ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Alleviating Cold-Start Problem in LARS* Using Hybrid Systems

Mili Mohan[1], Robert S[2]

M.Tech Student, Dept. of CSE, Marian Engineering College, Thiruvananthapuram, Kerala, India[1]

Assistant Professor, Dept. of CSE, Marian Engineering College, Thiruvananthapuram, Kerala, India[2]

ABSTRACT: Number of people who uses internet and websites for various purposes is increasing at an astonishing rate. More and more people rely on online sites for purchasing rented movies, songs, apparels, books etc. The competition between numbers of sites forced the web site owners to provide personalized services to their customers. So the recommender systems came into existence. LARS* is a location-aware recommender system that uses location based ratings to produce recommendations. LARS* supports a taxonomy of three novel classes of location-based ratings, namely, spatial ratings for non-spatial items, non-spatial ratings for spatial items, and spatial ratings for spatial items. The item based collaborative filtering used for generating recommendations in LARS* suffers from cold start problem. In cold start problem, the recommenders cannot draw inferences for users who are new to the system (new user problem) & for items which does not have sufficient ratings (new item problem). New user cold start problem can be resolved by utilizing the demographic data explicitly given by a user. Also the content based filtering does not suffer from new item cold start problem as it focuses on utility of item for user rather on its ratings. This paper proposes a hybrid recommender system which exploits the demographic details of user and content based filtered data of items for alleviating cold start problem and thereby improving the efficiency of recommendations generated by LARS*.

KEYWORDS: Location Aware Recommender System, Collaborative filtering, cold-start problem, demographic filtering, content based filtering, Hybrid Systems

I. INTRODUCTION

With the rapid advancements in the field of position localization techniques, people are allowed to share their locations and location related contents through different social networking sites. Location data bridges the gap between the physical and digital worlds and enables a deeper understanding of user preferences and behavior. Recently, advances in location-acquisition and wireless communication technologies have enabled the creation of location-based social networking services, such as Foursquare, MovieLens etc. In such a service, users can easily share their geospatial locations and location related content in the physical world via online platforms. For example, a user with a mobile phone can share comments with his social network about a restaurant at which he has dined on an online social site.

LARS*[1] is a location aware recommender system which is built to generate high quality location based recommendations. It is a single framework consisting of three types of location based ratings:

1. Spatial ratings for non-spatial items represented as a 4-tuple (user,ulocation,rating,item).
2. Non-spatial ratings for spatial items represented as a 4-tuple (user,rating,item,ilocation)
3. Spatial rating for spatial items which is represented as a 5-tuple (user,ulocation,rating,item,ilocation)

Here ulocation & ilocation represents the user location and item location respectively.

LARS*[1] produces recommendations using spatial ratings for non-spatial items, i.e., the tuple (user, ulocation, rating, item) by employing a user partitioning technique that exploits preference locality. This technique uses an adaptive pyramid structure to partition ratings by their user location attribute into spatial regions of varying sizes at different hierarchies. For a querying user located in a region R, it applies an existing collaborative filtering technique [3] that utilizes only the ratings located in R. It produces recommendations using non-spatial ratings for spatial items, i.e., the tuple (user, rating, item, ilocation) by using travel penalty, a technique that exploits travel locality. This technique penalizes recommendation candidates the further they are in travel distance to a querying user. To produce recommendations using spatial ratings for spatial items, i.e., the tuple (user, ulocation, rating, item,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

ilocation).LARS* employs both the user partitioning and travel penalty techniques to address the user and item locations associated with the ratings.

Even though LARS* is a novel framework for generating location based ratings for spatial items and user, it suffers from cold start problem as LARS* is using an item based collaborative filtering technique for generating recommendations. The cold start problem arises when a user or a content item does not have sufficient historical data known to the system (or none at all), which makes it impossible to recommend content for new users or to recommend new offers. Main purpose of this paper is to improve efficiency of LARS* by alleviating the cold start problem by using a hybrid system. Here hybridization of demographic filtering and content based filtering is used.

