

Click Through Rate Prediction for Display Advertisement

Avila Clemenshia P.
Research scholar
PSGR Krishnammal College for Women
Coimbatore- 641004

Vijaya M. S.
Associate Professor
PSGR Krishnammal College for Women
Coimbatore-641004

ABSTRACT

Computational Advertising is the currently emerging multidimensional statistical modeling sub-discipline in digital advertising industry. Web pages visited per user every day is considerably increasing, resulting in an enormous access to display advertisements (ads). The rate at which the ad is clicked by users is termed as the Click Through Rate (CTR) of an advertisement. This metric facilitates the measurement of the effectiveness of an advertisement. The placement of ads in appropriate location leads to the rise in the CTR value that influences the growth of customer access to advertisement resulting in increased profit rate for the ad exchange, publishers and advertisers. Thus it is imperative to predict the CTR metric in order to formulate an efficient ad placement strategy. This paper proposes a predictive model that generates the click through rate based on different dimensions of ad placement for display advertisements using statistical machine learning regression techniques such as multivariate linear regression (LR), poisson regression (PR) and support vector regression (SVR). The experiment result reports that SVR based click model outperforms in predicting CTR through hyperparameter optimization.

Keywords

ad campaign metrics, click through rate, display advertisements, prediction.

1. INTRODUCTION

Online advertisement in the current marketing industry is a widely considered component that acts as a major resource provider for web users. There is a tremendous focus of advertisement on web search and is a key factor of revenue for business community. Web advertisement is broadly categorized into sponsored search, display advertisements or banner ads, and contextual advertising. In sponsored search, ad is selected based on search query and in context search the ad is selected based on the match with the page content. Selection of ad in display advertisements is decided by the ad exchanges and it is a tedious task. The main goal of display advertisement is to present brand advertisements and messages to the visitors of the site. Publisher's present the ad to the user by displaying the advertisement in their web pages. Click through rate is the percentage at which the user clicks the ad out of the impression. The website or application the user visits impacts the CTR of the ad. The placement of ad plays a vital role in increasing the CTR of an ad. There is an ad server for advertisers, publishers and ad network that involves in administration of the ads and the way it is distributed in web sites. The ad server allocates them in different websites after they receive the ad files.

The ad is displayed to the users when they visit a web page or application. The task of predicting click through rate especially for display advertisement is quite a challenging issue. Various factors affecting the value of the CTR can be viewed in different dimensions. The features that are

generated contribute to evaluate the CTR metric in different disciplines through the analysis of these features. The prediction of advertisement click through rate is an important metric in the online advertising. An impression in the context of web advertisement is the count when an ad is fetched from its source each time. Click or non-click is not taken into account to measure the impression. CTR is a metric that measures the ad campaign performance. The number of times the ad is displayed to users is known as impression. CTR is used to measure the number of clicks advertisers receive on the ads per the number of times it is displayed.

The click-through rate of an advertisement is calculated as the number of clicks on an ad divided by the number of times the ad is shown which is the impression and is expressed as a percentage. This predicted CTR value provides the advertisement agencies with a vision to identify the most viewed advertisements and hence can create ads similar in quality and place them appropriately. This prediction model deals with the identification of rate at which an ad is viewed in the online sites based on the placement and relative ads. Predicting the future response of the ads helps in generating ads with better quality and improves the revenue.

Many researches were carried out on sponsored search where the search query features are taken into account. This research work is proposed for display advertisement focusing on features like the number of clicks, impressions and CTR of the site and application on which the ad is placed, the previous and next ad that has been viewed, the site domain and application domain where the ad has been located.

We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace the content with your own material.

2. LITERATURE SURVEY

Various analogous research works are reviewed and analyzed to understand the nature and circumstances of the work. The purpose has been well studied and the need for proposed work is identified based on the literature survey. Fang Wanga* et al., presented a model [7] that performs the task of Behavioral Targeting which is to predict the users who click the target ads based on the historical data. This was used for presenting the relevant ad to the users. The dataset was downloaded from track2 of the KDD Cup 2012 datasets that are provided by a leading Internet company in China. The value predicted was the CTR of ads in web search engine using the Multiple Criteria Linear Programming Regressions (MCLPR), Support Vector Regression (SVR) and Logistic Regression and they concluded that MCLPR provided better results in performing behavioral targeting tasks.

