ADVERTISING ANALYTICS AND RECOMMENDATION STRATEGY

Aditi dube [1], Sachin Dhodi [2] , Suman Chitta [3] , Savita Sangam [4]
B.E Student [1-3], Head of Department [4]
[1-3]Department of Computer Engineering, [4]Department of Information Technology
[1-4]Shivajirao.S.Jondhale College of Engineering, Dombivli,India
[1-4]University of Mumbai

Abstract- Advertising analytics refers to the process of quantifying impressions, clicks and conversions that digital ads generate. The proposed system is a web-based application whose goal is to provide real-time and static on-demand services for advertisers and publishers to decide where to place advertisements so as to achieve maximum marketing values. This overcomes the setbacks of the existing systems like OLX, Magic Bricks where the advertisements are placed based on intuitions irrespective of the location. The system will perform data analytics to enable advertisement companies to move away from intuitive advertising to data-based decision making by collecting advertising data from various social network sites along with demographic information at different locations, and establish an innovative model for advertising and trend prediction. The advertisement companies and others who want to gain insights with geographical distribution will be attracted towards the application as it will reduce their cost of placing advertisements while improving their effectiveness. This will be done by using a recommendation engine that deploys collaborative filtering and map reduce methods for location recommendation. This is an effort to analyze advertisement data and make recommendation of locations to advertisers telling them where to place an advertisement by developing an "analytics service" which allows advertisers and publishers to achieve "maximum marketing values at minimal costs".

Keywords- Collaborative filtering, Location Recommendation, Location based services, Recommendation system, Advertising analytics, Big Data, Social network.

I. INTRODUCTION

In today's day to day life advertising has become a necessity for everybody be it seller who wants to increase their sales by promoting brand or customers who purchase products/ services after being aware about it. Also, recommendation systems are powerful technology for extracting additional value for a business from its user database. Recommendation system benefits users by enabling them to find items they like at same time help business by generating more sales. Currently, Recommendation systems are crucial tool in E-commerce and web.

Advertisement play an important role in customer's life as it is a best medium to communicate with them. Advertising helps to inform the customer about variety of products useful to them along with various brands available in market.

The work in this project is an effort to analyse advertisement data and make recommendation of location to advertiser like where to place advertisement by developing an analytics service which allows advertisers and publishers to achieve maximum marketing values at minimal costs.

For this we will use the following technologies:

1. Collaborative Filtering

Collaborative filtering is a method of making automatic predictions (**filtering**) about the interests of a user by collecting preferences or taste information from many users (**collaborating**) which in our case are the ads the user is or will be interested in viewing.

2. Map-reduce

MapReduce is a programming model for processing large **data** sets with a parallel distributed algorithm on a cluster. **Map Reduce** when coupled with HDFS can be used to handle **big data**. It has an extensive capability to handle unstructured **data** as well.

3. Cosine similarity computation

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.

4. Naive Bayesian classifier

Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

II. LITERATURE SURVEY

Following is discussion on previous research work on analytics and location-based recommendation systems:

Behavioural targeting [1] is a marketing method that uses web user information to strengthen advertising campaigns. The technique involves gathering data from a variety of sources about the potential customer's online browsing and shopping behaviours. This information helps create ads that are relevant

to that specific user's habits and interests, which the publisher can then display in that visitor's web browser.

The primary purpose of this technique is to deliver advertising messages to the behavioural target markets that have shown the most interest in them. The process involves compiling web searches, purchase histories, frequently visited websites and other information to create a full user profile, revealing what your audience wants, avoids and purchases. Using these data points, companies can formulate ads that align with the individual consumer's trackable preferences and needs, without conveying messages the viewer would find unappealing or irrelevant.

Amir Gandomi, Murtaza Haider, in their work [2] presents a consolidated description of big data by highlighting the metrics used to define characteristics of big data. They have explained various analytic methods to gain valuable and valid insights from big data like text summarization for text analytics, phonetic based systems for audio analytics, server-based architecture for video analytics and statistical method for predictive analytics.

Vangie Beal [3] defined Analytics as follows: Analytics refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning. Analytics focuses on developing new insights and understanding of business performance based on data and statistical methods.

