STAT430 Final Project

Rick Yabuki-Soh

Professor: Nathaniel Stevens

Executive Summary

In these hard pandemic COVID-19 times, I am sure many of you are spending more time on Netflix. But how much time do you spend trying to find what you want to watch? In this experiment, we will investigate the average browsing time through Netflix menu of movies and TV shows before a viewer decides which one to watch. Minimum browsing is what Netflix hopes to achieve because it increases the assurance that a viewer will patronize a movie and not exit Netflix. This is a desirable outcome because when a viewer watches a movie, Netflix earns revenue.

The factors we are interested in are “Tile Size”, “Preview Size” and “Preview Length”. Through factor screening, we found that “Preview Size” and “Preview Length” were the important factors that affected browsing time. We found that the optimal browsing time lied in the vicinity of 95 seconds (preview length) and 0.5 (preview size). A more precise methodology determined that the minimum browsing time was estimated at 15.06768 minutes when preview length was 87.69546 seconds and preview size was 0.6497187. But these values are not practically feasible, so we rounded preview length to 88 seconds and preview size to 0.65 which achieves a browsing time of 15.1 minutes.

Introduction

This experiment investigates how long users take to decide which movie to watch in Netflix. The objective of our research is to minimize the browsing time to decision. The Netflix homepage displays TV shows and movies as tiles, and each row represents a different genre. The factors we are interested in are “Tile Size”, “Preview Length” and “Preview Size”. Tile size is the ratio of a tile’s height to the overall screen height. Preview length is the duration, in seconds, of a TV show’s or movie’s preview. Lastly, the preview size is the ratio of the preview window’s height to the overall screen height.

Our finding is that preview length and preview size significantly influences the browsing time. In our data set the preview length ranges from 30 seconds to 120 seconds and the preview size ranges 0.2 to 0.8. We find that the minimum browsing time of 15.10293 minutes occurs when the preview size is 0.65 and the preview length is 88 seconds.

Our design involves only 3 factors and this is an over simplified version of the more complex movie selection behaviour process of a Netflix viewer. However, it is a starting point and we can build it up for future experiments with a few more relevant factors: good possible choices are the gender and age groups of the viewer. Naturally, we can stereotype that young males may go directly to the action genres and thus lower their browsing time prior to selection.

Effective experimentation is sequential. Information gained in one experiment can help to inform future experiments. The goal of this response surface methodology is that we first do a factor screening which helps us conduct the method of steepest descent which leads us to the response optimization. Throughout the experiment, you will notice how the conclusion in the first phase (factor screening) introduces a new question. We answer this new question in the second phase (using the method of steepest descent). And at the conclusion of phase 2 we introduce yet another question. And that is answered by the third phase (the response surface design). The conclusion of the response surface design answers the big question of our experiment: the response optimization.

Factor Screening

The **question** we want to answer is "Which factors significantly impacts browsing time on Netflix". Hence, the metric of interest is the browsing time on Netflix, and we want to know how can it be minimized. In our **plan**, we let the response variable be the "continuous time until a Netflix user watches a show" (i.e. stops browsing). The design factors that we will be investigating are "Tile Size", "Preview Size" and "Preview Length". We aliased *Preview.Length* with the interaction of *Tile.Size* and *Preview.Size*, and performed a 2^{3-1} fractional factorial experiment. We are conducting a 2^{3-1} fractional factorial experiment instead of a 2^3 factorial experiment because we are tight on cost (and cannot afford to conduct many experimental conditions) but still want to test all three factors. We will also be working with a significance level of $\alpha = 0.05$. The experimental units are the Netflix users, and we sampled 250 users in each condition. Our **data** was *randomly* collected through a simulator via design matrix shown below. Otherwise, if the data was not random, we would conduct an A/A test to make sure that our data was collected randomly.