II RELATED WORKS

Recommender systems are nowadays used in a large variety of application setting, ranging from online stores, music and movie recommendation, to social media recommender and many more. Each of these applications has its particular characteristics, with greatly differing temporal dynamics or volatility, amounts of available data, use of explicit or implicit indicators, etc. Recommendation algorithms are best known for their use on e-commerce Web sites, where they use input about a customer's interests to generate a list of recommended items. Amazon.com [3] uses recommendations as a targeted marketing tool in many email campaigns and on most of its Web sites' pages, including the high traffic Amazon.com homepage. The demographic-based and collaborative filtering approaches hybridization had been introduced by researches for improving the recommendation quality rather than solving "cold-start problem". A group of researchers have applied a hybrid model-based approach on movie domain using user demographic data to enhance the recommendation suggestion process, it classified the genres of movies based on user demographic attributes, such as user age (kid, teenager or adult), student (yes or no), have children (yes or no) and gender (female or male). Additionally, other researchers modified user similarity calculation method to employ the hybridization of demographic and collaborative approaches. Different combinations of content based filtering and collaborative filtering are also available.

This paper is organized as follows: section III gives an overview of recommendation techniques, section IV deals with Hybrid system, section V explains the proposed system, section VI gives the evaluation of proposed system and section VII concludes the paper.

III RECOMMENDATION TECHNIQUES

Recommendation technique deals with the method used by the system to generate recommendations to the user. Most commonly used are the filtering techniques.

A. Collaborative Filtering

Collaborative filtering (CF) [3] [4] assumes a set of n users and a set of m items. Each user expresses opinions about a set of items. These ratings can be either numeric or unary and are represented as a matrix with users and items as dimensions. CF [1] generates top-k recommendations by applying cosine similarity or any other similarity computing mechanism on these matrixes. For that a similarity score is computed for each item that has at least one common rating by same user.

B. Demographic Filtering

The demographic filtering [9] generate recommendations based on the demographic profile (age, job, gender etc) of the user. This technique uses a co-relation between people like collaborative ones but using different items. The advantage of this technique is that it doesn't need a history of user item as in collaborative ad content based filtering. This recommender obtains group of user having similar demographic attribute(s) forming a neighborhood from which newly recommended items are generated as discussed in [9].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

C. Content based Filtering

Content based filtering [7] generate recommendations based on the utility of the item for the user which is estimated by utilities that are assigned by the user for the similar item. Systems implementing a content-based recommendation approach analyze a set of documents and/or descriptions of items previously rated by a user, and build a model or profile of user interests based on the features of the objects rated by that user. The profile is a structured representation of user interests, adopted to recommend new interesting items. The recommendation process basically consists in matching up the attributes of the user profile against the attributes of a content object. The result is a relevance judgment that represents the user's level of interest in that object. If a profile accurately reflects user preferences, it is of tremendous advantage for the effectiveness of an information access process.

D. Comparison of Recommendation Techniques

This section compares the different recommendation techniques mentioned in section III. The advantage and disadvantage of different recommendation are shown in Table 1.

IV HYBRID SYSTEM

In this system [10], different independent recommendation mechanisms are hybridized together to remove the limitations of the individual techniques. Commonly collaborative is combined with content based system [5] to build an efficient recommender system having better performance. There are different techniques which can be used in hybridization as explained in [11], [12], summarized as:

- Weighted : The ratings of several recommendation techniques are combined together to produce a single recommendation
- Switching: The system switches between different recommendation techniques depending on the current situation.
- Mixed : Recommendations from several different recommenders are presented at the same time
- Feature combination : Features from different recommendation data sources are thrown together into a single recommendation algorithm
- Cascade : One recommender refines the recommendations given by another
- Feature augmentation : Output from one technique is used as an input feature to another
- Meta- level : The model learned by one recommender is used as input to another.