Rosebt blog [4] provided information regarding different types of analytics which companies can use to learn from and better engage with their customers. They are:

Descriptive Analytics*:* Looks at data and analyse past events for insight as to how to approach the future.

Diagnostic Analytics: Monitor past performance to understand what happened and why.

Predictive Analytics: Uses data to determine the probable future outcome of an event or a likelihood of a situation occurring.

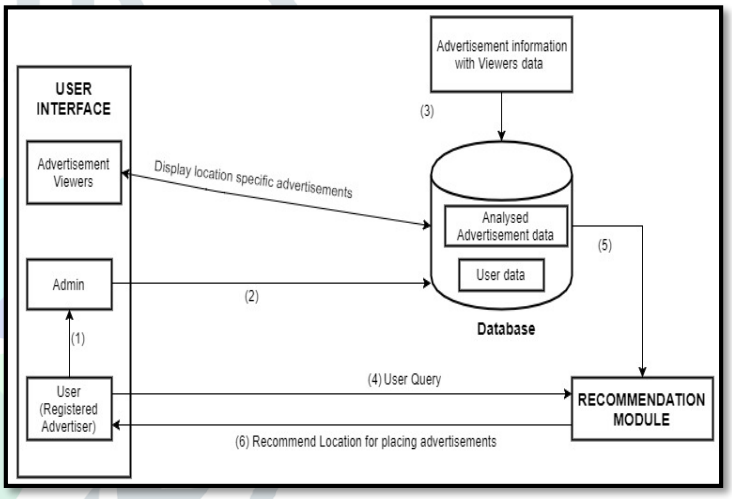
Prescriptive Analytics: Synthesizes big data, business rules, and machine learning to make predictions and then suggests decision options to take advantage of the predictions.

Lei Deng, Jerry Gao [5], have introduced a novel recommendation engine for advertisers by developing big data analytics service in advertising and marketing. The system aims at providing on demand services to advertisers and publishers to decide when, what, where and how to place advertisements by analyzing advertisement data on social media sites like Yelp and Twitter using MapReduce framework and Synthesis Index location recommendation strategy so to reduce their costs which improving ad-placement effectiveness.

J.Bao, Y.Zheng, and M.F.Mokbel [6], have built a location based and preference-aware recommender system that offers a particular user, a set of venues (like restaurants, shopping malls, movie theatres etc.) with in geospatial range by considering user preferences and social opinion regarding that venues from local experts. The proposed recommendation system, comprise of two parts: *offline modeling* and *online recommendation* where the *offline modeling* part discovers the local experts in each city as well as extracts user's preference from their location history and the *online recommendation* part applies preference-aware candidate selection algorithm to select local expert based on user's preference so as to make recommendation by combining local experts reviews with user's preferences.

Chi-Yin Chow, Jie Bao, and Mohamed F. Mokbel, have proposed GeoSocialDB [7] named social networking system which implement three services like location-based news feed, location-based news ranking and location-based recommendation based on users personalized spatial and social preferences as query operators inside a database engine by highlighting research challenges that needs to be addressed so to optimize query processing performance.

M.-H.Park, J.-H.Hong, and S.-B.Cho [8] proposed a map-based personalized recommendation system which reflects user's preference modeled by Bayesian Networks (BN). This system consist mainly of three sections, where first section include collecting and preprocessing content information like time, location, weather, and user location from mobile device. Second section involve training Bayesian network model using collected data and last section involves mapping content and location database for recommending user a preferred item by displaying it onto mini-map.



John S. Breese David Heckerman Carl Kadie in their work [9] introduced collaborative filtering along with its type: Memory based and Model based. They performed empirical analysis on prediction algorithms related to collaborative filtering like correlation, vector similarity of Memory based and Cluster model, Bayesian network of model based by using MSWeb dataset. The observation shows that memory-based prediction algorithms outperform model-based algorithms.

Badrul S, and John R [10] introduced an item-based collaborative filtering algorithm which overcomes data sparsity issue related to user based collaborative filtering algorithm by analyzing item-item matrix instead of user-item matrix for making recommendation. This work introduces techniques used for computing item-item similarities like correlation and cosine similarity followed by prediction computation techniques like weighted sum and regression model for obtaining recommendations. Results suggest that item-based algorithm provide high-quality recommendation and better performance than user-based algorithm.

