

Paper 124-28

Multistage Cross-Sell Model of Employers in the Financial Industry

Kwan Park and Steve Donohue
The Principal Financial Group

ABSTRACT

This paper details the steps to develop a multistage cross-sell model of employers in the Financial Services industry. This model can be used to score the likelihood of an employer to purchase multiple products. Topics covered include data preparation for data mining, several advanced modeling techniques, and multistage cross-sell model score comparison. Strong non-linear curvatures among the input variables were found during the modeling process and examined. The modeling techniques used were Decision Tree, Neural Network, Regression and Memory Based Reasoning in SAS Enterprise Miner (version 4.1) of Windows 2000 system.

INTRODUCTION

In the Financial Services industry, the cross-selling and retaining of customers have become very important issues. These issues have been addressed via the development of many predictive models. These models have been designed to identify customers having a high likelihood of purchasing multiple products. However, the vast majority of these models have been done at the business-to-consumer level. The independent variables for these models have generally included transactional data: transaction frequency, amount of transaction and purchase sequence, and demographic data (age, gender, income, geographic location, etc.)

Alternatively, this paper explains an approach for multistage cross-sell modeling at the business-to-business (or employer) level. For example, it examines selling Group Medical Insurance to employers Group Long-Term Disability plans. The independent variables used were employer level transactional data (years of relationship, number of products owned, total premium values, and firmographic data (total number of employees, residential population, sales volume, standard SIC code, legal status, region, etc.) The object of the modeling being to score the employers on likelihood to become multiple products owners and predict the sequence of product purchase.

Three sets of products were analyzed: Group Insurance, Pension, and Executive Benefits. Included in Group Insurance were Medical/Health, Dental, Vision, Long-Term Disability, Short-Term Disability, and Life. Within the Pension set were 401(K) plans and Defined benefit cases. As for Executive Benefits, 19 product lines were utilized. Dun and Bradstreet employer data was employed for the firmographic data. SAS Enterprise Miner was used for modeling and the modeling techniques were decision trees, neural network, logistic regression and memory based reasoning. For the development procedure, SAS's

SEMMA methodology was applied. Strong non-linear relationships among the input variables were found and these non-linear curvatures contributed to the models in different ways.

OBJECTIVE

The objectives for this modeling are the following:

1. Identify employers with a high likelihood of becoming multiple product owners.
2. Determine the product purchase sequence.
3. Assign a score according to likelihood to purchase identified product.

DATA

This study deals with two types of data for each of the three employer groups:

Transactional data – product coverage, length of relationship, total contract amount, premium amount, number of active employees under contract, etc.

Firmographic data – geographic region, metropolitan area, total number of employees, sales volume, years in business, number of customers, legal status, gender of CEO, private/public company, SIC industry code, etc.

(The firmographic data was extracted from Dun and Bradstreet and matched with tax-id.)

MODELING

The first object function of the modeling efforts was to predict the multiple product employers.

1. Data was sampled to insure the same size of the prior target, and partitioned as training and validation data set of 60% and 40%, respectively, with strata information of prior target probability.
2. Data Mining Database Node was used to produce only one data mining database, which optimized the performance of analytic nodes.
3. Missing values were replaced by 'tree imputation with surrogates' method. This method is identical to tree imputation, with the addition of surrogate splitting rules. A surrogate rule is a back up to the main splitting rule. When the main splitting rule relies on an input whose value is missing, the first surrogate rule is invoked. If the first surrogate also relies on an input whose value is missing, the next surrogate is invoked. If missing values prevent the main rule and all the surrogates from applying to an observation, the main rule assigns the observation to the branch assigned to receive missing values.

4. Decision Tree, Neural Network, Logistic Regression and Memory Based Reasoning technologies were applied to predict the potential multiple products owners.

Among the Least Squares, Gini reduction and Entropy reduction, Gini with Model assessment measure of 'Total leaf impurity' showed the best model.

The Gini index is interpreted as the probability that any two elements of a multiset chosen at random with replacement are different. A pure node has a Gini Index of 0, as the number of evenly distributed classes increases, the Gini index approaches 1.