Technique	**Advantages**	**Disadvantages**
Collaborative Filtering	A. Can identify cross-genre niches. B. Domain knowledge not needed. C. Adaptive: quality improves over time. D. Implicit feedback sufficient	I. New user ramp-up problem J. New item ramp-up problem K. Gray sheep problem L. Quality dependent on large historical data set. M. Stability vs. plasticity problem
Content Based Filtering	B, C, D	I, L, M
Demographic Filtering	A, B, C	I, K, L, M, N. Must gather demographic information

Table1. Comparison of different Recommendation Techniques

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

V PROPOSED SYSTEM

A. Overview:

LARS* has ability to produce location-aware recommendations using each of the three types of location-based rating within a single framework as mentioned in section I. The filtering method used in LARS* is item based collaborative filtering. From section III, it is clear that collaborative filtering depends on previous history of user or how many times a location/item has been rated by a user. So what about a new user or new location which doesn't have any previous ratings? This scenario is called Cold Start Problem [6].

Cold start problem is of two types: new user problem and new item problem. In new user problem, the recommender system cannot draw accurate inferences for users who are new to the system as they have no historic data to gather. In new item problem, the items which does not have sufficient ratings or a cold item will not be recommended to the user. New user cold start problem can be resolved by utilizing the demographic data explicitly given by a user. Also the content based filtering does not suffer from new item cold start problem as it focuses on utility of item for user rather on its ratings.

This paper is proposing a hybrid recommender system which exploits both the demographic details of user and the content based filtered details of locations for alleviating cold start problem and thereby improving the efficiency of recommendations generated by LARS*. The schematic diagram of hybrid system used in LARS* is shown in Figure 1.

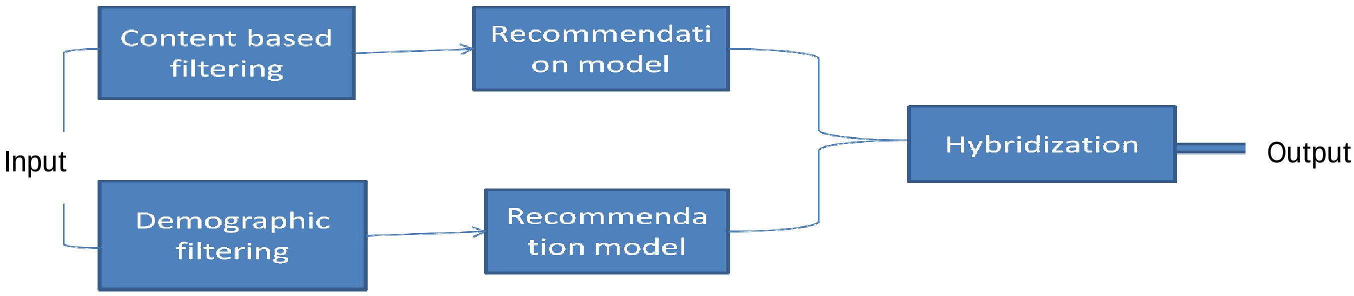

Figure 1: Schematic Diagram of Hybrid System in LARS*

B. Objective:

The proposed system is designed with the objective of improving the scope of recommendation of LARS* by alleviating the cold start problem. Here we are developing a sample client- server application of travel recommendation system where client is an Android Smartphone user.

C. Data Structure:

The data structure used is the pyramidal quadtree structure [13] having height h=6. The root cell of the Quadtree is constructed with all the ratings in the system given by all users (h=1). The root cell boundary is determined with the help of latitude and longitude. Latitude angles can range up to *+90 degrees* (or *90 degrees north*), and down to -90 degrees (or *90 degrees south*) and Longitude angles can range up to +180 degrees (180 degrees east), and down to -180 degrees (180 degrees west). At level h=2, the root cell is divided in four equal parts and ratings for locations that are within that boundary limit is inserted into respective cells. This process continues till quadtree reaches the level h=6. Whenever a user *u* requests for the recommendation, the system retrieves *u*'s latitude and longitude and checks in which cell *u*'s position falls and extract all the ratings in that cell and is fed to the hybrid system to generate recommendation.

