Behavioral Targetting

Introduction

The amount a given game is shown is referred to as impressions. We define I_q as the relative probability that the q_{th} game is shown. Since the total probability must add to one and given that we have a platform that contains N games we can write

$$\sum_{q=1}^{N} I_q = 1 \tag{1}$$

If H is the total number of impressions we are able to provide then we can write HI_q as the total impressions we provide for game q.

For each game we are able to make a measure of events, X (X_q is a measure of X on the q_{th} game). Also suppose that this measure X is a function of the impressions I that we provide for it (as in $X(I)$). One common example of a metric that depends on the impressions provided for a game is the "clicks" that take a user from the impressed game to the game itself. The relationship between the users clicking and the impressions provided is the "click through rate" (referred to as ctr for short) which describes the number of users clicking (or total clicks) per impression provided. For example, if we impressed a game 1000 times on a set of users and 200 users clicked, it would have a "click per impression" (in this paper we will refer to as CPI) value of 0.2.

Continuing, suppose we measure that a game q has a CPI value of CPI_q. Then, the total number of users clicking that game is proportional to how much it is impressed, namely: HI_qCPI_q is the total number of users clicking on that game. In other words, by increasing I_q we will increase the number of users that click on it. Total number of users clicking across all games, T would be

$$H\sum_{q=1}^{N} I_qCPI_q = T \tag{2}$$

The question we ask in this discussion is how do choose the frequency in which we impress each game, referred to as i $\mathbf{I}$, to maximize T. For example, the reader could imagine that impressing games with a high CPI would generate more clicks than allocating impressions for the games with the lowest CPI. We can define how much better (or worse) one impression vector, $\mathbf{I^a}$ in generating total clickers, T versus another vector $\mathbf{I^b}$ by $Z^{a,b}$ defines as

$$Z^{a,b} = \frac{H\sum_{q=1}^{N} I_q^aCPI_q}{H\sum_{q=1}^{N} I_q^bCPI_q} \tag{3}$$

Recognizing that the H's cancel and the sum $\sum_{q=1}^{N} I_q^zCPI_q$ can be written as a dot product $\mathbf{I^zCPI}$, we can condense it is

$$Z^{a,b} = \frac{\mathbf{I^aCPI}}{\mathbf{I^bCPI}} \tag{4}$$

Generic Optimization

One obvious approach to optimizing a metric, such as "users clicking" is to show the games that have a higher CPI more. An easy way to do this is define the q_{th} component of the impression vector proportional to its CPI_q value:

$$I_q = L * CPI_q \tag{5}$$

where L is a normalization constant defined as:

$$L = \frac{1}{\sum_{q=1}^{x} CPI_q} \tag{6}$$

Let us refer to this impression vector in the general case as $\mathbf{I_Q}$.

Optimizing on a Behavior Specific Basis

The CPI value of a game may be strongly dependent on the behavior of the user, especially the game the user is currently playing. For example, from a category perspective, users who are playing word games may be more likely to click on other word games while users on puzzle games may be more likely to click on other puzzle games, etc. Or it may be more subtle: users who play Bouncing Balls may have a higher ctr value on Crazy Cabbie than users playing other games.

In Figure 1 and 2, we show a graphical representation of CPI on a from specific game to specific game basis. This data was accumulated by measuring all the impressions for a game p shown on a game q as well as all of the users clicking on a game p given that they were playing on a game q for an entire week's worth of data.

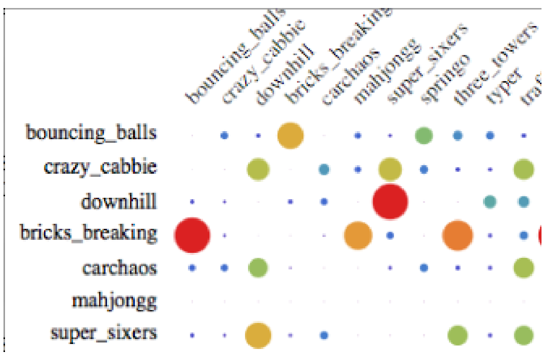

Figure 1: CPI transition matrix for some of our top games. The source of the clicker is in the column and while the destination is in the row. For example, users who are playing bricks breaking have a high CPI to other games in that row.

