Advertisement Mining using Hidden Markov Model

K. Sathiyamurthy[1], P.Dhivya [2], DJ.Panimalar[3]

Assistant Professor, Department of CSE, Pondicherry engineering College, Puducherry, India[1]

Student, Department of CSE, Pondicherry Engineering College, Puducherry, India[2,3]

Abstract: Predicting the individual's web-browsing behavior on the internet is more important for advertisement mining. This makes the advertisers or the publishers to successfully interact with users in providing relevant advertisement, many intelligent interfaces requires a method for recognizing, analyzing and predicting user behavior actions. Initially we need to collect a log of datasets with user behavior attributes. Hidden Markov Model is used to derive a pattern for predicting the behavior of the users on the web. Also, it helps in analyzing the performance of user behavior profile and estimating the advertising cost to perform Mapreduce jobs based on user's search query. With these collected information, optimized advertisement is deployed to the user using Bigquery mechanism.

Keywords : Hidden Markov Model, behavioral targeting, Mapreduce, Advertisement mining, Bigquery.

I. INTRODUCTION

As the Internet has now increasingly become an essential tool for marketing and advertising, the effectiveness of online advertising towards consumers beliefs and characteristics remains crucial yet debatable issue among various advertisers and publishers. Online advertising mining is a form of marketing and advertising that uses the Internet to promote the products to users. This can be implemented by Hidden Markov Model (HMM).HMM is a statistical Markov model having hidden and visible states.

Hidden Markov model (HMM) is good at information extraction and has the ability to guess the hidden states of each observation symbol with high accuracy, thereby adapting the concepts of HMM to investigate the user's browsing behavior.HMM is used to compute the likelihood of behaviors with its appropriate probability. The problem of predicting a behavior of user on a web-site has gained more value due to the rapid growth of the world wide web(WWW) and the need to personalize and influence the user's browsing behavior. The necessary knowledge of user browsing behavior history is helpful to predict the future sequences which are likely to be visited by the user in future. There is a huge scope for advertisers to design prediction model based on user's browsing page sequences. The main objective of these prediction models is to attain more prediction accuracy. In Hidden Markov model the term hidden refers to sequence of state through which the model get ahead of [1]. In general, Markov model has functionality to predict future action based on the results of previous actions. In reference of proposed system for user behavior prediction process, hidden Markov model is used to predict relevant advertisement based on history of previously visited webpage as well as the browsing behavior of user which is collected from cookies information. The rest of the paper is organized as follows: first, we review relevant literature towards online advertising using HMM model, second describes about the proposed system framework, third presents experimental setup with its conclusion.

II . RELATED WORK

Jianqing Chen et.al.,[2] have analyzed the economic implications of behavioral targeting for advertisers and online publishers. They have also recognized two effects associated with behavioral targeting, the competitive effect and propensity effect which affects the online publisher's revenue in opposite directions. The relative strength of these two effects determines whether the revenues of publisher is positively or negatively affected. The dominant advertiser is worse off under behavioral targeting when they have   significant competitive advantage over their competitors because under traditional advertising, he/she would then grab a larger group of users and still recognize a decent payoff.

The problem of web prediction is proficiently handled by Ajeetkumar S. Patel et.al.,[3] by the fusion of the both genetic algorithm and hidden Markov model. Their fusion system successfully increases the average overall prediction accuracy by 5%.To achieve their desired results, experiments were designed and achieved using UOFS data set with parameters such as number of N-grams, labeling, and partition percentage. Their comparative results also shows that smaller N-gram models performance is better than higher N-gram models in terms of prediction accuracy due to the reduction in number of sessions.

Kun Liu et.al.,[4], have attempted to evaluate the predictive power of social data across 60 consumer domains on a large online network of over 180 million users in a period of two and a half months. Also, they have done a comprehensive study of social data in the context of behavioral targeting on such an unprecedented scale. Their analysis offers interesting insights into the value of social data for developing the next generation of targeting services. They have developed a wide-array of supervised and unsupervised approaches to control social data for Behavioral Targeting.

