Advertising Coverage Prediction Based on grp-reach Dual Parameter Model

Ting Jiang[1, #], Siben Li[2, #], Yujia Chen[3, #, *]

[1]School of Computer and Cyber Sciences, Communication University of China, China, 100024

[2]School of Data Science and Media Intelligence, Communication University of China, China, 100024

[3]School of Economics and Management, Communication University of China, China, 100024

*Corresponding author: chenyujia@cuc.edu.cn

#These authors contributed equally.

Abstract. In today's era, the intelligent development trend of media is unstoppable. The rich media scattered the users' viewing media contacts, leading to the complexity of media advertising exposure and related effect evaluation and prediction. In this paper, this study use as simple a model as possible to predict the advertising coverage in the case of cross media platform data. On the basis of data processing, this paper proposes a prediction model of advertising coverage —— grp-reach dual parameter model. This model can better predict the advertising coverage. Compared with other traditional models, the goodness of fit increases by 0.08 on average, which is helpful to predict users' media consumption behavior. The model can also infer the investment required to increase another reach percentage point when the advertising cost is known, which can provide new ideas and strategies for advertising in sub media delivery, and provide decision-making reference for advertising delivery of different advertising companies.

Keywords: Advertising Coverage, Dual Parameter Model, L2 Regularization.

1. Introduction

Advertising exists in the daily life of the public in a variety of forms. As an important tool of marketing, it often imperceptibly affects the consumer behavior of users. The degree to which advertising media can grasp users' attention is an important criterion for evaluating the success or failure of advertising [1-2]. With the rapid development of the advertising media industry, the media consumption behavior of users has important reference significance for marketing advertising and media content dissemination.

This paper mainly solves the prediction problem of user media consumption behavior in complex environment. Reach-curve can reveal the relationship between the exposure number of an advertisement and the number of people watching the advertisement to a certain extent, and can measure the advertising effect on mobile and ott. However, in practice, it is difficult to obtain the complete curve of reach-curve. In this paper, we establish the reach-curve fitting model of sub media, and give the prediction function of advertising coverage.

In this paper, considering the relationship between grp and reach and the situation of multiple media platforms, a dual parameter model is established according to Poisson distribution and logit inverse function, and L2 regularization is introduced to solve the model using gradient descent method.

2. Grp-reach Dual Parameter Model

2.1. Data

The data set of this study provides the total exposure ratings grp and user coverage reach of different advertisements on different platforms and media. The data is preprocessed.

2.2. Model preparation

2.2.1 Functional Relationship Analysis

Set the function relationship between grp and reach to $r_k(g)$, and introduce the variable k to the number of times the ad is reached.

For the batch of ads corresponding to each campaign, two points are known:

$r_k(0) = 0$, when grp=0, there is naturally reach=0.

$r_k(G) = R, R \in [0,1]$, this time corresponds to the end of advertising. Where, G is the final grp and R is the corresponding reach.

According to the above analysis, assuming that there is a linear relationship between grp and reach, we can get $r_k(G) = R/G \cdot g$. Observe the scatter plot of multiple types of data, as shown in Figure 1. Compared with the linear regression, it is easy to know that the data trend presented is not a completely linear trend [3]. Every time a grp is added, reach will not grow evenly, and the growth rate will gradually decrease, and the maximum value of reach will approach a limit value, which cannot fully reach 100% [4-5].