Tile.Size	Preview.Size	Preview.Length
0.1	0.3	90
0.3	0.3	30
0.1	0.5	30
0.3	0.5	90

Analysis:
This is our linear regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i3}$$

where

x_i1 is the main effect of *Tile.Size*
x_i2 is the main effect of *Preview.Size*
x_i3 is the main effect of *Preview.Length*

The three hypotheses we want to test are:

$$H_0: \beta_1 = 0 \text{ v.s. } H_A: \beta_1 \neq 0$$
$$H_0: \beta_2 = 0 \text{ v.s. } H_A: \beta_2 \neq 0$$
$$H_0: \beta_3 = 0 \text{ v.s. } H_A: \beta_3 \neq 0$$

The results of the analysis of variance table below was used to test our hypotheses:

AVOVA TABLE					
Source	Sum of Squares	DF	Mean Square	F-value	P-value
Tile.Size	533.7	1	533.69	57.129	9.239e-14
Preview.Size	669.3	1	669.27	71.641	9.137e-17
Preview.Length	839.7	1	839.27	89.880	1.790e-20
Residuals	9304.6	996	9.34		

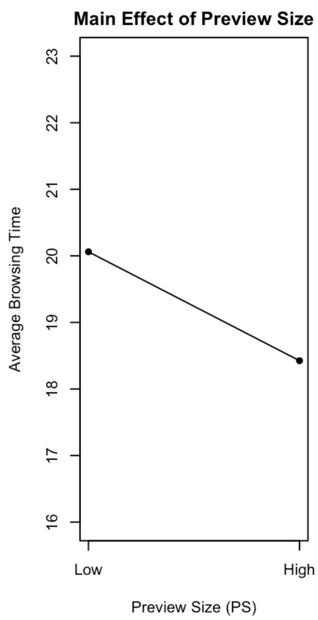
(Table 1)

Looking at 'Table 1' we can see that all the p-values are less than $\alpha = 0.05$, but note that *Preview.Length* is aliased with *Tile.Size* and *Preview.Size* and aliasing anywhere causes confounding everywhere.
Hence we have:

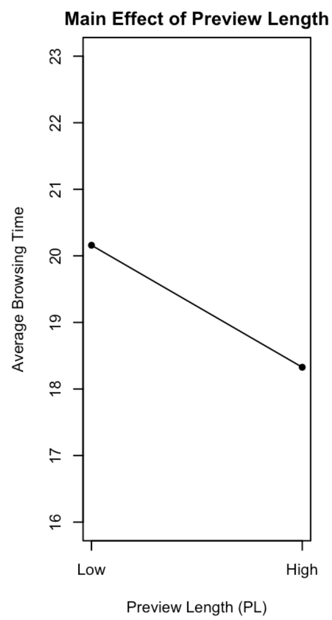
$$\textit{Tile.Size} \times \textit{Preview.Size} = \textit{Preview.Length}$$
$$\textit{Tile.Size} \times \textit{Preview.Length} = \textit{Preview.Size}$$
$$\textit{Preview.Size} \times \textit{Preview.Length} = \textit{Tile.Size}$$

So we do not know if the above three factors are significant due to the main effect or the interaction effect. We do have the support of the principle of effect sparsity, but we are still unsure. Notice that *Preview.Size* and *Preview.Length* have the smallest p-values, so we will mainly work with these factors. Let us assume that the significant p-value of *Tile.Size* is due to the interaction effect of *Preview.Size* and *Preview.Length*.

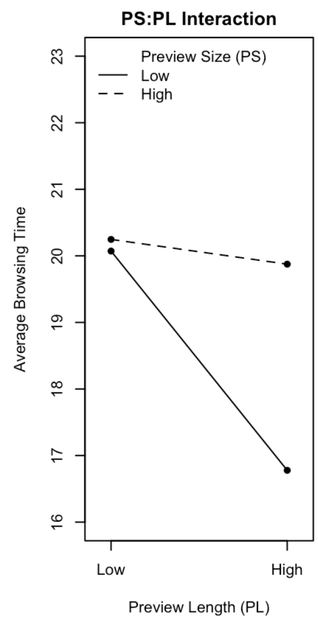
To get a better insight on our factors, we plot the main effects and interaction effect:



(Figure 1)



(Figure 2)



(Figure 3)

We can see from the main effect plots, that if we set *Preview.Size* and *Preview.Length* to their high levels, it significantly decreases the average browsing time. From the interaction plot, we see that if *Preview.Length* is at its low level, there isn't much difference in average browsing time, whether *Preview.Size* is at its low or high. But when *Preview.Length* is at its high level, there is significant difference in average browsing time whether *Preview.Size* is at its low or high level. We also know from 'Table 1' that the p-values of *Preview.Size* and *Preview.Length* are significant, and we assumed that the significance of *Tile.Size* is due to the interaction effect of *Preview.Size* and *Preview.Length*.

Hence, we **conclude** that the main effect of *Preview.Size* and *Preview.Length* significantly influences average browsing time.

