Pricing Based Framework for Benefit Scoring

Nitesh Chawla
University of Notre Dame
Notre Dame, IN 46556
nchawla@nd.edu

Xiangning Li
University of Notre Dame
Notre Dame, IN 46556
xli3@nd.edu

ABSTRACT

Data mining models often carry the final objective of maximizing profit or minimizing cost. This problem becomes even more profound in financial applications that can have multiple constraints, such as interest rate, score cut-off, and the loan amount to allocate. In this paper, we present a pricing framework for discovering the total profit from a probabilistic model, given a benefit function.

1. INTRODUCTION

An important application of data mining in the financial industry is "scoring" the customers for loans. Credit scoring methods, typically, apply a cut-off paradigm of accepting or declining potential customers. The cut-off score is derived from a model learned on past consumer characteristics, and influences not only the accept or decline decision but also the loan amount and accompanying interest rate. Our goal is to evolve the scoring decision with a pricing scheme that drives the overall utility of the model. The utility is defined in terms of the economic benefit or the profit from approving loan for a customer. We want to able to structure the loan amount and the interest rate for a customer based on the propensity to default.

Our work builds upon the cost-sensitive learning literature [5, 11, 4, 7, 12] and the relevant literature from finance [10]. Stein [10] extends a cut-off score based approach to a pricing approach resulting in a more flexible and profitable model. Using the ROC curve quantities, he formulates the net present value of taking a lending decision and the corresponding benefit of a true negative.

We formulate a pricing framework based on the probability of default assigned by a scoring model and a benefit matrix. The benefit matrix specifies the benefit (positive or negative) from making a prediction. For example, a true negative will result in higher benefit as it is a positive return on the investment (a non-defaulting customer is correctly predicted to be a non-default). The profit from a customer is conditioned on the propensity to default (or not default).

UBDM'06 August 20 2006, Philadelphia, Pennsylvania, USA

Table 1: Benefit Matrix

	Actual Non-default	Actual Default
Predict Non-default	b_{00}	b_{01}
Predict Default	b_{10}	b_{11}

We want to be able to construct a pricing scheme that compensates the risk of the customer, and accordingly proposes an interest rate and a loan amount. Thus, we propose the interest rate and loan amount that can be assigned to an approved customer. We illustrate the workings of the pricing scheme using case dependent benefit functions. We discuss that the calibration of the model and the resulting quality of the estimates is more important than the resulting rank-ordering. We hasten to point out that, while this work serves as a preliminary proof-of-concept, our ongoing goal with the work is to demonstrate the applicability of the proposed approach using multiple simulations. We want to eventually compare multiple models using ROC curves, probability loss functions, and profits in dollars.

2. PRICING SCHEME

The key utility of applying data mining in a business model is the objective of maximizing profit or minimizing cost. The profit is related to the accuracy of the default probability predicted by a model and the case dependent benefit function. A typical benefit matrix can be defined as in Table 1 [5]. The benefit matrix elements reflect the benefits from assigning the loan, as per the corresponding probability of default. Morever, we assume that the benefits can be different for individual customers as they will be conditioned on the loan amount and the interest rate. The benefits should be tunable for the different levels of risk as reflected by the probability of default. This can be achieved by generating a different pricing function. Furthermore, we will assume that a customer, k, is accompanied with a probability of default of P_k. In our empirical analysis, we will assume that $b_{10} = b_{11} = 0$, since there is no benefit or cost from not offering a loan to customer.

Then, the profit from predicting a customer as a non-default (ND) (making the loan offer), can be calculated as:

$$B_{ND} = (1 - P_k)b_{00} + P_k b_{01} \quad (1)$$

and the benefit from predicting the customer as default (declining the application) can be calculated as:

$$B_D = (1 - P_k)b_{10} + P_k b_{11}. \quad (2)$$

To result in a higher profit from issuing the loan than refusing it, we should ideally have $B_{ND} > B_D$. This will form the restricting condition for allocating loan to a customer, and the basis of our pricing scheme. In a rigid competitive environment, it is being able to price and accept the B_D customers that gives the critical advantage.

