Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Investigating purchasing patterns for financial services using Markov and MTD models

A. Prinzie & D. Van den Poel
Department of Marketing, Ghent University, Belgium

Abstract

Over the past two decades, the financial markets have become more competitive resulting in diminishing profit margins and blurring distinctions between banks, insurers and brokerage firms. Hence, nowadays a small number of large institutions offering a wider set of services dominate the financial-services industry. These developments stimulated the implementation of Customer Relationship Management (CRM). Given the increasing customer-acquisition cost, marketers realize that the best prospects for the sales of current and new financial services are the current customers. So cross-sell actions are created to motivate existing customers to use additional services from the firm. A great opportunity lies in cross-selling insurance products to bank clients and vice versa. In this study, we investigate purchase patterns of financial services from an international financial institution to identify cross-selling opportunities. We introduce the Mixture Transition Distribution model (MTD) as a parsimonious alternative to the Markov model enabling the estimation of high-order Markov chains and facilitating the interpretation by providing a much smaller transition matrix and lag parameters. Our results are in favor of the MTD model.

1 Introduction: sequence analysis for cross-selling purposes

Sequences have been the subject of research in archaeology, biology, computer sciences, economics, history, meteorology, psychology and sociology. In marketing, sequence analysis has been applied in choice modelling. Consumer behavior can be seen as a *sequence*: i.e., a succession of events (in casu decisions/choices). The rationale behind the existence of typical purchase sequences is two-fold: 1) the logical succession due to complementary goods [1] and 2) utility maximisation under budget constraints [2].

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

Several authors found evidence for the existence of a *typical* acquisition pattern for consumer durables [3, 4, 5] as well as for financial services [6, 7]. Such a priority pattern allows the marketer to predict which service a customer will acquire next, given the position of the latter in the acquisition pattern. Kamakura and Ramaswami [7] are the first authors using the discovered priority pattern to identify cross-sell opportunities.

Cross-selling pertains to efforts to augment the number of services a customer uses within a firm. Given increasing acquisition costs, current customers are the best prospects for the sales of current and new financial services. Moreover, selling additional products to a customer positively influences the buyer-seller relationship [8] increasing the customers' lifetime value [9] and decreasing his churn chance. Somewhat surprisingly, cross-selling has received limited attention in the academic literature. Kamakura and Ramaswami [7] identify prospects for cross-selling by predicting the likelihood an investor, given his financial maturity, would acquire a financial service, given its' acquisition difficulty.

In this paper, we analyze acquisition sequences for financial services to identify cross-sell opportunities. Developments in the financial markets aroused bank assurance companies' interest in Customer Relationship Management (CRM) and, more specifically, in cross-selling. Since most financial-services groups originated from mergers between bank and insurance companies, cross-selling insurance products to bank-only customers and bank products to insurance-only customers is a major concern. An in-depth analysis of the acquisition sequences is the foundation of an efficient cross-sell strategy. In this study we compare n-th order Markov models with n-th order Mixture Transition Distribution Models (MTD) [10], as different techniques for sequence analysis.

2 Methodology

The Mixture Transition Distribution Model (MTD) was introduced by Raftery [10] as a parsimonious model for high-order Markov chains with a finite statespace. Although Markov chains are well suited for the representation of high-order dependencies between successive observations of a random variable, they result in a large number of parameters to estimate. As the order l of the chain and the number m of possible values increase, the number of independent parameters increases exponentially and becomes rapidly too large to estimate efficiently [11]. On the other hand, the MTD model involves only one additional parameter for each extra lag [10]. Besides being more parsimonious, the MTD model is also attractive from a managerial perspective. Firstly, whereas the transition matrix of a high-order Markov model is hard to interpret, the MTD model overcomes this caveat by giving a short form $m \times m$ transition matrix capturing the overall tendencies. Secondly, the MTD model provides phi-weights representing the importance of each lag on the current state. This section is based essentially on Raftery [10] and Berchtold and Raftery [11].

2.1 Markov chains

The Markov chain is a probabilistic model representing dependencies between successive observations of a random variable. In this paper, we consider a discrete-time random variable X_t taking on values from the finite set N={1, ..., m}. We want to explain the value taken by X_t as a function of the values taken by previous observations of this same variable.