D. System Architecture:

The architecture of the proposed system is shown in figure 2. The functioning of hybrid system in the figure is elaborated using figure 1. Here whenever a new user login, similar users are found who demographically matches with user and thus recommend top k locations that are rated by the similar user. Thus help in removing the new user problem. For new location that are not yet rated, system will check any previous ratings by user

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

matches with new location (content based). If so, new location is recommended to the user. For a new user there won't be any content based recommendation, as new user has no historic rating to match with. The locations recommended from both filtering techniques are combined and send to user.

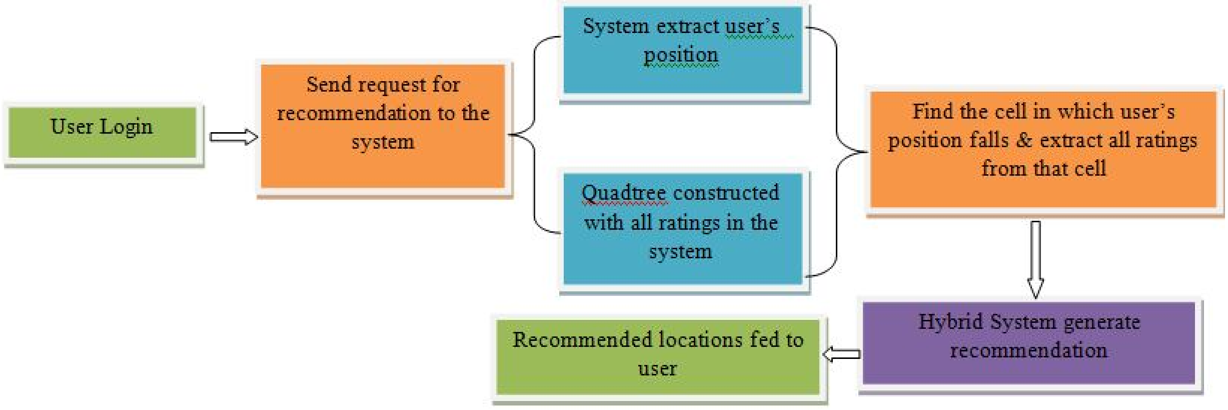

Figure 2: The proposed architecture of LARS* with hybrid system

VI EXPERIMENTAL ANALYSIS

This section provides experimental analysis on the proposed system based on actual system implementation using J2EE and Android. We compare LARS* with standard item based collaborative filtering technique along with LARS* using hybrid system.

Foursquare: a real data set consisting of spatial user ratings for spatial items derived from Foursquare user histories. Foursquare doesn't publish each "check-in" for a user and also the demographic details of the user. So we are manually generating some users and their ratings. Suppose we have 7 users *U*={ user1, user2, user3, user4, user5, user6, user7} where user7 is a new user. Suppose each filtering technique is recommending top 3 recommendations. So the hybrid system will recommended 6 locations to the client.

A. Comparison of LARS & LARS* using hybrid system:*

Here we are comparing LARS* and LARS* using hybrid system based on the number of locations recommended for each user. Comparison is shown in figure 3. From figure, we can conclude that proposed system is recommending more locations than that of existing LARS*. Proposed system is generating locations for the new user *user7* whereas existing LARS* is not recommending any location for the new user.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

