A High-Dimensional Set Top Box Ad Targeting Algorithm including Experimental Comparisons to Traditional TV Algorithms

Brendan Kitts, Dyng Au, Brian Burdick
PrecisionDemand
821 Second Avenue, Suite 700, Seattle, WA. 98104. USA
bkitts@precisiondemand.com

Abstract—**We present a method for targeting ads on television that works on today's TV systems. The method works by mining vast amounts of Set Top Box data, as well as advertiser customer data. From both sources the system builds demographic profiles, and then looks for media that have the highest match per dollar to the customer profile. The method was tested in four live television campaigns, comprising over 22,000 airings, and we present experimental results.**

Keywords-television; advertising; targeting; set top box

I. INTRODUCTION

TELEVISION is the largest advertising category with 65 billion dollars spent in 2011 [19]. Advertising placements have proliferated from 20,000 advertising placements in 1950 to over 20 million possible placements per month today [26]. This presents a formidable problem for advertisers. Which programs, or time slots, should they place their ad into?

Making this problem more difficult, television has in the past been uniquely challenged for data collection. Most television targeting methods rely on the age-gender of a 25,000 person panel by Nielsen Corporation [27], [28].

This paper will present a system for automated television (TV) ad targeting using Set Top Box data. The paper also reports experimental results from four live television campaigns. As a result of the new data and methods we believe that TV advertising can be made considerably more relevant for consumers than ever before.

II. PRIOR WORK

Advertisers have been trying to target their advertisements since the 1940s. Nielsen was one of the first companies to set up a panel of viewers who manually recorded their TV viewing behavior. Advertisers used this by selecting TV programming with a high proportion of the age-gender that they believed was their target demographic [11]. Unfortunately age-gender is very coarse for targeting. The panel is also small (25,000 people) leading to small sample size extrapolation challenges [3], [27], [28]. Finally, Nielsen does not provide any solution around what should be their correct age-gender target.

Smyth and Cotter, Chorianopoulos, et. al. and Bozios, et. al. have investigated methods for TV ad targeting to individuals [7], [9], [29]. However in almost all cases current TV hardware lacks the capability to target ads to individual persons.

Tellis et. al. presented a system that used historical Toll-Free-Phone-Number (TFN) ad performance to estimate responses from a placement in the future [31]. One difficulty is the method relies upon phone numbers in the ad. However the greater practical problem is the number of historical airings are small compared to the millions of possible placements, and it can be expensive to generate data for the technique and effectively identify placements.

Canning et. al. provided a method of defining a target profile, although it was still panel based [8]. They described an approach in which a panel of users are tracked extensively including their viewing habits and purchases. This means that if the advertiser wants to target cereal, they can simply look up the viewing of cereal buyers and target those programs that rate highly for cereal buyers. Their technology is only available in 370,000 homes and any expansion may raise insurmountable privacy challenges [15].

Balakrishnan et. al. discuss a method for buying TV ads which attempts to either minimize Cost per Impression (CPI), or maximize unique persons reached [6]. However the method does not have any concept of targeting – reaching persons who are interested in the offer and are most likely to convert.

Despite a flicker of work in this area [6], [8], [14], Set Top Box approaches are still far from being widely used for real TV campaigns. The TV industry almost universally continues to use Nielsen age-gender breaks which have been available since the 1950s. We will show in this paper that there are considerable benefits to employing a "Big Data" approach of using the direct behavior Set Top Box viewing data for TV ad targeting, and we show experimental results from four large television campaigns, comprising over 22,000 television commercial airings.

III. THE TV AD TARGETING PROBLEM

The problem that we want to solve is to select a segment of TV media to purchase to insert an ad, such that advertiser value per dollar is maximized. Let M_i be a contiguous segment of time in the TV MPEG video stream that a station is offering for sale, $CPI(M_i)$ be the cost per impression of the timeslot, and $r(M_i)$ be the value to the advertiser per impression. The objective is to maximize:

$$M_i : \max \frac{r(P, M_i)}{CPI(M_i)} \quad (1)$$

According to Nielsen Competitive Data, there are over 20 million possible TV placements per month in the United States [26].

In the sections that follow, we will describe the machine learning methods that we are using to predict the value per impression $r(P, M_i)$ and cost $CPI(M_i)$ of each media instance, and how this is used to create an automated targeting system.

IV. Ad Targeting Empirical Research

The Consumer Research field has spent decades conducting experimental studies into consumer response to advertising [21]. One of the key findings from these studies is that ad effectiveness is enhanced by a match between the traits in the advertisement and those of the consumer. For example, Aaker, Brumbaugh and Grier showed that magazine ads featuring African American actors produced 9% higher responses from African American viewers [1]. They also found similar lifts for Homosexual viewers watching ad content with Homosexual actors (46%) and White viewers watching White actors (3%). This finding has been replicated over a wide range of demographics [12], [33], [34]. Aaker, Brumbaugh and Grier also showed negative lift for ads featuring one group being targeted to a different group. For example, if African American actors are shown to White audience, there is a decrease in response. Therefore the need for similarity between ad and audience seems to be symmetric with greater similarity producing greater positive response, and greater dissimilarity producing negative consumer response.

Which traits should be targeted for effective advertising? With regard to demographics, there is evidence that people exaggerate the importance – or weight more heavily - rare demographic differences between themselves and the general population, possibly due to psychological factors about personal identity [16]. For example, McGuire, McGure, Child, and Fjoki found that of the majority White students in an American school, only 1% spontaneously mentioned their ethnicity in describing themselves, where-as 14% of minority Hispanic students and 17% of African American students did so [23]. This finding has also been replicated in studies on height, wearing glasses [22], hair color, weight, birthplace and gender [10]. As a result, rare traits are believed to be more salient for consumers in terms of identifying the sub-group to which they belong.

Based on the above 30 years of research we can put forward two hypotheses on ad targeting:

H1: Advertising is more effective if media is viewed by persons with demographics similar to people who have purchased the advertiser's product [1],[12],[21],[33],[34],.

H2: Advertising is more effective when the demographics that match between advertiser product and viewing population are rare [10],[16],[22],[23].

V. Traditional TV Targeting Approaches

We've outlined the research on ad targeting. We will now discuss how ad targeting has been performed traditionally on television.

