

Targeting Customer Groups using Gain and Cost Matrix: a Marketing Application

J-H. Chauchat¹, R. Rakotomalala¹, M. Carloz², and C. Pelletier²

¹ ERIC Laboratory - University of Lyon 2
5, av. Pierre Mendes-France
F-69676 BRON - FRANCE
chauchat,rakotoma@univ-lyon2.fr

² CCF Bank
Marketing Opérationnel
4, place de la Pyramide
PARIS LA DEFENSE - FRANCE
mcarloz@ccf.fr

Abstract. This paper introduces gain-cost classification matrix for targeting customer groups. A real life application was realized on a customer database of CCF Bank containing more than 400,000 instances. Results shows that scoring is almost insensitive to different cost-gain hypothesis; on the other hand, the optimal targeted group size and the expected global benefit rely to cost-gain hypothesis.

Keywords: cost-sensitive learning, customer targeting, decision tree, marketing application

1 Cost-Benefits Aspects of Targeting in Marketing - Introduction.

Identifying target groups for marketing purposes is one of the most common applications of Data Mining. It aims at identifying the group of clients most likely to respond positively to a new product. This targeting is often done using classical machine learning techniques such as decision trees. The underlying assumption is that clients contacted in vain (false positive) and those who might have bought had they been contacted (false negative) can be treated in a similar manner. In real life situations, the hypothesis is not verified, as the two cases are not symmetrical. In the first case, the company spent resources with no return on its expense; in the second case, there is a potential income loss; moreover, correctly targeted clients represent substantial gains for the company.

In this paper, an example of targeting a customer group using decision trees will be examined, using a 400,000 rich customer base of Crédit Commercial de France (CCF). The originality of this paper lies in the introduction of a cost-gain matrix during the construction of the prediction model to identify the group the least expensive to target. In section 2, the determination of costs and benefits (gains) in the context of a direct marketing campaign is developed and

the maximization of the expected gains is showed. Related works are covered in section 3, concentrating on authors who suggested ways to include costs in the scoring models and on the solution developed here. The CCF case study is detailed in section 4; hypotheses about costs are tested, the sensitivity of the scores (the estimated probabilities of purchase) is studied as well as the optimal target size for a maximized expected benefit. The proposed decision model is applied to an actual targeting problem with respect to the advertising campaign of CCF Bank financial product using a database of more than 400,000 clients. The paper concludes with an evaluation of the strategy and suggested improvements currently under further development.

2 The general problem of targeting in marketing. The question of cost-benefits.

Usually, error minimization algorithms are used to target a group of clients or prospects for a given commercial campaign [1] and the problem can be reduced to the simplistic case of the campaign cost matrix:

		Decision	Decision
		Targeted	Not Targeted
Reality	Does (or would) buy	No Error, Cost = 0	Error, Cost = 1
Reality	Does not (or wouldn't) buy	Error, Cost = 1	No Error, Cost = 0

In reality, the matrix is different: when a targeted client buys (= true positive), there is a benefit and some cost if he doesn't (= false positive); if no targeting is done, there is no benefit, or gain, nor cost save the absence of a benefit (= false negative).

Working in a business context, it is better to use a "gain matrix" than a "cost matrix".

		Decision	Decision
		Targeted	Not Targeted
Reality	Does (or would) buy	Gain = +a	No benefit, no cost
Reality	Does not (or wouldn't) buy	Error, Cost = -b	No benefit, no cost

2.1 Accounting for costs and benefits in targeting.

The following aspects of introducing costs and benefits are of interest in this paper : estimating the expected unit cost and unit benefit; the role these play in determining the optimal target (marginal gain and total gain); and a fresh model for estimating the expected total benefit with respect to the quality of the estimated probability of purchase and cost/gain hypotheses.

