Abstract # 007-768

Saving Money by Using Statistical Models Before Launching a New Product

Alexandre T. Mafra
Escola Adm. Empresas (EAESP)
Fundação Getulio Vargas (FGV)
Av. 9 de Julho, 2029
São Paulo, SP, Brazil
ZIP code 01313-902
alexandre.mafra@fgv.br

Wagner Bronze Damiani
Marriot School of Management
Brigham Young University (BYU)
540 TNRB
Provo, Utah
ZIP code 84602
damiani@byu.edu

POMS 18th Annual Conference

Dallas, Texas, U.S.A.

May 4 to May 7, 2007

Abstract

A reasonable amount of time and money is usually spent in the launching process of a new product or service. This paper covers a real issue faced by one of three major credit card players in the world. The sales results can be improved while reducing time and cost by using neural network models. The revenue forecasting process for a new product or service deployment among the current customers is supported by information technology. This is even done before starting to spend the corporate budget to advertise it. The efficacy of the application of clustering methodology with on line analytical processing technology to the profiling of potential buyers for new products and services is demonstrated. The financial analysis and simulation process can be largely enhanced through the path presented in this paper. The available literature is also briefly revised.

1. Introduction

The development of a new product is always a major challenge for organizations. Product development represents the expectations of a company to explore new markets, to expand its presence with new product lines within not so new market segments, or to increase or regain sales within well known segments [1]. Development is considered to be the first stage in the product life cycle, and thus a critical one: all subsequent stages depend on a good job done at this point [2][3].

Product development consists of two interdependent, yet parallel, processes: product design and market analysis. This paper focuses on the market analysis for a new product of a major credit card company.

This new product was targeted at small business professionals such as lawyers, engineers, veterinarians, and other professionals who run small companies, and that would benefit from a specialized version of their personal credit cards to pay for expenses that are related to their professional activity. This product would allow the separation of personal and professional transactions, have improved financial management capabilities, and a package of benefits associated with using the credit card to pay for services and supplies associated to such expenses.

The company suspected that there was an important subset of their personal credit card accounts that would match the target business profile. The adoption of the new product by these customers could lead to an expected increase in the use of the credit card as their preferred payment method therefore, increasing the overall revenue and profits of the credit card company. At this point, two fundamental questions are posed:

- What are the spending profiles of these small business professionals?
- How many customers in the personal cards database match these profiles?

These are quite non-trivial questions due to their speculative nature. The answers for them do not arise from standard Business Intelligence (BI) query and aggregation techniques [4][5]. More sophisticated techniques must be employed to build up the spending profiles and to query for these profiles in the personal cards database. It is also necessary to use a data source capable of separating business and personal transactions.

2. Development of customer profiles

2.1. Data source for the profiles

The corporate card transactions database was chosen as the source for the business transactions in which the spending profiles were developed. This choice is supported by two reasons. First, the accounts in the corporate card transactions database are supposed to respond positively to the package of benefits offered by the corporate card,which are similar to the benefits to be offered with this new product. Second, the corporate database has a long history of transactions, long term corporate customers, high data quality, and was readily available.

2.2. Variable selection

The variables used should reflect the spending habits of each account. Each transaction, either in the corporate or personal cards database, is associated with a commercial establishment, which is classified according to its business segment. Certain business segments were selected as the more representative of their spending habits.

The number of transactions in each business segment was computed in both databases. All business segments holding more than 1% of the total number of transactions in its database were selected. The most frequent segments from both databases were merged together, yielding to 15 very unique business segments.

The total spending along with the number of transactions over a one year period, in each segment, was computed for every account in both databases. These totals were divided by the number of active months within the one year period. Accounts with less than 6 months of activity were discarded. Accounts that did not use the card at least once in the last 2 years were also discarded as inactive. Accounts that did not make at least two transactions in any of the selected segments were also discarded for having too few observations.

The 30 variables (15 spending + 15 frequency) computed for each account were packed as an attribute vector representing the spending habits of the account.