Jun Yan et.al.,[5] have provided an empirical study on the click-through log of advertisements collected from a commercial search engine. From the experiment results

over a period of seven days, they have drawn three important conclusions: (1) Users who clicked the same ad will truly have similar behaviors on the Web; (2) Click-Through Rate (CTR) of an ad can be averagely improved as high as 670% by properly segmenting users for behavioral targeted advertising in a sponsored search; (3) Using short term user behaviors to denote users is more effective than using long term user behaviors for BT. They have also directed statistical t-test which verified that all conclusions drawn in the paper are statistically important.

III . PROPOSED METHOD
A . ARCHITECTURE DIAGRAM

The proposed system architecture of our system is shown in Figure1, as it demonstrates the Data analysis for mining advertisement by building HMM model, thereby generating Log-likelihood values for predicting pattern. Secondly, uploading those collected patterns into the app engine in order to perform MapReduce programming to generate key value pairs for each advertisements and examining the efficiency of advertising cost in HDFS framework. Finally retrieving relevant advertisement by executing Bigquery.
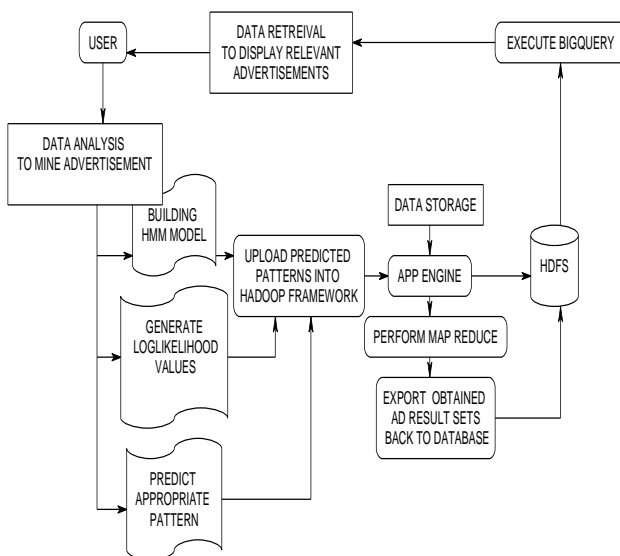

Fig. 1. Proposed System Framework

Figure 1 also illustrates the framework and the modules included in proposed approach.
1.Data Analysis
2. Data Storage
3. Data Retrieval

B . DATA ANALYSIS

Analyzing the log of datasets using Behavioral targeting technique in order to increase the effectiveness of their advertising using information collected on an individual's Web-browsing behavior. Thereby, displaying targeted advertisement by utilizing these information. Later, Applying this click stream data to hidden markov model to uncover consumer browsing patterns and serving targeted ads matched to an individual.

1. HIDDEN MARKOV MODEL TO PREDICT USER'S BROWSING BEHAVIOR
Hidden Markov model (HMM) is an excellent method in extracting behaviour of user and predicting appropriate pattern .The hidden and visible states along with observation sequence can be implemented with high accuracy, to analyze the user's browsing behavior.HMM is used to represent and to compute the log- likelihood values of user behaviours with its appropriate probability.

The problem of predicting a user's behaviour on a web-site has acquired more importance due to the fast growth of technology and the need of personalizing and influencing a user's browsing experience. Hidden Markov models and their variations have been found well suited for addressing this problem.

It involves the following steps, constructing HMM model with given advertisement related datasets, generating sequences of observations and symbols using the above constructed HMM and learning these parameters of a HMM using Log-likelihood values.

2. GENERATING LOGLIKELIHOOD VALUES USING PARAMETERS AND PREDICTING PATTERNS.
After building HMM model, log-likelihood values has to be estimated using Baum-Welch (forward-backward) algorithm in order to derive the pattern for retrieving appropriate advertisement. Thereby, deriving an proficient procedure for estimating the parameters from unlabeled data using Hidden states, visible states and observation sequence.

C . DATA STORAGE
In MapReduce programming framework for mining advertisement, collected data and user behavioral patterns were processed in isolation by mappers and reducers tasks. The obtained pattern for user behavior profile is imported to the datastore (Hadoop distributed File System) and Appengine. The images allied to each keyword or for common classifier are stored in the data store (child table) as a blobkey. Blob key is the reference to the file on the blob store. Then the images for the particular keywords are imported to the blob store using map reduce.

