 79, 85],[105, 20, 26, 3, 39, 4, 40, 44, 49, 5, 50, 64, 74, 89, 90, 94, 95, 96, 97] [107, 109, 114],[84, 87],[32, 43, 68, 72, 77, 81, 83, 9, 92]	1,2,3,4,5,6,7,8,9,10,11,13,15,16,17,18,19,20,21,23,24,25,26,27,28,29,30,31,32,33,34,35,37,38,39,40,41,42,43,44,45,46,48,49,50,52,53,54,55,56,57,58,60,62,63,64,65,66,68,69,70,71,72,73,74,75,77,78,79,80,81,82,83,84,86,88,89,90,91,92,93,94,95,96,97,100,103,105,106,107,108,109,111,112,114,116,117,119,120, repcount is = 99
5	100	6	6	[1, 10, 100, 102, 103, 104, 105, 106, 107, 108, 109, 11, 111, 112, 113, 114, 115, 116, 117, 119, 120, 13, 15, 16, 17, 18, 19, 2, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 3, 30, 31, 32, 33, 34, 35, 37, 38, 39, 4, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 5, 50, 51, 52, 53, 54, 55, 56, 57, 58, 6, 60, 62, 63, 64, 65, 66, 68, 69, 7, 70, 71, 72, 73, 74, 75, 77, 78, 79, 8, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 9, 90, 91, 92, 93, 94, 95, 96, 97] [100, 102, 103, 104, 106, 108, 120, 99],[110, 111, 112, 113, 115, 116, 117, 120, 14, 38, 48, 6, 73, 88],[101],[105, 107, 109, 114],[119, 67]]	1,2,3,4,5,6,7,8,9,10,11,13,15,16,17,18,19,20,21,23,24,25,26,27,28,29,30,31,32,33,34,35,37,38,39,40,41,42,43,44,45,46,48,49,50,52,53,54,55,56,57,58,60,62,63,64,65,66,68,69,70,71,72,73,74,75,77,78,79,80,81,82,83,84,86,88,89,90,91,92,93,94,95,96,97,100,103,105,106,107,108,109,111,112,114,116,117,119,120, repcount is = 99
6	120	2	1	[1, 10, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 11, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 12, 120, 13, 14, 15, 16, 17, 18, 19, 2, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 3, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 4, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 5, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 6, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 7, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 8, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 9, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99]	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120, repcount is = 120

From the above table it can be observed that the increase in value of k (for top k query) is able to initialize more representative points and less number of clusters. From this result it can be identified that the optimum selection of k results in optimum clusters.

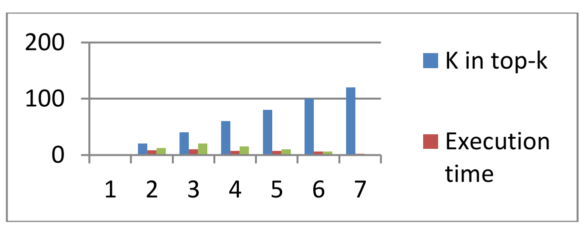

Fig. 4 Relation among k, number of clusters and execution time.

From the above figure it is clear that increase in k is resulting in decrease in number of clusters and execution time.

The following is the set of cases where k value is fixed and other parameters are varying.

Table 4 Result set 2

Serial No	Theta value	Execution time Seconds	No.of Clusters Obtained
1	0.1	7	12
2	0.2	8	15
3	0.4	10	21
4	0.6	12	26
5	0.8	13	44
6	1.0	21	51

From the above table it can be observed that the increase in value of theta (for similarity) is able to increase in the number of clusters and needs more execution time. From this result it can be identified that the optimum selection of theta results in optimum clusters.

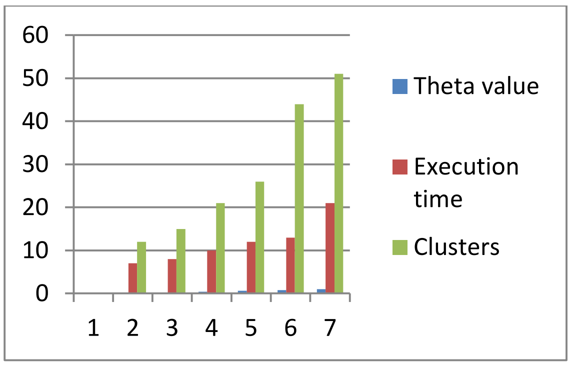

Fig. 5 Relation among theta, number of clusters and execution time

From the above figure it is clear that increase in theta is resulting in increase in number of clusters and execution time.