Estimating the expected unit cost and unit benefit The expected unit gain from selling one additional unit must be estimated. For example, with an

investment, the discounted gain over the period must be estimated. Now, to estimate expected unit costs, the direct cost of one commercial action and the marginal cost of targeting an additional client are distinguished. Fixed costs comprise the concept, forms and leaflets design, staff training, etc. Marginal (or variable) costs comprise what an additional targeted customer costs. For example, printing one more copy of a form letter, stuffing the envelope, postage stamp, etc. would be some of the marginal costs of a large scale mailing. In a telephone marketing campaign, the marginal costs would comprise the cost of one phone call, the salary of the phone operator, equipment rental, management of call-backs, etc.

If estimating is difficult, hypotheses about costs and benefits can help. Such hypotheses would be incorporated into the estimated probability of a purchase for each client on the database which are to be used for estimating the expected benefit from each customer and for drawing marginal and total gain curves (see § 4.1 and figure 6).

Determining the optimal target In order to determine the optimal target, individuals, or homogenous groups of individuals, are sorted by decreasing probability of a purchase $P(i)$, thus $P(i)$ is a decreasing function of i .

Then, the expected marginal benefit can be computed as:

$$G_m(i) = a \times P(i) - b \times (1 - P(i))$$

This marginal benefit is equal to "a" if individual "i" buys with certainty and equal to "-b" if individual "i" does certainly not buy. Now, an optimal target could be the set of clients for which the expected marginal benefit is positive, that is:

$$\begin{aligned} G_m(i) &= a \times P(i) - b \times (1 - P(i)) \\ &= (a + b) \times P(i) - b \end{aligned}$$

or

$$P(i) > \frac{b}{a + b}$$

Figure 1, shows the marginal benefit per targeted client as a function of the proportion targeted on the database, with a false positive cost $b = 200$ and a true positive benefit $a = 2,000$ euros. Note that $P(i)$ is a decreasing function of i and G_m is a decreasing function of $P(i)$, hence $G_m(i)$ is a decreasing function of i .

On this example, the marginal benefit is zero for $P(i) = 200 / (2,000 + 200) = 0.09$ that is to say $i/N = 18\%$.

Alternatively, the optimal target corresponds to the point where the total gain curve is maximum. Summing benefits of individuals from $i = 1$ to $i = x$, the total benefit is the sum of the marginal benefits minus fixed costs:

$$G_T(x) = \sum_{i \leq x} [a \times P(i) - b \times (1 - P(i))] - \text{FixedCosts}$$

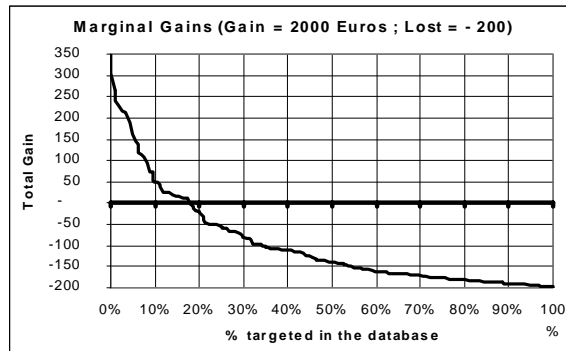


Fig. 1. Marginal benefit per targeted client, given a false positive cost of -200 euros and a true positive benefit of 2000 euros

As the marginal benefit is the derivative of the total benefit, and since the latter is a decreasing function, the two conditions are equivalent.

Figure 2 shows the Cumulated Total Benefit as a function of the proportion targeted on the database. It corresponds to the previous chart showing that the maximum total gain is attained for $P=18\%$.

2.2 Estimating the target

A scoring model is usually evaluated using a ROC curve [15] or a TQC coefficient [6] which Piatetsky-Shapiro and Steingold [14] renamed "L-quality measure".

TQC (Targeting Quality Coefficient) is a coefficient used to compare tracking procedures. The coefficient is similar to a Gini coefficient in statistics [11], lift charts used in marketing, ROC curves from signal theory [8] or health sciences [17]. The coefficient can help comparing trees, or any score model, used in tracking procedures.

In general, classifiers are compared using the "test error rate", that is the proportion of "misclassified" among a sample independent of the learning sample [12]. For the situation at hand (tracking rare groups), the usual error rate is ill adapted. Rather than looking for the most likely class of an individual given his characteristics, the probability of having a rare characteristic is estimated: disease, fraud, breakdown, tele-purchase, ...