2.1 Making Non-default Prediction Optimal

An ideal pricing scheme will be governed by the condition that no customer is ever turned down. Rather, a loan product is always priced for the customer, in conjunction with the interest rate, such that the resulting decision leads to a higher profit. That is the loss given default should not exceed the benefit ensuing from granting the loan.

The benefits are obviously example dependent. They are functions of the loan amount (x). The benefits (or losses) for true negatives and false negatives can be defined as a function of x: $b_{00}(x)$ and $b_{01}(x)$. The benefits for false positives and true positives typically are regarded as constants: b_{10} and b_{11}. Since we want $B_{ND} > B_D$, for an optimal non-default prediction, the price $b_{00}(k, x)$ asked for a randomly chosen customer scoring $S(k)$, should be

$$(1 - P_k)b_{00}(k, x) > (1 - P_k)b_{10} + P_k b_{11} - P_k b_{01}(x). \quad (3)$$

The above formulation, arising from equations 1 and 2, compensates for the default risk — the benefit from accepting a customer should be greater than the chance of losing due to the default. Hence, to grant the loan of x for a customer k, the resulting benefit should be:

$$b_{00}(k, x) > \frac{(1 - P_k)b_{10} + P_k b_{11} - P_k b_{01}(x)}{(1 - P_k)}. \quad (4)$$

To do further analysis, we made some simple assumptions: we suppose that the benefit from a non-defaulting customer (true negative) is a function of the loan amount x, the interest rate r and the risk premium $g(x)$ ($b_{00}(x) = g(x) + rx$), particularly we assume that $g(x) = ax$, here a is a very small constant (for example, $a = 0.005$)); while the loss from a false negative is $b01 = cx$, where c is also a constant and $-1 < c < 0$, and a reasonable guess for the value of c could be -0.35 [10]. Then, the interest rate asked for a specific customer can be derived from equation 4 as follows

$$r > \frac{(1 - P_k)b_{10} + P_k b_{11} - P_k cx}{(1 - P_k)x} - a, (r \geq 0) \quad (5)$$

This above equation is a limiting condition for making the loan offer an optimal choice. Figure 1 shows the interest rate asked for the customers to appropriately compensate the default risk that the bank has taken with the assumptions: $b_{10} = b_{11} = 0$, $c = -0.35$, and $a = 0.005$. This result indicates that for any customer we could generate the interest level corresponding to the default rate. But for high default rate customers, we find out that the interest rate is too high to be realistic, so a limit for the interest rate becomes necessary. Now if we define the maximum interest rate r_{max} allowed in the real practices, then the interest rate asked for any customer should satisfying

$$r \leq r_{max}. \quad (6)$$

Note that r_{max} can be fixed prior, depending on the bank practices and the risk appetite. In practice, the following

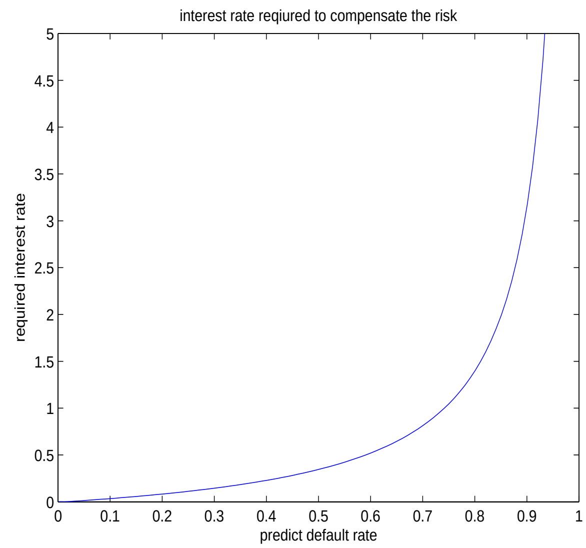

Figure 1: Interest level asked for compensating the default risk ($b_{10} = b_{11} = 0$, $c = -0.35$, and $a = 0.005$).

equation should be satisfiable:

$$r_{max} > \frac{(1 - P_k)b_{10} + P_k b_{11} - P_k cx}{(1 - P_k)x} - a, \quad (7)$$

Rearranging the above equation, we have

$$((a + r_{max})(1 - P_k) + P_k c)x > (1 - P_k)b_{10} + P_k b_{11}. \quad (8)$$

We can now define the range of the loan amount x at different values of P_k, while charging the customer at the possible maximum interest rate r_{max}.

