TOWSON UNIVERSITY

MASTER'S THESIS

Aiding Linear Television Media Planning Through Bayesian Inference and Forecasting

Author:
Matthew TIGER

Supervisor:
Mr. Jason MUHLENKAMP

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

at

Towson University

April 19, 2018

Declaration of Authorship

I, Matthew TIGER, declare that this thesis titled, "Aiding Linear Television Media Planning Through Bayesian Inference and Forecasting" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

TOWSON UNIVERSITY

Abstract

Department of Mathematics

Master of Science

Aiding Linear Television Media Planning Through Bayesian Inference and Forecasting

by Matthew TIGER

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too…

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor…

Contents

List of Figures

List of Tables

List of Abbreviations

LAH	**L**ist **A**bbreviations **H**ere
WSF	**W**hat (it) **S**tands **F**or

For/Dedicated to/To my…

Chapter 1

Introduction

The world of advertising today consists of

We are trying to provide better linear tv media plans.

We will use the theory of Bayesian statistics to develop a model that will forecast future telecast airings' audience impression concentration distributions corrected for small sample sizes present in the measurement training data.

1.1 Background

TV sellers have airtime available for sale to be used to air advertisements. TV buyers purchase airtime from sellers in order to air their desired advertisemensts.

TV buyers want to buy ad airtime that will help them share their message with a target audience.

A media plan consists of an allocation of a number of equivalized units to a collection of selling title / weeks.

Audience measurement companies provide measurements television viewership. Some companies take these measurements using a statistical sample of the TV viewing population and tracking the viewership of that sample and extrapolating to the population at large.

TV sellers bundle together airtime into a unit called a selling title. This unit contains logically grouped airings of content. Sometimes the content within the selling title is specific, other times it's a blend of content packaged together.

An example of a piece of content available is *Funny Animated Cartoon* and an example of a selling title woudl be *Funny Animated Cartoon M-F 8PM*.

TV sellers accept media plans allocated at the selling title/week level and then as the flight date approaches determine the actual ad schedule that will air. This is known as trafficking the ads.

The planning is done based off of demo estimates, which are impression estimates of a particular age/gender demographic. The TV seller provided these estimates using their substantive knowledge. The planning for strategic targets is done by forecasting a percentage. The resulting forecast is then the percentage multiplied by the forecasted demo estimate

1.2 Motivation

The sample sizes used for the majority of telecasts in the measurement source are too small to be trusted on their own. This is the major challenge of providing forecasts for media plans; using the data on its own will produce inaccurate forecasts that do not take into account the small sample size.

(Show picture of sample sizes in networks)

A motivating example is related to evaluating the performance of baseball players: players are judged by their batting average but this metric is not informative with few at-bats. With more information about the league and past historical performances we can use bayesian inference to come up with a better estimate that takes such factors into account as well as the observed at-bats.

Simiarly, we hope to use the data from past airings to perform better inference on telecasts with small sample sizes.

1.3 Challenges

Low sample size must be addressed. Variance within selling-title/week. Impression estimates are tied to demo forecasts.

For the purposes of this thesis we will be addressing the first issue alone. We will assume for testing that we know the outcome of the demo with prescience and that we can traffick the ads wherever we want.

1.4 Problem Description

Summarize the challenges and what we are trying to solve. Specifically, given the "noisy" measurement data and the supplier data, can we create model to forecast the in-target impressions for future selling title airings with some degree of accuracy.

Chapter 2

Data

We make use of several datasets in order to provide the necessary forecasts.

2.1 Audience measurement data

The audience measurement data contains the programs that aired and who watched them according the audience measurement source.

Each program in the data contains certain covariates such as the hour in which a program began and on which day of the week etc.

Each program contains the list of panelists that watched the program and how much viewing is associated to the program. We filter down to only commercial viewing since this is what advertisers are most interested in.

We are going to use the measurement source data for three networks labeled BCST for a broadcast network that has large reach, ETMT for a cable entertainment network, and SPTS for a cable sports network.

This data is stored at the minute-level and stores which respondents were watching each minute of a given telecast.

