Estimating Campaign Benefits and Modeling Lift

Gregory Piatetsky-Shapiro
Knowledge Stream Partners
Boston, MA 02109
gps at kstream.com

Brij Masand
GTE Laboratories
Waltham, MA 02451-1128
brij at gte.com

Abstract

In assessing the potential of data mining based marketing campaigns one needs to estimate the payoff of applying modeling to the problem of predicting behavior of some target population (e.g. attriters, people likely to buy product X, people likely to default on a loan, etc). This assessment has two components: a) the financial estimate of the campaign profitability, based on cost/benefit analysis and b) estimation of model accuracy in the targeted population using measures such as lift.

We present a methodology for initial cost/benefit analysis and present surprising empirical results, based on actual business data from several domains, on achievable model accuracy. We conjecture that lift at ***T*** (where ***T*** is the target frequency) is usually about sqrt(1/***T***) for a good model. We also present formulae for estimating the entire lift curve and estimating expected profits.

Keywords: database marketing, estimation, lift

1. INTRODUCTION

Direct marketing is a common area for applying data mining [1, 4,5]. The goal is to predict a specific behavior of the customer, such as buying a product, attriting (churning) from a service, defaulting on a loan, etc. If a company can identify a group of customers where the target behavior is more likely, (e.g. group more likely to churn), then the company can conduct marketing campaigns to change the behavior in the desired direction (e.g. decrease churn). If targeting criteria are well chosen for a direct mail campaign, the company can mail to a much smaller group of people to get the same number of responses. The increased concentration of the 'right targets' (e.g. churners or responders) in such targeted campaigns enables increased ROI.
In assessing the potential of such campaigns one may want to estimate certain financial parameters such as profitability or relative increase in the concentration of the targets (lift) in the targeted group. In this paper we ask whether and when it is possible to quickly estimate parameters such as lift based on the problem features before attempting the complex task of modeling. We define guidelines for when one might consider deploying marketing campaigns and propose estimation formulae for lift at an important fixed point and over the entire lift curve.

In sections 2 and 3 we define the basic terminology and guidelines for conducting such an assessment. In section 4 we investigate the relationship between lift and the target frequency. In section 5 we analyze how the lift decreases with respect to increasing subsets of population and in section 6 we use the lift curve and cost estimates to derive an estimate for expected maximum profit. We conclude with a discussion of the limits of these heuristics and possible directions for extending this inquiry.

2. ESTIMATING CAMPAIGN PAYOFF

The key parameters for the assessment of a targeted marketing campaign are:

- ***N*** -- the total number of customers (e.g. 1,000,000)
- ***T*** -- the fraction of target customers, (e.g. 0.02) who have the desired behavior (e.g. response to mail offers)
- ***B*** -- Benefit of an accepted offer ***A*** by a customer correctly identified as a target (e.g. $50)
- ***C*** -- Cost of making an offer ***A*** to a customer, whether a target or not, e.g. $5)

Let ***Profit***(***P***) be the profit of making offer to ***P*** percent of all customers (where $0 < P < 1$). The profit for making an offer to all customer is

(1) ***Profit(1.0) = NTB - NC = N (TB - C)***

Using values for ***N***, ***T***, ***B***, and ***C*** as above, the offer has an estimated profit of -$400,000, i.e. making an offer to everybody is unprofitable.

Of course, costs and especially benefits are not uniform, see [9], but this method is useful for an initial estimate. Also there may be a probabilistic aspect to the offer (whether it is accepted or not) which can be folded into the estimated benefit or cost values for simplicity.

Note that whether the profit is positive does not depend on N, B, or C but only on whether $TB - C > 0$ or

(2) $TB / C > 1$

i.e. only on the benefit/cost ratio and the target frequency.

We also note that even when contacting everybody is profitable, it may not be the most profitable course. One could improve the profit by selecting a subset of customers, as examined in the next section.

The Promised Land

The promise of data mining is that it can find a subset P_2 (of size $N P_2$) of the population where the fraction of targets T_2, is ficiently higher than in the overall population so that making an offer to P_2 is *more* profitable than making an offer to all.

We can use the formula (1) to estimate the profit of making an offer A to P_2

(3) $Profit(P_2,T_2) = NP_2 (T_2 B - C)$

The standard measure for comparing the target frequency of a subset with the target frequency in the entire population is the lift, defined as $Lift (P_2) = T_2 / T$.

We can rewrite (3) as

(4) $Profit(P_2,T_2,) = NP_2 (T^* Lift(P_2) B - C)$

Again we note that whether making an offer to a subset P_2 is profitable depends only on whether $T^* Lift(P_2) B > C$. We can formulate it as

Campaign Profitability Condition:

(5) $Lift(P_2) > C/BT$

This gives us a rule of thumb for estimating the lift we need to achieve to make the campaigns profitable.

