Demand Cannibalization in Cinema Scheduling

Sangam Verma SANGAM.VERMA@OPERASOLUTIONS.COM
Advanced Solutions, Express Trade Towers, Film City, Noida 201301

Pawas Gupta PAWAS.GUPTA@OPERASOLUTIONS.COM
Advanced Solutions, Express Trade Towers, Film City, Noida 201301

Abstract

This paper addresses the problem of demand cannibalization experienced by a show because of similar type of shows being scheduled around the same time. We propose a new model for this cannibalization which not only helps in generating profit optimal schedules which are closer to the business understanding but is also mathematically more accurate than the existing model.

1. Introduction

The motion picture industry is a prominent economic activity generating total worldwide box office revenue of $38.6 billion in 2016. Movie forecasting and programming in practice tend to be associated with intuition rather than formal analysis and characterizes the tradition of decision-making in the film industry. For each week's movie program, management must determine what movies will be shown, on which screens, on which days, and at what times. The analytical approach towards movie scheduling for an individual movie theatre consists of three interconnected stages: Demand Prediction, Sample Space Generation and finally; Optimization with Constraints satisfaction. The job of optimization process is to find the optimal selection from given sample space, which maximizes the profit while adhering to given linear and nonlinear time constraint set. To generate a movie schedule, optimization process(optimizer) must consider impact of several factors at play; age and duration of movie, prints available for a movie, format of movie (2D, 3D) etc. Although the approaches for above-mentioned stages have been sufficiently explored and implemented with their literature in place(J. et al., 2009a), the phenomenon of the interaction between two similar shows is still untouched. It is observed that when two similar kind of shows (shows with same or different format of same movie) are placed near each other, they tend to eat into each others attendance. For a movie session playing on any given day, at any given site and on any given screen at any minute, demand is predicted independent of the sessions nearby. The prediction of demand in such a manner does not include the effect which surrounding shows may induce on the given show for which demand is being predicted. Hence, to incorporate such behaviour, another factor needs to be taken into consideration while scheduling. For instance, A show of *Star Wars* scheduled at 8:00 PM on certain screen of Manchester site on Sunday carries a predicted attendance of 200 as per demand models. Now, if another show of the same movie is scheduled at 8:00 PM on another screen of same cinema site on same day, the overall predicted attendance at 8:00 PM should get distributed between these two shows. If the two screens are identical in nature, then ideal attendance distribution should be equal. But, the problem becomes complex when multiple formats of various movies interact at multiple intervals. Without the existence of demand cannibalization (the phenomenon explained above), the movie with highest predicted attendance will get scheduled in all the screens with total attendance of that movie linearly increasing with every incremental show, which in life scenario is not the case.

The aim is to discuss an approach which sufficiently explains demand cannibalization behaviour and then implement it as part of scheduling. The task of deciding an approach becomes challenging due to certain considerations which should be taken in account: 1) Dynamic calculation for the effective cannibalization factor while scheduling. 2.) Unavailability of a defined benchmark for measuring the performance of the ap-

The Fifth Internal Technical and Analytical Conference of Opera Solutions (TACOS 2017), October 26th, 2017, San Diego, USA.

proach 3.) Difficult to isolate the impact of a single component on overall scheduling. The approach discussed here incorporates all the above listed factors to help generate a solution which shows greater promise than its existing counterpart and is closer to the business understating we are trying to achieve, which in turn results in better profit numbers. This paper is organized as follows: In Section-2, existing demand shift approach is briefly discussed, along with its limitations. Section-3 provides detailed description of the new approach and how it overcomes problems associated with the older approach. Section-4 contains result of our new model and conclusion is presented in Section-5.