The authors, Zhipeng Fang et al., generated a Bayesian Network model [16] to determine CTR of new ads. Initially they have developed a keyword Bayesian networks using the

keywords of the ads in certain domain. Then they have formulated an algorithm to generate inferences of probability distribution approximately for the known keywords and rendered the suggestions for predicting the CTR of new ads. The dataset was derived from KDD Cup 2012-Track 2. Ling Yan et al., developed a model [15] from the logs of the advertisements displayed in one of the most famous C2C e-commerce web sites in China, www.taobao.com. The problem of CTR prediction was formulated as a binary classification task and proposed coupled group lasso (CGL), for CTR prediction in display advertising. Hashing, subsampling and distributed learning techniques were utilized to generate the model.

Deepayan Chakrabarti et al., framed a model [6] for Contextual advertising where user experience and revenue shared by the site publisher and the ad network depend on the relevance of the displayed ads to the page content. The model was mapped to standard cosine similarity matching. Haibin Cheng et al., developed the framework [8] for the personalization of click models in sponsored search. They developed user-specific and demographic-based features that reflect the click behavior of individuals and groups. The features were based on observations of search and click behaviors of a large number of users of a commercial search engine. The results demonstrate that the personalized models significantly improve the accuracy of click prediction in sponsored search.

In this paper the authors Thore Graepel et al., rendered the model [14] for web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine. They have proposed a new Bayesian click-through rate (CTR) prediction algorithm based on a probit regression model that has mapped discrete or real-valued input features to probabilities. They have compared it to a calibrated Naïve Bayes algorithm. Kushal Dave et al., provided with a model [9] to predict the Click-Through Rate for rare or new ads. In this model the prediction was done for sponsored search ads. The dataset was collected for the US market and had 12 days search log of a popular search engine. The regression model was derived using gradient boosted decision trees (GBDT) and the click information of semantically related ads. Dawei Yin et al. put forward a model [5] based on the contextual factors for click modeling in sponsored search. Using these contextual factors that are explored they had presented the click model for sponsored search.

From the background study it was perceived that most of the works were based on sponsored search and contextual type of advertisements and the related problems were modeled as binary classification. Also attributes considered for generating models were demographic features, context features and content features. This emphasizes the need of more research on display advertisements by designing and deriving the most effective features paving way for generating the new model.

3. DESIGN - CTR PREDICTION MODEL

Click through rate prediction is a sustainable model that is essential in assessing the performance of display advertisement. The significance of the CTR prediction task in the field of display advertisement led to the proposal of this work with greater reliability. The research is initiated by data acquisition from an ad Server and it contains click log details. The dataset is analyzed to identify the contributing features to predict the CTR on the selected advertisements. The training

dataset is prepared from the raw dataset through data transformation and feature extraction. The CTR models are build by employing regression algorithms namely linear regression, poisson regression and support vector regression. The architectural framework of the proposed model and the various phases involved in this research are illustrated in Figure.1.