For example, if T is 0.02, B is \$50, and C is \$5, then we need to select a subset with lift > 5 to achieve profit.

We should also note that condition (5) is the ***minimum*** condition of profitability, which defines the largest subset which is still profitable. It does not define the optimal subset with the highest profit -- such a bset is the one that maximizes equation (4).

The following sections discuss estimating of lift and selection of optimal subset.

3. ESTIMATING LIFT

To compute the actual lift data miners go through the lengthy process of examining the data available, selecting the data, obtaining permissions, transferring the data, obtaining the meta-data, cleaning the data, modeling etc. The initial steps of data preprocessing and cleaning is often very time consuming [2] and may take a couple of months before even the first model appears.

Can we say something about the expected lift for typical business problems such as attrition and cross-sell before we do the modeling?

In general, the answer is no, since every problem is different and has different underlying patterns. Furthermore, different modeling algorithms generally produced different results [6]. However, we examined a number of actual business from the financial and telecommunications arena and found a surprising result – it appears that there is a heuristic formula for estimating the lift.

4. LIFT CONJECTURE FOR LIFT AT *T*

In this section we present an empirical observation about the typical lift one can see in campaign-type data mining problems, where we need to classify customers into two types: target and non-target population. We are not claiming that every model will have a lift similar to the formula below – one can easily come up with a poor or random model with worse results. What we are proposing is that for a certain class of target marketing problems there is a rough estimate for the *best* lift one can expect.

4.1 Lift vs accuracy

First we address the question: Why do we focus on lift and not overall accuracy?

A modeling method, such as neural network or a decision tree typically build a model on a training set to predict the target probability and is further evaluated on a separate validation set. The validation set is then sorted in descending order by the target probability. The frequencies of actual and predicted target values are then measured at some set intervals, usually in 5% increments. Generally speaking, if the modeling method is successful, we should find a higher concentration of targets at the top of the list and this higher proportion of targets can be measured in terms of “lift” to see how much better than random the model-based targeting is. While accuracy measures correct predictions for the whole population, lift measures the increased accuracy for a targeted subset, e.g. the top part of a model-score ranked list. Therefore it is very possible [5] that a modeling method with lower accuracy can result in higher lift at the top of the list. To produce a better lift, one doesn't need to have higher accuracy over all targets but only over a sufficiently large subset.

Since the goal of the target marketing is not to predict the customer behavior for everybody, but find a good *subset* of customers where the target behavior has a higher proportion, we are interested in models that maximize lift and not the overall accuracy. Analyzing the lift curve and lift values is thus very important for target marketing applications.

4.2 Definition of lift for a subscriber list ranked by model-score

Let $\mathit{Lift}(P,T)$ be the lift obtained by selecting the top P percent of the customer list sorted by a model score. For example if T is 0.02 there are 2% targets in the overall population. Then if the top 5% of the list has a 10% concentration of targets (or 0.1 as a fraction) then $\mathit{Lift}(5\%)$ is 0.1/0.02 = 5. Alternately lift can also be calculated by looking at the cumulative targets captured up to P% as a *percentage of all targets* and dividing by P%. For example if the top 5% of the list captures 20% of *all targets* then lift is 20/5 = 4. Lift is a common measure of how well model selection does compared to a random selection.

Note: If we don't use percentage after P, then we are using P as a fraction. So $\mathit{Lift}(5\%,T)$ is the same as $\mathit{Lift}(0.05,T)$

Table 1 shows a representative lift table from the KDD-Cup 97 [7], generated by co-winner BNB [3]

Pct	Records	Cum Hits	Hits%	CumPHits %	Lift
5	7531	448	5.95	21.9	4.38
10	15062	708	4.70	34.6	3.46
15	22592	897	3.97	43.9	2.92
20	30123	1012	3.36	49.5	2.47
25	37654	1158	3.08	56.6	2.26
30	45185	1265	2.80	61.8	2.06
35	52716	1348	2.56	65.9	1.88
40	60246	1433	2.38	70.0	1.75
45	67777	1539	2.27	75.2	1.67
50	75308	1610	2.14	78.7	1.57
55	82839	1673	2.02	81.8	1.49
60	90370	1721	1.90	84.1	1.40
65	97900	1782	1.82	87.1	1.34
70	105431	1833	1.74	89.6	1.28
75	112962	1870	1.66	91.4	1.22
80	120493	1910	1.59	93.4	1.17
85	128024	1948	1.52	95.2	1.12
90	135554	1994	1.47	97.5	1.08
95	143085	2017	1.41	98.6	1.04
100	150616	2046	1.36	100.0	1.00

Table 1. Kdd-Cup 97 data for BNB program

For example, the second line in this table means that at 10% of the model-score sorted list, there are 15062 records, of which 708 are "hits" – targets correctly identified. The cumulative accuracy in the first 10% is 708/15062= 4.70%. 708 hits are also 34.6 percent of all targets, giving a lift of 34.6/10 = 3.46. Here N = 150616, targets= 2046, T=1.36%.