Individuals with a predicted probability of belonging to the rare group of at least $x\%$ are tracked; by varying x with respect to cost and expected benefits of ensuing actions, a larger or smaller set of individuals "at risk" is selected.

Hence, the quality coefficient must depend on the predicted probabilities given by the classifier: it ranges from 0 for a random classification (i.e. all predicted probabilities are equal to p , the global probability of having the rare characteristic), to 1 if the classifier recognizes perfectly the members of both

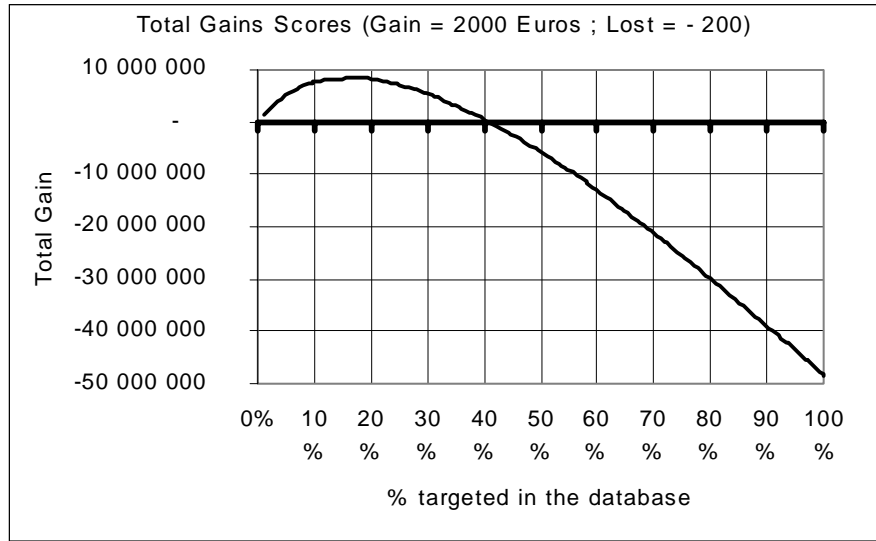


Fig. 2. Cumulated Total Benefit as a function of % targeted on the database, false positive cost $b = 200$ euros and true positive benefit $a = 2,000$ euros

classes (in this case, the predicted probability is 1 for members having the rare characteristic, and is set to 0 for the other ones).

Table 1 shows how the TQC coefficient is constructed. Individuals are sorted by decreasing predicted probabilities; then, two cumulative functions of the relative frequencies are computed:

1. the cumulative proportion of individuals in the population,
2. the cumulative proportion of individuals with the rare characteristic.

Computations from the decision tree built on a validation file of size $N = 1000$ individuals, with $A = 100$ bearing the rare characteristic, are displayed in Table 1. For a given individual, the predicted probability is the proportion of "rare" individuals among the individuals on the same leaf of the decision tree. For example, selecting the 4 individuals with the largest predicted probabilities (that is $x\% = 4/1000$ of the population), $F_i = 2/100$ of the "rare" individuals are expected to be covered.

The two cumulative distributions are linked in Figure 3 : the proportion of selected population on the horizontal axis and the estimated proportion of targeted individuals on the vertical axis. The true curve must lie between two extremes:

1. Perfect targeting, displayed as two straight segments joining three points: $(0;0)$ where no one is selected and no one is covered, $(\frac{A}{N}; 100\%)$ exactly

Rank = i	% Total Population	Pred. Prob = P_i	Class	% Cumulative Class "1" = F_i	Surface element = $(1/N) * (F_{i-1} + F_i)/2$
1	$1/N = 1/1000$	100 %	1	$1/A = 1/100$	$(1/N)*(1/A)/2$
2	$2/N = 2/1000$	100 %	0	$1/A = 1/100$	$(1/N)*(1/100+1/100)/2$
3	$3/N = 3/1000$	70 %	0	$1/A = 1/100$	$(1/N)*(1/100+1/100)/2$
4	$4/N=4/1000$	70 %	1	$2/A = 2/100$	$(1/N)*(2/100+1/100)/2$
...
N	$N/N = 100%$	0 %	0	$A/A=100 %$	$(1/N)*(F_{N-1}+1)/2$
SUM =	—	—	A	—	$Area$