- if $P_k > \frac{r_{max}+a}{r_{max}+a-c}$,
$$x < \frac{(1 - P_k)b_{10} + P_k b_{11}}{(r_{max} + a)(1 - P_k) + P_k c}, \quad (9)$$
- if $P_k < \frac{r_{max}+a}{r_{max}+a-c}$,
$$x > \frac{(1 - P_k)b_{10} + P_k b_{11}}{(r_{max} + a)(1 - P_k) + P_k c}. \quad (10)$$

Based on the above analysis, for a specific customer, we can decide not only the amount of money to loan but also the interest rate. Note that these values are derived using both the probability of default and the benefit matrix.

2.1.1 *Making Non-negative Profit*

We have discussed conditions to make non-default prediction to be optimal under various circumstances. However, these will not guarantee a positive profit, when $B_D < 0$. To ensure non-negative profit from each customer, $B_{ND} \geq 0$ is also necessary. So the true non-default prediction benefit function $b_{00}(k, x)$ should satisfy

$$(1 - P_K)b_{00}(k, x) \geq -P_k b_{01}(x). \quad (11)$$

From equation 8, we can derive the following conditions: if $P_k = 1$ (see equation 8), we should have $b_{01}(x) \geq 0$, which derives ($x \leq 0$) implying that no money should be loaned;

when $P_k = 0$, the above equation 11 is always true, so always pass the loan at a minimal interest rate; and for $0 < p_k < 1$, the required interest rate should be

$$r(k, x) \geq \frac{-P_k c}{1 - P_k} - a. \tag{12}$$

Again if there exists a maximum interest rate r_{max} allowed, then it should satisfy

$$r_{max} \geq r(k, x) \geq \frac{-P_k c}{1 - P_k} - a. \tag{13}$$

So, $P_k = \frac{r_{max}+a}{r_{max}+a-c}$ will lead to zero profit with r_{max}. $P_k > \frac{r_{max}+a}{r_{max}+a-c}$ will lead to negative profit even with r_{max}, thus implying that the loan request should be declined. We can gain non-negative profit, only for the following condition: $P_k < \frac{r_{max}+a}{r_{max}+a-c}$, with an interest rate $r > \frac{-P_k c}{1-P_k} - a$.

Combining the results we have now, we get the following conditions for making non-negative profit from each customer:

- For customers with $P_k = 1$, always decline the application.
- For customers with $P_k = 0$, always accept the application.
- For customers with $0 < P_k < 1$, the interest rate should satisfy equation 4 and 11 simultaneously.
  - if $\frac{r_{max}+a}{r_{max}+a-c} < P_k < 1$: results negative profit, decline the application.
  - if $P_k = \frac{r_{max}+a}{r_{max}+a-c}$: zero profit, decline or accept decision can be taken on the basis of profit from resulting underwriting charges, etc.
  - if $0 < P_k < \frac{r_{max}+a}{r_{max}+a-c}$, decide the interest rate by equation 5 and 6 for the loan, or determine the amount of money should be lent by equation 10.

We would like to point out that similar analysis can be easily conducted for any changes in benefit functions, thus becoming a useful tool for loan practices.