4.3 Relationship between Lift and T

By definition lift at 100% is always 1, i.e. $\mathit{Lift}(100\%,T) = 1$. We also observe that the maximum possible lift is $1/T$.

Since the data mining problems are all different it is very difficult to make general statements about the lift curve behavior. However, based on our experience in several business domains and observations of similar results obtained by others, we can make several heuristic observations about the lift curve in targeted marketing problems.

(1) As P increases from 0 to 100, lift is usually monotonically decreasing with increasing P

$$\mathit{Lift}(P_1,T) > \mathit{Lift}(P_2,T), \text{ for } P_1 < P_2$$

This is truer for neural net models, and is less so for decision tree models, especially at the top (first 5-10%) of the lift curve.

(2) As P increases from 0 to 100%, lift is decreasing at a slower than linear rate, i.e.

$$\mathit{Lift}(P,T) < 2\,\mathit{Lift}(2P,T)$$

(3) We also observe comparing lift at a fixed point for two problems with different T, we usually find that the lift is higher for problems with smaller T.

$$\text{Usually if } T_1 < T_2 \text{, then } \mathit{Lift}(P, T_1) > \mathit{Lift}(P, T_2)$$

Typically, in literature lift is reported in 5% increments. Since we previously observed that lift at a fixed percentage like 5% depends heavily on T, we decided to neutralize the effect of varying T by measuring lift at T %. If the model was perfect, than all the TN targets would be concentrated in the initial $T\%$ of the list, giving this segment the lift of $1/T$. In practice the lift at T is usually much lower than that. We have collected data from a number of problems from our actual experience in building models in telecommunications and financial.

4.4 Analysis of Lift from Real Data

Table 2 shows a few examples from our recent practice. In these cases, lift is computed on a separate validation (25-40% of all data) set that was not used in any way for training and which has the same proportion of the targets as the full data set.

Since ***T*** is usually not a round number, precise lift at ***T*** % was usually not available and was estimated by linear interpolation of lifts at lower and higher percentages,

i.e. if $P_1 < T < P_2$, and lift at P_1 is L_1 and lift at P_2 is L_2, then lift at ***T*** is estimated as

$$L_1 + (L_2 - L_1)(T - P_1) / (P_2 - P_1)$$

For almost all of the problems below the lift was computed in 1% increments, which reduces lift interpolation error.

We have experimented with neural net packages (Predict from Aspen Technology), decision trees (C4.8, C5.0, and CART), and several version of Bayesian methods. In general, lifts are comparable within the rough guidelines we are looking, but for decision trees and bayesian models the lifts at the top of the list are usually smaller than for neural nets. We explain this by decision trees being less sensitive to precise patterns with smaller support, which neural nets can pick up and which can influence lift at the top of the list.

Since the lift curve is decreasing at less than the linear rate, we looked at the relationship between ***Y*** = log10(***Lift***) and ***X*** = log10(1/***T***) (or log10(100/***T***%) if ***T***% is expressed as a percentage) and computed a regression between these terms.

The regression produced a formula (with $R^2 = 0.86$)

$$(6) \quad \log10(Lift(T,T)) = -0.0496 + 0.518 \log10(1/T)$$

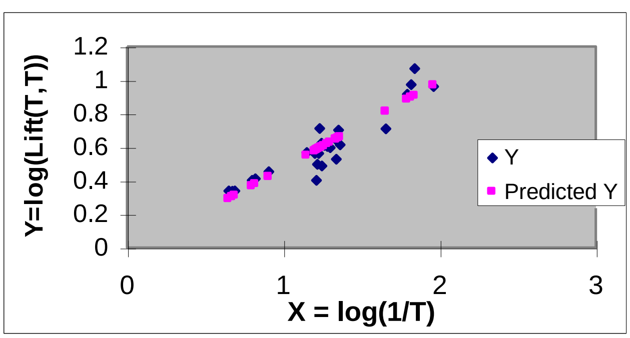

Fig 1. log(1/T) vs log(Lift(T,T))

Figure 1 shows the actual and estimated (using formula 6 above) values of log10(***Lift***(***T***,***T***)) for the cases in table 2. The log-log regression shows a good fit, especially for smaller values of lift.