2. Previous Demand Shift Methodology

At a given session, the overall demand shift affecting it (β) takes into consideration impact of surrounding sessions of same Film (2D and 3D formats of the same movie), playing anywhere in 120-minute radius of session start. The effective demand shift is incorporated as part of attendance calculation and the formula for attendance calculation is given as:

$$Attendance(A) = \alpha \cdot \beta \cdot \gamma \cdot s.f.$$

α provides the day level demand of the movie and considers how performance of a movie varies with different types of days, sites, film cluster, age of the movie, size of the movie, rain, snow etc.γ determines how the day level demand of the movie splits at different time slots within the day. It is a coefficient realized between 0 to 1. β then considers the impact on attendance of a session because of the same movie scheduled close by (within +/- 2 hours). The screen factor manually adjusts for the impact, different kinds of screens have on attendance. Together these provide an estimation of attendance for a session. The cannibalization(β) can be broadly categorized under two types:

- **Self-Cannibalization (Direct Beta):**It considers the impact of a session on session of the same movie and format such as impact of 3D session on other 3D session of the same movie.
- **Cross-Cannibalization (Cross Beta):**It considers the impact of session on other sessions of the same movie but different format such as impact of 2D session on 3D session of the same movie.

Demand shift for a given show is calculated as:

$$\beta = \beta_{direct} \cdot \beta_{cross}$$

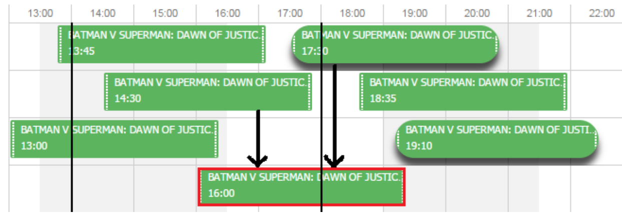


Figure 1. Calculation of effective demand shift

Each β curve for (both direct and cross) demand shift comprises of 49 coefficients β_{120b},β_{115b},....,β_{5b},β_0,β_{5a},...,β_{115a},β_{120a} (b=before and a = after), attributing to shows of the same film placed at +/-120 minutes, with respect to a given show. Moreover, both direct and cross β are evaluated by using following formula:

$$\beta = \prod_{120b}^{120a} (\beta_i)^{x_i}$$

where, x_i denotes the number of films of the same format for direct beta (and a different format for cross beta) at time i.
To make it clear, we demonstrate it by showing an example. Suppose we want to calculate demand shift on 16:00 show of *Batman Vs Superman 2D* as shown in Figure 1. The beta curve for Saturday is given in Figure 2. There are only two shows of the same movie in 2 hours range. One show is starting at 14:30 which is 2D (so, Direct $\beta_{90b} = 0.93$) and other show is starting at 17:30 which is 3D (so, Cross $\beta_{90a} = 0.98$). Therefore, overall demand shift on show at 16:00 is 0.9114 (0.93*0.98). This described approach however fails to

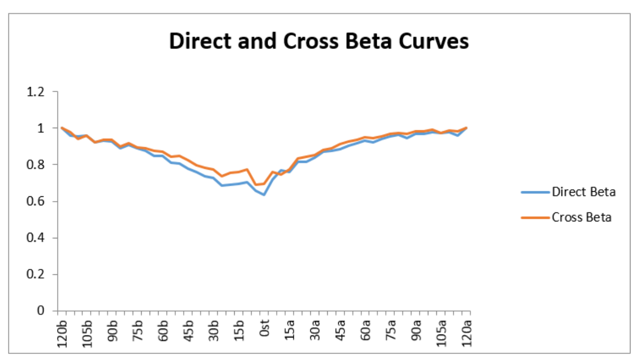


Figure 2. Example of beta curves

capitalize because of the following reasons:

- Dip in Incremental admits: It is observed that when multiple shows of a movie interact with each other, the overall attendance tends to decrease with an increase in number of shows. This is counter intuitive and hence the current beta model is unable to capture the multiple shows scenario.

- Low scheduling of 3D shows: In a real scenario, at a cinema site, the 3D screens are fewer in comparison and movies with 3D formats can only be played on screens which support 3D. Hence, limited space is available for scheduling 3D movies. This factor combined with the severity of the current demand shift leads to very low scheduling of 3D shows. For a certain week, whether sufficient number of 3D shows for a movie have been scheduled using existing demand shift approach can be estimated from demand ratio of its 3D and 2D format in that week and from the % of 3D shows for that movie that actually went on sale(in that week).