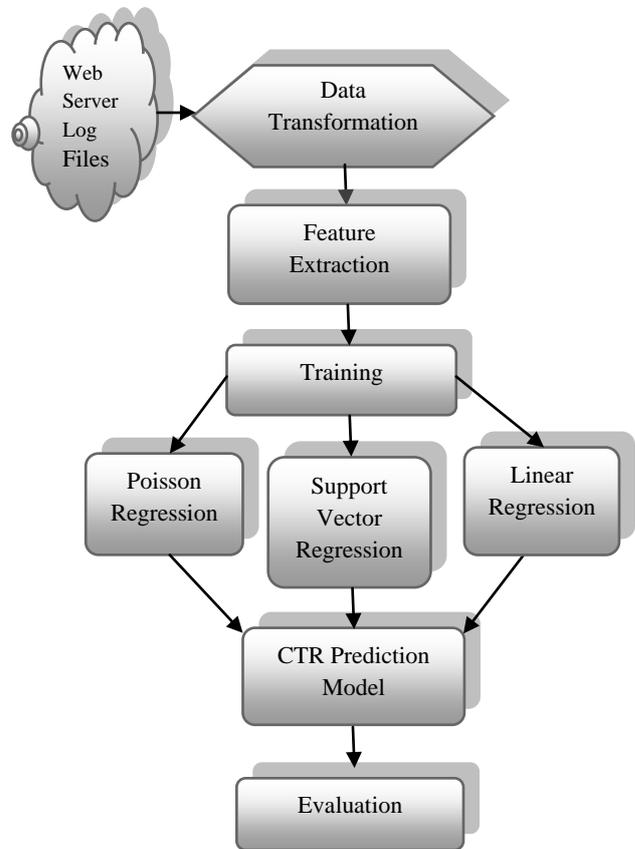


Fig 1: Architecture of proposed CTR prediction model

3.1 Dataset

The dataset¹ is acquired from an ad server of a digital marketing agency. The dataset contains 10 days of click log data ordered chronologically with attributes id, click, hour, banner position, site id, site domain, site category, app id, app domain, app category, device id, device ip, device model, device type, device connection type. The attributes with header C14 to C21 are described as anonymized categorical variables. The dataset contains 40 M instances but not appropriate for building the CTR prediction model. Hence data transformation and feature extraction are performed to attain the maximum prediction accuracy.

3.2 Data Preprocessing

Data Preprocessing is necessary to ensure meaningful data mining results. Initially the data collected for research are well studied to make certain that they are within the span of the problem framed. The irrelevant features such as device id, device ip, device model, device type, device connection type, hour and record id are found to be out of the scope of the anticipated model and hence removed. In order to extract

¹ <https://www.kaggle.com/c/avazu-ctr-prediction>

features for each unique ad which rely on the previous ad, next ad, site_id, app_id, site_domain, app_domain, these attribute are coded from hexadecimal to numeric format. This coding will enable a well organized feature extraction process. The coding involves two steps. In the first step, the one-to-one mapping of the index value to the corresponding attribute value is performed. Then each of the occurrences of the unique value is replaced by the index value in the next step.

3.3 Feature Extraction

The feature extraction phase takes the vital responsibility in building the CTR model and efficiently predicting the CTR metric. Features are extracted from the various dimension of the ad location such as web site, application, site domain, application domain information and neighboring ads like previous ad and next ad in each web page. These components related to an ad in a web page are found to be most contributive features towards the generation of perfect results. The metrics like number of clicks, impression and CTR are computed for the above seven components. These three metrics of seven components forms features of the training dataset for the proposed model and the process of computing the features are described below.

3.3.1 Component 1: Web Site and its Metrics

The website on which the advertisement is placed has to be taken into account. If a site has been visited more often, then it is certain that the ad is displayed to the users many times and this increases the number of times an ad being viewed and clicked. The site CTR highly stimulates the ad CTR. So the number of clicks, impression and the CTR of the site where the ad is placed is contemplated in the feature vector. The site id and the related click attribute in the dataset are processed to derive the number of clicks, impression and CTR on website. This is computed by identifying the websites on which the ad has been placed and evaluating the above mentioned features as described in Table 1.

Consider an ad with id 15001 which is placed in different websites with site id as 86, 89, 90, 149, and 183. The impressions and number of clicks are found from the corresponding click attribute values. CTR is calculated from the formula,

$$CTR = \text{Number of clicks} / \text{Impressions.}$$

Table 1. Features for Website

Site id	Impression	Click	CTR
86	13	7	0.538
89	56	34	0.607
90	9	5	0.555
149	88	60	0.681
183	230	198	0.860
Mean	79.2	60.8	0.648

3.3.2 Component 2: Application and its Metrics

Similarly the application in which the ad is placed equally contributes to the prediction of CTR. Higher the application CTR greater is the CTR to be predicted. The three metrics based on the applications are thus obtained for the ad placed. The application id and the related click attribute in the initial

dataset are processed to derive the number of clicks, impression and CTR based on the application and computed.

3.3.3 Component 3: Site Domain and its Metrics

The several website under the same domain invokes frequent visitors. The advertisement that is displayed on these websites influences the repeated guests of a domain to click on the ads. The number of clicks, impression and the CTR of the site domain where the ad is placed is included in the feature vector on this motive. The site domain id and the appropriate click attribute in the preprocessed dataset are computed and the number of clicks, impression and CTR are extracted.