Method of Steepest Descent

Since we determined the significant factors, we now want to determine **where** the optimum lies. As mentioned in the previous step, we will focus on *Preview.Size* and *Preview.Length*. The **plan** is to use a 2^2 factorial experiment to estimate a first order response model and find the vicinity where the optimum lies. The **data** was collected from the simulator via design matrix shown below.

Preview.Length	Preview.Size
30	0.3
90	0.3
30	0.5
90	0.5
60	0.4

Analysis:
The linear regression model with linear predictors:

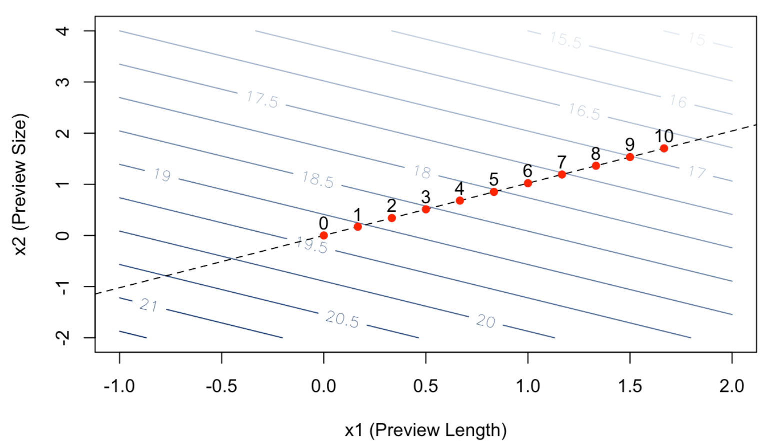
$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{PQ} x_{PQ}$$

Summary Table				
	Estimate	Standard Error	t value	Pr(>|t|)
Intercept	19.88085	0.19167	103.727	0
Preview.Length (x_1)	-0.74996	0.09583	-7.826	1.072e-14
Preview.Size (x_2)	-0.76624	0.09583	-7.996	2.924e-15
x_{PQ}	-0.70722	0.21429	-3.300	9.931e-04
$x_1 : x_2$	-0.71757	0.09583	-7.488	1.322e-13

(Table 2)

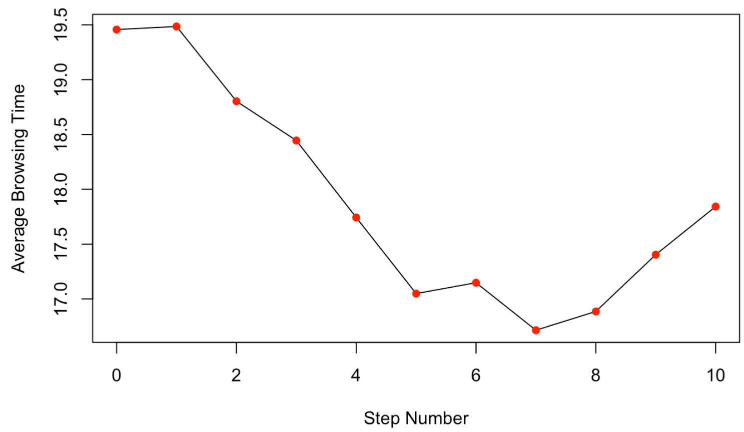
Testing for curvature by $H_0: \beta_{PQ} = 0$. From 'Table 2' we see that the p-value is 9.931e-04 which is less than $\alpha = 0.05$. So we reject the null hypothesis and conclude that there is significant curvature. Now, we reorient ourselves to the optimum. The step size we chose was $\lambda = \frac{1/6}{|\beta_1|}$ where the value 1/6 was chosen to ensure steps of 5 seconds in *Preview.Length*.

Visualizing the path of steepest descent starting at (0,0):



(Figure 4)

Step Number versus Average Browsing Time



(Figure 5)

Usually, when the average browsing time increases, we stop and move back one step. But just to be on the safe side, we took 11 steps. From ‘Figure 5’ we see that step 7 corresponds to the lowest observed average browsing time. So we will conduct another test of curvature in this region. Letting (95, 0.5) [i.e. Step 7] be the new center point, we have the following summary table:

Summary Table				
	Estimate	Standard Error	t value	Pr(>|t|)
Intercept	16.47886	0.19296	85.398	0
Preview.Length (x_1)	0.52239	0.09648	5.414	7.374e-08
Preview.Size (x_2)	-1.05074	0.09648	-10.890	1.909e-26
x_{PQ}	1.09087	0.11422	9.551	6.549e-21
$x_1 : x_2$	0.30703	0.09648	3.182	1.498e-03

(Table 3)

Testing for curvature by $H_0: \beta_{PQ} = 0$. From ‘Table 3’ we see that the p-value is 6.549e-21, which is less than $\alpha = 0.05$. So we reject the null hypothesis and **conclude** that there is significant curvature. Hence, we are in the vicinity of the optimum.