The Application of cluster information

The resulting clusters guide the user by suggesting the suitable recommendation of products which are similar to the preference set given by the user. As the clusters are formed based on the top set of preferences as weights the resulting cluster information is quite dissimilar from the traditional k-means clustering with normal weights. Therefore the modified information can suggest more suitable products (mobiles) to users based on their preferences.

Table 5: Fixed thing (K value =60)

Total tuples =	120
Maximum weights =	745
K value	60
Maximum attributes =	11
vectorCommonThreshold	100

Comparison of Proposed Algorithm with Traditional Methods

K-means is a popular approach for data grouping particularly with numerical attribute data .This has been in use in recommender systems for several years. The k-means approach treats each attribute with equal preference and does not consider weights with respect to priority attributes. Giving weights to attributes is an improved approach. Other limitations of k-means algorithm include the scalability and the burden of prior decision on number of clusters. The traditional approach needs high computational effort. The proposed approach using R-Tree saves significant amount of computation time. The proposed approach used an optimization criterion to decide the initial cluster centers. The traditional approach needs comparatively more iterations for clustering than the proposed R-tree based method. The time complexity of the proposed approach is O (n*log (n)), whereas the traditional methods like k-means algorithm need O (n2) of time.

Comparison and analysis

The following are the comparative sets of results wherein one set is the result of weighted clustering with random number of clusters (traditional approach) and the another set is the result of weighted clustering(new approach) with improved representative initial clustering centers based on top weighted list. In column 3 the cell values represent the values obtained for traditional approach where the values showed in parentheses are the values obtained with the proposed new approach. Columns 4, 5 respectively represent the values obtained for traditional approach and the proposed new approach respectively.

Table 5 Traditional Vs New clustering result sets

Serial No	K in top-k	Execution time Seconds Traditional(new)	Clusters Traditional approach	Clusters New approach
1	20	21 (8)	20	12
2	40	38 (10)	29	20
3	60	58 (7)	34	15
4	80	63 (7)	36	10
5	100	79 (6)	39	6
6	120	83(1)	42	1

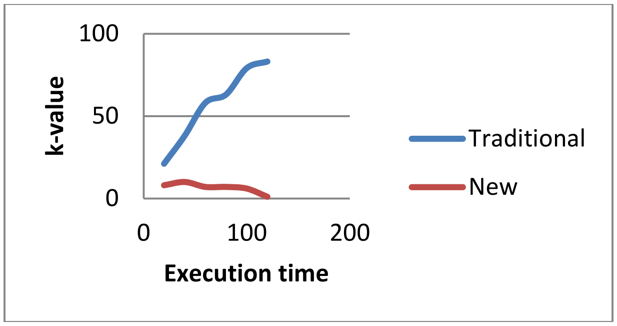

Fig. 6 Execution time

In the above figure the Y-axis represents the k-value of top k query. The X-axis scaled with the Execution time.

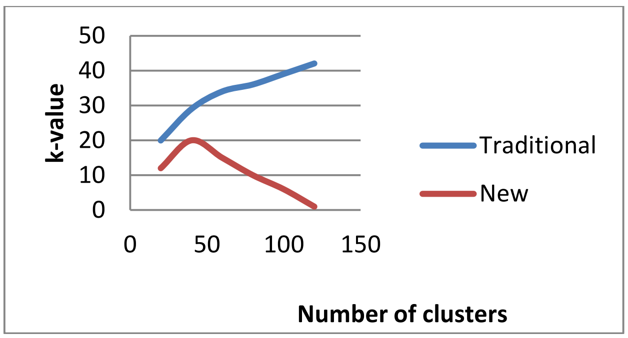

Fig. 7 Number of clusters

In the above figure the Y-axis represents the k-value of top k query. The X-axis scaled with the number of clusters.

From Table 5 and figures 6, 7 it can be observed that the new approach is better than the traditional approach in terms of clustering and execution time. With the increase in k value the execution time went up in traditional approach and came down with the new approach. The similar improvements are observed with respect to the number of clusters also.

Acknowledgement

We are very much thankful to all the participants who participated in the online survey by spending their valuable time.

8. Conclusion

The use of clustering approach for recommendation systems is a common practice for several years. Different clustering variants are tried in literature to improve the effectiveness of clustering. One of the major problems in traditional k-means and related algorithms of data partitioning is the requirement of choosing right clustering centers (k centers) at the process initialization. One more prevailing limitation of recommendation process is the incorporation of user preferences. This paper addressed these two problems by introducing a new way of embedding user preferences using top k query and followed a better way of deciding initial clustering centers. The ever burning problem of selecting right product (mobile) is undertaken to apply the proposed approach. The user interests towards mobile phone features were collected through an online survey. The survey revealed several interesting facts in terms of the awareness of the user about the features like security, advanced apps and so on. The proposed process can save time and able to provide optimal clustering. The results showed that the new approach of product recommendation based on the user voting is more effective than the existing approaches in terms of time and usability.