The definition of the number of impressions per equivalized unit airing during a telecast is the weighted average of the in-target commercial viewing seconds over the number of commercial seconds that aired during the telecast. Formally, let A be the set of panelists that are considered in-target out of a total of n possible panelists. Suppose that telecast i has t_i minutes of programming and let s_{ij} be the number of commercial seconds during minute j of the telecast. Further, let $p_{ijk} = 1$ if panelist k was watching during the j-th minute of telecast i and 0 otherwise, and let w_k be the weight assigned to the panelist by the measurement source. Then, m_i^A, the average commercial minute (ACM) of telecast i is

$$m_i^A = \frac{\sum_{k=1}^{n} w_k \sum_{j=1}^{t_i} s_{ij} p_{ijk} \mathbf{1}_A(p_k)}{\sum_{j=1}^{t_i} s_{ij}} \tag{2.1}$$

A related quantity is the unweighted average commercial minute, denoted r_i^A, which is derived from (2.1) by setting the weights $w_k = 1$ for all k and taking the ceiling of the resulting quantity, i. e.

$$r_i^A = \left\lceil \frac{\sum_{k=1}^{n} \sum_{j=1}^{t_i} s_{ij} p_{ijk} \mathbf{1}_A(p_k)}{\sum_{j=1}^{t_i} s_{ij}} \right\rceil$$

This quantity represents the average number of in-target panelists viewing commercials on telecast i.

The impression concentration of target A relative to target B where $A \subseteq B$ for telecast i is

$$c_i = \frac{m_i^A}{m_i^B}.$$

Similarly, the unweighted impression concentration for target A relative to target B for telecast i is

$$u_i = \frac{r_i^A}{r_i^B}.$$

As to be expected r_i^A and m_i^A are nearly perfectly correlated. Further, the Q-Q plots show that the distributions of c_i and u_i for a given network agree after a linear transformation.

(include QQplots)

This suggests that the distribution of unweighted impression concentrations is similar to distribution of impression concentrations. Thus, the ratio of in-target panelists to demo panelists is similar to the ratio of in-target average audience to demo average audience and we may use the raw sample sizes to estimate a population proportion that will be proportional to the actual impression concentration.

The purpose of the preceding paragraph is to justify the assumption that we can use the raw proportion of in-target to in-demo panelists to estimate the weighted impression concentration.

2.2 TV supplier data

TV suppliers provide us with a future schedule of selling title/content airings. They also provide the demo estimates for a variety of demos for future selling title/weeks

2.2.1 Linear Media Schedule

Here we will talk in more detail about the schedule and what it contains. We will also provide an example of such data. (include example record)

2.2.2 Forecasted Demographic Estimates

These are estimates of forecasted demo impressions per unit. This means the supplier expects to get a certain amount of demo impresions for each allocated unit. (include example record)

Chapter 3

Model

3.1 Preliminaries

In the following section we will provide preliminary information necessary to understand the model.

3.1.1 Units of analysis and observation

We are most interested in the content that airs.

the units of observation are selling title airings in the schedule. The outcome variable we are measuring for each unit i is r_i^A the average number of panelists associated to target A given the average number of panelists associated to target B where $A \subseteq B$. Of primary interest is the proportion of panelists in target A relative to target B.

3.1.2 Covariates

Using the conclusions of paper TV ratings, we include the following covariates into the model:

1. Broadcast Month - The broadcast month associated to the start of a selling title airing. "The key link between the broadcast and Gregorian calendars is that the first week of every broadcast month always contains the Gregorian calendar first of the month."
2. Day of Week - The day of the week associated to the start of a selling title airing. These are encoded from 0 - 6 with 0 being Monday and 6 being Saturday
3. Stratified Hour - We define the following groupings of hours as this covariate. These groupings are adapted from the measurement source. morning - airing start hour is in (6, 7, 8, 9), daytime - airing start hour is in (10, 11, 12, 13, 14), early_fringe - airing start hour is in (15, 16, 17, 18), prime_1900_2000 - airing start hour is 19, prime_2000_2100 - airing start hour is 20, prime_2100_2200 - airing start hour is 21, prime_2200_2300 - airing start hour is 22, late_fringe - airing start hour is in (23, 0, 1), graveyard - airing start hour is in (2, 3, 4, 5)
4. Content Id - The id denoting the content associated with the scheduled airing. e.g. Famous Funny Animated Cartoon show
5. Lead-in Content Id - The id denoting the content associated with the preceding airing.
6. first-run - Denotes if the airing is not a repeat airing. This is a 1 if it is not a repeat and 0 if it is a repeat. This information is obtained through the measurement source.

7. genre-id - Denotes the genre associated to the airing. This conveys the general flavor of the aired content. We will enumerate the different genre-ids and their meaning. This information is obtained through the measurement source.