Table 1. Building TQC, the Quality Targeting Coefficient, on an artificial example

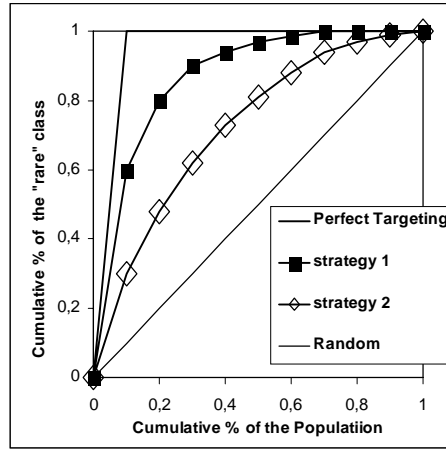


Fig. 3. Comparing various strategies: cumulative proportions scatter-plot for targeting clients most susceptible to respond.

- $\frac{A}{N}$ of the population is selected and it is the whole targeted group, and (100%; 100%) where every one is selected hence the target is attained;
- Random targeting, displayed as the diagonal: selecting $x\%$ of the population covers $x\%$ of the targeted group.

The coefficient TQC is defined as the ratio of two areas: the "Area" between the real curve and the diagonal, and the area between perfect targeting and the diagonal.

"Area" is calculated using the trapezium rule:

$$Area = \sum_{i=1, N} (1/N) * (F_{i-1} + F_i)/2$$

From Table 1 :

$$TQC = \frac{2 \times Area - 1}{1 - \frac{A}{N}}$$

Hence, $TQC = 0$ for random targeting, and $TQC = 1$ when targeting is perfect.

Neg.	Pos.	Leaf size	% Pos.	Leaf Gain	Marginal Gain	\sum Base	\sum Pos.	\sum Gains
0	0	0	0 %	0	0	0 %	0 %	0
1000	1500	2500	0.60	3,250,000	1,300.0	11 %	60 %	3,250,000
2500	190	2690	0.07	- 775,000	- 288.1	23 %	68 %	2,475,000
2690	180	2870	0.06	- 895,000	- 311.8	36 %	75 %	1,580,000
2530	160	2690	0.06	- 865,000	- 321.6	48 %	81 %	715,000
2760	160	2920	0.06	- 980,000	- 355.6	61 %	88 %	- 265,000
1940	100	2040	0.05	- 720,000	- 352.9	70 %	92 %	- 985,000
900	30	930	0.03	- 375,000	- 403.2	74 %	93 %	- 1,360,000
5680	180	5860	0.03	- 2,390,000	- 407.9	100 %	100 %	- 3,750,000

Table 2. Scoring 1 : Targeting from an 8-leaf tree - True positive Gain : a = 2500
False-positive Cost: b = 500

TQC may be negative if a very bad targeting procedure is used: few targeted instances would be selected first, and most of them at the end.

This coefficient is better suited to our study as recognizing individuals is not the goal; isolating those of interest is.

The chart of the maximum benefit. Here, charting the maximum benefit against the size of the target yields a more natural way of estimating the target: the benefit is zero when nothing is done; the curve reaches its maximum at the optimal target size and the expected benefit for the whole marketing effort is read off the y-axis; the curve drops below zero as soon as costs win over gains. Should the curve be all negative, then the maximum is at the origin and it is best to not launch the marketing operation.

Moreover, this total gain chart can be used to compare two scoring models preferably over usual charts and indices. Effectively, one model may have higher total gains even though the ROC curves cross.

Tables 2 and 3 and figures 3 and 5 compare two strategies where the *TQC* index is lower (0.62 compared to 0.71), while the total benefit is higher (3,250,000 F compared to 2,625,000 F).