3.3.4 Component 4: Application Domain and its Metrics

The various applications under the same domain have regular visitors and the ads are displayed to them more frequently. The users are encouraged to click on the delivered ads. The number of clicks, impression and the CTR of the application domain where the ad is placed is included in the feature vector for the prediction model. The application domain id and the corresponding click attribute are considered for generating these features.

3.3.5 Component 5: Previous Ad and its Metrics

In display advertisement, commercial message is conveyed and the specific traits are used to target the audience. The ads that the customers encounter during web browsing attract and motivate them to click on the ad and check on the details. The user who tends to click the ad has the high probability of viewing the neighboring ads. The ad placed in previous location to the selected ad is referred to as previous ad. The previous ad varies for every different website in which the ad is placed. Also there are more than one previous ad for the ad under consideration. The respective clicks, impression and CTR of the previous ad have serious impact on the selected ad. The previous ad metrics are thus generated by processing the click attribute of the source dataset with respect to the previous id. The mean values for number of clicks, impression and CTR of previous ads are derived forming the features for analysis.

3.3.6 Component 6: Next Ad and its Metrics

The customers accessing an ad during the regular browsing activity are mostly persuaded by the ad that is located next to the ad in focus. This might lead them to click on the ad and proceed for that ad service provided. This factor highly affects the increase in CTR metric which motivates the metrics namely number of clicks, impression and CTR of the next ad to be included as features for the proposed model. There are several ads that are in the next locality to the ad under consideration in different web sites. The click attribute of the source dataset is utilized with respect to the next ads and the mean values for number of clicks, impression and CTR are computed.

3.3.7 Component 7: Current Ad and its Metrics

The unique ad from the transformed dataset are found to be distributed in different setting and so grouped to obtain corresponding number of clicks, impression and the CTR which is the response variable is generated using the formula,

$$CTR = \frac{\text{Clicks}}{\text{Impressions}} \times 100\%$$

For instance, Current ad metrics for Ad id = 375,
Impression = 38

$$\begin{aligned} \text{Clicks on Current ad} &= 7 \\ \text{CTR} &= (7 / 38) * 100 = 18.4 \% \end{aligned}$$

The MATLAB code has been used to create the above 21 features and the training dataset with 394 instances has been developed. The process for computing the three metrics for previous ad is shown in Table 2.

Table 2. Features for Previous Ad

Previous ad id	Impression	Click	CTR
4578	1	0	0
6616	1	0	0
12000	1	0	0
15699	1	0	0
15701	3	2	0.667
15703	2	0	0
15704	4	1	0.25
15705	1	0	0
15706	2	0	0
15707	2	1	0.5
15708	2	1	0.5
16071	1	0	0
16208	1	0	0
17653	1	1	1
17654	1	0	0
17914	1	1	1
18987	1	0	0
18993	3	0	0
19771	1	0	0
19772	1	0	0
20153	1	0	0
20352	1	1	1
20366	2	0	0
20596	1	0	0
20634	1	1	1
20751	1	0	0
Total	38	9	5.9167

3.4 Model Generation

The CTR metric of the targeted ad is taken as the response variable and the three metrics of the seven significant components of the targeted ad are treated as independent variables for generating the prediction models using regression. The three variants of regression namely Linear Regression, Support vector Regression and Poisson Regression algorithms are implemented for building the models.

3.4.1 Multivariate Linear Regression

The type of Linear Regression that involves two or more predictors (x_1, x_2, \dots, x_n) is termed as multivariate linear regression. The regression line is visualized in n-dimensional space when more than one predictor is used. The line can be computed by replacing the single-predictor by the parameters for each of the predictors.

$$y = \theta_1 + \theta_2x_1 + \theta_3x_2 + \dots + \theta_nx_{n-1} + e$$

In this work the extracted features are the independent variables except for the CTR value of the current advertisement which is the dependent variable. The implementation is done in the R tool and the model is generated.

In linear regression, the regression line has a smallest distance between itself and each data point. The regression line does not touch some data points, but not others. The distances between the data points and the regression line are called error terms. There are many uncontrollable factors in the business world that leads to the occurrences of error. The least-squares method is used for finding the line of best fit.

The relationship that exists in a single predictor variable and the response variable where all the other predictor variables in the model are fixed is identified using a fitted linear regression model.