3. SIMULATIONS AND RESULTS

At this point, we have only conducted preliminary experiments on publicly available UCI datasets [1] — the covtype, and pima dataset. Since, we did not have a financial dataset readily available, we chose to use a moderately unbalanced datasets from the UCI repository. We randomly divided the dataset into 70% for training and 30% for testing. We used four different classifiers: C4.5 decision trees [9] (J48), NBTree [6], Bagging with J48, and Bagging with NBtree. Our eventual goal is to utilize multiple learning algorithms with a variety of datasets to validate the proposed pricing scheme. We believe that the ROC curve analysis is limiting for pricing schemes, as it does not allow one to evaluate the quality of P_K. Hence, we also use different loss measures that capture the quality of the probability estimates. The actual total profit is a function of:

$$\sum_i (I(y = ND)b_{00}(k_i, x_i) + I(y = D)b_{01}(x_i)) \tag{14}$$

in which k_i is the corresponding percentile for the ith customer, y is the actual class of the instance (k_i, x_i), and $I(\cdot)$ is

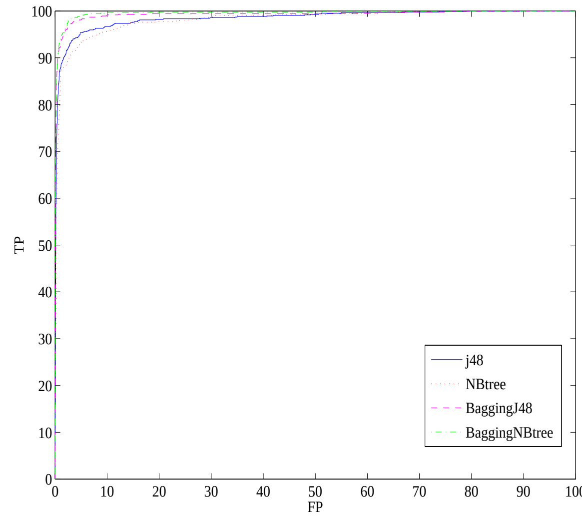

Figure 2: ROC curves for Covtype dataset from J48 (-solid), NBtree (dotted), BaggingJ48(dashed), and BaggingNBtree(dash-dotted)

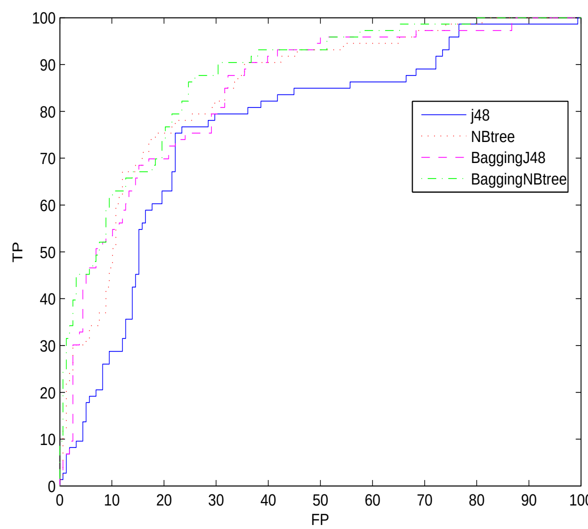

Figure 3: ROC curves for Pima dataset from J48 (-solid), NBtree (dotted), BaggingJ48(dashed), and BaggingNBtree(dash-dotted)

Table 2: Profit from Different Models ($r_{max} = 0.1$)

Dataset and Model	Probability Loss		Profit		
	average NCE	average brier	simulation 1	simulation 2	simulation 3
Covtype			$1.0e+5*$	$1.0e+6*$	$1.0e+5*$
J48	0.0268	0.0295	4.9074	0.5061	4.7703
NBtree	0.0369	0.0369	4.5403	0.4740	4.4527
BagJ48	0.0180	0.0203	6.6351	0.6815	6.7984
BagNBtree	0.0177	0.0204	9.7745	1.0131	9.8566
Pima			$1.0e+4*$		
J48	0.2380	0.3479	3.6396	0.9603	1.6686
NBtree	0.1924	0.2855	3.0532	0.7739	2.0492
BagJ48	0.1992	0.2950	3.2529	1.1815	1.6608
BagNBtree	0.1844	0.2724	4.7752	2.7599	2.9511