Problem	**Method**	**N**	**T (%)**	**Lift(T,T)**
ad-980819	Nnet	7091	2.20	5.11
ad-9810-4245	Nnet	38313	1.51	9.39
ad-9902-28757	Nnet	29375	1.44	11.70
ad-gm-990224	Nnet	11190	1.60	8.21
ad-sc-990224	Nnet	4937	1.09	9.13
bx-mo25	Nnet	1225	22.28	2.18
bx-mo75	Nnet	1285	20.39	2.17
Xs-E	DT	4636	12.40	2.96
Xs-T	DT	6187	21.15	1.99
Xs-Z	DT	4636	15.80	2.92
Xs-H	DT	11709	15.10	2.30
Xs-BC	Nnet	150000	5.70	5.50
sr1_m31	Nnet	16400	7.07	3.69
sr1_m35	Nnet	12200	4.67	3.36
sr1_m37	Nnet	32900	4.33	4.10
sr2_m03	Nnet	23900	4.59	4.39
sr2_m05	Nnet	10100	5.22	4.02
sr3_m03	Nnet	16600	6.05	3.14
sr3_m05	Nnet	10900	5.94	3.65
sr3_m10	Nnet	22500	5.47	4.07
sr4_m26	Nnet	20900	6.27	3.64
sr4_m32	Nnet	17700	5.84	5.13
sr5_m04	Nnet	26800	6.13	2.52
sr5_m06	Nnet	71100	4.43	5.01
sr5_m07	Nnet	86900	5.01	3.94
sr5_m08	Nnet	19900	5.67	3.07

Table 2: Lift Statistics from actual Targeted Marketing Problems

A simplification of this formula is

$$\log10(Lift(T,T)) = 0.5 \log10(1/T)$$

or

$$(7) \quad Lift(T,T) = \mathrm{sqrt}(1/T)$$

We tested the simplified formula on the same cases, with results shown in Table 2b. Where

Estimated lift at ***T*** = sqrt (1/***T***) = sqrt(100/***T***%)

Error = Actual - Estimated Lift

Relative error = (Error/Actual Lift), in %.

Overall, we see a very good fit, with correlation of actual lift to sqrt(1/***T***) of 0.918. We also observe that the simplified formula is usually within 20% of the actual lift. This led us to

GPS Lift Rule of Thumb:

For targeted marketing campaigns, a good model lift at T, where T is the target rate in the overall population, is usually sqrt(1/T) +/- 20%.

Figure 2 shows a graph of actual lift at ***T*** vs. estimated by the sqrt(1/***T***) formula.

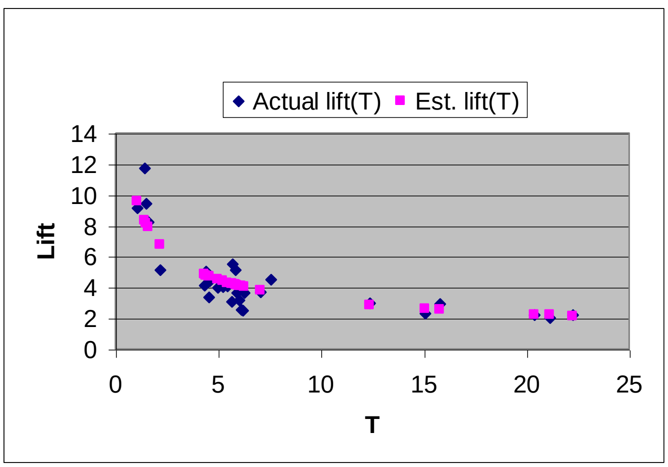

Figure 2. Actual lift at T vs Estimated

4.5 Discussions of possible sources of Exceptions

The exceptions to these rules fall into at least four categories.

First, there are poor modeling methods that produce bad results and lower lifts.

Second, there are classes of problems where either target behavior is too random (e.g. predicting lottery numbers) or good data is not available. Almost all of our examples deal with predicting customer behavior using previous examples of same or related customer behavior, e.g. predicting attrition using previous examples of attrition. We also had experience with models that try to predict customer behavior using purely external demographic data that is not related to targeted product or service and is not frequently updated. In such cases lift is typically less than the GPS lift of sqrt (1/***T***).

Third, we sometimes find a lift that is much higher. A frequent cause of very high lift is having information leakers. Those are fields that are functionally dependent on the target behavior but reflect customer behavior that happens after or at the same time as the target behavior. For example, suppose we are predicting which customers are likely to stop paying the mortgage. A field which records non-payment of insurance would be a leaker, since insurance and mortgage are frequently paid on the same bill.