3.4.2 Support Vector Regression

Support vector regression is the regression analysis which is derived from the natural extension of large margin kernel methods used in classification. It retains all the properties namely duality, sparseness, kernel and convexity that characterize maximal margin algorithms of support vector machines. It has become a powerful technique for predictive data analysis with many applications such as to model disease inception and level that are relying on the various behaviors and environmental condition in biological contexts. It has also been used for areas such as drug discovery, civil engineering, sunspot frequency prediction, image tracking, image compression, control system and system identification problems.

It is being vastly utilized in the area of metamodeling. These metamodeling techniques have been developed in different variation down the road to reduce the cost of computation for analysis and simulation of computer based codes. Metamodelling is the process of building a model of a model to provide a replacement for a computer code that is has volumes of computations. Metamodelling techniques such as response surface methodology, radial basis functions, are widely adopted also multivariate adaptive regression splines and kriging are commonly used.

\mathcal{E} -Insensitive loss function and regression

Regression is the problem of finding out a function to approximately map from an input domain based on training sample to the real numbers. This denotes the difference between the output of hypothesis and its training value as the output residual, an indication of the accuracy of the fit at this point. To decide on the importance of this accuracy is essential, as small residuals are very inevitable and there may be need to avoid large ones. This measure is determined by the loss function. The choice of different loss function would result in a different overall regression strategy.

SVM regression performs linear regression in feature space that is of high-dimension using \mathcal{E} -insensitive loss and also

enables to reduce model complexity by minimizing $\| \omega \|^2$. This can be done by using non-negative slack variables $\xi_i, \xi_i^* \ i = 1, \dots, n$, in order to measure the variation of training samples beyond \mathcal{E} -insensitive zone.

It is well known that SVM generalization performance or estimation accuracy depends on a good setting of C and \mathcal{E} that are the meta-parameters parameters and the kernel parameters respectively. SVM model complexity and consequently its generalization performance depends on all three parameters which complicates the problem of optimal parameter selection. Selecting a particular kernel and its function parameters type is generally based on application-domain knowledge and this also reveals how the input (\mathbf{x}) values of the training data are distributed.

The balance between the incompatible model complexity and the degree to which deviations larger than \mathcal{E} that is determined by parameter C which can be tolerated in optimization formulation for example, the only objective is to minimize the empirical risk if C is too large or close to infinity irrespective to model complexity part in formulating optimization[13].

3.4.3 Poisson Regression

Poisson Regression is a type of generalized linear models where the random component is specified by the way in which the response variable is Poisson distributed and it is a count. The log-linear model is equivalent to poisson regression model if all explanatory variables are discrete. The response variable in the proposed model is the CTR metric. This is suitable to the proposed work as the response variable is the click through rate. The explanatory variables can either be continuous or can be combined with categorical variables. The independent variables are continuous variables in this case of the click prediction. In poisson regression the explanatory variables can also be categorical. The poisson distribution is discrete and positive. Poisson regression is the model that fits the hypothesis formed by the underlying random process that creates at a rate determined by other variables in the model a small number of events. The basic Poisson regression model is given by

$$\ln(\lambda_i) = \sum_{k=0}^n \beta_k x_{ik}$$

The above is the regression equation that relates the natural logarithm of the mean or expected number of events for the case i which is $\ln(\lambda_i)$ to the product of sum of the products of each explanatory variable x_{ik} , and a regression coefficient, β_k .

3.4.4 Evaluation measures

Correlation is a standardized number that describes how closely the two variables are related. The correlation coefficient always lies between -1 and 1. A coefficient of 1 represents perfect positive correlation, 0 represents no correlation, and -1 represents perfect negative correlation. Correlation coefficient measures the statistical correlation between the actual values(a_i) and predicted values(p_i) using the formula,

$$\text{Correlation Coefficient} = \frac{S_{PA}}{\sqrt{S_P S_A}}$$

$$\text{where } S_{PA} = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{n-1}, S_P = \frac{\sum_i (p_i - \bar{p})^2}{n-1}, \text{ and } S_A = \frac{\sum_i (a_i - \bar{a})^2}{n-1}$$

Error term is the difference between the regression line and actual data points used to construct the line and root mean squared error is evaluated using,

$$RMSE = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}}$$

Higher Correlation and least error values are expected for the best prediction accuracy.