References

[1] A.Paramythis, S. Weibelzahl, and J. Masthoff, "Layered evaluation of interactive adaptive systems: framework and formative methods", Springer Science+Business Media B.V. 2010.

[2] A. Ojokoh, M. O. Omisore, O. W. Samuel and U. I. Eno, "Archetypal Personalized Recommender System for Mobile Phone Users", Research Methods: Concepts, Methodologies, Tools, and Applications Copyright: © 2015 ,Pages: 20.

[3] P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu and C. Newell, "Explaining the user experience of recommender systems", Journal of User Modeling and User-Adapted Interaction (UMUAI), vol. 22, 2011.

[4] Bozhi Yuan, Bin Xu, Tonglee Chung, Kaiyan Shuai, and Yongbin Liu, "Mobile Phone Recommendation Based on Phone Interest", B. Benatallah et al. (Eds.):WISE 2014, Part I, LNCS 8786, pp. 308–323, 2014. Springer International Publishing Switzerland 2014.

[5] Cuadrat Seix, "Ph.D. Thesis: Estudio sobre Evaluación de la Usabilidad Móvil y Propuesta de un Método para Tests de Usabilidad Cuantitativos basado en técnicas de eye-tracking", Universidad de Lleida, Espan&Tild;a, 2012.

[6] Chen, Daniel; Mocker, Martin; Preston, David S.; and Teubner, Alexander. 2010. "Information Systems Strategy: Reconceptualization, Measurement, and Implications," *MIS Quarterly*, (34: 2) pp.233-259).

[7] C. Webber, S. Pesty, and N. Balacheff, "A multi-agent and emergent approach to learner modelling," in Proceedings of the 15th European Conference on Artificial Intelligence (ECAI 2002), 2002, pp. 98–102.

[8] Deng-Neng Chen, Paul Jen-Hwa Hu, Ya-Ru Kuo, Ting-Peng Liang, "A Web-based personalized recommendation system for mobile phone selection Design, implementation, and evaluation", Expert Systems with Applications 37 (2010) 8201–8210.

[9] N. Chin, "Empirical Evaluation of User Models and User-Adapted Systems", User Modeling and User-Adapted Interaction, 2001.

[10] Do, T.M.T., Gatica-Perez, D.: By their apps you shall understand them: mining large-scale patterns of mobile phone usage. In: Proceedings of the 9th International

Conference on Mobile and Ubiquitous Multimedia, MUM 2010, pp. 27:1–27:10. ACM, New York (2010).

[11] Ephrati and J. Rosenschein, “Multi-agent planning as a dynamic search for social consensus,” in Proceedings of the 13th International Joint Conference on Artificial Intelligent, 1993, pp. 423–429.

[12] E.J. Salazar and O. Ortega, "Sistema de búsqueda personalizada y recomendación de documentación científica", Revista Iberoamericana de inteligencia artificial, vol. 30, 2010.

[13] J. Martin, "Top 10 lessons learned developing, deploying, and operating real-world recommender systems", ACM Recsys, 2009.

[14] Xue, C. Lin, Q. Yang, G. Xi, H. Zeng, Y. Yu, and Z. Chen, “Scalable collaborative filtering using cluster-based smoothing,” in Proceedings of the 28th annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2005, pp. 114–121.

[15] Horozov, T., Narasimhan, N., Vasudevan, V.: Using location for personalized poi recommendations in mobile environments. In: International Symposium on Applications and the Internet, SAINT 2006, p. 6. IEEE (2006).

[16] Huang, K., Zhang, C., Ma, X., Chen, G.: Predicting mobile application usage using contextual information. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 1059–1065. ACM (2012).

[17] Lang, “Logical preference representation and combinatorial vote,” Annals of Mathematics and Artificial Intelligence, vol. 42, pp. 37–71,2004.

[18] L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information Systems, vol. 22, 2004.

[19] Chen and H. K. Tsoi, "Users' Decision Behavior in Recommender Interfaces: IMPACT of Layout Design", RecSys'11, Workshop on Human Decision Making in Recommender Systems, 2011.

[20] J. Nielsen and R. Molich, "Heuristic Evaluation of User Interfaces", ACM CHI Proceedings, 1990.