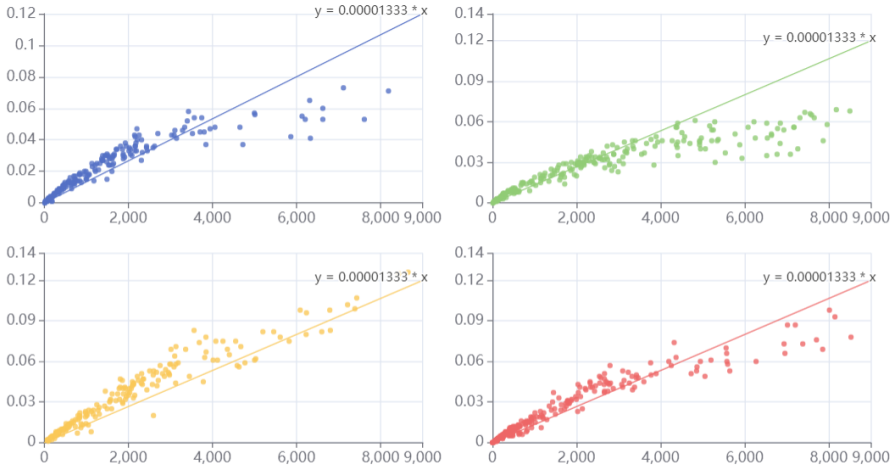


Figure 1. Linear Trend Chart

However, at a very small age of grp can be considered linear regression, i.e. $r_k(G) = g \cdot \iota$. where $\iota = \iota(k)$ represents the reciprocal of the expected impression required by the first user to see the k ad, and the minimum value of ι is $\frac{1}{k}$. When $k = 1$, the first user to see the k ad only needs 1 exposure. When $k > 1$, the first user to see the k ad needs to be exposed m times, and there must be $m \geq k$. Therefore, $\iota(k) = \frac{1}{m} \leq \frac{1}{k}$, using $\iota(k) = \frac{1}{k+log_2 k}$ at this time has a good fitting effect, and the limit of it can be obtained:

$$\lim_{G \to 0} \frac{\partial}{\partial g} r_k(g = G) = \iota \tag{1}$$

2.2.2 Reach-curve Model

It is easy to know from the above analysis that the data trend shown in the scatter plot is not linear, so simple linear regression is not suitable for fitting the nonlinear relationship of this data.

Set $X(X \geq 0)$ as the number of exposures. According to the conditional probability formula, this paper proposes the following relationship:

$$P(X \geq k, R) = P(R) \cdot P(X \geq k \mid R) \Leftrightarrow r_k = \rho \cdot p_k \tag{2}$$

Where, $\rho \in [0,1]$ is the maximum probability of possible contact, $p_k \in [0,1]$ is the probability of touch k times given the actual touch.

Therefore, this paper successfully decomposes the reach-curve model into two parts, and expresses each part with a different probability model. In this paper, the p_k part is parameterized in the model building part[6].

2.3. Model Establishment

2.3.1 $p_k(g)$ Parameterization

This article provides two methods for $p_k(g)$ parameterization.

(1) Poisson distribution of g:

The method has an exponential prior distribution of β on the Poisson rate λ, as shown in the following equation:

$$p_k(g) = 1 - \frac{\beta}{g+\beta} \tag{3}$$

The exponential prior distribution can be extended to the $\Gamma(\alpha,\beta)$ distribution, and the following can be obtained:

$$r_k(g) = \rho[1-(\frac{\beta}{g+\beta})^{\alpha}] \tag{4}$$

By derivation, we can get the growth rate of reach:

$$\frac{\partial}{\partial g} p_k(g) = \frac{\alpha}{\beta}(\frac{\beta}{g+\beta})^{\alpha+1} \tag{5-1}$$

$$\lim_{g\to 0}\frac{\partial}{\partial g} r_k(g) = \frac{\rho\alpha}{\beta} \tag{5-2}$$

It can be seen that the degree of freedom of equation (5-1) is 3, but only two data points can be used for a batch of advertisements. Therefore, one parameter must be fixed. According to the nesting structure of the exponential model, this paper sets $\alpha \equiv 1$.

