Click Through Rate Prediction of an Advertisement according to User's interest

Aditya Bahukhandi IIT2017142
Utkarsh Pankaj IIT2017082 Arsh Panghal IIT2017048

December 3, 2020

Contents

Abstract

Placing advertisements in appropriate locations on web pages leads to the rise in the CTR value which in turn influences the growth of customer access to advertisement resulting in more profit rate for the ad exchange, advertisers and publishers. Therefore it becomes important for a business enterprise to predict the CTR metric for formulating an efficient ad placement strategy and effectively targeting the potential customers. We will implement different machine learning and deep learning models including the likes of deep interest networks and piece-wise linear regression and compare the different predictions obtained using those models.

Keywords: Click Through Rate, Advertisement metrics, Prediction

1 Introduction

As we are progressing into the digital era, majority of the people have been using the internet for various purposes. With a billions of web pages visited by users in a single day, online advertising becomes very important and a major tool for the companies to meet its desired consumers and the consumers or users to meet their desired products. Hence the importance of online advertising can be understood from here. The advertisements can be broadly categorised into sponsored search, banner ads, display ads and contextual advertising. In a sponsored search, the advertisement is selected based on the search query. The display advertisements have a role to present brand advertisements and messages to the site visitors. There is a term called "Click Through Rate" that comes up in the field of advertisements. This represents the percentage at which the user clicks the ad out of the impression. The CTR of an ad is the number of clicks on an advertisement divided by the number of times the advertisement is shown which is then represented as a percentage. The factors that have a role on the CTR value is the website or application the user visits. Some examples of the organisations or companies that are prominent in this field and where research in such area is common are Google and Facebook. They act as a mediator between the companies which want to show their advertisements and the consumers or the visitors on their websites.

2 Motivation

With the advent of digital technologies and increasing usage of internet and millions of web pages visited by users, online advertising has become one of the most prominent ways for companies reaching out to their customers. This also helps the businesses to expand their consumer base and in turn increase their profits. Hence, it is also important on how and where the advertisements are placed on applications and websites so that the CTR percentage is high and so the advertisements are effective in increasing the customers for the businesses.

3 Literature Review

The paper by Avila Clemenshia P. and Vijaya M.S. [1] try to show that the placement of advertisements are a major factor in determining the CTR value and hence, their paper stresses on the importance of increasing the CTR value in order to increase the profits of the corporate world that pours money for advertisement. Therefore the paper presents a predictive model that produces the CTR value based on different placements of the advertisement on the website using statistical machine learning regression techniques such as Poisson Regression (PR), Linear Regression (LR) and Support Vector Regression (SVR). As a result of the paper, it is found that SVR based click model gives the best result in predicting CTR through hyperparameter optimization.

The paper by Guorui Zhou, Chengru Song, Xiaoqiang Zhu, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, Kun Gai (Alibaba Group) [2] proposes a novel model: A deep interest network is built here which creates a local activation unit which learns the user interests from his history with respect to a certain advertisement. This vector varies over different advertisements, improving the expressive ability of a model.

The paper by Kun Gai1 , Xiaoqiang Zhu1 , Han Li1 , Kai Liu [3] explores how the ctr prediction problem can be solved using large scale piecewise linear models i.e LS-PLM models, which overcomes the shortcomings of the LR model following the divide and conquer strategy. To accommodate for non-linearity, the feature space is divided into several local regions and a linear model is fitted in each region, such that depending on the number of divided regions any nonlinear function may be fitted on the data. The aforementioned model is also highly scalable and can be trained on multiple machines in parallel. This paper also proposes an optimization technique based on directional derivatives and the quasi-newton method for efficient model training.

The paper by Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu [4] focuses on employing a hybrid model solution to the click through rate prediction problem. This model combines the boosted decision tree model with the linear classifier which significantly increases the accuracy of the linear classifier. The paper explains that by first transforming the input features by binning the feature and treating the bin index as a categorical feature, the linear classifier can thus learn a piecewise nonlinear map for that feature, resulting in a significantly higher accuracy.

The paper by Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li and Xiuqiang He [5] focuses on DeepFM : Factorization-Machine based Neural Network for CTR Prediction. In this paper, they have illustrated as to how to derive an end-to-end learning model that majorly focuses on both low and high order feature interactions. The DeepFM model that is proposed here, combines the capability of factorization machines for recommendation and deep learning for feature learning in a new neural network architecture. Innumerous experiments and illustrations have been discussed in this paper which demonstrate the efficiency and effectiveness of deepfm over other models for CTR prediction.