Nielsen Corporation's Target Rating Points (TRPs) on age-gender demographics are generally regarded as the most widely used form of targeting on television. This form of targeting defines a Target Rating Point as the number of persons who match the advertiser's target demographics divided by total population in a targeted area. In order to convert this into a measure of precision, we will express this as number of persons who match the advertiser's demographics divided by total viewing persons and multiplied by 100. Therefore, 100 means that of the people watching a particular program, all of them were the desired target.

Let $\bar{P}_{d,v}$ be the demographics of the set of persons who the advertiser wishes to target. We will define a demographic-value $p_{d,v} \in \{0,1,MV\}$ to be a formal proposition about the person p of the form d=v, eg. income=\$50K..\$60K. The proposition $p_{d,v}$ equals 1 if it is true, 0 if false, and missing value (MV) if it is unknown. Let $Q(M_i)$ be a set of viewers who are watching TV media instance M_i and where this viewing activity is recorded by the Nielsen panel and $q_k \in Q(M_i)$. Let # be the cardinality of a set, $\#r_T$ be persons that match on all demographics. The TRPs for Media Instance M_i can be defined as follows:

$$TRP(P, M_i) = r(P, M_i) = 100 \cdot \frac{\#r_T(M_i, \bar{P})}{\#Q(M_i)} \quad (2)$$

where $q_j \in r_T(M_i, \bar{P})$ if $\forall d, v: q_{j,d,v} = \bar{P}_{d,v}$ and where no values can be missing. For example 50 means that 50% of the people are a match to the desired demographics.

This algorithm meets H1 as it attempts to target the demographics of the desired population. However it doesn't do as well with H2 – typically age-gender is used for everything on television – yet these may not be the most distinctive demographics that define the target group.

One possible reason for the use of age-gender could be limitations in the Nielsen panel itself. The Nielsen's panel only has 25,000 people distributed across 210 Direct Marketing Association areas, so about 119 people per area [28]. There have been media reports of major rating shifts due to a single African American panelist moving – which is possible given that on average there would only be 16 African Americans per area [27]. As a result, rare demographics may well have too few persons to be usable – age and gender may be the only demographics exposed because they're the only demographics with enough data to be reliable.

There are some further challenges with TRPs. The counting definition above is workable for advertisers using age and gender, as even with a single age-gender combination such as Male 25-34, the population is still 1.8% of the full population [32]. Therefore in a program with 100,000 viewers, 1,800 would match the target at random providing fairly good statistics. However the method no longer works if an advertiser attempts to target using a large number of demographics – with 3,328 demographics, almost no people will have the exact same demographic readings that the advertiser is trying to reach, and so the method will routinely report 0% in the target group or statistically unreliable numbers.

The problem is there is no concept of "similarity" in the Target Rating Points scheme – for example, 35 year old females are similar to 34 year old females, yet the 35 year olds are outside of the 25-34 target.

VI. HIGH DIMENSIONAL SET TOP BOX TARGETING

The algorithm that we describe utilizes Set Top Box data to create a deep demographic match with the advertiser's product. The method carries several specific advantages: (1) the amount of usable viewing behavior increases from 25,000 to as much as 12 million persons (according to [19] the number of available set top boxes in 2012 was 33.9 million not counting satellite companies), (2) with more viewers, even rare demographics are usable, (3) we enrich with third party data sources to obtain up to 3,328 demographics attached to each person, creating a much more refined match than age-gender, and (4) the method also automatically derives the ideal vector to target from advertiser sales records. The method is similar to look-a-like algorithms that are used in online advertising [36], but uses a look-a-like between the demographics of buyers and media (the population viewing a TV program) because TV fundamentally supports ad insertion into media and not individuals.

NetW	Person	DateTime	Mins	Program
ESPN	10195589	3/10/12 3:00 PM	22	College Basketball
SCIFI	10195589	3/10/12 3:00 PM	7	Survivorman
SCIFI	10195589	3/10/12 3:30 PM	4	Survivorman
ESPN	10195589	3/10/12 3:30 PM	26	College Basketball
ESPN	10195589	3/10/12 4:00 PM	30	College Basketball
ESPN	10195589	3/10/12 5:30 PM	12	College Basketball
ESP2	10195589	3/10/12 5:30 PM	17	NASCAR Racing
ESP2	10195589	3/10/12 6:00 PM	30	NASCAR Racing
ESP2	10195589	3/10/12 6:30 PM	7	NASCAR Racing
ESPN	10195589	3/10/12 6:30 PM	2	College Basketball
SCIFI	10195589	3/10/12 6:30 PM	21	Survivorman
ESPN	10195589	3/10/12 7:00 PM	3	College Basketball
ESP2	10195589	3/10/12 7:00 PM	22	NASCAR Racing
ESP2	10195589	3/10/12 7:30 PM	12	NASCAR Racing
NICK	10195589	3/10/12 7:30 PM	29	Victorious
ESPN	10195589	3/10/12 7:30 PM	18	College Basketball
NICK	10195589	3/10/12 8:00 PM	9	Big Time Movie

Fig. 1c. Actual Set Top Box viewing record for Pereson 10195589 showing station, program and date. The demographics for this viewer include "Male", "Owns SUV", "Age=44-45", "Interest in spectator sports" "motorcycle racing", "football", "baseball", "basketball".

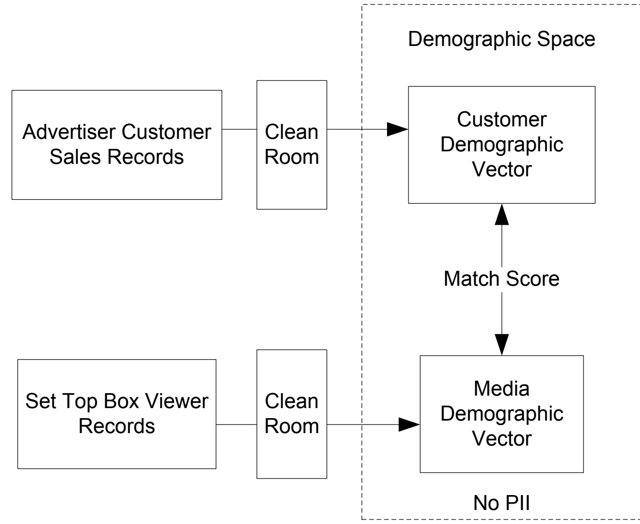


Fig. 1a. Matching process

Let P be a set of persons who have purchased the product – this set is known by the Advertiser. We will automatically use these advertiser sales records to infer the target. $p_j \in P$ is a person in the set to be targeted. We use commercially available consumer demographics [2] to enrich each person with up to D=3,328 demographics, so that $p_{j,d,v} \in \{0,1,MV\}$. Let $\#P_d$ be the cardinality of the set of persons who have the demographic d with any value v that is non-missing. We now calculate $\bar{P}_{d,v}$ for each demographic d,v – this is the probability of a demographic proposition d=v being true in the advertiser's set of purchaser.