(2) Expit Function Method:

The expit function is the inverse function of the logit function:

$$logit[p_k(g)] = \beta_0 + \beta_1 \cdot \log g \tag{6}$$

$$\begin{aligned} p_k &= expit\,\beta_0 + \beta_1 \cdot \log g(\beta_0 + \beta_1 \cdot \log g) \\ &= \frac{e^{\beta_0+\beta_1\cdot\log g}}{1+e^{\beta_0+\beta_1\cdot\log g}} \\ &= 1 - \frac{1}{1+e^{\beta_0}\cdot g^{\beta_1}} \\ &= 1 - \frac{e^{-\beta_0}}{1-e^{\beta_0}\cdot g^{\beta_1}} \end{aligned} \tag{7}$$

This equation is very similar to Equation (3), β_0 and β_1 represent intercept and slope respectively, taking $\beta_1 = 1$, the model obtained by the two methods can be sorted out, and the result is shown below:

$$r(g;\rho,\beta) = \rho(1-\frac{\beta}{g+\beta}) = \rho\frac{g}{g+\beta} \in [0,\rho] \tag{8}$$

2.3.2 Dual Parameter Model

β_0 controls the efficiency of grp. The smaller β makes reach grow faster, and the larger β makes reach grow slower. Taking its derivative, we can get the following:

$$r'(g;\rho,\beta) = \rho\frac{\beta}{(g+\beta)^2} \tag{9}$$

When $g = 0$, there is $r'(0) = \frac{\rho}{\beta}$, and from Equation (9) and the known point (G, R), we can get:

$$\frac{\rho}{\beta} = \iota \Rightarrow \rho = \beta\iota \tag{10-1}$$

$$\rho\frac{G}{\beta + G} = R \Rightarrow \rho = \frac{R(G+\beta)}{G} \tag{10-2}$$

For the simultaneous equation (10-1) and equation (10-2), we can find the parameters ρ and β as:

$$\hat{\rho} = \min(\frac{G \cdot R}{G - R/\iota}, 100) \tag{11}$$

$$\hat{\beta} = \begin{cases} \frac{\hat{\rho}}{\iota} = \frac{G \cdot R}{G - R/\iota}, \text{if } \rho < 100 \\ G\frac{100 - R}{R}, \text{if } \rho = 100 \end{cases} \tag{12}$$

2.4. Least square method with L2 regularization constraint

According to the establishment of the dual parameter model, this paper first uses the least square method to fit, which can minimize the square of error, and find the best matching function of the data. Due to the large noise of data, over fitting is easy to occur. In this paper, L2 regularization is used. The formula is as follows:

$$\begin{aligned} J(\omega) &= \sum_{i=1}^{n}(y_i - X_i\omega)^2 + \lambda\sum_{j=0}^{n}{\omega_j}^2 \\ &= \sum_{i=1}^{n}(y_i - X_i\omega)^2 + \lambda \| \omega \|^2 \end{aligned} \tag{13}$$

Where ω_j is the coefficient of the j feature out of the total p feature; λ is a hyperparameter used to control the penalty strength of ω_j, the larger its value, the simpler the generated model, and its ideal value is solved using the gradient descent method; $||\omega||^2$ is the L2 norm; $\lambda||\omega||^2$ is the contraction penalty term or regularized L2 norm that attempts to reduce the variance[7-8].

2.5. Model Optimization

This paper uses Python software to program and solve the regression model, mainly using scikit-learn, scipy and symphony libraries. For the dual parameter model, this paper considers the optimal value of the regularization term coefficient alpha, which can determine the penalty degree of the model[9]. In this paper, the optimal value of the multi-step search coefficient alpha is used to establish a pipline, its step size is set to 0.1, the results are obtained through the search, and the goodness-of-fit R^2 is used as the performance index in the process of changing the coefficient alpha, and the resulting relationship graph is shown in Figure 2.

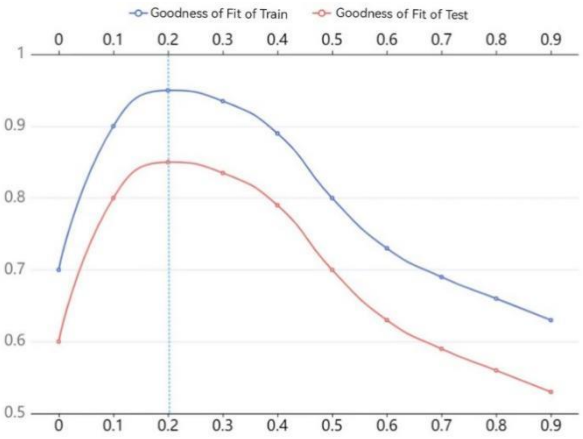


Figure 2. goodness of fit of different alpha

It can be seen from Figure 2 that the best fitting effect is obtained when the coefficient alpha=0.2.