Figure 4 compares the two strategies according to the Targeting Quality Coefficient (*TQC*); it is lower with "scoring 1" than "scoring 2" (0.62 compared to 0.71), while the total benefit is higher (3,250,000, compared to 2,625,000).

Figure 5 compares the two strategies according to the Total Gain; it is higher with "scoring 1" than "scoring 2" (3,250,000, compared to 2,625,000), while *TQC* is lower.

Neg.	Pos.	Leaf size	% Pos.	Leaf Gain	Marginal Gain	\sum Base	\sum Positive	\sum Gains
0	0	0	0 %	0	0	0 %	0 %	0
1000	800	1800	0.44	1,500,000	833.3	8 %	32 %	1,500,000
1400	600	2000	0.30	800,000	400.0	17 %	56 %	2,300,000
1000	330	1330	0.25	325,000	244.4	23 %	69 %	2,625,000
2530	430	2960	0.15	-190,000	- 64.2	36 %	86 %	2,435,000
2510	190	2700	0.07	-780,000	- 288.9	48 %	94 %	1,655,000
1800	100	1900	0.05	-650,000	- 342.1	56 %	98 %	1,005,000
2710	50	2760	0.02	-1,230,000	- 445.7	69 %	100 %	- 225,000
7050	0	7050	0.00	-3,525,000	- 500.0	100 %	100 %	- 3,750,000

Table 3. Scoring 2 : Targeting from an 8-leaf tree - True positive Gain : a = 2500
False-positive Cost: b = 500

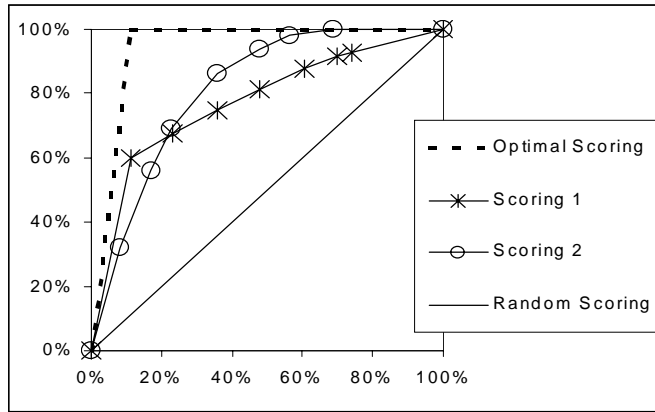


Fig. 4. Targeting Quality Coefficient

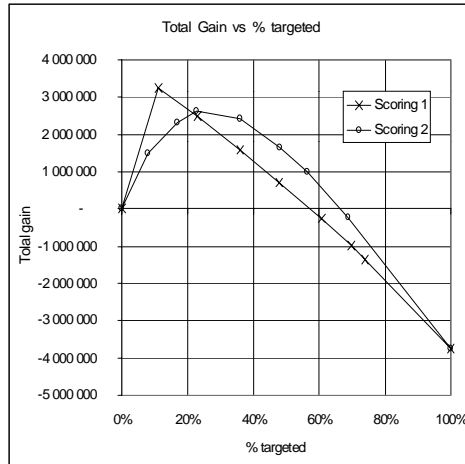


Fig. 5. Total gain vs. Percent Targeted

3 Cost-sensitive machine learning process

3.1 General framework of cost-sensitive machine learning

Most machine learning algorithms aim at minimizing classification errors, regardless of any other criterion [13]. Yet, in most real-life cases, various costs, hidden or not, influence classification model design and usage: collection and measurement of the predictive attributes, classification by a teacher, etc.. Turney details a classification of those various errors [21]. This paper deals mainly with the cost of a misclassification error during the learning phase. A cost matrix $C = (c_{ij})_{K \times K}$ is used; the entry c_{ij} is the cost associated with classifying an individual as a "j" while the correct classification should have been "i", and the dimension of the matrix, K, is the number of values for the class attribute. In most of the work on cost of learning, the diagonal of C is supposed null, that is to say that there is no cost associated to the correct classification of an individual; in our opinion, the diagonal should be negative, that is to say that there is a negative cost (i.e. a gain) associated with a correct decision.