Table 1 below showcases percentage of 3D shows scheduled with current demand shift approach and 3D demand along with the production percentage of shows for some of the previous blockbuster weeks:

Week	3D Demand	3D Production Shows	3D Shows
18-12-2015	30%	32%	13%
25-03-2015	35%	27%	12%
01-04-2015	34%	23%	11%
08-04-2015	35%	9%	10%
15-04-2015	38%	42%	8%
22-04-2015	42%	24%	12%
29-04-2015	33%	24%	13%

Table 1. Comparison of 3D% shows scheduled using beta approach with the production %

3. Lambda Approach

With older approach, the severity of effective demand shift multiplicatively increases with addition of every new show, which sometimes leads to decrease in overall admits (more overall admits before show was added). The demand shift factor should be such that its effect saturates after a certain threshold of shows. The new devised demand shift factor is given as:

$$\beta_{eff} = \frac{1}{1 + \sum_d \lambda_d + \sum_c \lambda_c}$$

Where, λ_d represents direct demand shift factors and λ_c represents cross demand shift factors. Similar to beta curve, there are 49 coefficients for both cross and direct demand shift attributing to shows in every 5-minute interval ranging from -120 min to +120 min. The lambda curve for Saturday is shown in Figure 3 below.

In Lambda approach to demand shift, the effect of

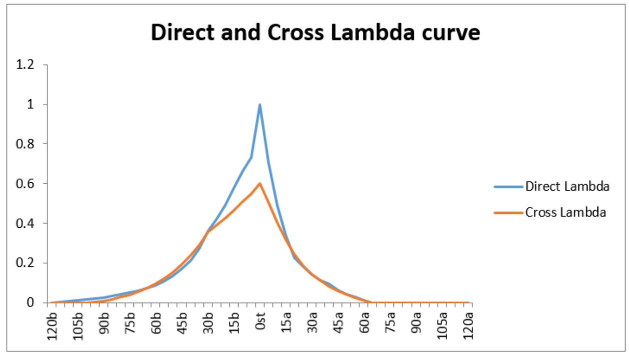


Figure 3. Example of lambda curve

lambda tends to saturate with every incremental show which poses an inherent advantage for this approach over previous one that it eliminates negative incremental admits issue completely. To illustrate the advantage of lambda approach over beta we consider the scenarios shown in Table 2. For the shows starting at same time the attendance should get equally distributed and hence the effective beta is 0.5 (β_0), which when transformed to lambda becomes 1(λ_0). Assume that the attendance of a *Batman Vs Superman 2D* show scheduled at 8:00 PM is estimated as 100.
It can be observed in 3 concurrent shows scenario that beta approach results in negative incremental admits where lambda approach results in 0 incremental admits which is more intuitive as per the business understanding. So, lambda based demand shift approach has helped solve the first issue poised by the beta based demand shift approach. The lambda based demand shift approach is further tweaked to address issue pertaining to low % screening of 3D shows, which is explained in further sections.

3.1. Bucketed lambda

It has been observed form historical data that the big movies (movies having high demand) experience much less severe demand shift as compared to small movies. This factor is however not considered while applying the effective demand shift during scheduling. Based on the 7 years worth of information on movie performance, with the help of multiple margins, movies have been classified into 8 separate buckets, with the movie size increasing with the increasing bucket level, where movie size is uncapped session attendance for