$$\bar{P}_{d,v} = \frac{1}{\#P_d} \sum_{p_j \in P} p_{j,d,v} \qquad (4)$$

$\bar{P}$ is an ideal customer demographic probability vector with D=3,328 elements. We need to target this profile on TV media.

An example set of Set Top Box viewing activity is shown in Figure 1c. We will enrich the Set Top Box viewing persons with the same D=3,328 demographics. We then aggregate each piece of media M_i (eg. Survivorman 3:00pm, 3/10/2012 in Figure 1c) into an identically sized D-dimensional demographic vector $\bar{M}_i$ based on the set of persons who viewed that television program as shown below:

$$\bar{M}_{i,d,v} = \frac{1}{\#Q_d(M_i)} \sum_{q_k \in Q(M_i)} q_{k,d,v} \qquad (5)$$

We next define similarity or "tratio" r between advertiser target $\bar{P}$ and media $\bar{M}_i$ as the correlation coefficient between the product and media demographic vectors.

$$tratio(\bar{P}, \bar{M}_i) = r(\bar{P}, \bar{M}_i) = \frac{\bar{P}^{+} \cdot \bar{M}_i^{+}}{\left|\bar{P}^{+}\right| \cdot \left|\bar{M}_i^{+}\right|} \qquad (6)$$

$$\bar{P}_{d,v}^{+} = \frac{\bar{P}_{d,v} - \mu_{d,v}}{\sigma_{d,v}};\ \bar{M}_{i,d,v}^{+} = \frac{\bar{M}_{i,d,v} - \mu_{d,v}}{\sigma_{d,v}} \qquad (7)$$

$u_{d,v}$ and $\sigma_{d,v}$ are the mean and standard deviation of the demographic from an unbiased US population. We exclude any demographics (convert them to missing) if they have fewer than B=25 people.

Figure 1a shows the process graphically – media and advertiser customers are both mapped into high dimensional demographic space, and then matching is performed.

The algorithm has many properties which address computational and operational requirements of this domain. Firstly the present algorithm has excellent scaling characteristics compared to other algorithms, and so can be deployed on the higher data volumes available with set top boxes and demographics. Given S=12 million set top boxes, M=750 million historical media placements, D=3,328 demographics per person, and C=1 buyer profile targets, the training time for calculating matches can be calculated in $M*S*D + C*S*D + C*M*D$ time, which is linear in M, Set Top Boxes, and demographics, so is highly efficient. This enables us to easily pre-compute scores for hundreds of different ads for every possible media placement.

The above matching algorithm also meets the Targeting Theory literature's hypotheses we introduced earlier - H1 is met as we are targeting the media with the most similar demographics as defined by the correlation coefficient. H2 is true as we have converted demographic probabilities into z-scores – so that traits that are unusual compared to the US population have relatively higher positive and negative z-scores.

The algorithm is able to offer strong support on privacy. Unlike other algorithms which require individual person behavior linking across domains [8], the present algorithm only needs to convert audiences into aggregated demographic vectors. As a result it is possible for the aggregation to demographics to be performed by the primary sources, enabling only aggregated demographic data to be transmitted and used for targeting.

One other useful characteristic is that because vectors are built within the primary sources, the full amount of data can be used in assembling each vector.

The algorithm is also easy for Human Media Buyers to use in practice. Human Media Buyers often have to decide whether to spend \$100,000 or more on a single TV spot. The reason why a program is being recommended can be shown by calculating (8):

$$\mathrm{Cov} = \left(\bar{P}_{d,v}{}^{+} \cdot \bar{M}_{i,d,v}{}^{+}\right) \qquad (8)$$

When sorted from highest to lowest, this identifies the demographics d,v with the highest covariance between the buying population $\bar{P}$ and media viewing population $\bar{M}_i$. This provides insight into what is driving the match, and can help Media Buyers understand why the media matches their target (Figure 1b) [35].

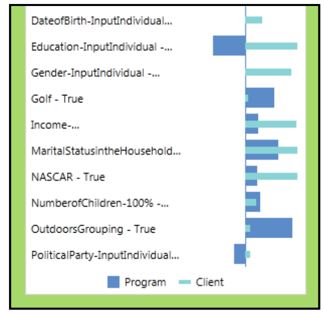


Fig. 1b. Detailed demographic match statistics on a particular TV program and its suitability for advertising a handyman product. The wider bars refer to the standardized demographic score for the program. The narrower bars are the standardized demographic reading for the advertiser's target. The particular product depicted above is a handyman tool which skews towards older males. This shows that the program matches on Marital Status, Nascar interest, Outdoor interest, etc.

VII. Cost Prediction

Recall that our objective is to find the best value per dollar media on television for our advertisement (1). In order to complete the value per dollar measure we now turn to predicting the second variable that we need to create (1) $CPI(M_i)$.

Cost per Impression $CPI(M_i)$ is predicted using the clearing prices of historical media to predict an estimated clearing price of the new media instance M_i.

$$CPI(M_i) = \sum_T h_T \cdot CPI(\bar{m}_{i,T}) \qquad (9)$$

where $CPI(\bar{m}_{i,T})$ is the average historical clearing price observed for media $m_{i,T}$. Because clearing prices for TV media are recorded by several industry groups including SQAD and Nielsen, it is possible to train the system to predicting the actual cost based on historical cost, ie. where $CPI^*(\bar{M}_\mathrm{i})$ is an actual clearing price for a media instance, and $CPI(\bar{M}_\mathrm{i})$ is an estimate.

$$\min \sum_\mathrm{i} \left[CPI(\bar{M}_\mathrm{i}) - CPI^*(\bar{M}_\mathrm{i})\right]^2 \qquad (10)$$

VIII. Experiments

In the sections that follow we will document three experiments measuring the quality of the system in targeting media.