2.3. Clustering of spending profiles

Once the data corpus of attribute vectors for each account in the corporate cards database was constructed, it was then decided to use a clustering algorithm to develop the spending profiles. Jain, Murty and Flynn present an extensive review on clustering algorithms [6], as does Berkhin [7]. This decision was supported by the belief on the existence of natural groupings of spending behaviors, confirmed later by the results.

Cluster analysis is the organization of a collection of attribute measurements into clusters based on similarity. Accounts within a valid cluster are supposed to have a spending pattern more similar to each other than they are to a pattern belonging to a different cluster.

2.4. Artfificial Neural Networks (ANN) and Self-Organizing Maps (SOM)

The clustering technique of choice was to execute Vector Quantization [8][9] of the spending attribute vectors using Self-Organizing Maps [10][11]. In vector quantization, a codebook of representative vectors is computed from observation vectors taken from a training database (the corporate cards in this case). Each observation in the training database is associated to its closest representative vector (code vector). The group of all observation vectors associated to each code vector is a cluster, and the position of the code vector constitutes the cluster center.

Self-Organizing Maps constitute a class of Artificial Neural Networks (ANNs) that are very efficient in the execution of vector quantization.

Artificial Neural Networks

According to Haykin [12], an Artificial Neural Network (ANN) is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process.
2. Inter-neuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

The design basis of an ANN is the computational model of the artificial neuron, known as *Processing Unit* (PU) or *Processing Element* (PE). The most used model is the non-linear neuron, represented in **Figure 1**. The three basic elements of a processing unit are:

1. A set of *connections* or *synapses*, each of which characterized by a *weight* or *strength*.
2. An *adder* that linearly combines the input signals x_i, weighted by the connection weights $w_{k,i}$.
3. An *activation function* $\varphi(\cdot)$ for limiting the amplitude of the output signal of the processing unit y_k.

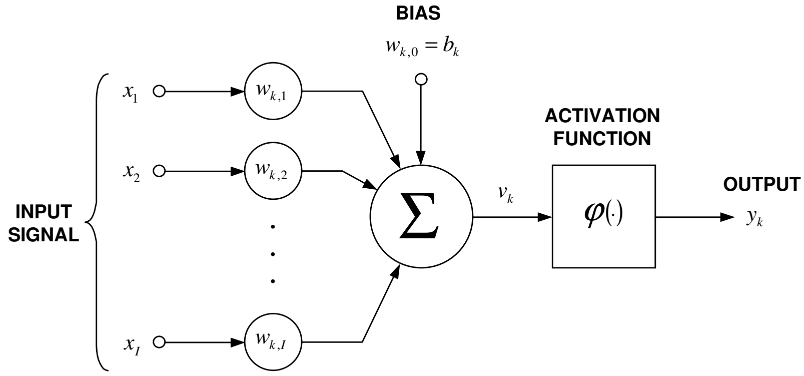

Figure 1: Non-linear neuron model.

Haykin presents a comprehensive source on Artificial Neural Networks [12], its architectures, training algorithms and applications.

Self-Organizing Maps

Self-Organizing Maps were formalized by Teuvo Kohonen in 1990 [10]. They are single layered ANNs, made of linear activation function units, fully connected to the input nodes, having feedforward signal propagation (**Figure 2**). As an additional feature, a normally uni- or bi-dimensional lattice is associated to a SOM, on which it is defined a discrete coordinate system. The SOMs units are assumed to be located at the lattice's nodes, thus having a position defined on this coordinate system. This feature is used to define relative positions between units in a SOM. Kohonen presents a comprehensive reference on Self-Organizing Maps, its properties, training algorithm and applications [11].