4. EXPERIMENT AND RESULTS

Experiments are carried out in R environment to build the CTR models for display advertisement by implementing machine learning technique. The dataset with 1,31,070 instances is collected from the ad server of a marketing agency. Preprocessing and feature extraction processes have been carried out as described on section 3. Linear regression, poisson regression and support vector regression are the three algorithms that are trained using the training dataset with 394 instances for model generation and the models are evaluated for their performance. The evaluation measures such as the root mean squared error and correlation are used for performance evaluation.

Table 3. Performance of CTR Models

Models	RMSE	Correlation Coefficient
LR CTR	0.0146	0.7255
PR CTR	0.311	0.7175
SVR CTR	0.013	0.7933

There are two important inferences made from the above results of correlation and RMSE. As the value of the correlation coefficient is a positive number it is the indication that they are positively related and because the resulting value is relatively far from indicating no correlation, the strength of the correlation is strong. The root mean squared error is less. These two factors imply that the model is efficient.

The correlation coefficient between the actual values and predicted values of CTR for the three regression models are analyzed along with RMSE and the comparative results are tabulated in Table 3 and illustrated in Figure 2.

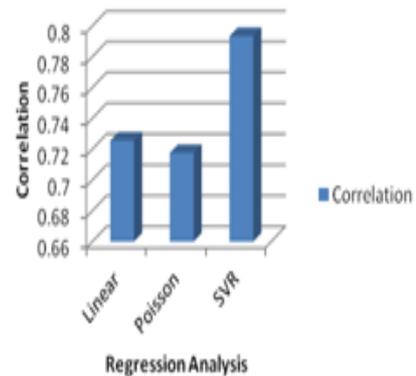


Fig 2: Performance evaluation of CTR models

It is evident from the estimates that the model generated using SVR renders better accurate results with reduced error and maximum correlation coefficient.

The hyperparameter optimization is performed on the SVR model for model selection which is the problem of choosing a set of hyperparameters for a learning algorithm and the goal is to optimize a measure of the algorithm's performance on an

independent data set. Cross-validation is used to estimate this generalization performance. The loss function on the training set alone is optimized. In effect, learning algorithms learn the parameters that model or reconstruct their inputs, while hyperparameter optimization is to ensure the model does not over fit its data by tuning.

The SVR based CTR prediction model is optimized by operating on a chosen range of parameters. Here the chosen range of parameters ϵ is 0 to 1 in the scale of 0.1 and cost in the range $2^{\wedge} (2 : 9)$. Training is performed for the different couples of ϵ and cost, and the best model is chosen. The error value and correlation for the best model derived is evaluated and the result is illustrated in Figure 3 & Table 4.

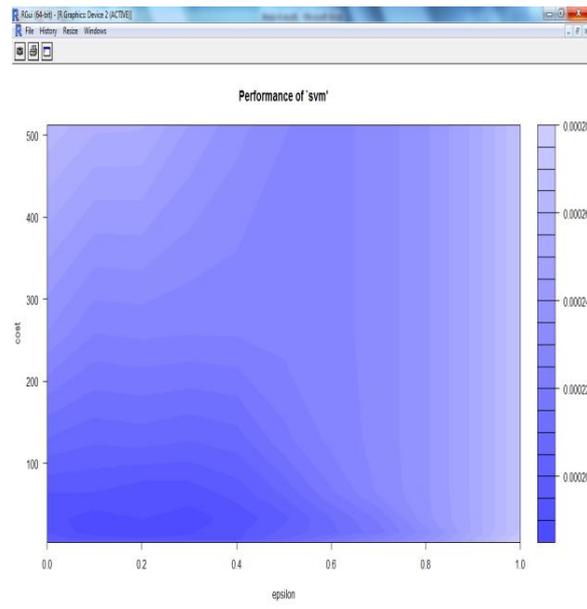


Fig 3: Optimization Performance

Table 4. Performance of Optimized Models

Models	RMSE	Correlation Coefficient
LR CTR	0.0146	0.7255
PR CTR	0.311	0.7175
SVR CTR	0.013	0.7933
Optimized SVR CTR	0.005	0.9086

5. CONCLUSION

This research work demonstrates the modeling and implementation of click through rate prediction for display advertisement. The contributive features are identified based on various component of the targeted ad in a web page and extracted to develop the model. Supervised machine learning paradigms namely linear regression, poisson regression and support vector regression are adopted to construct the models. The performance of the regression models are evaluated based on the error deviation and correlation between the actual and predicted values. It is observed from the experimental results that support vector regression model is more efficient than the other two models in predicting click through rate. The SVR based model is then optimized for best performance. As the scope for future research, the models can be generated using large dataset, more related features can be identified and the

models that were built can be integrated to create ensemble model so as to improve the prediction result.