III. PROPOSED SYSTEM

Existing advertisement solution like OLX is an global online marketplace [12] which facilitates buying and selling

services and goods such as electronics, household goods, furniture, cars, bikes etc. However, it has some shortcomings like:
Selling Things on OLX: Advertisers need to fill up advertisement form provided on OLX site which need information like Ad Title, its category and description along with photos or videos regarding product to be sold.

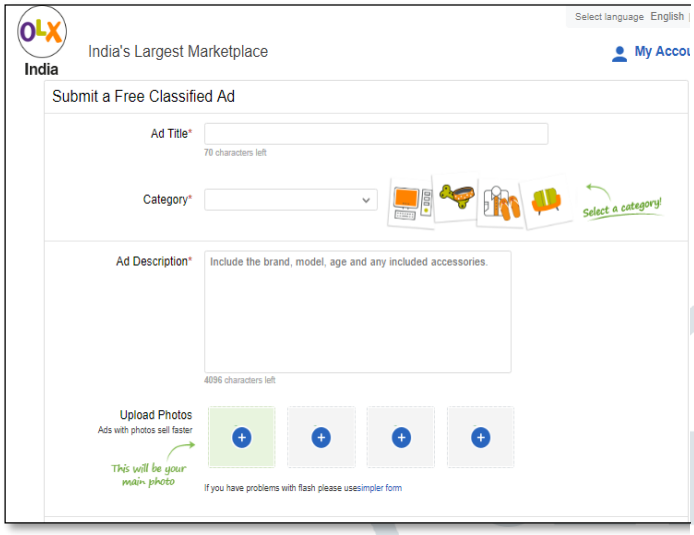


Fig: 3.1 Place advertisement on OLX website[14]

The work in this paper will overcome shortcoming of existing advertisement system like OLX where advertisers needs to make intuitions of locations for placing advertisements as well as enable viewers to view advertisements at specific location.

The proposed system is based on Predictive analytics whose architecture is depicted in Fig.3.2

Fig 3.2: System Architecture Diagram

The proposed system architecture has following three main modules. They are:

1. **User Interface**
2. **Analytics Module**
3. **Recommendation Module**

These three modules are explained in detail along with the roles and functionalities they provide.

1. User Interface

This module defines three main user roles for the system with well-defined operations to perform. They are:

- Administrator
- User (Registered Advertiser)
- Advertisement Viewers

2. Analytics Module

This module will analyse real-time advertisement data collected from Google analytics by using MapReduce API.

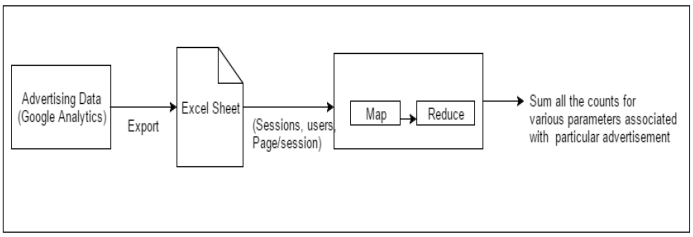


Fig: 3.3 Analytics flow

3. Recommendation Module

This module performs analysis of real-time advertisement data fetched from viewers at different location and using collaborative filtering algorithm recommend location to advertiser for placing new advertisement.
This is the main module of our system. We are using collaborative filtering in this module.
This module will perform two steps:

1. **Classification:**
   - Data from analytics module will be taken as input to define category for classifying advertisements.
   - When advertiser wants to place new advertisement, pre-processing technique will be applied on new advertisement description and then using Naive Bayesian Classifier it will be classified to appropriate category.

2. **Recommendation:**
   - After classifying new advertisement to appropriate category, advertising data of similar product along with their demographic information will be fed as input to collaborative based filtering recommendation engine for recommending a location where new advertisement needs to be placed.

