Click Through Rate Prediction

Mehdi Farhangian

Online advertising is a multi-billion dollar business producing most of the revenue for search engines. As an online advertiser, a key question to ask is how many customers are going to respond to your offer. One area drawing recent attention among both researchers and practitioners is prediction of Click Through Rate. Being able to predict this rate means being able to influence ranking, placement, filtering and pricing of ads. This can translate into improved sales, revenue, and customer engagement to meet your business targets.

The commonest form of internet advertising is the "pay per click" model with Cost Per Click (CPC) billing whereby advertisers are charged for each click on their ad. The strategies advertisers choose depends on the likelihood that an ad with a specific feature is clicked. An advertisement with the right features substantially increases the probability that a customer will see and click the ad. In addition, showing a relevant ad improves the satisfaction and overall experience of the customer. As a result, it is critical for success to have an accurate model that helps us predict the click through rate and also enables us to run if-then scenarios that discover the best features for a particular ad.

In this article we provide some guidelines for developing a model to predict click-through rate for ads without prior information. In addition, at the end of article we look at the respective model errors, considered with and without demographic features.

Predicting click through rate

Whenever an ad is displayed, there is a chance it will be viewed which is highly related to the position of the ad. Ads further down on the page have less chance to be viewed and consequently to be clicked. Thus, we assume that the probability of an ad being viewed depends on the probability that it is viewed multiplied by the probability that it is clicked. We also simplify our formula by assuming that the probability of a view (impression) is independent of the probability of a click.

$$p(click|ad, position) = p(click|impression) * p(impression|possition)$$

Based on this formula, we can predict the probability of a click on an ad based on its position. It is easy to predict the click through rate of ads that are frequently displayed, but subtle methods are required to predict the number of clicks for new ads. Historical data tells us the click through rate of frequently repeated ads. For instance, if an ad is viewed 1000 times and clicked 20 times the click through rate is 0.02. However, this estimation might vary and some 'Bayesianists' might disagree with us that this number is reliable. We can leave Bayesianists and Frequentists to their arguments while we simply estimate numbers as binomial MLE (maximum likelihood estimation), clicks / impressions, focusing in this article on estimation of newly created ads with no prior information. Changing the ad position significantly decreases the probability that a user will click on an ad.

A framework for Click Through Rate prediction

The aim here is to provide some guidelines to develop a model that improves an advertising system by predicting click through rate accurately. We look to predict this for ads with no prior information. Finding the right features is the most important part of prediction and should take up most of a data scientist's time.

The first part of building this model is gathering features. Each ad contains information such as, landing page, keywords, title, body, display URL, ad match type clicks, views and so on. For all the ads in our data set, we have the click through rate (CTR) that is equal to number of the clicks divided by the number of impressions. After gathering the data, we divide our data set into training set and test set. In this case, we choose 70% of data in the training set and 30% in the test set.

A logistic regression is ideal for problems where the dependent value is either 0 or 1. Using logistic regression, we predict CTR given a set of features:

$CTR = 1/(1 + e^{-z})$

Where

$Z = \sum w_i f_i(ad)$

Where $f_i(ad)$is the value of the ith feature for the ad and W_i is the learned weight for the feature. We should allocate the largest part of our time doing model development to feature selection. This is an ongoing task always open to improvement. A dtailed explaination of this approach with some examples can be seen in [this article] (https://www.microsoft.com/en-us/research/publication/predicting-clicks-estimating-the-click-through-rate-for-new-ads/).Also, I have provided more information here.

Based on my experience, the keyword alone is not adequate to predict the click-through rate. So, we need to find out the features that help us better estimate click through rate. To do this, we recommend exploration of four categories: Appearance, Attention Capture, Landing Page Quality and Relevance described as follows:

- Appearance explains if the ad is aesthetically pleasing or not. We used the following features from this category:

  - Number of words in title
  - Number of words in body
  - Contains Good Capitalisation or not
  - Contains too much punctuation or not

- Attention Capture : Explains to what extent the ad attracts the attention of its target audience and if the following features are used:

  - Whether title contains an action word or not
  - Whether the body contains an action word or not
  - Does the ad contain a number (about price or discount)