4 Problem Statement

CTR is an important factor through which we determine the effectiveness of the position of an advertisement. In this paper, we discuss ML and data mining techniques to create a CTR Prediction Model for an advertisement according to the user's interest.

5 Proposed Methodology

5.0.1 Dataset Description

The dataset that is used for subsequent model training is the Amazon dataset consisting of user review data and its associated metadata, using which Amazon prepared models to serve it's ads and recommendations on its website.

Containing over 50 million entries, the dataset is over 20 gb in size. We have used the subset of the data pertaining to the electronics category out of the plethora of products categories available in the dataset. Moreover, the data is in JSON format which needs a bit of processing before it's used. The reviews data consists of the following

- reviewerID - ID of the reviewer
- asin - ID of the product
- reviewerName - name of the reviewer
- helpful - helpfulness rating of the review
- reviewText - text of the review
- overall - rating of the product
- summary - summary of the review
- unixReviewTime - unix time of the review
- reviewTime - raw time of the review

The associated metadata on the other hand consists of descriptions, price, sales-rank, brand info, and co-purchasing links and comprises the following fields:

- asin - ID of the product
- title - name of the product
- price - price in US dollars
- imUrl - url of the product image
- related - popular related products consisting of “also bought”, “also viewed”, “bought together”, “buy after viewing” products
- salesRank - sales rank information
- brand - brand name
- categories - list of categories the product belongs to

5.0.2 Dataset Preparation

There is intensive data preprocessing involved in the Amazon dataset which consists of the user reviews and its associated metadata. The raw data available is in the json format and needs to be converted into pandas dataframes making it a lot easier to perform further data crunching. Therefore, first of all the required categorical variables of the review and metadata dataset namely asin, categories and reviewerID are mapped to unique numeric values which helps in easier processing later on in the training phase. The reviews of every user is also sorted according to the timestamp. Next up, the training and test sets are created using a 90:10 split. The ith entry for every user in the training set are his reviewer id, his history of reviewed i-1 books and the ith book he will review if label is put 1 which is otherwise the book he won't review if label is 0. Entries for the test set are also the same with the exception of the label which is not included. After these respective steps, the data is well suited for training the models.

5.0.3 Model Training

1. **Model 1 : lr+gbdt**: As the first model we used the gradient boosted Decision trees in combination with logistic regression wherein the index of the prediction leaf node is fed as a sparse input to the linear classifier. For the gradient boosted decision tree, the maximum tree depth was set as 4 and the learning rate used was 0.01. Bagging fraction, which is the parameter specifying the fraction of data to be used for each iteration, was kept 0.8 and the parameter for regularization lambda as 0.2. Moreover, AUC was used as the evaluation metric. For the logistic regression part l2 regularization was used with parameter c set as 1.0.

2. **deepfm**: The deepfm model comprises two components namely the FM component and the deep neural network component which are trained jointly on the same shared input. The Factorization Machine(FM) component proves to be a very effective model when working on sparse data, which is often the case in CTR data, and is able to effectively capture the appropriate order 2 feature interactions for a good prediction. The other component involved in the model is the deep neural network component wherein a feed forward neural network is used for learning the higher order interactions. An embedding layer is also added to the network that converts the input vectors into lower dimensional vectors

before feeding them to the dense layers, thereby dealing with the sparsity that pervades the ctr datasets. Moreover, the FM model serves as part of the overall learning architecture as it's latent feature vectors(V) are used as weights learned for the conversion of the input features to the embedding features which serve as the input to the hidden layers. A total of 3 dense layers with ReLU activations was used in building up the hidden layers of the network. A dropout of 0.5 was also used in all these layers to prevent overfitting. The output layer of the network; however, employs a sigmoid activation ultimately for click prediction. Additionally, the latent dimension of the FM is kept 10. For training the model we use the Adam optimizer which is an adaptive learning algorithm and a combination of Stochastic Gradient descent and RMS prop. The loss function that is used is the categorical cross entropy along with AUC as the evaluation metric.

3. **Deep Interest Network**: The deep interest model employs an architecture a lot similar to the deep learning component of the deepfm model. It also consists of an embedding layer followed by some hidden layers and a softmax output layer; however, it uses an adaptive activation function called DICE, instead of the more common ReLU, allowing it to adaptively calculate the user interest representation vector while taking into account their past interactions with respect to a given ad and ads similar to it which drastically improves the performance of the model. The embedding layer serves the purpose of generating lower dimensional data of a fixed length to efficiently deal with the sparsity of the dataset. Furthermore, L2 regularization is applied to handle the case of overfitting the model to the data. Mini Batch Gradient descent is used to train the model and AUC is again used as the evaluation metric.