The Gini Index formula is as follows:

$$1 - \sum_{j=1}^r p_j^2 = 2 \sum_{j < k} p_j p_k$$

Entropy is another measure of variability of categorical data. Consider r mutually exclusive events with probabilities p_1, p_2, \dots, p_r , then Entropy is defines as:

$$H(p_1, p_2, \dots, p_r) = - \sum_{i=1}^r p_i \log_2(p_i)$$

Chi-Squared test statistic is as follows:

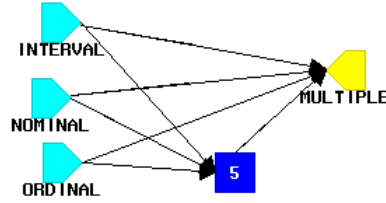
$$\chi_v^2 = \sum \frac{(O - E)^2}{E}$$

The artificial neural network was originally developed by researchers who were trying to mimic the neural system of the human brain. By combining many simple computing elements into a highly interconnected system, these researchers hoped to produce complex phenomena such as intelligence. While there is considerable controversy over whether artificial neural networks are really intelligent, there is no doubt that they have developed into very useful statistical models. More specifically, feedforward neural networks are a class of flexible nonlinear regression, discriminant, and data reduction models. By detecting complex nonlinear relationships in data, neural networks can help to make predictions about real-world problems.

Neural Network architecture was set as one layer of five hidden nodes, and ordinal, nominal and interval variables and hidden nodes were used an inputs to target variable. Hyperbolic Tangent function was used as activation function and general linear combination function was used. The tangent hyperbolic function is defined as

$$g(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$$

The network architecture was:



This network model is somewhat more complex than usual neural network model architecture. The choice of model complexity is always an issue to modelers. The above architecture was chosen by trial and error among many possible architectures

This multiplayer perceptron (MLP) is a feed-forward neural network that uses sigmoid activation functions. This MLP is one of the most common types of neural network used for supervised prediction. Empirically, it is often found that the tangent hyperbolic function give rise to faster convergence of training algorithms than the logistic functions.

Another alternative of the function is ordinary radial basis function (ORBF). ORBF networks are universal approximators in theory, but in practice they are often ineffective multivariate function estimators. In this modeling, ORBF networks were excluded.

The most common incorrect concept of a neural network is that it cannot be explained or cannot be represented as an equation. However, a neural network is an application of non-linear regression that can be represented as a single equation even though that equation is very complicated. The beauty of a neural network is that it can find very complicated non-linear relationship among variables under the condition that the errors are quite small.

As the target variable is a dichotomous variable, a logistic regression model was used, and for the memory based reasoning model, a default setup was used. Usually, many modelers favor the backward selection method because it evaluates all the input variables in relation to the target variable. In this modeling process, stepwise selection, backward selection and forward selection methods were applied to develop regression model, and stepwise method showed the best performance. The logistic regression equation is as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_n\beta_n$$

The last modeling technique applied in this modeling is memory-based reasoning (or case based reasoning). It is a process that identifies similar cases from history data or a database and applies the information that is obtained from these cases to a new