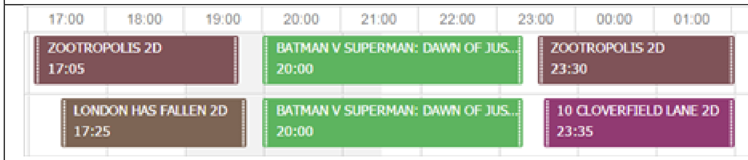
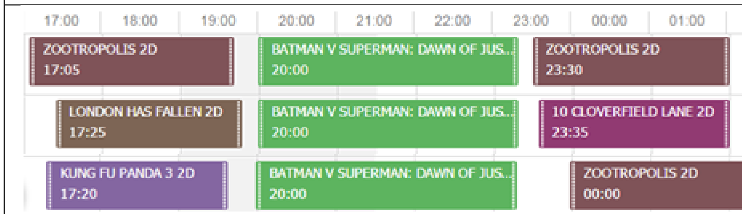
Scenario	Total admits with β	Total admits with λ
	$2*100*\frac{1}{2}$	$2*100*\frac{1}{1+1}$
	$3*100*\frac{1}{2}*\frac{1}{2}$	$3*100*\frac{1}{1+1+1}$

Table 2. Example scenario for incremental admits

a movie. For example, 8th bucket would have movies like Star Wars in its first week and 1st bucket would have a Bollywood movie in its second or third week of release. Multiple iterations were done for deciding the number of buckets. The margins were made such that that each bucket has a significant and preferably equal number of sessions from the historical data. The severity of demand shift decreases with increase in the bucket level. So, in the eighth bucket where big movies play, the effect of demand shift is mitigated, even though ample shows of them gets scheduled on a day. Figure 4 below shows an analysis of multiple shows being scheduled at some time intervals (15 min, 30 min etc.) and their impact on overall admits for that film.

From this analysis, it can be observed that the impact of demand shift on adding a show gets saturated after a threshold of shows have been scheduled which is not the case with beta where we are observing negative incremental admits. This inherently solves the issues present in the older approach. The 8 identified bucket levels are used to learn demand shift factors for each of the bucket separately, thereby forming a binary sparse matrix for each level. We treat each level as an attribute and the coefficients are learnt for each combination of level(Abhinn & Himanshu, 2016). Suppose we have a variable X which is dependent on factors a,b,c. We create a sparse matrix for X for all combinations of a*b*c and for each show S (each row in data corresponds to a show) we populate 1 in column C (represents one of the combination of a*b*c) if S is realized at C and 0 otherwise. In this way we transform the variables to binary form and set target as log of attendance. Demand shift is dependent on these two factors: Bucket level and nearby shows. For demand shift we create a sparse matrix M of size m*n for each day where m are number of shows in training data and n = 98, 49 (-120,-115,, 0,,+115,+120 minutes) for direct and 49 for cross. M[i,j] = x implies that for the show in i_{th} row there are x shows of the same film j*5 minutes apart on same site, date but different screen. The curves are calibrated at each 5-minute interval in +/-2 hours and hence we get 16 curves separately for each bucket type, cross and direct (8(bucket type) *2 = 16 curves). The final curves obtained represent if a given film f is scheduled at a site S, screen s1 and time t (in minutes), how would its demand fall if f is scheduled at site S, screen s2 and time t+/-120 minutes. Similarly, sparse features are formed for other factors as well in the attendance formula thus formed represents an independent variable. The intraday training equation is:

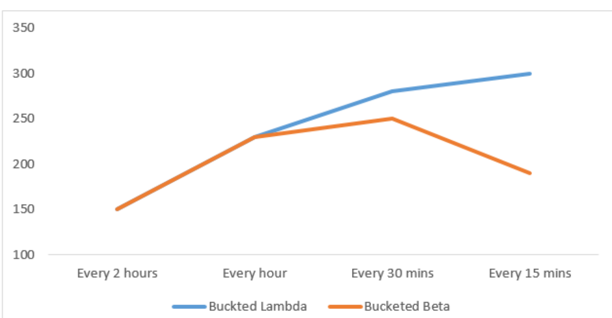


Figure 4. Total admits variation with increasing number of nearby shows

$$\log\left(Attendance\right) = M_{\alpha} + M_{\beta} + M_{\gamma} + M_{s.f.}$$

Here M_i is the sparse matrix for i^{th} intraday component. We learn the coefficients of these matrices, take their exponential followed by a pseudo matrix inversion (only for demand shift) and obtain our intraday components.