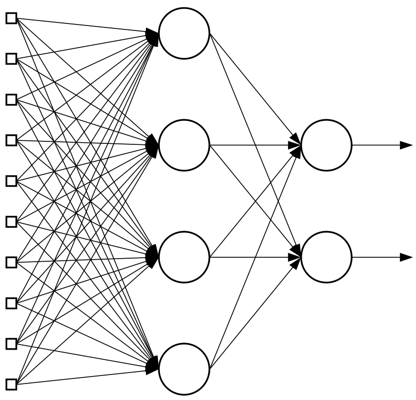
Figure 2: Single layer architecture in a SOM neural network.

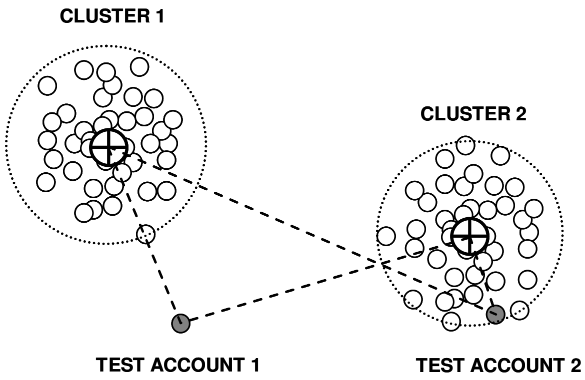

Figure 3: Two clusters, (1) and (2), with the associated corporate card accounts inside the dotted circles. The radius of the dotted circle is the maximum distance between the cluster center and an associated corporate account. Test account 1 is closer to the center of cluster (1), so it is associated to this cluster. Although it is associated to cluster (1), the account is not inside the dotted circle of cluster (1), so it is considered to be "too far" to match the spending pattern represented by the cluster center. Test account 2 is associated to cluster (2) and is close enough to be considered a good match for the spending pattern represented by the center of cluster (2).

3. Search of matching profiles

Having the set of cluster centers (code vectors) representing the spending profiles found in the corporate cards database, it is possible to search for similar profiles in the personal cards database. A matching criterion must be defined to accomplish this task.

The matching criterion employed was based on the distance between the attribute vector of the candidate account and the cluster centers. First, the account is associated to the cluster with the closest center. Then the distance between the candidate vector (personal card) and its associated cluster is compared to the furthest vector associated to the same cluster in the training (corporate cards) database: the candidate must not be farther than this maximum distance (**Figure 3**).

4. Results

The numbers demonstrated here were rescaled to protect the privacy of real data. The personal cards database is assumed to have exactly 1,000,000 candidate accounts satisfying the initial criterion, and all the other numbers were rescaled accordingly. In this scale, the corporate cards database has 25.346 accounts.

After some initial tests with a sample subset of the corporate database, it was decided to use 15 clusters, represented by 15 output neurons in the SOM network with a one-dimensional lattice. Other tests were made with 5, 10, 15, 20, 25, and 30 clusters, and the number of 15 clusters was chosen because it still presented a substantial reduction on the final quantization error when compared to 10 clusters; when 20 clusters were tested, the reduction on the final quantization error compared to the 15 clusters quantization error was much less noticeable.

The final results can be seen in **Table 1**. In general terms, 17.3% of the personal cards accounts have a candidate profile for a business targeted card. The absolute number of 173,364 accounts is almost 7 times bigger than the total number of 25,346 corporate accounts. This is very promising given the large universe for customer prospecting.

Cluster	# Corporate accounts	# Personal accounts	# Personal with corporate profile	% Personal accounts with corporate profile
1	3,950	344,949	69,447	20.1%
2	9,484	336,540	65,255	19.4%
3	866	66,457	8,267	12.4%
4	772	112,084	9,333	8.3%
5	436	15,515	1,748	11.3%
6	3,064	66,016	13,720	20.8%
7	323	9,221	662	7.2%
8	282	13,831	1,055	7.6%
9	241	4,235	113	2.7%
10	2,481	8,400	2,335	27.8%
11	798	5,038	322	6.4%
12	448	994	56	5.7%
13	910	5,493	383	7.0%
14	464	6,062	368	6.1%
15	828	5,165	302	5.8%
Total	**25,346**	**1,000,000**	**173,364**	**17.3%**

Table 1: Distribution of corporate and personal accounts according to their associated clusters, and the number and percentage of personal accounts that match the distance criterion to be considered to have a corporate profile.