In addition to the adjustment coefficient alpha, this paper also conducts cross validation on the dataset. According to the number of samples in the data set, this paper uses K fold cross validation, taking K=5 to ensure the training effect.

For the selection of evaluation indicators, this paper uses goodness of fit R^2 to measure the performance of the model during the training and testing of the whole model[10]. The value range of R^2 is [0,1]. The larger the value, the higher the degree of interpretation of the independent variable to the dependent variable, the higher the percentage of the change caused by the independent variable in the total change, and the denser the observation points near the regression line. The calculation formula of R^2 is:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} = \frac{\sum_i (f_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \tag{14}$$

3. Results

3.1. Model Comparison Experiment before and after L2 Regularization

Compare the models before and after L2 regularization, and observe the experimental results, as shown in Figure 3.

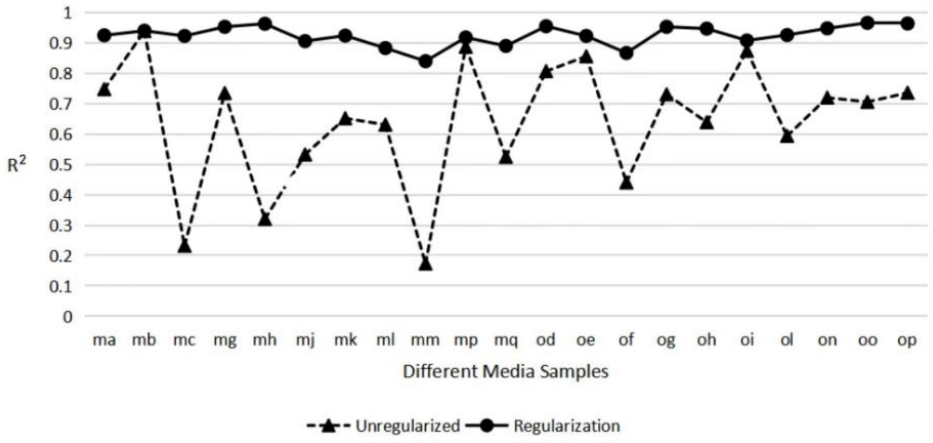


Figure 3. Comparison of model results before and after L2 regularization

It can be seen from Figure 3 that the difference between the goodness of fit of training data and the goodness of fit of test data of the model after introducing L2 regularization is smaller than that of the model without L2 regularization. This shows that the over fitting situation is restrained after the introduction of L2 regularization, and the generalization ability of the model is improved.

3.2. Comparison Experiment of Different Methods for Suppressing Overfitting

In this paper, L2 regularization and L1 regularization are compared to inhibit over fitting, and the results are shown in Table 1.

Table.1. Comparison of Results of Different Methods for Inhibiting Overfitting

Suppress Overfitting Method	R^2/%	RMSE
L1 regularization	94.3	0.258
L2 regularization	91.6	0.297

The evaluation index RMSE is the root mean square error. Calculate the mean value of the square sum of the corresponding sample point errors of the fitting data and the original data. The smaller the value, the better the fitting effect. The calculation formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_i (y_i - f_i)^2} \tag{15}$$

Table 4 shows that R^2 and RMSE of L2 regularization are better than L1 regularization.

3.3. Comparison Experiment of Different Models

This paper compares the goodness of fit of traditional linear regression, polynomial regression, decision tree, XGBoost, support vector machine, LASSO regression and dual parameter model, and the results are shown in Table 2.