Experiment I asks a preliminary question which is whether the system produces media targeting recommendations that match human judgment. A related question is whether human judgment could do as good a job at targeting as the system. We will show that that the answer to the first question is positive, and the second is negative.

Experiment II analyzes phone response prediction performance on two Phone Response TV ad campaigns. We ask whether the system produces phone response lift, and how it compares to traditional methods. Our results suggest a lift of about 35.7% using these methods.

Experiment III tests the method using a large Brand TV campaign. We examine how targeting improvements and cost savings can improve an existing campaign.

IX. EXPERIMENT I: HUMAN RELEVANCE

In order to measure the quality of the Ad targeting system, we used 5 human volunteers to score the relevance of TV programs for a Handyman product (Table 5, Campaign C). All volunteers were media buyers employed by PrecisionDemand, and so had some familiarity with television programs. Human volunteers scored TV Program relevance blind with a randomized order to avoid any list effects. Relevance was scored on a scale from 1 to 10. Each volunteer scored approximately 1,300 programs resulting in 6,500 human relevance scores in total.

Figure 2 shows the 25^{th} and 75^{th} distributions for ad targeting score for each labeled human relevance score. The targeting system appears to produce program targeting recommendations that match what humans would tend to recommend (R^2=0.34, $p<0.01$; Figure 3 HRS column).

The discrepancies between high dimensional match and human relevance were also interesting (Figure 4). Although some (eg. "Fresh Prince of Bell Air") appeared to be bad recommendations; the biggest discrepancy was on "Women's College Volleyball". The Media Buyers believed that this would not be relevant; however we think that handymen might actually like watching women's volleyball.

There are two conclusions from the data in Figure 3. C1: *High dimensional match between product and media is correlated with human relevance assessments of programs* (R^2=0.34, $p<0.01$; Figure 3 HRS column). This is consistent with H1 from the experimental literature, we have not seen this result reported in the literature and so we report it here.

C2: *High dimensional match is twice as predictive as TRP in terms of correlation with human relevance* (R^2=0.15, $p<0.01$ for TRP; R^2=0.34 for tratio; Figure 3). This is likely due to the richer demographics available, but it is useful to have the difference quantified so as to give some idea of the improvement if more variables are used.

The next question is, could humans score media well as the automated? In order to address this we analyzed human scores against buyer probability. Buyer probability is the probability of a known buyer of the advertiser's product being detected as watching the program being targeted. We find that both TRPs and High Dimensional Match out-perform humans on this task. We can therefore conclude that C3: *Both TRPs and High dimensional match out-perform human relevance judgments for identifying high probability of buyer media.* (R^2 for tratio= 0.25, R^2 for TRP = 0.17, R^2 for Human = 0.06, $p<0.01$; Figure 3 SourceViewPCT column). In other words, automated targeting provides additional value that human judgment alone would not capture. This is good news for the job security of data miners everywhere.

Furthermore we have an additional validation of targeting quality C4: *High dimensional match is 1.51 times more predictive than TRPs for identifying high probability of buyer media.* (R^2 for tratio= 0.25, R^2 for TRP = 0.17; Figure 3 SourceViewPCT column).

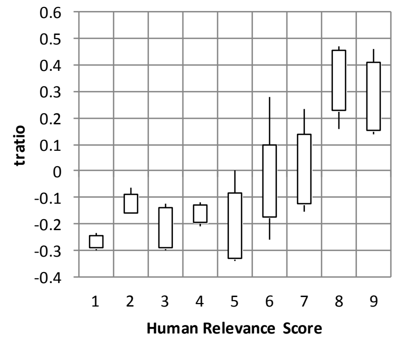


Fig. 2. Human relevance versus match score. Boxes show the 25^{th} and 75^{th} percentiles, and the needles show the 15^{th} and 85^{th} percentiles for variation in system targeting scores.

	HRS		SourceViewPCT	
Targeting method	AUC	R	AUC	R
TRP	0.5199	0.3798	0.5571	0.4086
High Dim Match (tratio)	0.5269	0.5798	0.5899	0.5025
Human Relevance	N/A	N/A	0.5728	0.2413

Fig. 3. Area Under the Curve (AUC) and Correlation Coefficient (R) for TRP versus High Dim Match in predicting either Human Relevance Score (HRS) or Set Top Box actual Buyer Probability (SourceViewPCT). For example, this shows that TRPs have a 0.3798 correlation coefficient wth human scores, where-as High Dim Match has a 0.5798 correlation coefficient.

Program name	High Dim Score	Rel rk	R rk	Df
WOMEN'S COLLEGE VOLLEYBALL	0.657378	3	9	-6
THE FRESH PRINCE OF BEL-AIR	0.38729	2	8	-6
THE NANNY	0.045733	1	5	-4
INSIDE EDITION	0.467253	5	9	-4

Fig. 4. Airings showing the largest discrepancy between human rating and automated targeting score. Humans rated the above media as having poor relevance, but the system rated it highly.

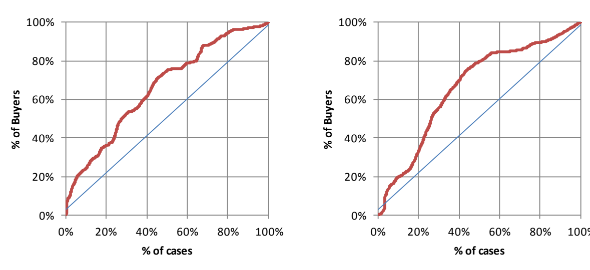


Fig. 5. Cumulative isolation of buyers (left) using human relevance, (right) using high dimensional match. The above graphs show the curves behind the Area Under the Curve calculations shown in Figure 3.

X. EXPERIMENT II: PHONE RESPONSE

In our next experiment we analyze performance on two live Phone Response Television campaigns. Phone Response campaigns embed a telephone number in the advertisement, as shown in Figure 6. The phone numbers are unique for each television station. When a viewer calls, they can be unambiguously attributed to a specific airing and station.

This is one of the few techniques that allows for an exact number of sales to be tracked for each airing, and so is an ideal test case for the system to see if it is effective at targeting to produce high phone response.