Taking a closer look at the results, it may be seen that clusters 1 and 2 concentrate approximately 78% of the total selected prospective consumer accounts (134,702 accounts). This concentration is quite remarkable, because the customers of these clusters may represent distinctive profiles to act upon.

When looking at the corporate accounts associated to cluster 1, it may be observed that the Supermarket and Gas Stations are by far the most frequent and value concentrating business segments. These segments do not characterize the use of the corporate credit cards to pay for professional expenses. It seems to be much more likely that these corporate cards are being used for personal expenses, and not the contrary.

On the other hand, cluster 2 has a much more business-like face. Transactions on airline services, lodging, restaurant establishments, and gas stations are the most common and valuable. This combination of business segments was naturally expected to be found on the profile of corporate users.

The other clusters show very specific spending patterns, such as in cluster 7, driven by transactions only at book stores, or in cluster 10, with transactions only with airline companies. As it happened with cluster 1, these could not be considered to be legitimate corporate uses of the credit card.

In summary, it can be stated that cluster 2 represents the spending profile that is adequate for a small business credit card: personal cards with sound corporate-like expenses. This represents a universe of 65,255 good prospective accounts, approximately 6.5% of the total personal cards database, a very positive perspective.

At the present moment, the small business product is already deployed however, it is still in the very early days of existence. Large investments are planned to promote the benefits of this product for its target customers.

5. Conclusions

The development of a new product can benefit a lot from good market analysis. Estimating the size of the target universe well can help anticipate the success or failure of the new product. This paper presented an Artificial Neural Network based technique, used in developing the spending profiles of corporate card users. The spending profiles arise from clustering the aggregated values and numbers of transactions on selected business segments. Once the profiles are defined, the personal cards database is scanned to find consumer accounts with spending histories that are similar to the corporate profiles. A matching criterion is then defined and the prospective accounts are analyzed.

The existence of natural groupings of accounts sharing similar spending patterns is confirmed. The results show that there is actually a universe of 6.5% of the consumer accounts that match a sound business driven profile. Surprisingly, the results also show that there is a large portion of the corporate accounts that are making "home use" of their credit cards. At the present moment, the product was already launched and is in its early days of sales.

6. References

[1] Booz, Allen, and Hamilton, "New Product Management for the 1980s", New York, 1982.

[2] George S. Day, "The product life cycle: analysis and applications issues", *Journal of Marketing*, vol. 45, no. 4, autumn 1981, pp. 60–67.

[3] Robert G. Cooper, *Winning at New Products: accelerating the process from idea to launch*, 3rd ed., Product Development Institute, 2001.

[4] Ralph Kimball, Margy Ross, *The Data Warehouse Toolkit: the complete guide to dimensional modeling*, 2nd ed., John Wiley & Sons, 2002.

[5] W.H. Inmon, *Building the Data Warehouse*, 2nd ed., John Wiley & Sons, 1996.

[6] A.K. Jain, M.N. Murty, P.J. Flynn, "Data Clustering: a review", *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, september 1999.

[7] Pavel Berkhin, “Survey of clustering data mining techniques”, *Accrue Software*, 2002.

[8] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic Publishers, 1992. (The Kluwer International Series in Engineering and Computer Science).

[9] T. Kohonen, *Learning vector quantization for pattern recognition*. Technical Report TKK-F-A601, Helsinki University of Technology, 1986.

[10] T. Kohonen, “The self-organizing map,” *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sept. 1990.

[11] T. Kohonen, *Self-Organizing Maps*, 3rd Ed., Berlin: Springer-Verlag, 2001 (Springer Series in Information Sciences, 30).

[12] S. Haykin, *Neural Networks: A comprehensive foundation*, 2nd Ed. Englewood Cliffs, NJ: Prentice Hall, 1999, 842 p.