Problem	**Num**	**T%**	**Act. lift(T)**	**Est. lift(T)**	**Error**	**Rel. Error**
ad-980819	7091	2.2	5.11	6.74	-1.63	-31.83%
ad-9810-4245	38313	1.52	9.39	8.12	1.27	13.54%
ad-9902-28757	29375	1.44	11.70	8.32	3.38	28.85%
ad-gm-990224	11190	1.61	8.21	7.88	0.33	3.99%
ad-sc-990224	4937	1.09	9.13	9.56	-0.43	-4.71%
bx-mo25	1225	22.29	2.18	2.12	0.06	2.89%
bx-mo75	1285	20.39	2.17	2.21	-0.04	-1.84%
Xs-E	4636	12.4	2.96	2.84	0.12	4.06%
Xs-T	6187	21.15	1.99	2.17	-0.18	-9.27%
Xs-Z	4636	15.80	2.92	2.51	0.40	13.84%
Xs-H	11709	15.10	2.3	2.57	-0.27	-11.89%
Xs-BC	150000	5.70	5.5	4.19	1.31	23.84%
sr1_m31	16400	7.073	3.69	3.76	-0.07	-1.84%
sr1_m35	12200	4.574	3.36	4.67	-1.31	-39.00%
sr1_m37	32900	4.33	4.1	4.81	-0.71	-17.32%
sr2_m03	23900	4.59	4.39	4.67	-0.27	-6.14%
sr2_m05	10100	5.228	4.02	4.37	-0.35	-8.75%
sr3_m03	16600	6.05	3.14	4.07	-0.93	-29.62%
sr3_m05	10900	5.94	3.65	4.1	-0.45	-12.33%
sr3_m10	22500	5.471	4.07	4.27	-0.20	-4.99%
sr4_m26	20900	6.278	3.64	3.99	-0.34	-9.37%
sr4_m32	17700	5.84	5.13	4.14	0.99	19.30%
sr5_m04	26800	6.13	2.52	4.04	-1.52	-60.32%
sr5_m06	71100	4.43	5.01	4.75	0.26	5.28%
sr5_m07	86900	5.01	3.94	4.46	-0.52	-13.17%
sr5_m08	19900	5.67	3.07	4.20	-1.13	-36.81%
Average		***7.589***	***4.513***	***4.599***	***-0.085***	***-7.05%***
StDev		***6.220***	***2.465***	***1.982***	***1.026***	***20.4%***

Table 2b: Actual and Estimated Lift Statistics

We found that such leakers are especially prevalent in cross-sell or cloning problems, when one tries to model look-alikes of customers with product X to find potential new customers for X. Frequently, having product X is related to several other fields in data, and those fields manifest themselves by contributing to very high lifts.

Finally, there are examples of truly predictable outcomes. A very strong rule like

if bill is not paid for 90 days,
then service is terminated with probability 100%

is likely to represent the company policy to terminate service for non-paying customers. Since such rules are usually known, data satisfying them should be excluded from modeling.

5. ESTIMATING THE LIFT CURVE

Here we extend the previous results from a point estimate for a lift curve to an approximate formula for the entire curve. The analysis is simplified by looking at ***CumPHits(P)***, an intermediate measure related to lift, but with nicer mathematical properties. ***CumPHits(P)*** is defined as the cumulative percentage of hits (targets correctly identified) in the first ***P*** percent of the model-sorted list and is related to lift by

$$Lift(P) = CumPHits(P) / P$$

Note: in the rest of this paper ***P*** denotes the percentage of the list expressed as a fraction (between 0 and 1). From the definition of ***CumPHits***, we observe that

$$CumPHits(0) = 0$$

$$CumPHits(1) = 1$$

From the observation that ***Lift(P)*** is decreasing monotonically with ***P*** increasing, but at a slower rate than ***P***, we infer that

For ***P*** increasing from 0 to 1, ***CumPHits***(***P***) is usually monotonically increasing with ***P***

Finally, based on the previous section, we are looking for a formula consistent with

$$Lift(T) \simeq \text{sqrt}(1/T) = T^{-0.5}$$

which is equivalent to

$$CumPHits(T) = T * Lift(T) \simeq T * T^{-0.5} = \text{sqrt}(T)$$

We examined a number of lift tables used in the previous section and compared regression of ***P*** vs ***CumPHits***, ***P*** vs log(***CumPHits***), log(***P***) vs log(***CumPHits***). Again, the best results were obtained from the log vs log regression.

Here are the results from performing regression on 15 of the problems above (for some problems we were unable to do the regression since we did not have the full lift table). We were looking for the regression

(8) $$\log10(CumPHits) = a + b \log10(P)$$

For the KDD-CUP-97 data (Table 1), log log regression gives $a = 0.027$, $b = 0.489$, and $R^2 = 0.988$.

Fig. 3 shows a plot of actual and estimated ***CumPHits*** for the KDD-CUP-97 data.