Response Optimization

Now that we know the vicinity of the optimum, we **want** to identify the location of the optimum. The **plan** is to run a two-factor central composite design to fit a full second-order response surface model. We intended to perform an axial condition at $a = \sqrt{2}$, but the corresponding *Preview.Length* and *Preview.Size* were at practically inconvenient levels, so we opted for $a = 1.4$. The **data** was collected from the simulator via design matrix shown below. For *Preview.Length*, we let the low level = 80 and high level = 110, and because we let 95 be the new center, and we need $a = 1.4$ to be less than 120 in the natural units. For *Preview.Size*, it was just a convenient low and high level.

Preview.Length		Preview.Size	
80	-1	0.4	-1
110	+1	0.4	-1
80	-1	0.6	+1
110	+1	0.6	+1
116	+1.4	0.5	0
74	-1.4	0.5	0
95	0	0.64	+1.4
95	0	0.36	-1.4
95	0	0.5	0

Analysis:
The full second order model we fit is:

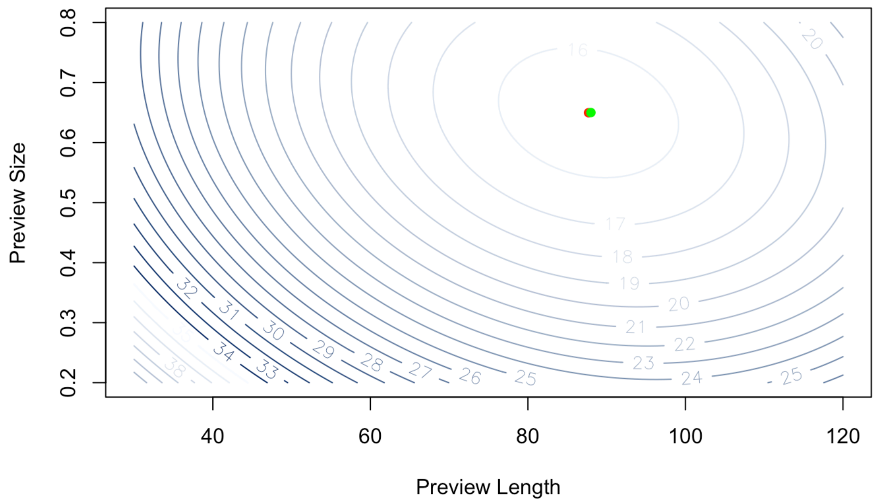
$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta\_22 x_2^2$$

Summary table shown below:

Summary Table				
	Estimate	Standard Error	t value	Pr(>|t|)
Intercept	16.53794	0.18694	88.465	0
Preview.Length (x_1)	0.52976	0.06645	7.972	2.467e-15
Preview.Size (x_2)	-1.22134	0.06645	-18.379	2.202e-70
$I(x_1^2)$	0.90026	0.11092	8.116	7.848e-16
$I(x_2^2)$	0.44557	0.11092	4.017	6.089e-05
$x_1 : x_2$	0.23179	0.09351	2.479	1.326e-02

(Table 4)

The contour plot is:



(Figure 6)

Using the estimates from ‘Table 4’, we find that the optimal point is (87.69546, 0.6497187), in red. However, this is not a practically feasible. So we are going to pick the sub-optimal point of (88, 0.65) in green. The sub-optimal point is less optimal but more practically feasible.

	Optimal	Sub-optimal
Point	(87.69546, 0.6497187)	(88, 0.65)
Estimate	15.06768	15.10293
95% Confidence interval	(8.694581, 21.440785)	(8.729829, 21.476034)

Therefore, we **conclude** that the feasible optimal point is (88, 0.65) which achieves an average browsing time of 15.10293 minutes with a 95% confidence interval of (8.729829, 21.476034).