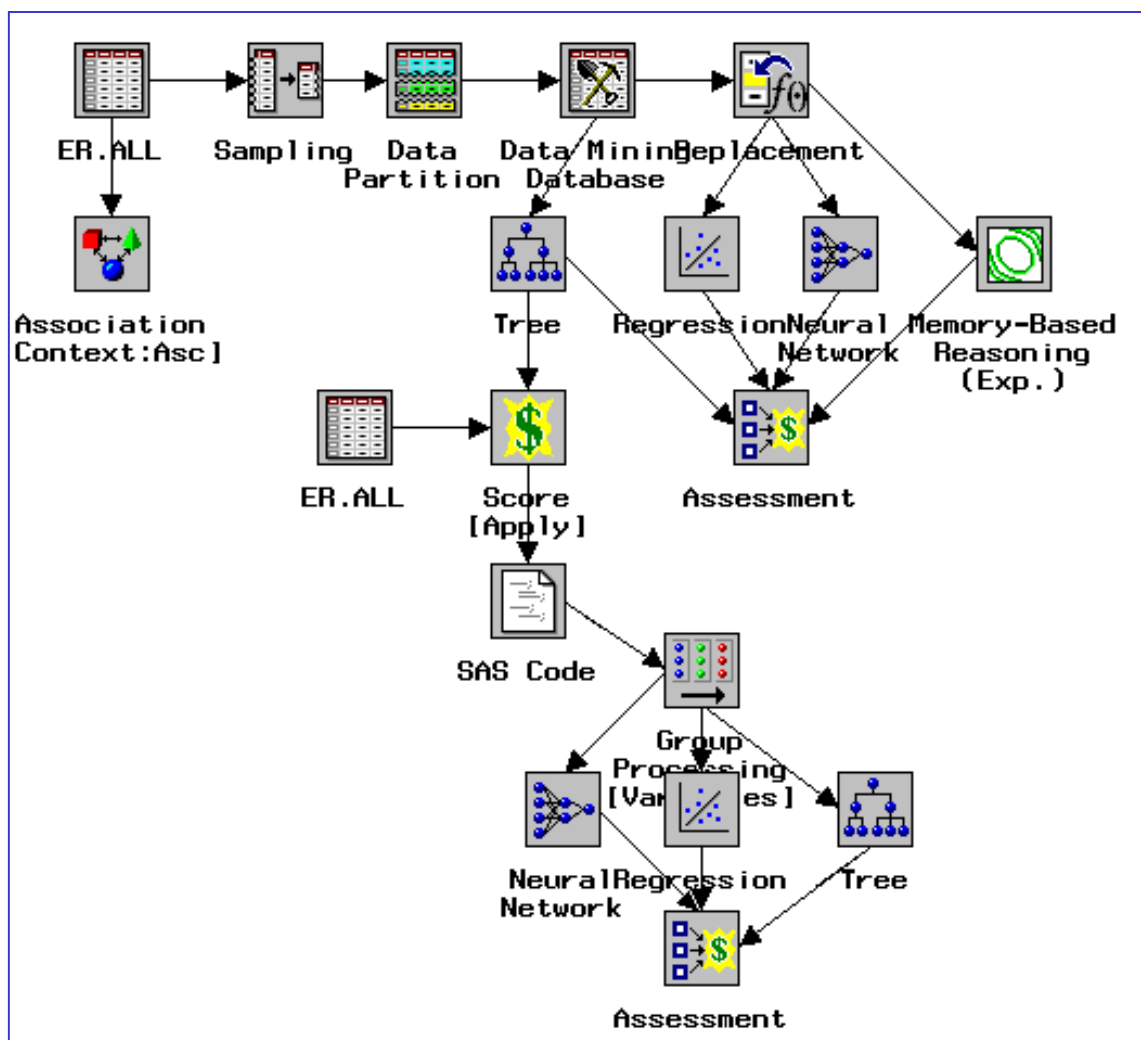
record. The memory based reasoning node in SAS Enterprise Miner is a modeling tool that uses a k-nearest neighbor algorithm to categorize or predict observations. The k-nearest neighbor algorithm takes a data set and a probe, where each observation in the data set is composed of a set of variables and the probe has one value for each variable. The distance between an observation and the probe is calculated. The k observations that have the smallest distances to the probe are the k-nearest neighbor to that probe. A default set up of memory based reasoning node was used in this modeling.

5. The decision tree showed the highest lift values within the top 30 percentiles. It produced nine segmentations and the following proc sql (statement is to compute the lift values of each nodes.)

```
proc sql;
create table scored as
select a._node_ , a.single, b.multiple
from (select _node_ , count(dunsno) as single
      from &_score
      where multiple='Single'
      group by _node_ ) a
left join
(select _node_ , count(dunsno) as multiple
 from &_score
 where multiple='Multiple'
 group by _node_ ) b
on a._node_ = b._node_ ;
quit;

data scored;
set scored;
if multiple=. then multiple=0;
total=single+multiple;
score=int(multiple/total*10000)/100;
run;
```