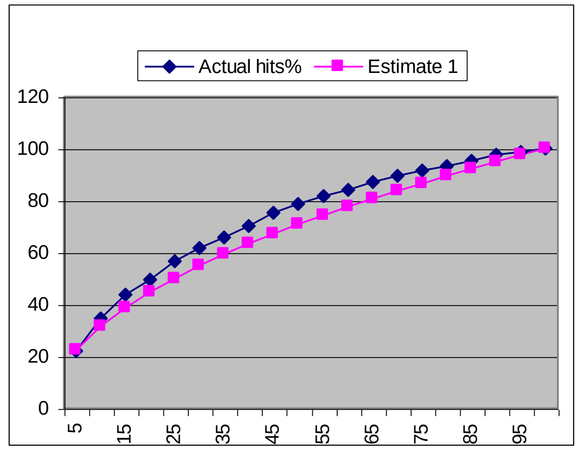

Fig. 3 Actual Cum hits% vs. Estimate

Problem	N	T%	a	b	R^2
ad-980819	7091	2.2	0.0868	0.5708	0.9592
ad-9810-4245	38313	1.516	0.0652	0.4466	0.9565
ad-990223-28757	29375	1.443	0.0505	0.3757	0.9287
ad-gm-990224	11190	1.609	0.0736	0.4365	0.8893
ad-sc-990224	4937	1.094	0.0061	0.4813	0.9923
bx-mo25	1225	22.286	0.0782	0.6911	0.9415
bx-mo75	1285	20.389	0.0740	0.7077	0.9423
kddcup-97-bnb	150616	1.358	0.0276	0.4891	0.9883
sr1_m31	16400	7.073	0.0210	0.5234	0.9967
sr1_m35	12200	4.574	0.0314	0.6271	0.9971
sr2_m03	23900	4.59	0.0433	0.5388	0.9880
sr2_m05	10100	5.228	0.0403	0.5344	0.9861
sr3_m10	22500	5.471	0.0066	0.5230	0.9947
sr4_m26	20900	6.278	0.0036	0.5115	0.9931
sr5_m06	71100	4.43	0.0145	0.5017	0.9984
sr5_m07	86900	5.01	0.0333	0.5746	0.9954
Average			**0.041**	**0.533**	
St. Dev			**0.0269**	**0.088**	

Table 3: Coefficients for regression of log10(CumPHits) vs log10(P%)

Next, we performed a similar regression for all 16 cases. The results are summarized in table 3.

Averaging the coefficients over 16 cases we get average R^2 = 0.97, average a = 0.041 and average b = 0.533.

We also did a further regression analysis of a and b versus T. We do not have enough space to present the details, but we did *not* find a significant relationship between a and T. We did find some correlation between b and T, with $R^2 = 0.66$

(9) $b = 1.13T + 0.467$

substituting this and $a = 0.04$ into eq. (8) we get

(10) $\log10(CumPHits(P,T))$

$= 0.041+(1.13T+0.467)\log10(P)$

or

(11) $CumPHits(P,T) = 1.1\, P^{0.467+1.13\,T}$

and $Lift(P) = 1.1\, P^{0.467+1.13T} / P = 1.1\, P^{-0.533+1.13\,T}$

However, for the purpose of getting approximate bounds we will use simpler estimates of $a = 0$ and $b = 0.5$, giving us

$\log10\,(CumPHits(P)) = 0.5\,\log10(P)$

or

(12) $CumPHits(P) = \text{sqrt}(P)$

and

$Lift(P) = 1/\text{sqrt}(P) = P^{-0.5}$

5.1 Ranges on lift curve

While the estimate $Lift(P) \simeq 1/\text{sqrt}(P)$ produces a reasonably close fit, we are also interested in getting a range for the lift. From Table 3, we see that the standard deviation on b is about 0.09 and b is between 0.4 and 0.6 most of the time. The constant term a is small and can be ignored for the purpose of initial approximation. Hence we can estimate

(13) $0.4\,\log10(P) < \log10\,(CumPHits(P)) < 0.6\,\log10(P)$

Since $P < 1$, all logs above are negative, and we can simplify to

(14) $P^{0.6} < CumPHits(P) < P^{0.4}$

or, dividing the above equation by P

(15) $P^{-0.4} < Lift(P) < P^{-0.6}$

Fig 4 shows the ranges the actual ***CumPHits*** curve for KDD-CUP-97 data and upper and lower bounds obtained from (15). We see that the actual curve generally falls between the bounds.