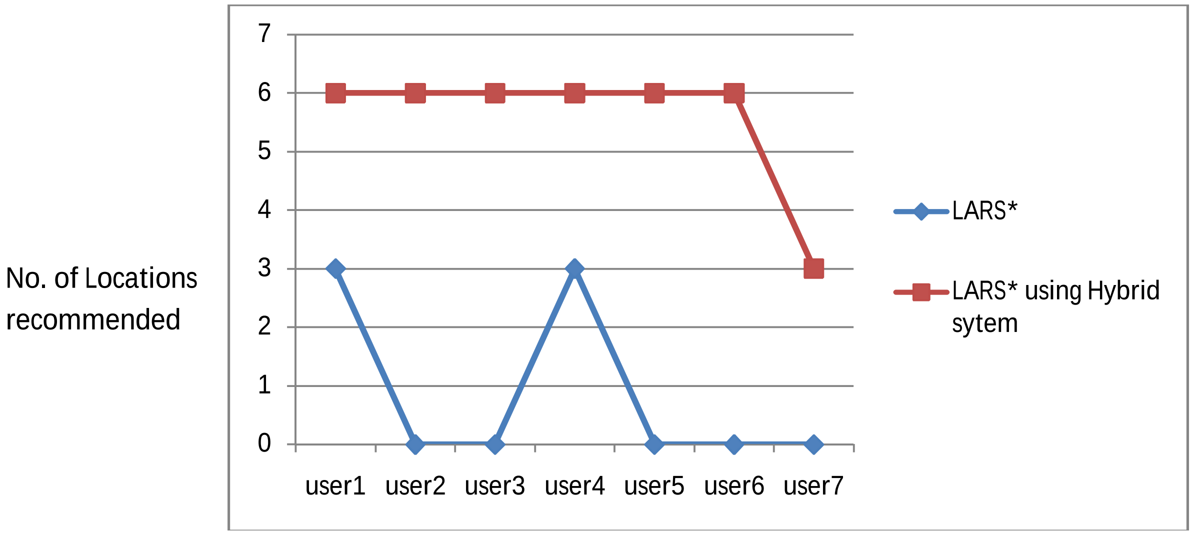

Figure 3: Comparison of LARS* & LARS* using hybrid system

B. Comparison of different filtering techniques:

For this comparison, we are comparing the filtering techniques- item based collaborative filtering, content based filtering, and demographic filtering with the hybrid system depending on the number of locations recommended. Each filtering technique is recommending top 3 locations from the list. From figure 4, it can be seen that for new user *user7* demographic filtering is generating recommendation whereas no locations are recommended for content based filtering and collaborative filtering as the user has no previous ratings. Also it can be seen that hybrid system is giving better results than independent functioning of these filtering techniques.