Figure 2: A larger view of the previous figure. There ar many patterns of games that have consistently high or consistently low for all sources of for a group of sources. However, there is also a large variety of CPI on a game by game basis

Let us define CPI_{pq} as the CPI value for users that are playing game p that click on game q. We are now able to define a new impression vector $\mathbf{I^{PQ}}$ with components I_{pq} as

$$I_{pq} = L_p * CPI_{pq} \tag{7}$$

where L_p is a normalization factor

$$L_p = \frac{1}{\sum_{q=1}^{N} CPI_{pq}} \tag{8}$$

Given that we are on game p how much better will we do by using the impression vector specific to that game $\mathbf{I^{PQ}}$ than the generic vector $\mathbf{I^{Q}}$. Namely what is $Z^{PQ,Q}$

$$Z^{PQ,Q} = \frac{\mathbf{I^{PQ}CPI}}{\mathbf{I^{Q}CPI}} \tag{9}$$

Shrinking the Inventory

Another approach to optimal impression allocation is to not show all games available in the inventory, bot only the top x games ranked by the CPI. Specifically, we can define a vector, $\mathbf{I_x}$ as

$$\mathbf{I_x} = \begin{cases} L_x * CPI_q & \text{if q <= x} \\ 0 & \text{if q > x.} \end{cases} \tag{10}$$

where L is a normalization constant defined as

$$L_x = \frac{1}{\sum_{q=1}^{x} CPI_q} \tag{11}$$

game	$Z^{PQ,Q}$	$Z^{PQ,Q}_{10}$
Snake	1.55	3.48
Kamikaze	1.56	2.94
Word Search	1.89	6.21
Data Worm	1.56	2.97
Crazy Cabbie	1.74	2.40
Parking Mania	1.54	2.78
Mini Epic	1.41	2.51
Bubble Spinner	3.56	5.25
Minifb Tower	1.99	3.96
Bouncing Balls	2.15	10.1
Wheel Burrow	1.48	2.48
Word Chaos New	2.89	3.73
Word Drop	7.56	4.27
Minifb Cube	2.23	3.21
Downhill Snowboard	1.76	2.49
Bricks Breaking	2.63	4.43
Qilox	1.95	2.45
Traffic	1.90	3.26
Carchaos	2.23	2.69
Hanger	4.98	5.11
Practice	2.29	3.02

Figure 3: A llist of some $Z^{PQ,Q}$ and $Z^{PQ,Q}_{10}$ for the most popular mini app games.Most games have a $Z^{PQ,Q}$ above 2 and a $Z^{PQ,Q}_{10}$ above 4.

In this language, $\mathbf{I^{Q}_{10}}$ would only show the top 10 games based on their overall CPI values. Also, $\mathbf{I^{QP}_{10}}$ would show the top 10 games based on their CPI values for each particular game p. In this way, we can see how much better $\mathbf{I^{QP}_{10}}$ would do than $\mathbf{I^{Q}_{10}}$ by writing the ratio.

$$Z_{10}^{PQ,Q} = \frac{\mathbf{I_{10}^{PQ}CPI}}{\mathbf{I_{10}^{Q}CPI}} \tag{12}$$

In Figure 3 we provide a list of $Z^{PQ,Q}$ and $Z_{10}^{PQ,Q}$ for some of the most popular games. If we compute $Z^{PQ,Q}$ for each game and then weight each $Z^{PQ,Q}$ proportional to how many gameplays game p receives, we can make an estimate of the overall increase in total click, T. Provided that we show all of the games weighted by their CPI given that the user is already on a game p, we can expect an increase in users clicking 2.13 times more. If we go even further and only show the top 10 CPI games for a specific game p, then we can expect 3.34 times more clicks.

Conclusion

In this paper we discussed a proposal for using the CPI specific to a user's current behavior: what game they are currently on in order to determine how we weight impressions. We saw that overall, we can get over 2 or 3 times more click through engagement by targeting user behavior. The astute reader may note that the CPI is not exactly the metric we want to optimize: just because users are clicking on a game more doesn't mean that is a good game to send them to. We need to take this information one step further and combine the CPI with theier activity once they get there such as their likelihood of publishing feed stories or their probability of buying virtual goods. Still, we can estimate that by targettng content on a behavioral level as described in this paper, we can also drive over 2-3x the number of feed stories published or virtual goods purchased for users that are sourced by the content display in consideration.

Sample References

Forthcoming Publication

Clancey, W. J. 1986a. Curious George.