Campaign A (“Medical”; Figure 5) was an advertisement for a discretionary medical procedure. The campaign comprised of over 11,761 airings placed between May 2012 and January 2013, producing 121,628 phone calls. The ad budget for Campaign A was 2.456 million dollars.

Campaign B (“Charity”; Figure 5) was an advertisement for a charity in Africa to provide medical care for millions of people. This campaign consisted of 1,663 ad airings placed between October and December, 2012. The campaign generated 785 phone calls and spent 145,363 dollars.

Measure	Campaign A (Medical)	Campaign B (Charity)	Campaign C (Tool)	Campaign D (Music)
startdate	14-May-2012	1-Oct-2012	23-Apr-2010	12-Dec-2011
enddate	17-Jan-2013	31-Dec-2012	15-Dec-2012	2-Jan-2012
airings	11,761	1,663	1,849	7,569
spend	2,456,615	145,363	3,561,815	1,010,380
impressions	679,937,622	46,226,663	781,897,374	302,546,484
CPM	5.27	3.72	3.27	3.34
mean tratio	0.21	0.14	0.039	0.30
stdev tratio	0.23	0.18	0.258	0.16
min tratio	(0.58)	-0.24	-0.448	-0.63
max tratio	1.00	1.00	0.574	0.64
phone responses	121,628	785	0	0
Imp per TVHH	5.91	0.40	6.80	2.64
Imp per TVHH wk	0.17	0.03	0.049	0.88

Fig. 5. Live TV Campaigns measured in this paper.

Fig. 6. Screenshot from a TV ad from Campaign A. The phone number is shown at the bottom of the screen.

Both ad buys were scored using age-gender Target Rating Points (2) as well as High Dimensional match scoring (6).

Figure 9 and 10 show the percent of impressions versus percent of responses, when scores are sorted in order from highest score to lowest, and Figure 7 summarizes the AUC statistics. The Area Under the Curve (AUC) for TRPs in both cases was 0.50 and 0.47. The AUC for tratio was 0.56 and 0.58 respectively (Figure 7).

There are two immediate findings. C5: *High dimensional match produces lift in Phone response*. Across both campaigns, 27% of responses (27.3% and 26.7%) were isolated in the top 20% of impressions, resulting in a lift of 35.7% in the top 20% of airings. This is consistent with H1 and consumer studies [1]. However we have not seen the result for television phone response reported in the literature before, and we also note it here [13].

We would also note the paradoxical finding: C6: *In the 13,000 airings that we tested, TRPs performed worse than random at predicting high phone response airings.*

The poor performance of TRPs is surprising. One explanation of this performance is that we were only able to measure phone responses on airings that actually ran on TV. The placements that were aired were biased - they were better targeted than average (tratio=0.21 and 0.14). For example, Product A individuals have high income, and become very high probability of conversion at the point of retirement based on analysis of their demograhics. The age-gender TRPs were male, 65+ individuals. This includes people who aren’t likely to want the procedure. Thus in this space of higher targeted ads, it appeared that the age-gender TRPs did not have enough variables to generate further discriminatory gains.

	AUC	
Metric	Campaign A	Campaign B
High Dim Match (tratio)	0.5625	0.5833
TRP	0.4953	0.4734

Fig. 7. AUC for Campaign A and B in terms of ordering phone responses.

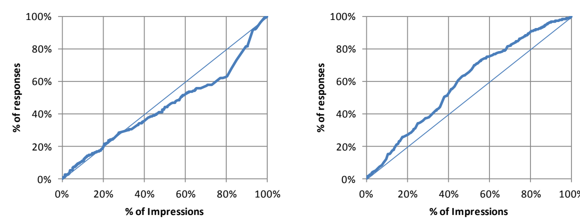


Fig. 9. Campaign B AUC for predicting phone response. (left) TRPs (right) tratio.

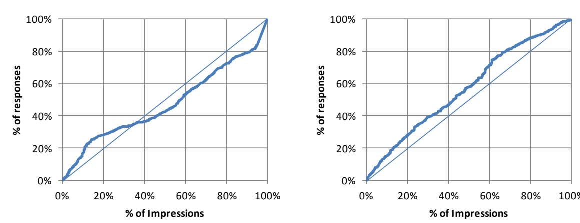


Fig. 10. Campaign A AUC for predicting phone response. (left) TRPs (right) tratio.

XI. EXPERIMENT III: BRAND

In our third experiment we used the method to target media for the Brand Television campaign for a large Internet Music Service company (Figure 5; “Music”).

Unlike the previous experiment, there was no phone response on these campaigns. However we were able to compare the media purchased during our deployment to previous media that had been run for this client.

The campaign ran between December 12 2011 and January 1 2012 and cost of $1.01 million dollars.

The internet company’s customers tended to be young adults, and the programs that were selected reflect that.

Figure 13 shows the demographics for the product buyers. Demographics that are high include “age=18-25”, “occupation=student”.

In order to measure the quality of targeting, we compared our automatically targeted media in December 2011 against the media that had been run by the client previously during the same month last year – December 2010.

Figure 9 shows the media that were purchased in the past (Plan I) versus new (Plan II). Previously SPIKETV, MTV, and COMEDY had occupied 25%, 20% and 18% of the buy respectively. These are young demographic stations, and so the previous targeting was thought to be reasonable.

Although similar on first glance, the Plan II had some surprising differences. Firstly the media was more diverse. We can define “Program Entropy” of the media plan as the entropy of the distribution of program impressions:

$$Ent(M) = -\sum_i \ln(b_i) \cdot b_i \text{ where } b_i = \frac{I(M_i)}{\sum_i I(M_i)} \quad (11)$$

Low entropy would suggest that most impressions are on a small number of large viewer programs such as “American Idol”. High entropy would suggest that a lot of different, small audience programs are being purchased. The program entropy for the automated buy was 4.43 versus 2.54 for the historical buy.

In addition the number of different programs purchased was also much higher – 823 distinct programs purchased in total versus 200 exactly in the previous plan. In addition the set of stations being purchased was also more diverse – programs from 30 national cable stations were purchased versus only 11 in the old plan.

Fig. 11. Screenshot from Campaign D TV ad.

