Click Through Rate Prediction of an Advertisement According to User's Interest

Aditya Bahukhandi IIT2017142
Arsh Panghal IIT2017048
Utkarsh Pankaj IIT2017082

Problem Statement

Calculation of Click Through Rate Prediction of an Advertisement according to the User's interest is our main objective in the project.

It is also of interest to see how the traditional machine learning models compare against newer deep learning based models in terms of speed and accuracy on a large dataset.

Click Through Rate

- Click Through Rate is a parameter through which the likelihood of an advertisement being clicked is measured and depends on the placement of an advertisement at a particular position in a particular site and the users' past history.
- More the percentage of CTR, more favorable is the outcome of the advertisement being clicked.

Motivation

- 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 thead exchange, advertisers and publishers.

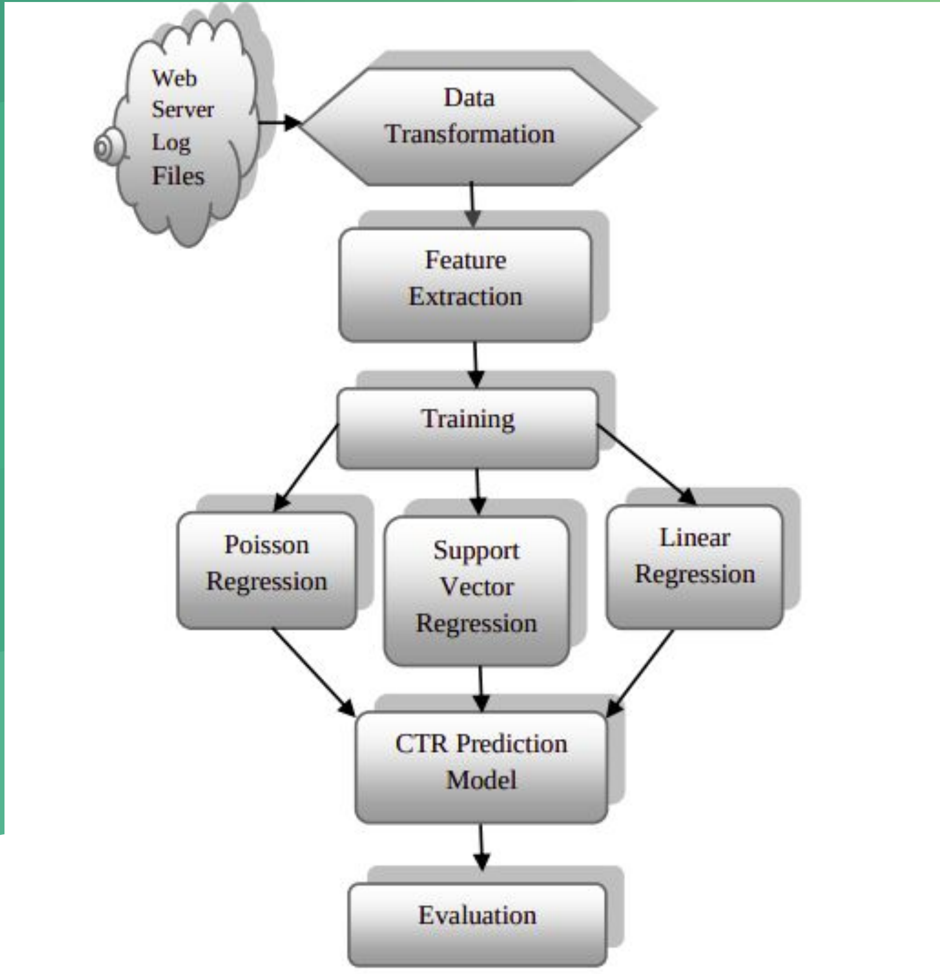
- Therefore it is important to predict the CTR metric for formulating an efficient ad placement strategy.

Literature Review

- The paper by Avila Clemenshia P. and Vijaya M.S. [1] 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.

Figure 1 : Proposed model for ML techniques



- 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 alinear 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.

- 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.

Proposed Methodology

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.