In a *first-order* Markov model the current value of X_t is fully explained by the value taken by that same variable at time t-1 (i.e., the first lag).
In a *l-th order* Markov chain, the present depends on *l* previous periods. The transition probabilities are:

$$\begin{aligned} &P\left(X_t = i_0 \middle| X_0 = i_t,\dots,X_{t-1} = i_1\right) \\ &= P(\left(X_t = i_0 \middle| X_l = i_l,\dots,X_{t-1} = i_1\right) \\ &= q_{i_l \dots i_0}(t) \end{aligned} \tag{1}$$

where $i_t,\dots,i_0 \in \{1,\dots,m\}$. The probability $q_{i_l i_0}(t)$ is assumed time-invariant: $q_{i_l i_0}$ (cf. a *homogeneous* Markov chain). The model is summarized in a transition matrix Q where the rows give the probability distribution of X_t given any possible value of X_{t-1}. When the order exceeds one, the transition matrix Q contains *structural zeros*, i.e., elements corresponding to transitions that cannot occur. By omitting these, the size of Q is reduced: i.e., *collapsed* or *reduced* form of Q denoted by R [12]. For m=2 and l=2 R is:

$$R = \begin{array}{cc} X_{t-2} & X_{t-1} \\ 1 & 1 \\ 2 & 1 \\ 1 & 2 \\ 2 & 2 \end{array} \begin{array}{c} X_t \\ \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} q_{111} & q_{112} \\ q_{211} & q_{212} \\ q_{121} & q_{122} \\ q_{221} & q_{222} \end{bmatrix} \end{array} \tag{2}$$

Each possible combination of l successive observations of the random variable X is a *state* of the model. The number of states is equal to m^l (in example: 2^2=4). Each row of the matrix Q contains (m - 1) independent probabilities. The total number of independent parameters to estimate is thus $m^l\,(m - 1)$. From the latter, it is obvious that each additional lag involves a substantial increase in parameters to estimate, which might prevent the modelling of high-order Markov chains. This is where the MTD model comes into the picture. Finally, the estimation of the $m^l\,(m - 1)$ parameters goes as follows. The maximum likelihood estimate of the transition probability i_l to i_0 is:

$$\hat{q}_{i_l \dots i_0} = \frac{n_{i_l \dots i_0}}{n_{i_l \dots i_0 +}} \tag{3}$$

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

and
$$n_{i_l \dots i_1 +} = \sum_{i_0=1}^{m} n_{i_l \dots i_0} \tag{4}$$

where $n_{i_l \dots i_1}$ denotes to the number of transactions of the type:

$$X_{t-l} = i_l, \dots, X_{t-1} = i_1, X_t = i_0 \tag{5}$$

The log-likelihood of the entire sequence of observations is:

$$LL = \sum_{i_l \dots i_0 = 1}^{m} n_{i_l \dots i_0} \log\left(q_{i_l \dots i_0}\right) \tag{6}$$

2.2 MTD model

The Mixture Transition Distribution (MTD) model was introduced to approximate high-order Markov chains with less parameters than the fully parameterised model. The effect of each lag upon the present is considered *separately*. As such, it is a simplification of the Markov model, because interaction effects between lags are not considered. The MTD conditional probabilities are a *mixture* of linear combinations of contributions in the past:

$$P\left(X_t = i_0 \middle| X_{t-l} = i_l, \dots, X_{t-1} = i_1\right) = \sum_{g=1}^{l} \lambda_g q_{i_g i_0} \tag{7}$$

The $q_{i_g i_0}$ are the probabilities of an $m \times m$ transition matrix R. The latter is much smaller than the transition matrix of a traditional Markov model ($m^l \times m$), which facilitates the interpretation. Moreover, additional insight is gained from the inspection of the lambda-parameters $\lambda = \{\lambda_1, \dots, \lambda_l\}$, which express the effect of each lag g on the present value of X (i.e., i_0). To ensure the results of the model to be probabilities the lag parameters are strictly positive and sum to one. The parameters λ and q of the MTD model, eqn (7), are estimated by maximising the log-likelihood of the model:

$$LL = \sum_{i_l, \dots, i_0 = 1}^{m} n_{i_l, \dots, i_0} \log\left(\sum_{g=1}^{l} \lambda_g q_{i_g i_0}\right) \tag{8}$$

The MTD model is far more parsimonious than the whole parameterised Markov model. The transition matrix R has only $m \times (m - 1)$ independent parameters (cf. m columns in R and m-1 independent parameters in each row of R). In addition, an l-th order model has $(l - 1)$ independent parameters (cf. lambda's sum to one). Hence, an l-th order MTD model has only $[m \times (m - 1)] + (l - 1)$ parameters to estimate, which is a lot less, given $l>1$, than the $m^l (m - 1)$ parameters of the l-th order Markov model. Moreover, notice that MTD requires only one additional parameter for each supplementary lag. See Table 2.