Demographics Name	Demographics Description	Index Vs Avg
DOB – Year	Born in the 1990s	3.535626
Age	18-25	2.543058
Occupation – Professional	Student	1.639695
Young Men's Apparel	True	1.497625
DOB - Year	Born in the 1980s	1.453135
Occupation	Student	1.332086
Personicx Classic	Collegiate Crowd	1.308201
Personicx Classic	Early Parents	1.290695
Personicx Classic	Rolling Stones	1.282595
Personicx Classic	Young Workboots	1.277156
Prefers to Shop Over the Phone	10 Least Likely to Buy via Phone	1.220277
Age	26-35	1.146012
Personicx Classic	First Digs	1.131693
Ethnicity	African American	1.113605
Expectant Parent	True	1.104495
Personicx Classic	Kids and Rent	1.041225
2nd Vehicle Year	2010 model	1.034739
Home Purchase Year	Home Purchased Between 2010-2014	1.012979

Personicx Classic	Resillent Renters	0.9758
Discretionary Income	Lower Discretionary Income Index (15-29)	0.969286
Personicx Classic	Mobile Mixers	0.914647
Personicx Classic	Urban Scramble	0.850848
Personicx Classic	Resolute Renters	0.830117
Occupation - Professional	Nurses Aide/Orderly	0.809397
Length of Residence - Less than 1 year	Less than 1 year	0.78623
Number of Children	5 Children	0.777976
Young Women's Apparel	True	0.755491
Prefers to Shop Over the Phone	09	0.722032
Personicx Classic	Career Building	0.719719
Male 2 and Under	True	0.708939

Fig. 12. Top z-score demographic variables for Music service

Plan I

ProgramName	HHImps	tRatio	tCPM	HHCPM
COMEDY CENTRAL MOVIE	82444265	0.2979	24.21	7.21
SOUTH PARK	74960611	0.2077	50.62	10.51
SPIKE TV MOVIE	72494862	0.184	27.23	5.01
C.S.I.	68310462	0.0659	48.16	3.17
COUNTRY MUSIC THEATER	57121249	0	Inf	3.92
ULTIMATE FIGHTER UNLEASHD	50889456	0.2444	17.58	4.3
TNA: IMPACT	48363354	0.1784	40.35	7.2
TRUE LIFE	46179462	0.358	26.4	9.45
THE ULTIMATE FIGHTER	44198298	0.1535	24.67	3.79
EVERYBODY HATES CHRIS	43170528	0.0652	48.09	3.13
TOSH.O	41897036	0.2836	45.07	12.78
BET MOVIE OF THE WEEK	41809858	0.0893	68.1	6.08
1000 WAYS TO DIE	40922416	0.1231	48.4	5.96
16 AND PREGNANT 2B	40858336	0.348	42.29	14.72
DUKES OF HAZZARD	39597335	0.073	51.51	3.76
DEADLIEST WARRIOR	35659716	0.1303	33.96	4.42
TOP 20 COUNTDOWN	35091005	0	Inf	4.4
MANSWERS	34866207	0.3723	7.75	2.89
COLBERT REPORT	34306631	0.1228	54.11	6.65
DAILY SHOW	31349461	0.07	104.35	7.3
FUTURAMA	30965179	0.2894	32.59	9.43
JERSEY SHORE 2	29915701	0.2488	87.6	21.8
JERSEY SHORE 2B	29414903	0.2467	121.71	30.03
MTV MOVIE	28319007	0.2885	44.55	12.85
GANGLAND	27914833	0.2288	11.51	2.63
JERSEY SHORE	27440822	0.2506	68.01	17.04
GEORGE LOPEZ	22733278	0.0315	99.14	3.12
SMASH HITS OF COUNTRY	21756475	0.2023	19.33	3.91
HILLS, THE	21451935	0.307	39.04	11.99
FAMILY MATTERS	17950800	0.1417	9.04	1.28
WALKER, TEXAS RANGER	16223193	0		2.4
NANNY	15776709	0.0678	62.3	4.22
COMEDY CENTRAL ROAST	15635185	0.2069	43.57	9.02
TEEN MOM 2	15632710	0.3063	59.55	18.24
COMEDY CENTRAL PRESENTS	15454652	0.2338	31.51	7.37
16 AND PREGNANT 2	14886124	0.3545	37.45	13.28
106 & PARK	13963540	0.1629	38.78	6.32
R&R PICTURE SHOW	13610421	0.254	22.05	5.6

Plan II

ProgramName	HHImps	tRatio	tCPM	CPM
MOVIE	40,212,194	0.3063	$15.59	$4.78
FAMILY GUY	28,019,283	0.2900	$7.98	$2.32
Show Information Not Available	26,523,501	0.2671	$10.73	$2.87
SPORTSCENTER	25,355,082	0.3373	$8.86	$2.99
AMERICAN DAD	7,619,082	0.2902	$10.86	$3.15
KOURTNEY & KIM TAKE NEW YORK	7,413,552	0.3634	$15.49	$5.63
TOSH.O	6,561,872	0.2836	$27.55	$7.81
COLLEGE FOOTBALL LIVE	5,455,771	0.2074	$0.88	$0.18
AQUA TEEN HUNGER FORCE	5,077,624	0.2736	$10.66	$2.92
WOMEN'S COLLEGE BASKETBALL	3,698,276	0.1421	$0.00	$0.00
CHELSEA LATELY	3,323,716	0.1594	$24.70	$3.94
TODDLERS & TIARAS	3,146,502	0.1990	$31.31	$6.23
ROBOT CHICKEN	3,074,566	0.3018	$5.12	$1.54
ALWAYS SUNNY IN PHILADELP	2,933,984	0.1481	$42.74	$6.33
TEEN MOM 2	2,784,380	0.2816	$6.71	$1.89
LOVE & HIP HOP	2,661,639	0.3729	$6.33	$2.36
COLBERT REPORT	2,623,649	0.1233	$18.48	$2.28
TOP 40 OF 2011	2,621,971	0.3875	$4.49	$1.74
CAKE BOSS: NEXT GREAT BAKER	2,620,429	0.2328	$27.21	$6.33
EVERYBODY LOVES RAYMOND	2,562,169	0.0000	Inf	$2.79
BLEACH	2,510,849	0.1768	$10.48	$1.85
THE DAILY SHOW WITH JON STEWART	2,389,397	0.3158	$14.58	$4.60
SCOUTED	2,376,899	0.3782	$13.27	$5.02
SCRUBS	2,328,237	0.3075	$16.71	$5.14
SUNDAY NFL COUNTDOWN	2,292,641	0.2046	$23.96	$4.90
SOUTH PARK	2,258,314	0.2077	$28.81	$5.98