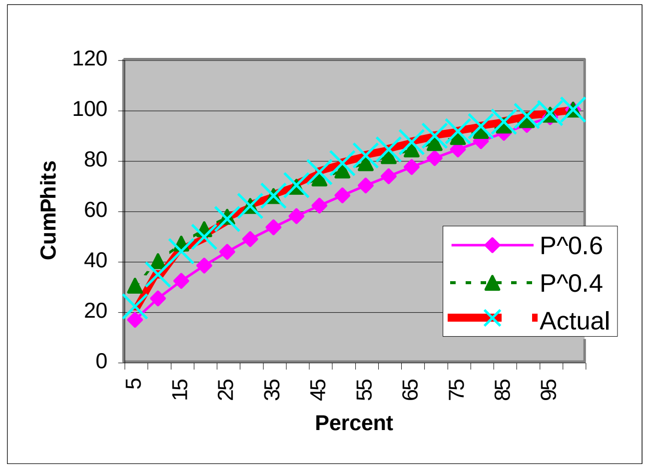

Fig 4. CumPHits for KDD-CUP-97 Data

6. ESTIMATING OPTIMAL PROFIT

In section 3 we derived a formula for estimating a profit of a campaign. We apply it to estimate the profit of selecting a subset of size $N_2 = NP$, with lift $T_2 = T * Lift(P)$ and get

(16) $Profit(P) = NP\,(TB*Lift(P) - C)$

Next, we substitute the estimate for lift

$Lift(P) \simeq 1/\text{sqrt}(P)$

derived in previous section, and get an estimate for the profit of selecting first P percent of the list:

(17) $Profit(P) \simeq NP((TB/\,\text{sqrt}(P)) - C) =$

$= NC((TB/C) * \text{sqrt}(P) - P)$

Let $K = TB/C$. This profit is maximized when

(18) $F(K,P) = K^*\,\text{sqrt}(P) - P$

is maximized, for $0 \leq P \leq 1$.

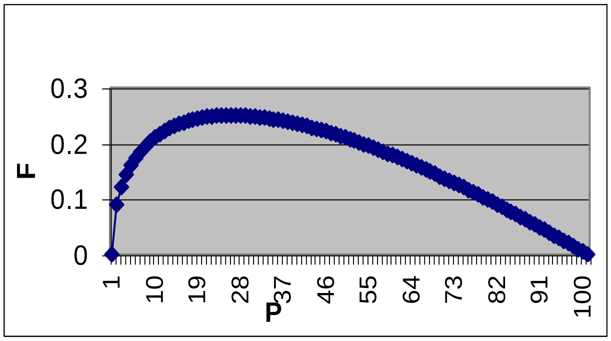

Fig 5. F(1,P)=(sqrt(P) – P) vs P

Figure 5 shows the curve of K*sqrt(P) -P for K = 1

We can find the maximum of $F(K,P)$ by finding when its derivative is zero. Since $dx^n/dx = nx^{n-1}$, the equation for derivative of $F(K,P)$ equal to zero is

$$d\,F(K,P)/dP = 0.5\,K\,P^{-0.5} - 1 = 0$$

or

$$P^{-0.5} = 2/K$$

or

(19) $$P = (K/2)^2$$

Indeed, in figure 5 we see that the maximum value of the curve for K=1 is achieved when $P = (K/2)^2 = 0.25$.

When $K = TB/C \geq 2$, maximum profit is achieved for $P \geq 1$, meaning that the best selection is the entire list. In this case modeling to select a subset of the list will not be useful. However, when $K = TB/C < 2$, maximum profit is achieved for $P < 1$, meaning that it is useful to perform modeling to select a subset of the population. We can state this condition as:

Data Mining Sweet Spot:

Selecting a subset to contact can increase profit when $K = TB / C < 2$

We can rewrite the Profit formula (17) as

(20) $$Profit(P) = NC(K^* \,sqrt(P) - P)$$

By substituting $P = (K/2)^2$ we get the maximum value

(21) $$MaxProfit(N,T,B,C) \simeq NC(K^*K/2 - (K/2)^2) = NCK^2 / 4$$

This formula gives an expected value of campaign MaxProfit. We can also estimate the variability in the profit value by using a range estimate for lift, based on (11)

$$Lift(P) = P^{-d} \text{ , where } 0.4 < d < 0.6$$

Then the profit from selecting a subset P to contact can be written as

(22) $$Profit(P) = NP\,(TBP^{-d} - C)$$

or, substituting $K = TB/C$

(23) $$Profit(P) = NCP\,(\,K\,P^{-d} - 1) = NC\,(K^*\,P^{(1-d)} - P)$$

By a similar reasoning,

$$F(K,P,d) = K^*P^{(1-d)} - P \text{ , where } 0 < d < 1$$

is maximized when

$$(1-d)\,K\,P^{-d} = 1 \text{ or } P^d = (1-d)\,K \text{ or}$$

(24) $$P_{MAX}(d) = ((1-d)K)^{1/d}$$

For example, when d=0.4, $P_{MAX}(d) = 0.279K^{2.5}$; for d=0.5, $P_{MAX}(d) = 0.25\,K^2$ and for d=0.6, $P_{MAX}(d) = 0.217K^{1.67}$.