3.2 Assumptions

We assume that the observations are exchangeable given the parameters of the model and the covariates of the selling title airing. We describe the definition of exchangeable here. Exchangeability is important since it allows us to use Bayes' Theorem to compute the posterior. The number of draws from the binomial aren't independent, but we use this as an approximation. this assumption is imperfect.

3.3 Description

See below for a formal description of the model. Since the data we observe is the average number of panelists in target A given the average number of panelists in target B with $A \subseteq B$, it is natural to model the likelihood as a binomial distribution with unknown probability parameter π.

$$\begin{aligned}
\theta_i &= \beta_0 + \sum_{j=1}^{m} X_i^{\mathsf{T}} \beta_j, \quad \beta_j \sim \mathrm{N}(0, 4^2) \\
\omega_i &= \mathsf{logit}^{-1}(\theta_i) \\
\kappa_i &\sim \mathrm{Exp}(X_i^{\mathsf{T}} \lambda_l), \quad \lambda_l \sim \mathrm{U}(1e-3, 1) \\
\pi_i &\sim \mathrm{Beta}\left(\omega_i \kappa_i + 1, (1-\omega_i)\kappa_i + 1\right) \\
y_i | n_i, \theta_i, X_i &\sim \mathrm{Bin}(n_i, \pi_i)
\end{aligned}$$

where $\mathsf{logit}^{-1}(\alpha) = \frac{\exp \alpha}{1+\exp \alpha}$.

3.4 Prior Distribution Choice

The coefficients are chosen so that the main mass of the distribution on the logit scale falls between 1e-5 and 1 - 1e-5.

We chose the lambda to be uniform on 1e-3 and 1 since the kappa variable is the concentration parameter which is roughly the number of trials in the binomial likelihood. For any given telecast airing we will have at most the number of trials be the number of panelists which cannot exceed 100K. This prior on lambda and the exponential reflects that.

We chose to separate variance based off of is-first-run since we noticed that this variable is indicative of how many trials there will be for an airing.

3.5 Inference

We perform the inference using `pymc3`. We will display convergence checks such as gelman-rubin statistics, effective number of parameters among other things to ensure that posterior distribution obtained are representative of the actual posterior distribution.

Show the calculation of rhat and neff and display summary statistics of values to show that inferences have approximate convergence and that trace objects are representative of

3.6 Validation

posterior predictive checking.

Given the posterior distribution obtained through inference, we make draws from the distribution to arrive at replicated data, y^{rep}. We check the distribution of the replicated data using test statistics comparing replicated data to the observed test statistic in the data.

- We check the resulting posteriors using test statistics of the mean, min, max, and std of the number of in-target panelists. This makes sure that the data generated by the model are consistent with the observed data.

- ordered discrete data check to see if data is under/overdispersed compared to replicated data.

- binned residual plots against observed in-target panelists to check that the regression fits make sense.

- The data was collected sequentially in time so we must check if residuals have autocorrelation.

- outlier check to see if any impressions conc predictions are greater than 1-2 std of impressions conc.

- realized residual plots for impressions conc

3.7 Model selection and Sensitivity Analysis

We use the WAIC to determine the model that has the most predictive accuracy. The WAIC is chosen because it is Bayesian and able to be compared across models with same likelihood.

Go through the definition of WAIC and explain how it is computed etc.

Here we will also check to see if the choice in prior is affecting inference.

- Remove the hierarchical nature on the variance parameter and evaluate how inference is affected - change normal distribution on coefficients to cauchy or student t to see if inference is affected

Chapter 4

Forecast Results

We want to compare the forecasts from the posterior predictive distribution on test data vs the industry standard forecasts per Tv ratings paper.

4.1 Predictive Accuracy

Compute probabilistic scores for bayes forecast and compute MAE for industry standard. Compare vs baseline of population proportion.

Show actual vs predicted plots of each airing

4.2 Media plans

4.2.1 Random Media Plans

Create media plans by randomly selecting airings and then compute HPD and mean of media plan where we assume that the demo estimates are known.

We want to check that outcomes are within are 95% HPD

4.2.2 Calibration

Show calibration plot of binning up all airings into 20 or so bins and then plotting mean of actual vs mean of predicted. If model is calibrated then "low stuff should perform low and the high stuff high."

4.3 New Content

We expect the model to perform better on new content than industry standard method. We should illustrate that here.

4.4 Inference Summary

We summarize the inferences using the mean of the posterior.

Chapter 5

Conclusion