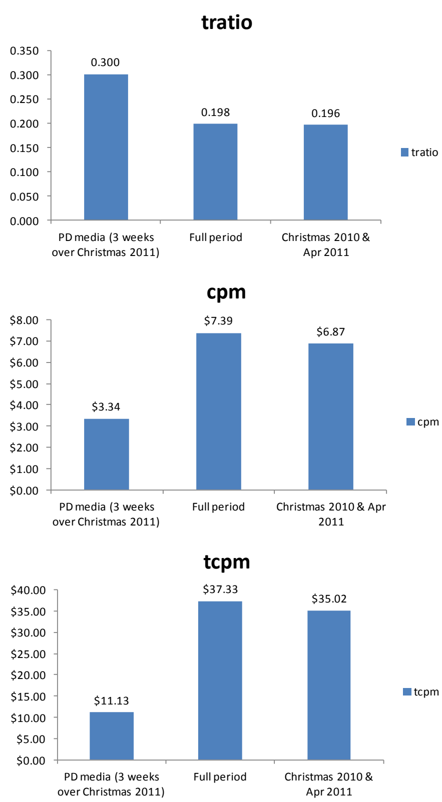
106 & PARK: BET'S TOP 10 LIVE	2,214,848	0.1680	$30.10	$5.06
HIGHLIGHT EXPRESS	2,174,786	0.2219	$11.78	$2.61
E! NEWS	2,109,953	0.2144	$26.04	$5.58
COMEDY CENTRAL PRESENTS	2,026,233	0.2338	$24.70	$5.77
YES, DEAR	1,944,001	0.1240	$37.31	$4.62
TWO & A HALF MEN	1,927,270	0.0173	$213.78	$3.70
FUTURAMA	1,911,040	0.2896	$19.13	$5.54

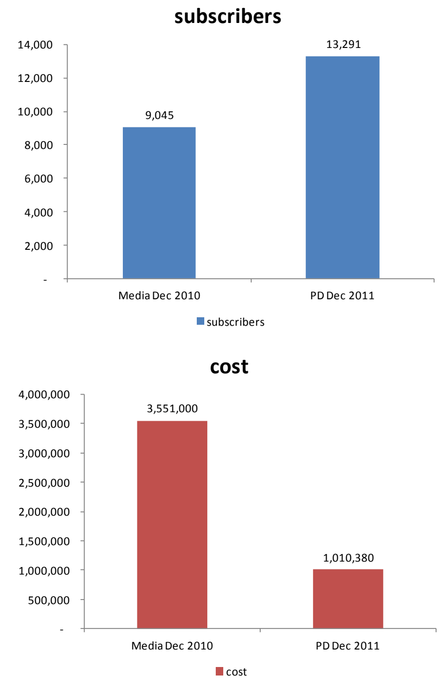
Fig.13. Media purchased under Plan I (TRPs) versus Plan II (High Dimensional match). HHImps = Household impressions, TImps = impressions * r(), tratio = *r*, tCPM = CPM/r, CPM = cost per thousand impressions, airings = number of times that an advertisement ran

We also examined the targeting score (6) and cost of the previous buy versus the campaign (Figure 16). The media that Plan II selected was 53% higher in targeting quality and 51% lower in Cost Per Impression (CPI).

These statistics indicated that the campaign was better targeted and far lower cost. However what overall effect did this have on revenue? In order to answer this question we used a local market treatment and control design which is beyond the scope of the current paper, but which is used by the advertiser for measuring lift from television and is described in detail in [20].

Subscribers generated in the targeted campaign increased from 9,045 to 13,291 (Figure 12). Cost was even more dramatic (and this is not a modeled quantity). The previous campaign cost over $3.55 million dollars. The new campaign cost only about $1.01 million.





media	Plan II Dec 2011	Plan I Dec 2010
subscribers	13,291	9,045
Cost	1,010,380	3,551,000
CPA	76	393
Seasonality	1.21	1.82
Imp/MHH/Wk	879	971
CPM	3.34	7.62
tratio	0.300	0.196
tCpm	11.13	38.09
Program count	200	823
Program entropy	2.54	4.43
Station count	10	30

Fig. 16. ROI statistics for two campaigns – Dec 2010 and Dec 2011

XII. Discussion

We have discussed a method for targeting television using new data available from Set Top Box. The targeting measure has been shown to be correlated with Human relevance (C1), Phone response per impression (C5) and Buyers per impression (C3). We have also compared the measure to existing targeting methods based on widely used Nielsen Target Rating Points (TRPs). Because of advantages in terms of the number of variables being used for matching and the number of Set Top Boxes, in our experiments the STB method out-performed TRPs on Human relevance (C2), Phone response per impression (C6), and Buyers per impression (C4).

TRPs are still a useful method of targeting and we have shown in our experiments that they are also predictive of human relevance and buyer probability. However our results suggest that the new methods that are possible due to Set Top Box data appear to be capable of generating higher amounts of lift (C2, C3, C4), and logically this seems likely due to the larger number of variables and persons. Given the rapid increase in the availability of Set Top Box data in the United States, we believe that Set Top Box targeting will become a common part of future TV campaigns.

REFERENCES

[1] J. Aaker, A. Brumbaugh, S. Grier, "Nontarget markets and Viewer Distinctiveness: The Impact of Target Marketing on Advertising Attitudes", Journal of Consumer Psychology, Vol. 9, No. 3, pp. 127-146. Lawrence Erlbaum Associates Inc. 2000.

[2] Acxiom Corporation, http://www.acxiom.com, 2013.

[3] C. Anders, "How the Nielsen TV ratings work — and what could replace them", IO9.com, http://io9.com/5636210/how-the-nielsen-tv-ratings-work--and-what-could-replace-them, 2010.

[4] L. Ardissono, L. Console, I. Torre, "An adaptive system for the personalized access to news", AI Communications, Vol. 14, No. 3, pp. 129-147. 2001.