Accounting for the cost of misclassification during the design of the classifier can be done in many ways; here are the four principal strategies:

- modifying the learning algorithm so that cost be part of the classification process [20]. This strategy requires modifying the learning algorithm which, in some cases, is not obvious [13];
- accounting for cost through the individuals' weights. This strategy can be implemented by replicating individuals whose relative weight should be increased [22]; this strategy is not suited to class-attributes with more than two values [18];

- encapsulating an algorithm in a general frame work which comprises the misclassification cost matrix, some meta-classifier [13]. This technique always work, but it is resource-consuming as it requires the repetition of the learning phase [3];
- finally, there is an adaptation of Boosting [10]. The original algorithm is modified so that the weights are modified according to their misclassification during the learning phase and to the misclassification cost [9][19]. This high performance strategy is also resource-consuming as a lot of classifiers must be built.

The main difficulty is to find an algorithm able to process quickly large volumes of data, given the number of models to be built corresponding to the cost matrices to be tested. The last two strategies are not well suited to this particular framework given the necessary computer resources. Modifying the weights, while suited to our 2-valued problem, appeared too complicated as introducing negative costs (gains) for correct classifications was of interest. The first strategy using decision trees [16] was retained for two main reasons: it was used successfully for a targeting on a large database [6]; there are many ways to adapt induction tree algorithms to incorporate misclassification cost [20].

When constructing a single tree using a learning set, misclassification costs can be introduced at two points : when computing the splitting measure during the growing process, or when computing misclassification cost during the pruning process [4]. The first seemed more appealing as it does not involve post-pruning, it is quicker and efficient for targeting problems [6]. Unfortunately, most work on induction trees show that the first option yields low-performance trees with respect to minimizing misclassification costs [7].

In light of all these constraints, it was decided to use an algorithm proposed by [2]. This strategy requires only minor adaptations of the original algorithm (C4.5 [16]).

3.2 Cost-sensitive decision tree for the CCF project

Induction trees are constructed stepwise : first, the growing process, during which the tallest possible tree is compiled by choosing at each node the best attributes as indicated by the splitting measure ; the second step, pruning, is the reduction of the tree by evaluating the precision of the leaves when classifying individuals. Authors appreciate differently this second step : some seem to think it is fundamental [4]; others think it can bring a smaller tree with no substantial gain of accuracy [23].

Nevertheless, if it seems more interesting to use some criterion of purity close to the roots, working on the bottom of the tree, the splitting criterion ought to be related to the purpose of the learning, namely minimizing misclassification error [5], and in the present case, minimizing misclassification costs.

The algorithm used here [2] is quite simple. It is a variation of Quinlan's C4.5 algorithm [16] with the two following modifications :

- Class probability distributions on a node is smoothed using Laplace's correction. Specifically, if a node has n instances, n_k of which are from a given class k , in a K -class problem, instead $\frac{n_k}{n}$, the probability assigned to the class is $\frac{n_k+1}{n+K}$. This estimation is more stable than traditional frequency because it is less extreme on smaller sample sizes.
- Pruning is then based on minimizing misclassification costs. Pruning leaves at a node can be done by comparing the expected misclassification cost at the node with the expected misclassification cost of its subtree (the expected misclassification cost of a subtree is the average of the expected misclassification cost of the leaves). Expected misclassification cost is computed in the following way: given a probability distribution (after Laplace correction), and a misclassification matrix, we look for the class with the expected minimal cost by applying the probability distribution to the misclassification cost matrix, expected minimal cost on a node corresponds to expected misclassification cost of the node.

4 The " CCF " case study

An example of targeting for a direct mailing will be reviewed below. The product of interest is a life insurance policy. The product is interesting for the bank as it guarantees steady income for the duration of the contract (8 years in France, for taxation reasons).

The methodology of cost-benefit estimation process will be reviewed for this particular marketing campaign.