Table 3: Profit from Different Models ($r_{max} = 1/3$)

Dataset and Model	Probability Loss		Profit		
	average NCE	average Brier	simulation 1	simulation 2	simulation 3
Covtype			$1.0e+5*$		
J48	0.0268	0.0295	5.1169	5.0477	5.0875
NBtree	0.0369	0.0369	5.9022	5.8350	6.3156
BagJ48	0.0180	0.0203	6.8557	7.0939	7.0119
BagNBtree	0.0177	0.0204	9.7755	9.9511	9.8868
Pima			$1.0e+4*$		
J48	0.2380	0.3479	5.8322	3.0318	3.0332
NBtree	0.1924	0.2855	6.2662	5.2977	2.8935
BagJ48	0.1992	0.2950	7.0204	3.9391	3.0269
BagNBtree	0.1844	0.2724	7.3293	5.5422	4.5443

the indicator function that has value 1 in case the argument is true and 0 otherwise (ND= Non-Default, D=Default).

The profit derived from an individual consumer can potentially be negative, due to bad loaning decisions, but the model ideally should result in aggregated positive profit. The weighted benefit of true negative should overcome the cost of false negative.

To compare the performance from different models, the benefit function b_{11} and b_{10} are assumed to be zero, the loan amount x asked by the customer is generated by *-raylrand(n)* in *matlab*, which generates random numbers with Rayleigh distribution, with $n = 10,000$. But the actual loan amount granted is determined by discussion in previous sections. We ran three different simulations with the random loan amounts generated from the Rayleigh distribution. We also report the losses on the probability estimates using the negative cross entropy (NCE) and brier score measures [3, 2].

$$NCE = -\frac{1}{n}\{(\sum_{i|y=1} \log(p(y=1|x_i)) + \sum_{i|y=0} \log(1 - p(y=1|x_i)))\}$$

$$Brier = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2$$

The limit of interest rate is set to be $r_{max} = 1/10$ and $r_{max} = 1/3$. Tables 2 and 3 show the results including the loss measures and overall profit for three different set of simulations and at different maximum interest rates. We notice that the quality of the estimates as indicated by the NCE or Brier score, particularly the NCE, agrees with the overall profit. This is not surprising as the benefit is weighted on the probability. Figures 2 to 3 show the ROC curves. We see that the ROC curves overlap for a large region of the ROC space. However, if we were to fit the ROC convex hull [8], it would potentially capture the models that result in maximal profits. For instance, bagged NBTree model has the highest profit for both covtype and pima.

We would like to point out that ROC curves are certainly limited in their evaluation of the quality of probability estimates. A classifier can achieve a high rank-ordering, but that is not necessarily indicative of the underlying calibration of the model [3]. Moreover, the benefit function is dependent on the quality of probability estimate. A weakly calibrated model can result in less economic utility, as evident from our results too. We advocate the use of appropriate loss measures, indicative of the calibration of the model for evaluating models utilizing a pricing scheme for loan decisions.

The empirical results are limited to be able to draw convincing observations. Nevertheless, there are strong trends that spell out. There is a strong relationship between quality of the probability estimates and the resulting profit from the model. Figure 4 shows the negative cross entropy computed per test example and the corresponding profit for the two dataset. It is clear that as the negative cross entropy increases, indicating poor quality of estimates, the profits decrease.