6 Implementation Plan

The plan for implementation based on the proposed methodology is formulated as a milestone approach with associated tentative deadlines.

Milestone	Objective
M1	Reading all the relevant research papers regarding already made algorithms for the CTR Prediction.
M2	Read about all the concepts related to the topic and the research papers.
M3	Pre-processing the Criteo Labs dataset.
M4	Construct different models based on the dataset and compare the results.
M5	Final optimizations, cleaning code and making the project submission ready.
M6	Final Documentation and evaluation preparation.

7 REQUIREMENTS

7.1 Hardware Requirements

- Any machine that supports the Google Chrome web browser
- GPU for training and testing the machine learning model

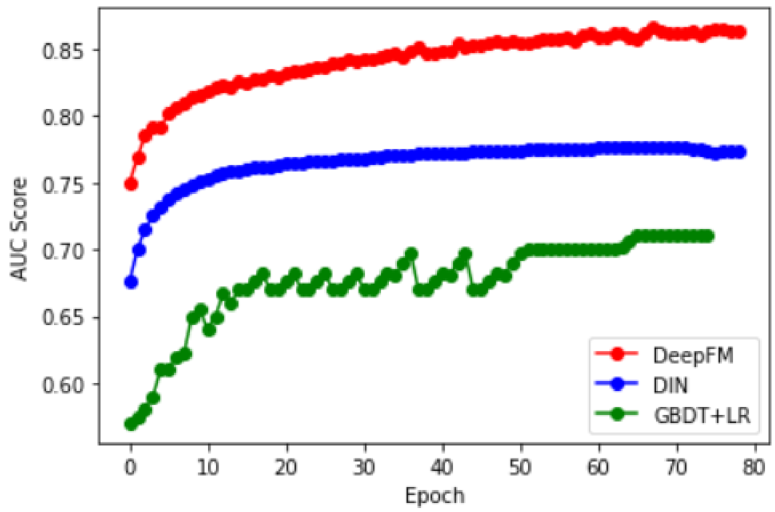
7.2 Software Requirements

- Google Chrome
- VM with more than 30 GBs of memory and google collaboratory
- Python 3.6
- Tensorflow v1.15 with keras API
- Pandas, Numpy, Sklearn, pickle, python libraries
- matplotlib and seaborn libraries for plotting purposes
- Spyder IDE or Jupyter Notebook

8 Results

After training the models we find the deepfm and deep interest network models easily outclass the gbdt + lr ensemble model and although the deep interest network is almost at par with the deepfm model on the training set, it outperforms the deepfm model on the test set. Both are trained for a total of 80 epochs on the dataset reaching an AUC score of over 0.8 for the deepfm and almost 0.8 for the DIN which is in stark contrast to the AUC score obtained by the gbdt+lr model which is almost 0.56. The AUC scores obtained from the models on the training and test sets can be summarised by the table below:

Model	Training Set AUC Score	Test Set AUC Score
Gbdt+lr	0.7100	0.5560
DIN	0.7734	0.7810
DeepFM	0.8672	0.8650



9 Conclusion

Hence, in this paper we have discussed about various methods that have been implemented in the past which were used to determine the CTR after which we have discussed methods that we had implemented and the results obtained in them. This CTR value varies from advertisement to advertisement and also from user to user. It is very useful to determine what type of ads to display to the user and where to place them on the screen so as to achieve maximum probability of user clicking on that advertisement.

References

[1] Avila Clemenshia P., Vijaya M. S., PSGR Krishnammal College for Women Coimbatore Click Through Rate Prediction for Display Advertisement 2016

[2] Guorui Zhou, Chengru Song, Xiaoqiang Zhu Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, Kun Gai Deep Interest Network for Click-Through Rate Prediction Alibaba Group September 2018

[3] Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, Ralf Herbrich, Stuart Bowers, Joaquin Quiñonero Candela Practical Lessons from Predicting Clicks on Ads at Facebook

[4] Kun Gai1 Xiaoqiang Zhu1, Han Li1, Kai Liu2†, Zhe Wang3 Alibaba Inc Yao Saint, Zhe Wei, Ying Ren, Meng Chang. Learning Piece-wise Linear Models from Large Scale Data for Ad Click Prediction 2017.

[5] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, Xiuqiang He DeepFM: A Factorization-Machine based Neural Network for CTR Prediction