4.1 Estimating Costs and Gains

Each subscriber signs a specific contract; for the sake of simplification, and since all the details of each contract are not stored in the data warehouse, the duration and average coverage of the life insurance policies subscribed over the last two years were used, as well as the two most popular types of investment tools (French or foreign shares, obligations, fixed interest, ...) associated with those contracts.

Gains were estimated under hypotheses concerning stock market returns and interest rates for the eight coming years, and accounting for "entry fees" (a part of a client's initial investment) and management fees (a part of the invested capital levied annually).

Three categories of costs were estimated. Direct monetary costs are easiest to estimate: design and printing costs of the mailed documents, cost of the address list, stuffing envelopes, postage. Then, the cost of training the commercial staff for this very marketing programme, negotiation and case management must also be estimated. Thirdly, the "psychological" or opportunity cost of a misdirected targeting must be estimated. A client being offered products that don't interest him can become dissatisfied or annoyed with the bank, and may tend to disregard other mailings that could be of interest. This psychological cost is much more

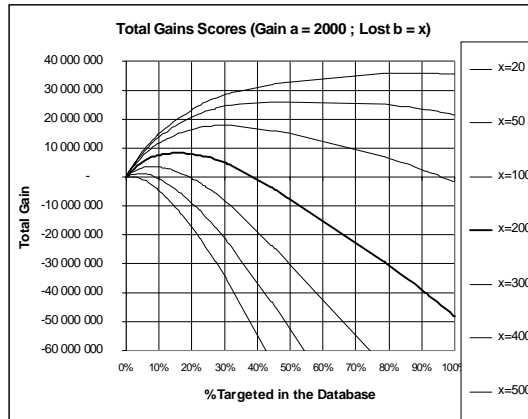


Fig. 6. Total gains scores

difficult to estimate, but it must be quantified as it is a true cost; disregarding this aspect could cause to flood clients with parasite offers.

There are several ways to estimate these elements of cost and gain. The sensitivity of the results to the various estimates should then be studied.

4.2 Total Gains : Sensitivity to the cost and benefit choice

For the work reviewed above, several cost-benefit matrices were tested. For confidentiality reasons, and given that the shape of the results basically depends on the ratio $b/(a+b)$, results are shown for the unit costs $b = 10, 50, 100, 200, 300, 400$ and 500 , and an arbitrarily fixed gain $a = 2,500$.

The a and b parameters play a role in the decision tree scores as seen in §3.2, yet the distribution of the scores appears rather robust to changes of the b values.

For both sets of scores, the optimal targeting proportion is almost the same, for a given ratio $\frac{b}{a+b}$. The expected total gain can be read off the figure 6, and the gain can be optimized for the chosen values of a and b .

We present the results obtained from a learning sample of 97,600 clients, test sample of 390,300 clients; Total Gains are computed for a 487,900 client-population:

The upper curve corresponds to $b = 10$; as the false-positive cost is very low, it is profitable to target 80 % of the population and the estimated total benefit would be very high, up to 36,000,000.

On the opposite, the lower curve corresponds to $b = 500$; as the false-positive cost is very high, one must be selective; it is optimal to target only 1% of the population and the estimated total benefit would be very low, 160,000.

A median situation corresponds to $b = 200$; the optimal target size is between 14 % and 18 % of the population and the estimated total benefit would be 8,300,000.

5 Conclusion

Studying cost-gain aspect of targeting in marketing, one must estimate a score for each client or prospect, that is to say his, or her, probability of a purchase. In this paper, we note that working on the CCF database, the score estimated using cost-gain matrix or not, are almost the same. But, the optimal target size and the global benefit expected depend directly on the cost-gain matrix used.

So, the main problem remains the estimation of each kind of cost and gain, and many hypothesis must be tried. The paper proposes a way to compute and to visualize easily expected global benefit for each score model, each cost-gain matrix, and to find the optimal customer target size.

Acknowledgements

This work was done in collaboration with the marketing service of CCF. The authors thank Mr. Claude Mathieu and Mr. Nicolas Fraisse for their support. We also thank Mr. Jean Dumais, Section Chief, Methodology Branch, Statistics Canada and guest lecturer at Université Lumière for his careful proof reading of the manuscript and his suggestions.