3.2. Modified cross beta for higher format

The limited availability of 3D screens results in low % 3D shows which remains a problem even with the updated lambda approach. In general, it is observed that 3D format of a movie has lower demand share (30%) as compared to 2D format of the same movie(70%). Even though cross demand shift is less severe as compared to direct demand shift, the magnitude of demand shift being applied to a 2D show and a 3D show are always same and since the demand of a 3D movie is comparatively low, it is not able to overcome the competition from the 2D format, which leads to low 3D show scheduling. For example, cross beta for same slot shows is 0.6. Assume that the raw attendance of a 2D show is 100 and 3D show is 50. If a 3D show is scheduled at the same slot as that of a 2D show the overall attendance of the system become 100*0.6+50*0.6=90. This is less than the raw attendance of a 2D show.Hence the loss in attendance because of demand shift should be different for two interacting shows when they are of different formats. The 3D cannibalization factor needs to be re-calibrated to take this factor into account. The updated cross demand shift formula is given below:

$$\beta_{eff} = \beta + (1-\beta) \cdot \frac{Max(R, \frac{1}{R}) - 1}{Max(R, \frac{1}{R})}$$

where,

$$R = \frac{\alpha_{2D}}{\alpha_{3D}}$$

The above equation is transformed for lambda approach using the β-λ relationship. It helps in solving the above mentioned issue by adjusting the cannibalization factor on the basis of the demand ratio.

3.3. Bucket assignment on movie level

As observed from data, the 3D format of a movie generally has around 30% share of the total demand. So, for a particular movie the 3D format would fall in a smaller bucket as compared to the 2D format of the same movie. This will result in a severe demand shift for the 3D format as compared to the 2D format because in lower buckets demand shift is more severe. A decreasing trend in the popularity of 3D movies is seen over the years which is also backed by the historic data. For example, past 2 years data suggest that 15 percentile of 3D shows fall in buckets 7/8 as compared to 25 percentile for 2D shows. This trend is much severe if more recent data is considered. However, since the bucket margins were trained on the basis of whole data, the model is not able to capture this change. The bucket levels even though are calculated on the basis of the attendance any differences introduced because of different formats of a movie needs to be handled. The bucket assignment for all the formats is now modified to the bucket of the format which has the highest demand (2D in most cases).

3.4. 3D Boosting algorithm

Schedule for each screen is generated using a Dynamic Programming Based Optimization Algorithm (Pawas & Praveen, 2016), applied to each screen in a sequential fashion. Once a screen has been completely scheduled, the schedule against it remains fixed. Hence, the order in which screens are scheduled impacts the quality of schedule. This sub-optimality in the schedule generation process cannot be tackled due to complexities involving time and space. This is one of the reasons for under scheduling of 3D movies, as while ordering the screen preference is given to high capacity screens most of which supports 3D formats. Thus, the 3D screens get occupied by 2D shows due to their high demand and the 3D shows remain under scheduled. Using the updated optimization Algorithm 1,

Algorithm 1 3D boosting

```
Input: Initial schedule Sch after the usual optimization, Screens 1 to s, Number of iterations n
Initialize iter = 0
repeat
  for i in 1 to s in the order of scheduling do
    Sch_New = Rescheduled i in Sch assuming shows in other screens of Sch as fixed
    Sch = Sch_New
  end for
  iter = iter + 1
until iter < n
```

when 3D shows were unable to be scheduled into a 3D screen in step 1 due to the order in which screens were scheduled, they have an opportunity to be scheduled in step 2. The number of such iterations also have a significant role to play. The number of iterations has a direct correlation with the improvement in quality of schedules. However, the incremental improvement tends to saturate after a certain number of iterations and increasing the number iterations negatively impacts the run time of the process. This trade-off between optimality and run time led to a constraint on the number of iterations. Currently, the number of iterations is fixed to 3.