D . DATA RETRIEVAL
Targeted advertisement can be portrayed to the browser using Bigquery from HDFS after performing MapReduce tasks. The obtained pattern for user behavior is imported to the datastore, then the images for the particular keywords are imported to the blobstore using map reduce. After performing this process, next step is to execute Bigquery mechanism. In order to apply behavioral targeting technique and to ensure the working of mechanism, the keyword is fed to the query processor and the query is instigated. MapReduce takes the request and initiates the job process. Thereby, appropriate advertisement is mined from the database related to the particular keyword and is deployed to the user.

IV. EXPERIMENTAL SETUP

The dataset used for the experimental setup contains the advertisement related attributes such as session, location, item, cost, CTR, CPC, CR, CPM that are collected from cookies[6] for examining individuals web browsing behavior. These datasets are taken into consideration for the following simulation results.

The HMM model for mining advertisement has to built using the attributes such as ,

Given: Visible states={Item ,Location, Session, Cost, CTR, CPM , CR, CPC}

Hidden states={Low, Medium, High}

Initialization: set $(\lambda=A,B,\pi)$ with random initial conditions. The algorithm updates the parameters of λ iteratively until convergence, following the rules below[7]:

$$\overline{\pi}_i = \gamma_i(1)$$

$$\overline{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_{ij}(t)}{\sum_{t=1}^{T-1} \gamma_i(t)}$$

$$\overline{b}_i(k) = \frac{\sum_{t=1}^{T} \delta_{O_t,O_k} \gamma_i(t)}{\sum_{t=1}^{T} \gamma_i(t)}$$

Based upon these modeling parameters, estimating a pattern in relevant to existing user behavior profile repository. After building HMM model, modeling cost has to be estimated using log-likelihood values using Baum-Welch (forward-backward) algorithm in order to derive the appropriate pattern for retrieving targeted advertisement. Thereby, estimating the parameters using Hidden states, visible states and observation sequence until it reaches convergence.

V . CONCLUSION

It has been proved that Hidden Markov Mode is an exceptional method in the field of recognizing pattern for user behavior. We have utilized log datasets with user browsing behavior characteristics as training data to construct our model. HMM model is used to guess the browsers intentions in online. In addition, this model helps to analyze the behavior of user to increase CPC rates. After estimating the individuals browsing pattern, we can deploy them with relevant advertisement using Bigquery mechanism.

ACKNOWLEDGEMENT

REFERENCES

[1] M. Awad and I. Khalil, Prediction of Users Web-Browsing Behavior: Application of Markov Model, IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 42, no. 4, pp.11311142, Aug. 2012.

[2] Jianqing Chen, Jan Stallaert," An Economic Analysis of Online Advertising Using behavioral targeting", *in American Society for Engineering Education*, vol. 6(4), pp .19-22,1996.

[3] Mr. Ajeetkumar S. Patel, Prof. Anagha P. Khedkar,**"**Genetic Algorithm and   Statistical MarkovModel Fusion for Predicting User's Surfing Behavior Using Sequence Patten Mining",*in International Journal of Emerging Trends & Technology in Computer Science (IJETTCS),*August 2014.

[4] Kun Liu, Lei Tang, "Large-Scale Behavioral Targeting with Social Twist", in *Proceedings of the CIKM'11,* October 24–28, 2011, Glasgow, Scotland, UK.

[5] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen," How much can behavioral targeting help online advertising?", in *Proceedings of the 18th International Conference on World Wide Web,* pages 261-270, 2009.

[6] Matthew Richardson, Ewa Dominowska, Robert Ragno,"Predicting Clicks: Estimating the Click Through Rate for New Ads",in Proceedings of the *International World Wide Web Conference Committee (IW3C2),*May,2007

[7] Paul M. Baggenstoss,"A Modified Baum–Welch Algorithm for Hidden Markov Models with Multiple Observation Spaces" in *Proceedings of IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, May 2001.