4. CONCLUSION

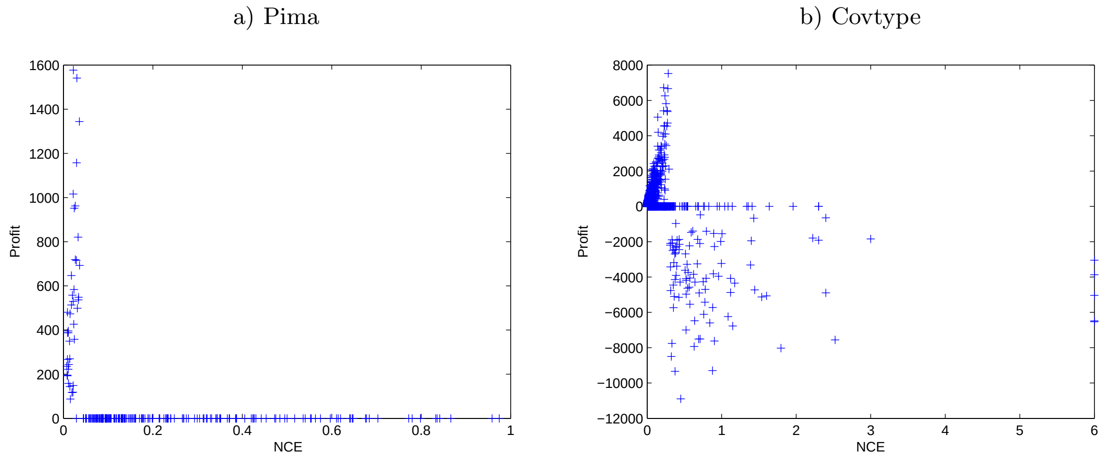

Figure 4: NCE vs Profit Plots

We discussed a pricing scheme for loan practice based on credit scoring models (classifiers) and benefit matrix. The framework enables pricing the loan for each individual customer conditioned on the loan amount and the corresponding interest rate. The two important conditions for the scheme are: 1. granting loans is more profitable than declining it; 2. ensuring a non-negative profit from loans. The theoretical framework can be encompassed to utilize the different policies as set by the bank. One factor that we did not explicitly consider is the amount of downpayment made by a consumer. That is a potential risk mitigator and can influence the interest rate.

As part of future work, we are investigating applications of the proposed scheme to other domains that require the parameterization of costs and benefits. The parameters are replaceable by the domain specifics. One application that comes to mind is medical informatics. For example, consider the case of diagnosing a patient with a disease. There are obvious benefits and costs associated with that domain. The benefits include correctly diagnosing and predicting the patient with disease, and the losses include the costs from mis-diagnosis.

Acknowledgments

We are grateful to Roger Stein for the pointer to his work. We would like to thank the anonymous reviewers for their helpful comments on the work.

5. REFERENCES

[1] C. Blake and C. Merz. UCI Repository of Machine Learning Databases http://www.ics.uci.edu/~mlearn/~MLRepository.html. Department of Information and Computer Sciences, University of California, Irvine, 1998.

[2] G. Blattenberger and F. Lad. Separating the Brier Score into Calibration and Refinement Components: A Graphical Exposition. *The American Statistician*, 39(1):26–32, February 1985.

[3] N. V. Chawla and D. A. Cieslak. Beyond error: Calibration and power of probability estimation decision trees and naive bayes. Technical Report 2006-09, University of Notre Dame, 2006. `www.cse.nd.edu/research/tech_reports`.

[4] P. Domingos. Metacost: A General Method for Making Classifiers Cost-sensitive. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 155–164, San Diego, CA, 1999. ACM Press.

[5] C. Elkan. The foundations of cost-sensitive learning. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 973–978, 2001.

[6] R. Kohavi. Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 202–207, 1996.

[7] D. Margineantu. Class probability estimation and cost-sensitive classification. In *Proceedings of the Thirteenth European Conference on Machine Learning*, pages 270–281, 2000.

[8] F. Provost and T. Fawcett. Robust Classification for Imprecise Environments. *Machine Learning*, 42/3:203–231, 2001.

[9] J. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, 1992.

[10] R. M. Stein. The relationship between default prediction and lending profits: Integrating roc analysis and loan pricing. *Journal of Banking and Finance*, 29:1213–1236, 2005.

[11] B. Zadronzy. One-benefit learning: Cost-sensitive learning with restricted cost information. In *Proceedings of the Workshop on Utility-Based Data Mining (UBDM) at the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (UBDM@KDD'05)*, 2005.

[12] B. Zadrozny and C. Elkan. Learning and making decisions when costs and probabilities are both unknown. In *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (KDD'01)*, 2001.