IV. EXPERIMENTATION RESULTS

Item based collaborative filtering is a model-based algorithm for recommender engines. In item based collaborative filtering similarities between items are calculated from rating-matrix. And based upon these similarities, user's preference for an item not rated by him is calculated. The steps are as follows:
Step 1: Write the user-item ratings data in a matrix form.
Step 2: We will now create an item-to-item similarity matrix whose idea is to calculate how similar an item is to another item using cosine similarity measure and Jaccard's coefficient measure.[13]

Cosine Similarity Computation:

- Identify all distinct words in both texts.
- Identify the frequency of occurrences of these words in both text & treat it as vector.
- Apply cosine similarity function.

Jaccard's coefficient Computation:
Jaccard's coefficient is a statistic measure used for comparing the similarity between textual information.
Collaborative filtering algorithm steps specific to this project work:

- **Step 1**: Create advertisers- advertisements ratings data in matrix format
- **Step 2**: Create advertisement-advertisement similarity matrix whose aim is to calculate similarity between various advertisements based on their advertisement description.
- Cosine similarity and Jaccard's coefficient is applied on advertisement description to examine which similarity measure provide better result.

Example: Advertisement1 Description: school bags, hand purse.
Advertisement 2 Description: Hand bags.

Cosine Similarity Computation

Count number of times each word appears in each description:

Words	Count	
	Ad description 1	Ad description 2
School	1	0
Hand	1	1
Bags	1	1
Purse	1	0

Table 3.1: Count of words in advertisement description.

Two vectors are:

v1={1,1,1,1} and v2={0,1,1,0}

- Cosine Similarity Computation:

cos(v1, v2) = (v1 · v2) / ||v1|| ||v2||
=(0+1+1+0)/ [sqrt(4)*sqrt(2)]
=0.7071

- Jaccard's Coefficient Computation:

p = 2, q = 4, r = 2, s = 0
s_{ij} = p/ (p+q+r)
= 2/6 = 0.333

- **Step 3**: For each advertiser who wish to place new advertisements similar to existing advertisements, location will be recommended based on above matrix and clicks collected for existing similar advertisements from database.

Thus, as per calculation advertisement 2 will be recommended same location as advertisement 1.

Similarly we calculated the cosine similarity and jaccard's coefficient of the following commodities:

Sr.No	Advertisement Descriptions of different category	Cosine Similarity	Jaccard Coefficient
1	Fashion	0.25	0.1111
2	Hotels	0.5773	0.2
3	Cars	0.4772	0.1818
4	Hotels	0.8164	0.2857
5	Fashion	0.3535	0.1428
6	Cars	0.37	0.1428
7	Real Estate	0.5	0.25
8	Cars	0.3162	0.1333
9	Mobiles	0.4082	0.1666
10	Real Estate	0.9486	0.25
11	Mobiles	0.5	0.2
12	Real Estate	1.3416	0.25
13	Fashion	0.25	0.111
14	Cars	0.5773	0.222
15	Mobiles	0.7071	0.25
16	Footwear	0.4082	0.1666
17	Mobiles	0.3535	0.1428
18	Fashion	0.5	0.1666
19	Footwear	0.8071	0.66
20	Fashion	0.2886	0.125

Table 3.2. cosine similarity and jacard's coefficient for different comodities.

By using real-time clicks collected for advertisements at different locations and collaborative filtering based location recommendation algorithm, system provide location prediction.

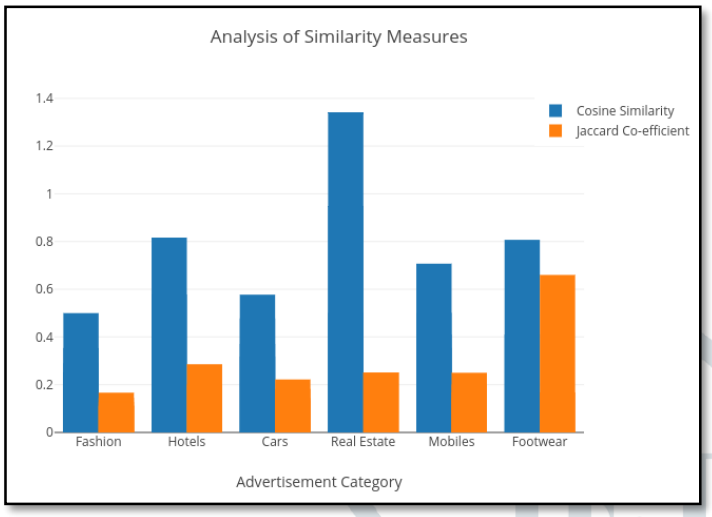


Fig 3.4. Analysis of similarity measures.