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

3 Financial-services application

Markov and MTD models are applied on the data warehouse of a major International Financial-Services Provider (referred to as IFSP) serving approximately 50 million customers. We try to gain insight into the sequence by which financial services are acquired to identify cross-sell opportunities. Like its' competitors, the IFSP went through a wave of mergers between bank and insurance companies resulting in a financial services group. Therefore, special attention is paid to the transition from insurance-only or bank-only customers to financial-services customers. For this study, we had purchase history for 3,520,921 Belgian customers. We randomly selected a training sample and hold-out sample of each 50,000 customers with no overlap between the two groups.

3.1 Sequence-construction procedure

To keep the number of states as low as possible, the sequences contain purchase behavior at the *product group* level, hence not at the *product* level. Following the internal policy of the IFSP, nine product groups are distinguished based on the characteristics of the services. The first six are banking-product groups, whereas the last three are insurance-product groups (see Table 1).

Table 1: Product Groups of the IFSP.

	Product Group	Description
BANKING	1	Savings and investments with low interest rates and no time horizon (e.g., savings accounts)
	2	Savings and investments with fixed low to medium interest rates and time horizon (e.g., forward accounts and bonds)
	3	Long term high-risk investment products (e.g., strategic funds)
	4	Medium-risk investment products without time horizon (e.g., structured funds)
	5	Short and long-term credit (e.g., mortgages and loans)
	6	Checking accounts
INSURANCE	7	Fire insurance
	8	Car insurance
	9	Other types of insurance (e.g., hospitalisation, familial, accident and life insurance policies)

Each sequence consists of *all* subsequent purchase events, even when the customer buys consecutively in the same product group. Purchases within several product groups at one purchase event are reduced to one product group, either the product group triggering the purchases in the other product groups or the product group fitting best the objectives of the IFSP.

4 Results

4.1 Model fit comparison: Markov, MTD and MTDg

We estimated up to third-order Markov and MTD models. Fourth-order models were not run, because the number of observations (50,000) does not allow for the estimation of a fourth-order Markov model (cf. 52,488 parameters!), and, hence, no comparison with the same order MTD model would be possible (cf. only 75 parameters). The random variable X indicates in which product group a customer buys: N={1,2,3,4,5,6,7,8,9}. With m=9 the number of independent parameters to estimate increases rapidly.

We compare the fit of the estimated Markov and MTD models using the Bayesian Information Criterion (BIC) [13]. The BIC balances the desire for a better fitting model against the desire for a model with as few parameters as possible, as can be seen from eqn (9):

$$BIC = -2LL + p\log(n) \tag{9}$$

where *LL* is the log-likelihood of the model, *p* is the number of independent parameters and *n* is the number of components in the log-likelihood. We do not consider the parameters estimated to be zero, which is in line with the convention in counting the degrees of freedom for models on categorical data. To be able to compare the Markov and MTD models using BIC, we did not consider the first three observations of each purchase sequence, i.e., maxorder=3. The model with the lowest BIC is chosen, which is approximately the same as choosing the model with the highest posterior probability [11]. We prefer the BIC to the AIC (i.e., Akaike's Information Criterion) statistic, because the former is a more consistent estimator of the true order of the Markov chain.

Table 2 summarizes our results. The independence model is worse than any other model. Among the Markov models, the second-order model has the lowest BIC value (BIC= 441,290). This is even slightly lower than the BIC value for the second-order MTD model. Nevertheless, with 647 independent parameters, the Markov model is far less parsimonious than the MTD (73) model. Overall, the best result is achieved by the third-order MTD model (BIC= 433,010).

Table 2: Comparing up to third-order Markov and MTD models.