References

1. M.J.A Berry and G. Linoff. *Data Mining Techniques: For Marketing, Sales, and Customer Support*. John Wiley and Sons, 1997.
2. J. P. Bradford, C. Kunz, R. Kohavi, C. Brunk, and C. E. Brodley. Pruning decision trees with misclassification costs. In Claire Nédellec and Céline Rouveirol, editors, *Proceedings of the 10th European Conference on Machine Learning (ECML-98)*, volume 1398 of *LNAI*, pages 131–136, Berlin, April 21–23 1998. Springer.
3. L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.
4. L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees*. California : Wadsworth International, 1984.
5. Carla E. Brodley. Automatic selection of split criterion during tree growing based on node location. In *Proc. 12th International Conference on Machine Learning*, pages 73–80. Morgan Kaufmann, 1995.
6. J.H. Chauchat, R. Rakotomalala, and D. Robert. Sampling strategies for targeting rare groups from a bank customer database. In *Proceedings of the Fourth European Conference - PKDD'2000*, pages 179–188, Lyon, France, 2000. Springer Verlag.
7. C. Drummond and R. Holte. Exploiting the cost of (in)sensitivity of decision tree splitting criteria. In *Proc. 17th International Conf. on Machine Learning*, pages 239–246. Morgan Kaufmann, San Francisco, CA, 2000.
8. J.P. Egan. *Signal Detection Theory and ROC Analysis*. Series in Cognition and Perception. Academic Press, New York, 1975.

9. W. Fan, S. Stolfo, J. Zhang, and P. Chan. AdaCost: misclassification cost-sensitive boosting. In *Proc. 16th International Conf. on Machine Learning*, pages 97–105. Morgan Kaufmann, San Francisco, CA, 1999.
10. Y. Freund and R.E. Schapire. Experiments with a new boosting algorithm. In *Proceedings of the 13th International Conference on Machine Learning*, pages 148–156, 1996.
11. C.W. Gini. Variabilita e mutabilita, contributo allo studio delle distribuzioni e relazioni statistiche. Technical report, Studi Economico-Giuridici della R. Universita di Caligiari, 1938.
12. T.M. Mitchell. *Machine learning*. McGraw Hill, 1997.
13. P.Domingos. Metacost: A general method for making classifiers cost-sensitive. In *Proceedings of the 5-th International Conference on Knowledge Discovery and Data Mining, KDD-99*, pages 155–164, New York, 1999. ACM Press.
14. G. Piatetsky-Shapiro and S. Steingold. Measuring lift quality in database marketing. *SIGKDD Explorations*, 2(2):76–80, 2000.
15. Foster Provost and Tom Fawcett. Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. In David Heckerman, Heikki Mannila, Daryl Pregibon, and Ramasamy Uthurusamy, editors, *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, page 43. AAAI Press, 1997.
16. J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, 1993.
17. J. Swets. Measuring the accuracy of diagnostic systems. *Science*, 240:1285–1293, 1988.
18. K. M. Ting. Inducing cost-sensitive trees via instance weighting. *Lecture Notes in Computer Science*, 1510:139–145, 1998.
19. Kai Ming Ting. A comparative study of cost-sensitive boosting algorithms. In *Proc. 17th International Conf. on Machine Learning*, pages 983–990. Morgan Kaufmann, San Francisco, CA, 2000.
20. P. Turney. Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm. *Journal of Artificial Intelligence Research*, 2:369–409, 1995.
21. P. Turney. Types of cost in inductive concept learning. In *Workshop on Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning (WCSL at ICML-2000)*, Stanford University, California, 2000.
22. I.H. Witten and E. Frank. *Data Mining: practical machine learning tools and techniques with JAVA implementations*. Morgan Kaufmann, 2000.
23. D.A. Zighed and R. Rakotomalala. *Graphes d'Induction - Apprentissage et Data Mining*. Hermes, 2000.