We calculate the accuracy of both of these values by using the following formula where an accepted value of V_A and an observed value V_O, are used to calculate accuracy.

$(V_A - V_O)/V_A \times 100$ = percent accuracy (1)

Considering hand bags for example, rounding off value of Cosine, we get 0.5773= 0.6 and Jaccard =0.2.
If we have to get the value to 1, we have 0.2*5 by Jaccard and in a similar case by Cosine we have 0.6*2 ≈ 1.2.
Now, if we consider the extra work done by these two methods, we get the ratio of 4:1.

[(4-1)/4]*100=75%

From above results, it is clear that cosine similarity provides 75% accuracy compared to Jaccard's Coefficient measure. Thus, collaborative filtering algorithm with cosine similarity measure is used for location recommendation to advertisers.
Collaborative filtering algorithm implemented two similarity measures namely: Cosine similarity and Jaccard co-efficient measure on advertisement description as shown.

V. CONCLUSION

This project involves analysing real time advertising data collected at different locations using item based collaborative filtering algorithm with cosine similarity measure which provides 75% accuracy compared to Jaccard's co-efficient measure to make effective trend predictions and advertisement recommendation in future. The core business activity affected will be advertising and marketing as the proposed system will allow advertisers and publishers to achieve maximum marketing values by placing advertisements at suitable locations. Also, it will enable viewers to view advertisement which are placed in their vicinity so that they can have information regarding various products and their brands available to fulfill their necessity.

As the project's main focus is analysing advertisements response at different locations, so currently advertisements placed for viewers are in simple textual format. Thus, future work can include:

- Displaying banner advertisements or video advertisements to viewers such that they will be persuaded to watch advertisements.
- In e-commerce sites, the proposed system can be included with behaviour targeting to provide location recommendation to advertisers for ad-placements.

VI. REFERENCES

[1] https://www.lotame.com/what-is-behavioral-targeting/

[2] Amir Gandomi, Murtaza Haider, "Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management 35 (2015) 137–144

[3] https://en.wikipedia.org/wiki/Business_analytics.

[4] Overview of types of analytics from *http://www.rosebt.com/blog/predictive-descriptive-prescriptive-analytics.*

[5] Lei Deng, Jerry Gao, "An Advertising Analytics Framework Using Social Network Big Data", 2015 5th International Conference on Information Science and Technology (IClST), 2015 IEEE.

[6] J.Bao, Y.Zheng, and M.F.Mokbel, "Location-based and preference aware recommendation using sparse geo-social networking data" in Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, 2012, pp. 199-208.

[7] Chi-Yin Chow, Jie Bao, and Mohamed F. Mokbel, "Towards Location-based Social Networking Services" in The 2nd Workshop on LBSN, 2010.

[8] M.-H. Park, J.-H. Hong, and S.-B. Cho, "Location-based recommendation system using bayesian users preference model in mobile devices" in Ubiquitous Intelligence and Computing. Springer, 2007, pp. 11301139.

[9] John S.Breese, David Heckerman, Carl Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", Technical Report MSR-TR-98-12, Microsoft Research, October1998

[10] Sarwar, B, "Item-based collaborative filtering recommendation algorithms" in WWW10 Conference, pp. 285–295 (2001)

[11] Kenneth Wai-Ting Leung, Dik Lun Lee, Wang-Chien Lee, "CLR: A Collaborative Location Recommendation Framework based on Co-Clustering", *SIGIR'11,* July 24–28, 2011, Beijing, China. ACM 2011

[12] OLX information: https://en.wikipedia.org/wiki/OLX

[13] Wael H. Gomaa , Aly A. Fahmy ,"A Survey of Text Similarity Approaches", International Journal of Computer Applications (0975 – 8887) Volume 68– No.13, April 2013

[14]OLX website link: https://www.olx.in/