2. Dataset Preparation

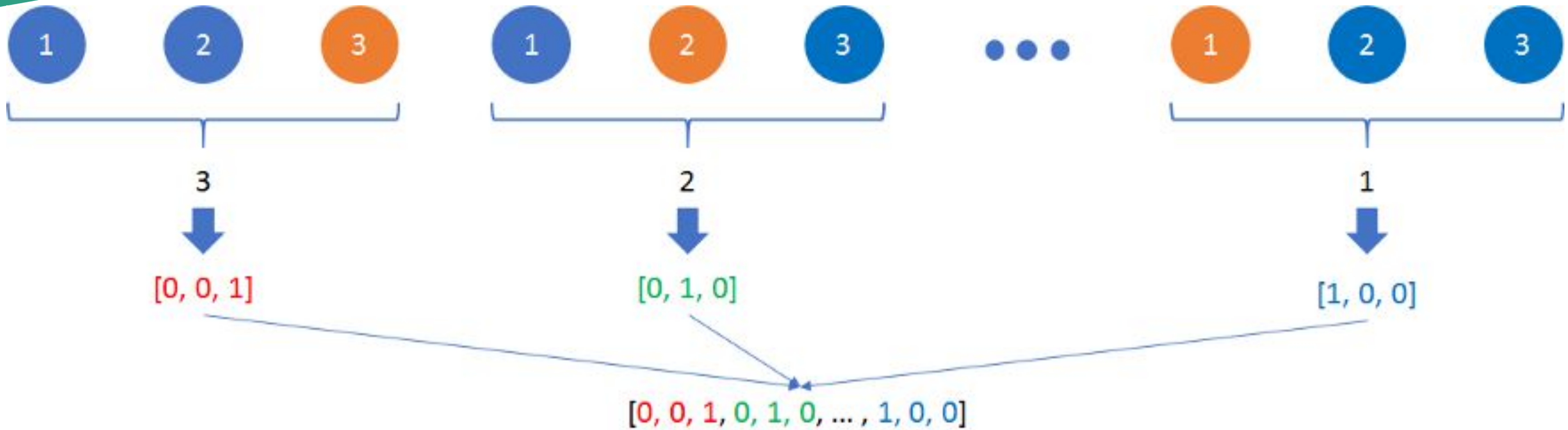
1. First of all the required categorical variables of the review and metadata dataset namely as in, categories and reviewerID are mapped to unique numeric values which helps in easier processing later on in the training phase.
2. The reviews of every user is also sorted according to the timestamp.
3. Next up, the training and test sets are created using a 90:10 split.
4. The ith entry for every user in the training set are his reviewer id, timestamp, his history of reviewed and browsed i-1 products and the ith product he will review if label is put 1 which is otherwise the book he won't review if label is 0.
5. Entries for the test set are also the same with the exception of the label which is not included.

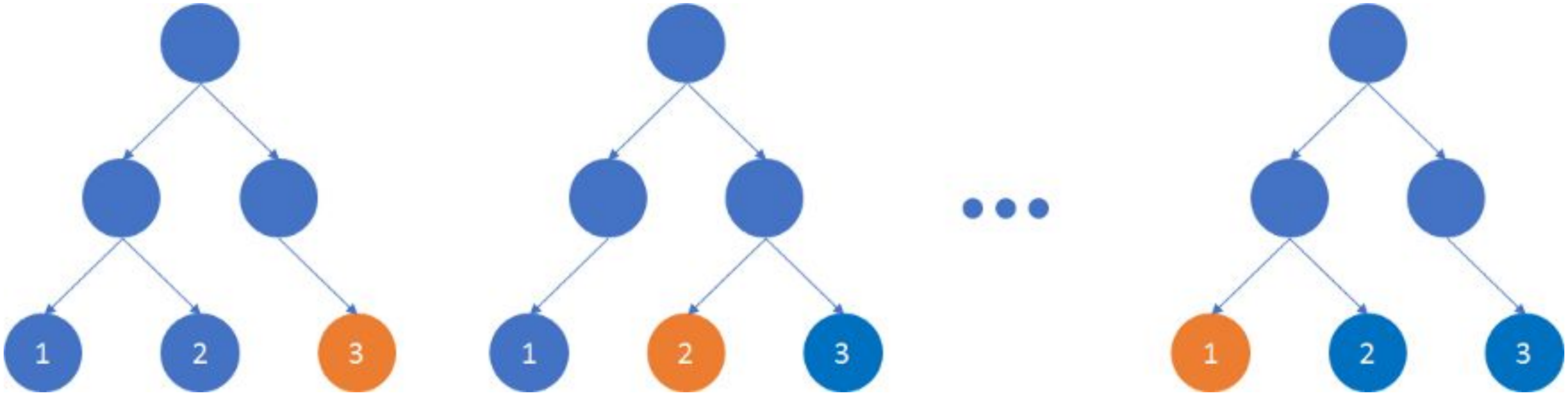
After these respective steps, the data is well suited for training the models.