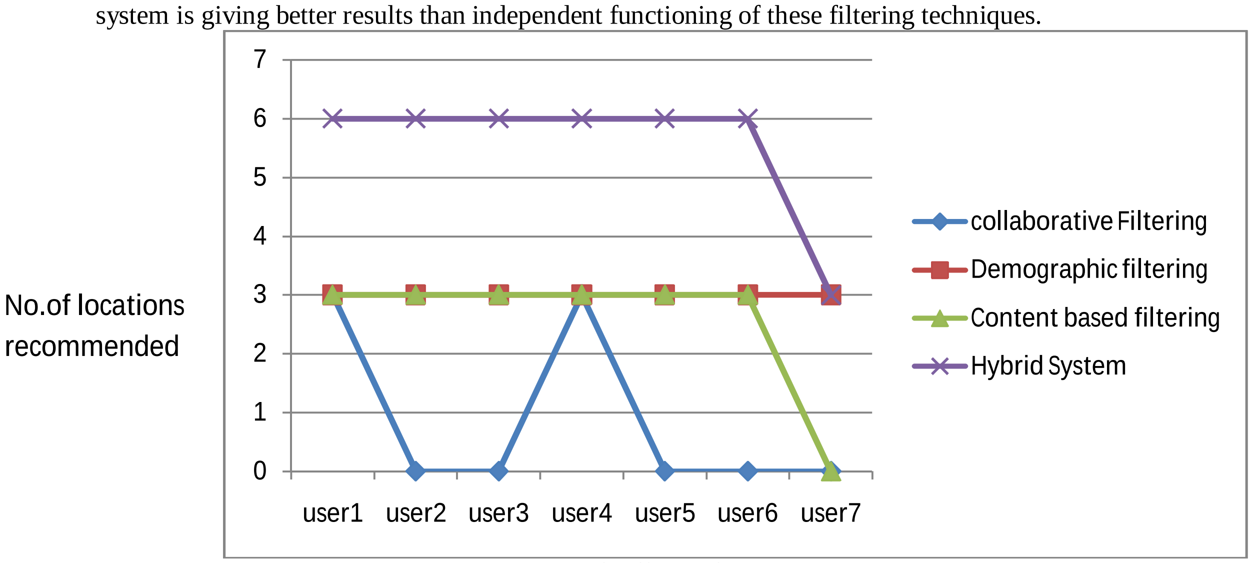

Figure 4: Comparison of different filtering techniques with hybrid system

C. Error Analysis on Proposed system:

Error analysis on proposed system is done with the help of WEKA tool. Here mean absolute error and root mean square error are calculated by WEKA automatically when the output of hybrid system is fed to it. It can

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

be seen that mean absolute error is less than 0.1 and root mean square error is less than 0.23. As the error rate of proposed system is low, the system is efficient as shown in figure 5.

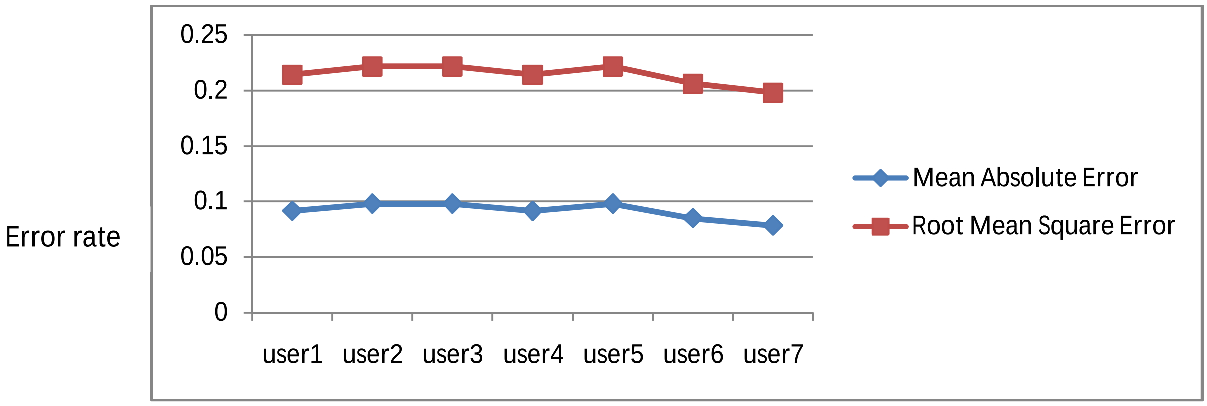

Figure 5: Error Analysis on Proposed system

D. Comparison of Accuracy:

Accuracy of the system is calculated with the help of confusion matrix and number of correctly classified instances by WEKA.

Accuracy (%)= Percentage of correctly classified instances

Using confusion matrix,

Accuracy (%) = (sum of all diagonal elements/sum of all elements in matrix) * 100

It can be also calculated with the help of precision in WEKA classification based on J48 classifier. From figure 6, it can be accuracy of proposed system is much higher than that of existing LARS*. Here accuracy is calculated for the number of locations the system can recommend for each user.