- Reputation covers the reputation of the brand. To measure quality of the ad, measure the following features:

  - Does the display URL end with .com
  - How long is the display URL
  - How many segments are in the Display URL
  - Does it contain a dash or number

- Landing Page Quality namely the page viewed by clicking the ad and, we assume, the quality of the landing page affects the decision of returning users to click or not. The selected features include:

  - Does the page contain flash?
  - What fraction of the page is covered with an image?
  - Is it covered with ads

- Relevance explains how the ad and search query are relevant and consequently the following features are considered:

  - What fraction of keywords appear in the title
  - What fraction of keywords appear in the body
  - Personalizing Click Through Rate prediction

Adding the above-mentioned features, increases the accuracy of click through prediction. However, one of the shortcomings of common models is ignoring the fact that users with different demographic backgrounds have different opinions about different ads. Hence a level of complexity must be added to the current models to cover these differences and provide a personalised prediction. This personalisation not only benefits the users by providing ads that are more relevant to them, but also helps the advertisers by connecting with users who are more engaged with their ads.
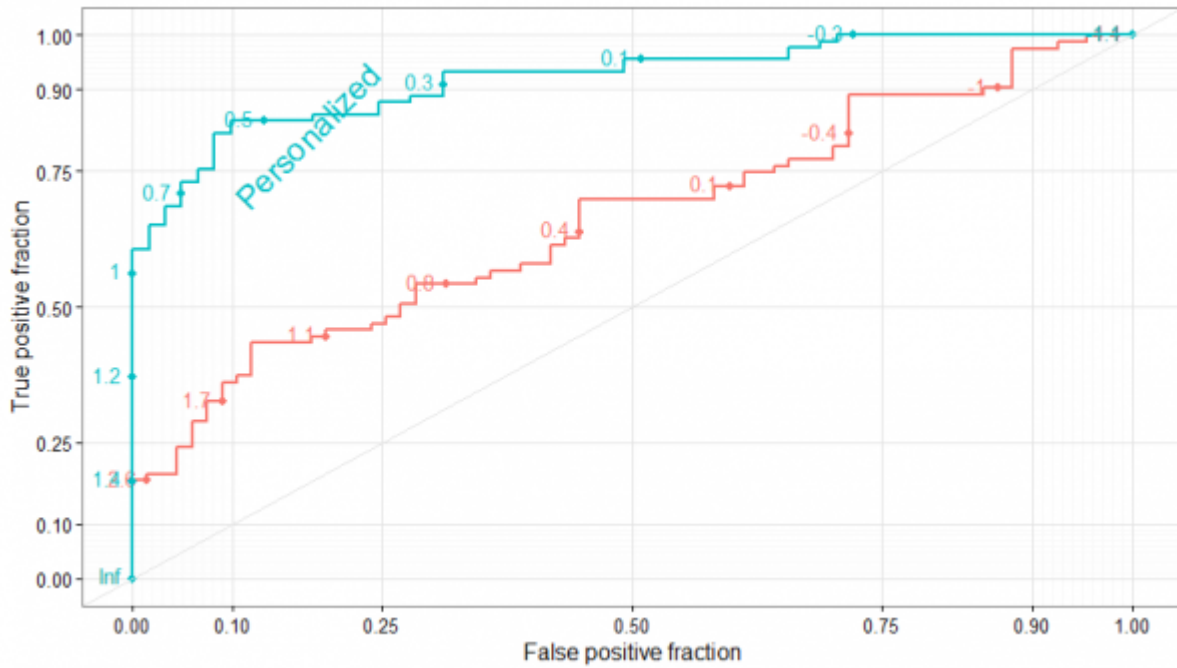
A user-specific feature captures the individual users' interactions and an AdWords API enables us to access this data at a granular level. In order to make our prediction more personalised, we include the following features in the model:

- Gender
- Age
- Location
- Interests
- Parental status

The following graph of sample data compares the accuracy of our model without personalisation features (the red line), with the blue line which represents where personalisation features are included in the model. We can clearly see the accuracy of the the model increases markedly with the addition of personalisation features.