Enterprise Miner Diagram



6. Node Description

The following table explains the node definitions and lift values

Segment	Lift
Years of Business >100 years AND Public/Private IS NOT MISSING AND Sales Volume IS : [0] [1M,2M] [100M,500M]	1.99
Headquarter Location AND Sales Volume IS : [2M,100M]	1.62
Single Location AND Sales Volume IS : [2M,100M]	1.49
Years of Business IS : (5,100] AND Public/Private IS NOT MISSING AND Sales Volume IS : [0] [1M,2M] [100M,500M]	1.15
Years of Business IS : (0,5] AND Public/Private IS NOT MISSING AND Sales Volume IS : [0] [1M,2M] [100M,500M]	0.95
Sales Volume IS : [300K,400K] [500K,1M] [500M+]	0.56
Sales Volume IS : [1,300K] [400K,500K]	0.36
Location Status IS MISSING AND Sales Volume IS : [2M,100M]	0.00
Public/Private IS MISSING AND Sales Volume IS : [0] [1M,2M] [100M,500M]	0.00

The lift values of the top four nodes are greater than 1, and the employers in these four nodes were selected to find the sequence of the next products. For this analysis SAS Enterprise Miner's Association node was applied. This is the second objective function of this modeling. The product purchase date is needed

to run the association node. The next table is the top 15 product sequences. Product A → Product B showed the highest frequency of product purchasing sequence followed by Product B → Product A.

Product Sequence	Frequency	Percent
Product A ==> Product B	1,109	5.71
Product B ==> Product A	874	2.13
Product A ==> Product C	531	2.73
Product C ==> Product A	422	1.84
Product A ==> Product D	343	1.77
Product A ==> Product E	313	1.61
Product B & Product C ==> Product A	280	1.89
Product A ==> Product F	279	1.44
Product A ==> Product B & Product C	279	1.44
Product F ==> Product A	244	1.21
Product A ==> Product B & Product E	243	1.25
Product A ==> Product B & Product D	243	1.25
Product E ==> Product A	240	2.77
Product D ==> Product A	238	3.68
Product B ==> Product C	234	0.57

7. The second stage models are to predict the employers who will be Product B among current Product A customers and to predict the employers who will be Product B customers among current Product A customers, etc. Since so many models are required,

only the top six nodes were chosen. Each was then scored for each individual employer. Next, the cross-sell scores were computed. The following table is an example of the scores.

Employer ID	Cross sell score	Product A → Product B	Product B → Product A	Product A → Product C	Product C → Product A	Product A → Product D	Product A → Product E
3201767	8.72	6.79	0.00	0.23	0.00	5.26	5.10
1905289	8.72	3.25	0.00	0.62	0.00	7.98	6.65
6706752	8.72	0.00	4.85	0.15	6.23	0.62	1.52
1008754	7.13	1.65	4.85	0.23	3.50	5.26	1.52
2709015	7.13	2.76	3.21	4.52	0.00	3.31	1.52
Etc.							

APPLICATION

The application of this approach is to aid the cross-sell marketing efforts at the employer level. Employers can be selected and appropriately targeted based on their propensity to purchase selected products. Marketing campaigns can be designed to focus on those employers that are most likely to purchase. Thus, this methodology can greatly impact the efficiency of marketing campaigns.

Steve Donohue
The Principal Financial Group
711 High Street
Des Moines IA, 50392
Work Phone: (515)362-2860
Fax: (515)283-5332
Email: Donohue.Steve@principal.com
Web: www.principal.com

CONCLUSION

This paper describes employer level cross sell modeling to identify the likelihood of purchasing multiple products and the product purchase sequence. It should also be noted that campaign results of employer level cross sell can take more effort and time compared to employee level campaign. Tracking must be established to identify the impact of any marketing efforts.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

REFERENCES

- Bishop, C.M. (1995), *Neural Networks for Pattern Recognition*, New York: Oxford University Press.
- Breiman, L., Friedman, J. H., Olshen, R.A., and Stone, C. J. (1984), *Classification and Regression Trees*, Chapman and Hall.
- Enterprise Miner Software, Online Tutorial (SAS V8).
- Ripley, B.D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press.
- Rud, Olivia Parr (2001), *Data Mining Cookbook*, New York: John Wiley & Sons.
- Zahavi, J. and Levin, N. (1997), "Applying Neural Computing to Target Marketing," *Journal of Direct Marketing*, 11, 5-22.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Kwan Park
The Principal Financial Group
711 High Street
Des Moines IA, 50392
Work Phone: (515)247-5647
Fax: (515)283-5332
Email: Park.Kwan.Soo@principal.com
Web: www.principal.com