6. REFERENCES

- [1] Abirami, R and Vijaya, M. S. 2012. *Stock Price Prediction using Support vector regression* CCIS 269, pp 588 – 597@ Springer Verlag Berlin Heidelberg 2012.
- [2] Agarwal, D., Agrawal, R., Khanna, R.. and Kota, N. 2010. *Estimating rates of rare events with multiple hierarchies through scalable log-linear models*. In Proceedings of the ACM SIGKDD Knowledge Discovery and Data Mining, pages 213–222.
- [3] Agarwal, D., Agrawal, R., Khanna, R.. and Kota, N. 2010. *Estimating rates of rare events with multiple hierarchies through scalable log-linear models*. In Proceedings of the ACM SIGKDD Knowledge Discovery and Data Mining, pages 213–222.
- [4] Agarwal, D., Chen, B.-C. and Elango, P. 2009. *Spatio-temporal models for estimating click-through rate*. In Proceedings of the International World WideWeb Conference, pages 21–30, New York.
- [5] Dawei Yin, Shike Mei, Bin Cao, Jian-Tao Sun Brian and Davison, D. 2014. *Exploiting Contextual Factors for Click Modeling in Sponsored Search*. ACM 978-1-4503-2351-2/14/02.
- [6] Deepayan Chakrabarti, Deepak Agarwal and Vanja Josifovski. 2014. *Contextual Advertising by Combining Relevance with Click Feedback*. Proceedings of the 31 st International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32, 2008, Beijing, China. ACM 978-1-60558-085-2/08/04.
- [7] Fang Wang*, Warawut Suphamitmongkola and Bo Wang. 2013. *Advertisement Click-Through Rate Prediction using Multiple Criteria Linear Programming Regression Model*. Elsevier B.V, Procedia Computer Science 17 p. 803 – 811.
- [8] Haibin Cheng and Erick Cantú-Paz. 2010. *Personalized Click Prediction in Sponsored Search*. ACM 978-1-60558-889-6/10/02.
- [9] Kushal Dave and Vasudeva Varma. 2010. *Predicting the Click-Through Rate for Rare/New Ads*. Report No: IIIT/TR/2010/15, Centre for Search and Information Extraction Lab International Institute of Information Technology.
- [10] Menon, A. K., Chitrapura, K.-P., Garg, S., Agarwal, D. and Kota, N. 2011. *Response prediction using collaborative filtering with hierarchies and side-information*. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 141–149.
- [11] Olivier Chapelle and Ya Zhang, 2009. *A dynamic bayesian network click model for web search ranking*. WWW '09: Proceedings of the 18th international conference on World wide web, pages 1–10, New York, NY, USA. ACM.
- [12] Richardson, M., Ewa Dominowska and Robert Ragno. 2007. *Predicting clicks: Estimating the click-through rate for new ad*. In Proceedings of the International World Wide Web Conference, pages 521–530.

- [13] Soman, K. P., Loganathan, R. and Ajay, V., *Machine Learning with SVM and other kernel methods*. Prentice Hall, 2009.
- [14] Thore Graepel, Joaquin Quinero Candela, Thomas Borchert and Ralf Herbrich. 2010. *Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine*. Proceedings of the 27th International Conference on Machine Learning, ICML 2010.
- [15] Ling Yan, Wu-Jun Li, Gui-Rong Xue and Dingyi Han. 2014. *Coupled Group Lasso for Web-Scale CTR Prediction in Display Advertising*. Proceedings of the 31st International Conference on Machine Learning, Beijing, China, JMLR: W&CP volume 32.
- [16] Zhipeng Fang, Kun Yue, Jixian Zhang, Dehai Zhang and Weiyi Liu. 2014. *Predicting Click-Through Rates of New Advertisements Based on the Bayesian Network*. Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2014, Article ID 818203, 9 pages.