[5] N. Arora, J. Hess, Y. Joshi, S. Neslin, J. Thomas, "Putting one-to-one marketing to work: Personalization, customization, and choice", Marketing Letters, Vol. 19, pp. 305-321. 2008.

[6] S. Balakrishnan, S. Chopra, D. Applegate, S. Urbanek, "Computational Television Advertising", Twelfth IEEE International Conference on Data Mining, pp. 71-80. 2012.

[7] T. Bozios, G. Lekakos, V. Skoularidou, K. Chorianopoulos, "Advanced Techniques for Personalized Advertising in a Digital TV Environment: The iMEDIA System", Proceedings of the eBusiness and eWork Conference, 2001.

[8] B. Canning, P. Bochman and Z. Faro, "Using Consumer Purchase Behavior For Television Targeting", US Patent 8,060,398 B2, http://www.google.com/patents/US8060398, 2009.

[9] K. Chorianopoulos, G. Lekakos and D. Spinellis, "The Virtual Channel Model for Personalized Television", Proceedings of the 1st EuroITV conferece: From Viewers to Actors pp.59-67. 2003.

[10] A. Cota and K. Dion, "Salience of gender and sex composition of adhoc groups: An experimental test of distinctiveness theory", Journal of Personality and Social Psychology, Vol. 50, No. 4, pp. 770-776. 1986.

[11] C. Dawson, "Television Advertising: In Need of Reinvention?", International Journal of Marketing, Vol. 16, No. 4. 1996.

[12] R. Deshpande and D. Stayman, "A tale of two cities: Distinctiveness theory and advertising effectiveness", Journal of Marketing Research, Vol. 31, pp. 57-64. 1994.

[13] Direct Response Academy, "Advanced DRTV Media: Direct Response Television Media Management", Training Course materials on Direct Television media buying, http://www.directresponseacademy.com, 2008.

[14] Engage Research, "Inside the Cave: An In-Depth Look at the Digital, technology and Analytics Operations of Obama for America", 2012.

[15] J. Fisher, "WPP Sues TRA Global Over Analysis System Claims infringement of proprietary method", AdWeek, June 17, 2011, http://www.adweek.com/news/technology/wpp-sues-tra-global-over-analysis-system-132633

[16] M. Forehand and R. Deshpande, "What We See Makes Us Who We Are: Ad Typicality as a Source of Elicited Distinctiveness", Journal of Marketing Research Vol. 38, August. 2001.

[17] I. Hawkins, R. Best, K. Coney, Consumer behavior: Building marketing strategy, Irwin/McGraw-Hill, NY. 1998.

[18] S. Harizopoulos, D. Abadi, P. Boncz, "Column-Oriented Database Systems", VLDB 2009 Tutorial. 2009.

[19] IAB Platform Status Report: Interactive Television Advertising, http://www.iab.net/media/file/ITV_Platform_Status_Report.pdf, 2012.

[20] B. Kitts, D. Au and B. Burdick, "Real-time Television ROI Tracking using Mirrored Experimental Designs", Data Mining Applications in Industry and Government Workshop, Proceedings of the Seventeenth Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2013.

[21] P. Kotler, G. Armstrong and R. Starr, Principles of marketing, Englewood Cliffs, NJ: Prentice Hall. 1991.

[22] W. Mcguire and C. McGuire, "Search for the Self: Going beyond self-esteem and the reactive self", in. R. Zucker, R. Aronoff, and A. Rabin (eds.), Personality and the prediction of behavior, pp. 73-120, Academic Press, New York. 1984.

[23] W. McGuire, C. McGuire, P. Child, T. Fujioka, "Salience of ethnicity in the spontaneous self-concept as a function of one's ethnic distinctiveness in the social environment", Journal of Personality and Social Psychology, Vol. 365, pp. 511-520. 1978.

[24] Nielsen Corporation, "A2/M2 Three Screen Report: Television, Internet and Mobile Usage in the United States", 4th Quarter 2008.

[25] Nielsen Corporation, "Americans watching more TV than ever", Nielsen Wire, May http://blog.nielsen.com/nielsenwire/online_mobile/americans-watching-more-tv-than-ever/ 2009.

[26] Nielsen Corporation, Nielsen Competitive Ad Occurrence Data.

[27] M. Schneider, "Fox wants answers from Nielsen", http://www.variety.com/article/VR1118003924?refCatId=14 , *Variety*, May 18, 2009.

[28] A. Segal, "Nielsen Ratings: An Inaccurate Truth Out of date television ratings system exposed", http://cornellsun.com/node/23180, The Cornell Daily Sun, April 26, 2007.

[29] B. Smyth and P. Cotter, "A Personalized Television Listings Service", Communications of the ACM, Vol. 43, No. 8, August 2000, pp. 107-111. 2000.

[30] W. Spangler, M. Gal-Or, J. May. "Using data mining to profile TV viewers", Communications of the ACM archive, Volume 46, Issue 12, pp. 66 – 72. December, 2003.

[31] G. Tellis, R. Chandy, D. MacInnis and P. Thaivanich, "Modeling the Microeffects of Television Advertising: Which Ad Works, When, Where, for How Long, and Why?", Marketing Science, Marketing Science, Vol. 24, No. 3, pp. 351–366, 2005.

[32] US Census Bureau, http://www.census.gov/2010census/ 2010

[33] T. Whittler, "The effects of actors' race in commercial advertising: Review and extension", Journal of Advertising, Vol. 20, pp. 54-60. 1989.

[34] J. Williams and W. Qualls, "Middle-class Black consumers and intensity of ethnic identification", Psychology and Marketing, Vol. 6, pp. 263-286. 1989.

[35] B. Kitts, D. Au, B. Burdick, J. Borchardt, A. Powter and T. Otis, "Demand Finder: Set Top Box Television Ad Targeting using a Novel Interactive Data Visualization System", Proceedings of the Thirteenth IEEE International Conference on Data Mining Workshops, December, Dallas, TX, IEEE Press. 2013

[36] S. Miles, "6 Tools for Targeting Brand Influencer Look-Alikes", Street Fight Magazine, August 26, 2013. http://streetfightmag.com/2013/08/26/6-tools-for-targeting-brand-influencer-look-alikes/

[37] A. Abramson, The History of Television, 1942 to 2000. Jefferson, NC: McFarland & Co.