4. Experimental results

Comparison of the lambda approach with beta approach needs to done both on the basis of the overall quality of the schedules and the forecasting accuracy of the models. Since its very difficult to quantify the quality of schedules a good measure is to compare the overall % of 3D sessions to 3D demand % and the % 3D shows which actually went on sale. Since the impact of demand shift is more critical for weeks involving big movies, tests were performed on weeks where blockbuster movies were released. A comparison for both of the approaches on the selected weeks is given in Table 3: To evaluate performance, we use following metrics: Mean Error by Mean (MEByMean) and Mean Absolute Error by Mean (MAEByMean) which are given as follows:

Week	3D Demand	3D Production Shows	3D Shows(β)	3D Shows(λ)
18-12-2015	30%	32%	13%	29%
25-03-2015	35%	27%	12%	26%
01-04-2015	34%	23%	11%	23%
08-04-2015	35%	9%	10%	20%
15-04-2015	38%	42%	8%	38%
22-04-2015	42%	24%	12%	26%
29-04-2015	33%	24%	13%	29%

Table 3. Comparison of 3D% shows scheduled using β and λ approach with the production %

$$MEByMean = \frac{mean(forecast - actual)}{mean(actual)}$$

$$MAEByMean = \frac{mean(|forecast - actual|)}{mean(actual)}$$

Here, Actual stands for the actual attendance reported for the session. The forecasted attendance is the attendance which is predicted using the respective models. Since, the demand shift is an integral part of the attendance formula any change in the approach results in overall change in forecasting models. To have a one to one comparison with the beta approach both the models were retrained with the same training data (05-01-2007 to 18-09-2014). A final comparison between the forecasted accuracy for both models is a good indicator of the overall performance of the approach. The error numbers calculated for selected weeks as well as for overall period of 52 weeks are reported in Table 4.

Weeks	MEByMean		MAEByMean	
	β	λ	β	λ
18-12-2015	-6%	1%	45%	43%
25-03-2015	24%	14%	59%	52%
01-04-2015	16%	5%	49%	43%
08-04-2015	17%	-3%	58%	45%
15-04-2015	5%	16%	53%	48%
22-04-2015	-17%	-14%	48%	43%
29-04-2015	-10%	-13%	44%	41%
52 weeks	8%	4%	51%	46%

Table 4. Forecast Error using β and λ

5. Conclusion

We presented a new model which attempts to rectify current gap in literature by modelling the demand cannibalization at a more granular level. This model outperformed the existing models both in terms of quality of schedules and accuracy of the forecasting process. We also created significant business impact with an estimated 2-4% increase in profit numbers. The cinema chain has a profit share of around 24%, however for big movies the share is 15%. With this new approach we aim to reduce this difference in profit shares by significantly improving the process for the big weeks which generate more profits than normal weeks. Apart from cinema chains demand cannibalization is observed in almost every industry which involves scheduling. The proposed approach can be extended to other industries as well, such as scheduling of Flights, allocating the hospital beds using demand prediction and scheduling of TV shows.

Acknowledgments

We would like to thank Ritesh Aggarwal, Raman Thakral, Suvam Ray and the entire reviewing committee for their guidance in making significant improvements to the paper. We would also like to thank Himanshu Sharma, Abhinn Kothari, Priyanshu Jain, Suyash Baderiya, Abhishek Jindal, Priti Haldar, Reema Kumari and Naveen Gupta for their support in building the new demand shift model.

References

Abhinn, Kothari and Himanshu, Sharma. A binary log linear approach for forecasting movie attendance. In *Technical And Analytical Conference of Opera Solutions*, 2016.

J., Eliashberg, Q., Hegie, J., Ho, D., Huisman, S.J., Miller, S., Swami, C.B., Weinberg, and B., Wierenga. Demand-driven scheduling of movies in a multiplex. In *International Journal of Research in Marketing*, volume 26(2), 2009a.

Pawas, Gupta and Praveen, Omar. Dynamic programming for movie scheduling. In *Technical And Analytical Conference of Opera Solutions*, 2016.