Model Training

Model 1 : gbdt+lr

- 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.


1
2
3
1
2
3
1
2
3
3
2
1
[0, 0, 1]
[0, 1, 0]
[1, 0, 0]
[0, 0, 1, 0, 1, 0, ... , 1, 0, 0]


1
2
3
1
2
3
1
2
3

Model 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.

Model 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.

Requirements

- Hardware Requirements
  1. Any machine that supports the Google Chrome web browser
  2. GPU for training and testing the machine learning and deep learning models
- Software Requirements
  1. Google Chrome with access to Google Colab
  2. A VM with Ubuntu 18.04 image and more than 30GB memory
  3. Python 3.6
  4. Tensorflow v1.15 with keras API, lightgbm
  5. Numpy, Sklearn python libraries
  6. matplotlib library for plotting purposes
  7. Jupyter Notebook or Spyder IDE

Implementation Plan

Before MidSem

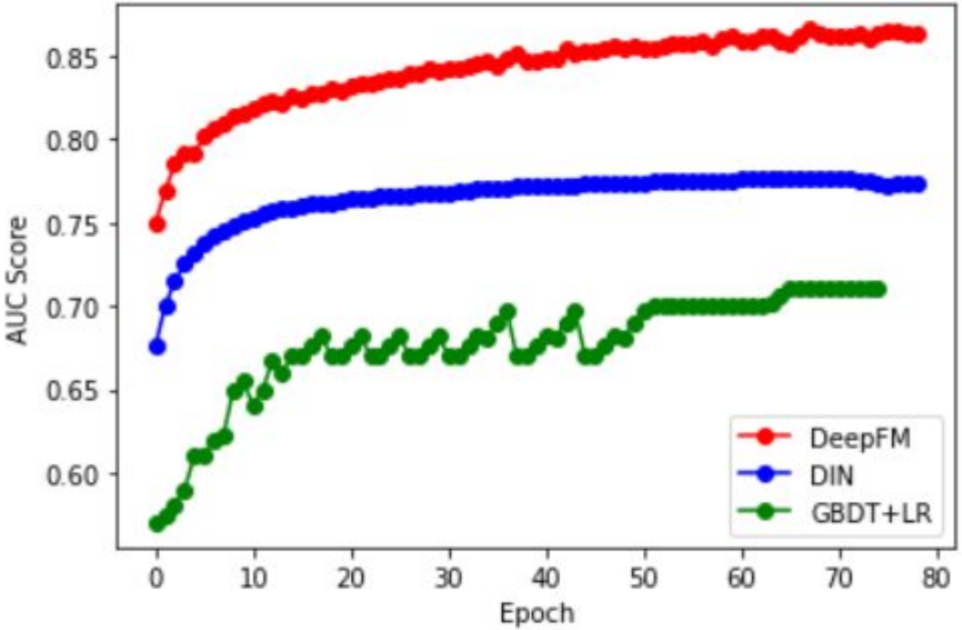
Milestone	Objective
M1	Reading all the relevant research papers regarding already made algorithms for the CTR Prediction
M2	Read about all the concepts related to the topic and the research papers
M3	Pre-processing the Amazon dataset

Milestone	Objective
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.

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 both the training set and the test sets, but we see that the deepfm model is better.

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



AUC Score
0.85
0.80
0.75
0.70
0.65
0.60
0
10
20
30
40
50
60
70
80
Epoch
DeepFM
DIN
GBDT+LR

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, XiaoMa, Yanghui Yan, Junqi Jin, Han Li, Kun Gai Deep Interest Networkfor 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 Quĩ noneroCandela Practical Lessons from Predicting Clicks on Ads at Facebook

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

Image References

Fig [1] : Clemesia, P. Avila (2019). *Click Through Rate Prediction for Display Advertisement* [Online image]. Proposed Model. https://www.ijcaonline.org/research/volume136/number1/clemenshia-2016-ijca-908332.pdf