[21] J. Nielsen, Usability Engineering, Academic Press, pp. I-XIV, 1-358, 1993.

[22] J. Nielsen, "10 Heuristics for User Interface Design", 1995.

[23] J. Nielsen, J, "First Rule of Usability? Don't Listen to Users", On line. Available: http://www.useit.com/alertbox/20010805.html.

[24] Joydeep Das, Partha Mukherjee, Subhashis Majumder, and Prosenjit Gupta, “Clustering-Based Recommender System Using Principles of Voting Theory”, 4http://en.wikipedia.org/wiki/Voting system.

[25] J. Rieman, M. Franzke and D. Redmiles, "Usability Evaluation with the Cognitive Walkthrough", ACM-CHI, 1995.

[26] Julia Y. Arana-Llanes, Juan C. Rendon-Miranda, Juan G. Gonzalez-Serna and Hugo O. Alejandres-Sanchez,”Design and User-Centered Evaluation of Recommender Systems for Mobile Devices - Methodology for User-Centered Evaluation of Context-Aware Recommender Systems” International Conference on Computational Science and Computational Intelligence (CSCI), 2014 IEEE.

[27] E. Alva-Obeso, Ph.D. Thesis: "Metodología de Medición y Evaluación de la Usabilidad en Sitios Web Educativos", Universidad de Oviedo, Spain, 2005.

[28] O’Connor and J. Herlocker, “Clustering items for collaborative filtering,” in Proceedings of the ACM SIGIR Workshop on RecommenderSystems, 1999.

[29] Nasraoui, O., Soliman,M., Saka, E., Badia, A., Germain, R.: A web usage mining framework formining evolving user profiles in dynamic web sites. IEEE Transactions on Knowledge and Data Engineering 20(2), 202–215 (2008).

[30] E. Jacobsen and B. E. John, "Two Case Studies in Using Cognitive Walkthrough for Interface Evaluation", CMU-CS-00-132, Computer Science Department School of Computer Science, Carnegie Mellon University, 2010.

[31] Pinyapong, S., Kato, T.: Query processing algorithms for time, place, purpose and personal profile sensitive mobile recommendation. In: 2004 International Conference on Cyberworlds, pp. 423–430. IEEE (2004).

[32] R. Mukherjee, P. Dutta, and S. Sen, “Movies2go-a new approach to online movie recommendation,” in Proceedings of the IJCAI Workshopon Intelligent Techniques for Web Personalization, 2001.

[33] Srivastava, J., Cooley, R., Deshpande, M., Tan, P.N.: Web usage mining: discovery and applications of usage patterns from web data. SIGKDD Explor. Newsl. 1(2), 12–23 (2000).

[34] Ting-Peng Liang, Paul J. H. Hu, Y. R. Kuo, D. N. Chen, “80. A Web-Based Recommendation System for Mobile Phone Selection”, 11th Pacific-Asia Conference on Information Systems.

[35] White, R.W., Bailey, P., Chen, L.: Predicting user interests from contextual information. In:Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, pp. 363–370. ACM, New York (2009).

[36] Xu, J., Liu, H.: Web user clustering analysis based on kmeans algorithm. In: 2010 International Conference on Information Networking and Automation (ICINA), vol. 2, pp. V2-6 –V2-9 (October 2010).

[37] X. Jiang, W. Song, and W. Feng, “Optimizing collaborative filtering by interpolating the individual and group behaviors,” in APWeb, 2006, pp.568–578.

[38] Yang, F., Wang, Z.: A mobile location-based information recommendation system based on gps and web 2.0 services. Database 7, 8 (2009).

[39] Yuan, S.T., Tsao, Y.W.: A recommendation mechanism for contextualized mobile advertising. Expert Systems with Applications 24(4), 399–414 (2003)

[40] Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative filtering meets mobile recommendation: A user-centered approach. In: AAAI, vol. 10, pp. 236–241 (2010).

Y.Subba Reddy received M.Sc (Computer Science) degree from Bharathidasan University, Tiruchirapalli, TN and M.E degree in Computer Science & Engineering from Sathyabama University, Chennai, TN. He is a research scholar in the Department of Computer Science, Sri Venkateswara University, Tirupati, AP, India. His research focus is on Data Mining in Clustering and Similarity measures.

P.Govindarajulu, Professor, Department of Computer Science, Sri Venkateswara University, Tirupathi, AP, India. He received his M. Tech., from IIT Madras (Chennai), Ph. D from IIT Bombay (Mumbai). His area of research are Databases, Data Mining, Image processing, Intelligent Systems and Software Engineering