Substituting (24) into formula for MaxProfit, we see that the estimated maximum profit is

(25) $$MaxProfit(N,T,B,C) = NC\,(K\,P_{MAX}(d)^{(1-d)} - P_{MAX}(d)) =$$

$$NC\,P_{MAX}(d)\;(K\;(((1-d)K)^{1/d})^{-d} - 1) =$$

$$NC\,P_{MAX}(d)\;(K/\,(1-d)K\; - 1) =$$

$$NC\,P_{MAX}(d)\;(1/(1-d) - 1) =$$

$$NC\,P_{MAX}(d)\;\;d\,/(1-d)$$

The table below shows the range of $P_{MAX}(d)$ and MaxProfit for different values of ***d.***

d	$P_{MAX}(d)$	Max Profit
0.4	$0.28\,K^{2.5}$	$0.186\,NCK^{2.5}$
0.5	$0.25\,K^2$	$0.250\,NCK^2$
0.6	$0.22\,K^{1.67}$	$0.326\,NCK^{1.67}$

Table 4. Variation of Maxprofit and $P_{MAX}(d)$ with d

While these are only estimates, we hope that they will be useful in providing initial ranges of profit values for different settings.

7. RELATED WORK

Paper [4] provides an overview of typical issues related to modeling for direct marketing. In [5] there is a discussion of maximizing payoffs using different modeling approaches. [1] investigates how to maximize lift for a specific decile or percentage (e.g. mailing depth for a campaign) of the sorted list. [9] investigates comparison of classifiers when dealing with skewed class distributions and non-uniform costs, which is the case with most applications of targeted marketing.

8. DISCUSSION

We present heuristics for deciding when to consider applying data mining, how to estimate lift, how to model the lift curve itself and how to estimate expected profits.

While we find that the heuristic formulae give a reasonable agreement (see section 5.4 for a discussion of exceptions), with the data we looked at from telecom and financial domains we need to expand our set of cases. Even though these approximations may be valid only for specific domains and applications similar to attrition and cross-sell , they may still be useful as rule of thumb estimates.

In our discussion we have assumed a sorted, ranked list of subscribers where the subsetting may happen by choosing a cutoff of model score or choosing a fixed percentage of the list. However with methods such as induced rules from a decision tree one can consider subsets independent of order, to which the rules in section 2 and 3 can still be applied.

9. FURTHER WORK

Apart from expanding the data from different domains and evaluating the robustness of the proposed formulae for estimating lift, the lift curve and expected profits, another direction can be to model the distribution of the target class in the whole population and in sub-populations (may be ordered) selected by a model.

Other interesting avenues include examining the empirical distribution of target density function, from which the cumulative density function could be obtained by integrating and testing these results on additional data.

10. REFERENCES

[1] Bhattacharya, S. Direct marketing response models using genetic algorithms. In Proceedings of KDD-98 (New York, NY, Aug 1998), AAAI Press, 144-148.

[2] Brachman R. and Anand T., The Process of Knowledge Discovery in Databases: A Human centered approach, In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R. (Eds.), 1996. Advances in Knowledge Discovery and Data Mining, AAAI Press.

[3] Elkan, C., Boosting and Naive Bayesian Learning. UCSD Technical Report No. CS97-557, September 1997.

[4] Ling, C.X., and Li, C., Data mining for direct marketing: problems and solutions, In proceedings of KDD-98 (New York, NY, Aug 1998), AAAI Press, 73-79.

[5] Masand B. and Piatetsky-Shapiro G. A comparison of different approaches for maximizing business payoff of prediction models, In proceedings of KDD-96 (Portland, OR, Aug 1996), AAAI/MIT press, 195-201.

[6] Michie, D., Spiegelhalter, D.J. and Taylor, C.C., (eds.), Machine Learning, Neural and Statistical Classification, Ellis Horwood, 1994.

[7] Parsa, I. KDD-CUP-97 results, www.epsilon.com/KDDCUP/index.htm .

[8] Piatetsky-Shapiro, G., et al. An overview of issues in developing industrial data mining and knowledge discovery applications, In proceedings of KDD-96 (Portland, OR, Aug 1996), AAAI/MIT press, 89-96.

[9] Provost, F. and Fawcett, T., Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions, In proceedings of KDD-97 (Newport Beach, CA, Aug 1997), AAAI Press, pp 43-48.