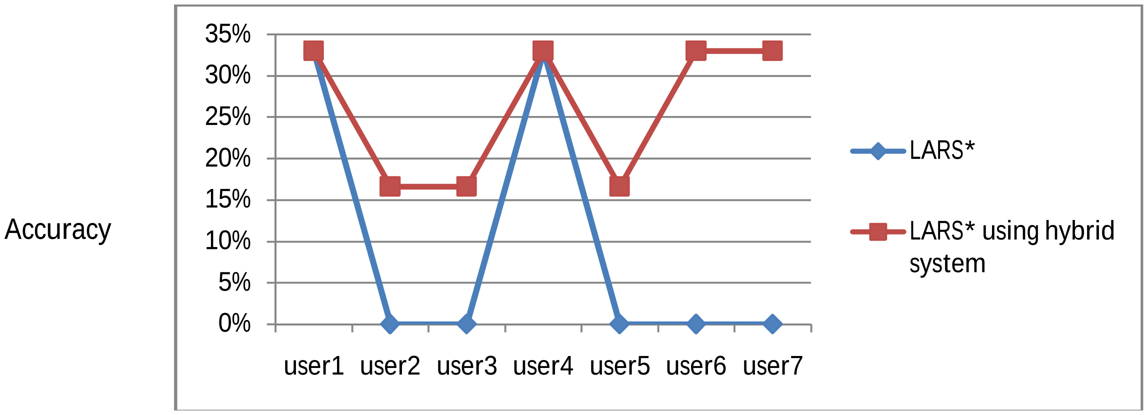

Figure 6: Accuracy Comparison

E. New Item Problem:

From analysis of the proposed system, new item i.e. new location problem is removed with the help of content based filtering. New locations are recommended to the users whose previously rated location's content matches with it. So we can conclude that new item problem is successfully removed with the help of content based filtering.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

VII CONCLUSIONS & FUTURE WORK

Even though LARS* is a novel recommender system, it has a drawback of cold start problem due to item based collaborative filtering. In order to remove it, a hybrid system of demographic and content based filtering is used. Experimental analysis on data set shows that proposed LARS* using hybrid system is more efficient than existing LARS* and is successfully removing the cold start problem. We had performed the experimental analysis on a small data set, in future it can be extended to large complex data and can be developed as an Android application similar to Foursquare.

REFERENCES

[1] Mohamed Sarwat, Justin J. Levandoski, Ahmed Eldawy, and Mohamed F. Mokbel, “LARS*: An Effcient and Scalable Location- Aware Recommender System, "IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 6, June 2014

[2] 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. Knowl. Data Eng., vol. 17, no. 6, pp. 734-749, Jun. 2005.

[3] G. Linden, B. Smith, and J. York, “Amazon.com recommendations: Item-to-item collaborative filtering," IEEE Internet Comput., vol. 7, no. 1, pp. 76{80, Jan./Feb. 2003.

[4] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,"in Proc. Int. Conf. WWW, Hong Kong, China, 2001.

[5] James Salter and Nick Antonopoulos,”CinemaScreen Recommender Agent: Combining Collaborative and Content-Based Filtering "in IEEE Computer Society, jan/feb ,2006.

[6] Eduardo Castillejo and Aitor Almeida and Diego Lopez-de-Ipina , “Social Network Analysis Applied to Recommendation Systems: Alleviating the Cold-User Problem," in Ubiquitous Computing and Ambient Intelligence,Lecture Notes in Computer Science Volume 7656, 2012, pp 306-313

[7] Pasquale Lops, Marco de Gemmis and Giovanni Semeraro, “Content-based Recommender Systems: State of the Art and Trends, "in Springer Science+Business Media, LLC 2011

[8] M. Pazzani, “A Framework for Collaborative, Content-Based, and Demographic Filtering,"Arti_cial Intelligence Rev., pp. 393-408, Dec. 1999.

[9] Laila Safoury and Akram Salah, “Exploiting User Demographic Attributes for Solving Cold- Start Problem in Recommender System," Lecture Notes on Software Engineering, Vol. 1, No. 3, August 2013

[10] Robin Burke, “Hybrid Recommender Systems: Survey and Experiments,"1997

[11] Jens Grivolla, Toni Badia,Diego Campo, Miquel Sonsona, Jose-Miguel Pulido, “A hybrid recommender combining user, item and interaction data,"European Union's Seventh Framework Programme managed by REA-Research Executive Agency http://ec.europa.eu/research/rea

[12] Chris Anderson, “Recommender systems for e-shops, "Business Mathematics and Informaticspaper,2011

[13]R. A. Finkel and J. L. Bentley, “Quad trees: A data structure for retrieval on composite keys,” Acta Inf., vol. 4, no. 1, pp. 1–9, 1974.