Table.2. Comparison of average R^2 of different models

Models	Linear Regression	Polynomial Regression (3rd degree)	Decision Tree	XGBoost	Support Vector Machines	LASSO Regression	Dual Parameter Model
Average R^2	0.802	0.824	0.858	0.770	0.817	0.789	0.931

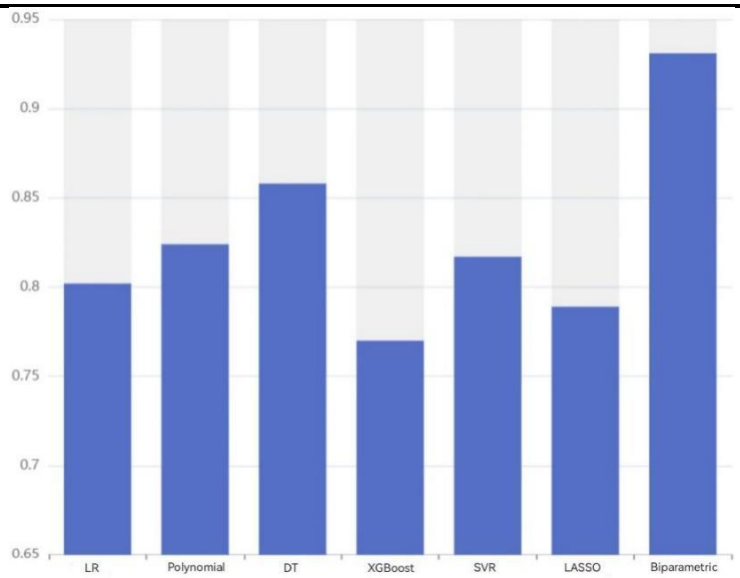


Figure 4. Comparison of average R^2 of different models

According to Figure 4 and experimental data, compared with linear regression, polynomial regression, decision tree, XGBoost, support vector machine and LASSO regression models, the dual parameter model established in this paper has greater advantages in performance, and its R^2 is 0.08 higher than the above six models on average.

4. Conclusions

The purpose of this paper is to find out the relationship between grp and reach variables in the sub media. In this paper, we first customize the dual parameter model for fitting, and then introduce L2 regularization and use gradient descent method to solve the problem, so as to improve the fitting accuracy and avoid over fitting.

The function form of the dual parameter model proposed in this paper is easy to understand and fits the relationship between grp and reach. This model can infer the investment required to increase another reach percentage point when the advertising cost is known, and can provide new ideas and strategies for advertising in sub media delivery, which is helpful to predict users' media consumption behavior.

References

[1] Jin Y, Shobowale S, Koehler J, et al. The incremental reach and cost efficiency of online video ads over tv ads [J]. 2012.

[2] Jin Y, Koehler J, Goerg G M, et al. The optimal mix of tv and online ads to maximize reach[J]. 2013.

[3] Li W, Jiang M, Zhan W. Why advertise on short video platforms? Optimizing online advertising using advertisement quality [J]. Journal of Theoretical and Applied Electronic Commerce Research, 2022, 17(3): 1057-1074.

[4] Rebecca J B, Ramalingam V V, Sugumaran V, et al. Predictive analysis of online television videos using machine learning algorithms [J]. Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications, 2022: 237-257.

[5] FULGoni G A M. Is the GRP really dead in a cross-platform ecosystem?: Why the gross rating point metric should thrive in today's fragmented media world [J]. Journal of Advertising Research, 2015, 55(4): 358-361.

[6] Chapelle O, Manavoglu E, Rosales R. Simple and scalable response prediction for display advertising[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2014, 5(4): 1-34.

[7] Aalam S S, Victorie T A D A. A Probabilistic Approach of Coverage prediction in Mobile Ad hoc Networks [J]. International Journal on Computer Science and Engineering, 2011, 3(12): 3824.

[8] Sun Ying, Mao Zhizhong. A prediction model for the number of advertisement responders based on audio-visual data [J]. Journal of Northeast University (Natural Science Edition), 2010, 31 (06): 765-768.

[9] Pashkevich M, Kharin Y S. Robust estimation and forecasting for beta-mixed hierarchical models of grouped binary data[J]. SORT-Statistics and Operations Research Transactions, 2004: 125-160.

[10] Goerg G M. Estimating reach curves from one data point [J]. 2014.