Model	**Order**	**LL**	**BIC**	**Number of parameters without structural zeros**
Independence		-278,030	556,150	8
Markov	1	-230,590	461,330	72
	2	-216,800	441,290	647
	3	-207,250	467,760	4485
MTD	2	-220,670	442,200	73
	3	-216,070	**433,010**	74

The third-order MTD model has lag parameters λ_1=0.4626, λ_2=0.2917 and λ_3=0.2457 and following R matrix:

$$R = \begin{bmatrix} \mathbf{0.3540} & 0.1963 & 0.0413 & 0.0338 & 0.1616 & 0.1110 & 0.0412 & 0.0451 & 0.0157 \\ 0.0469 & \mathbf{0.8365} & 0.0165 & 0.0290 & 0.0088 & 0.0094 & 0.0167 & 0.0250 & 0.0113 \\ 0.0924 & 0.1694 & \mathbf{0.4959} & 0.1758 & 0.0239 & 0.0158 & 0.0090 & 0.0138 & 0.0040 \\ 0.0766 & 0.2391 & 0.1259 & \mathbf{0.4410} & 0.0267 & 0.0077 & 0.0260 & 0.0486 & 0.0084 \\ 0.0914 & 0.0162 & 0.0091 & 0.0028 & \mathbf{0.5937} & 0.0835 & 0.0560 & 0.0646 & 0.0828 \\ 0.2775 & 0.0547 & 0.0144 & 0.0025 & \mathbf{0.4277} & 0.0417 & 0.0622 & 0.0524 & 0.0669 \\ 0.0389 & 0.0704 & 0.0106 & 0.0106 & 0.1059 & 0.0211 & 0.1426 & \mathbf{0.3861} & 0.2163 \\ 0.0201 & 0.0487 & 0.0100 & 0.100 & 0.0435 & 0.0146 & 0.2104 & \mathbf{0.5195} & 0.1285 \\ 0.0238 & 0.0746 & 0.0041 & 0.0041 & 0.1327 & 0.0287 & 0.2444 & \mathbf{0.3121} & 0.1764 \end{bmatrix}$$

Notice that the third-order Markov transition matrix would be 729*9. The lag parameters indicate a positive diminishing influence of the purchase history on the purchase at moment t. The previous purchase (i.e., the first lag) determines largely the product group in which a customer will buy next. A first look at the transition matrix teaches us that if a customer bought in a certain product group at time *t-1* (or *t-2* or *t-3*), this increases his probability of buying in that product group at moment *t* (cf. positive effect of the lags), especially for product group 2.

4.2 Discussion of the transition probabilities for the *bank* product groups

We observe a relationship between product groups 1, 2, 5 and 6. If a customer bought at moment *t-3*, *t-2* or *t-1* in **product group 1** (i.e., a savings and investment product at low interest rate and without time horizon), he has a probability of 64.6% to buy in another product group, with as biggest cross-sell opportunities product groups 2, 5 and 6, mentioned along descending transition probabilities.

With a repurchase probability of more than 80%, **product group 2** (i.e., savings and investment products with fixed low to medium interest rates and time horizon) is not an initiator of cross-product group purchases. The fixed time horizon of the products might explain the latter. The relationship between product group 2 and 4 is more intense than with product group 3.

Product group 3 (i.e., long-term high-risk investments) is related to product groups 4 (17.58%) and 2 (16.94%). Some of the customers who purchased a long-term high-risk investment in the past still buy a simple savings account later (i.e., product group 1, 9.24%). By the way, it is easier to cross-sell product group 1 when the customer bought a high-risk investment over his last three purchases than reverse.

Product group 4 (i.e., medium-risk investments without time horizon) is more related to product group 2 than product group 3 is. The latter is revealed by the higher transition probability from 4 to 2 (23.91%) than from 3 to 2 (16.94%). Product group 4 is also slightly related to product group 3 (12.59%). However, it is more likely to buy in product group 4 first, and to invest later in a long-term high-risk investment (i.e., product group 3) than the other way around. Product group 4 is also related to product group 1. Again, rather logical, it is more

Transactions on Information and Communications Technologies vol 29, © 2003 WIT Press, www.witpress.com, ISSN 1743-3517

straightforward to cross-sell a simple savings product to someone who already bought a medium-risk investment without time horizon than vice versa.

Product group 5 (i.e., short- and long-term credits) is related to bank product groups 1 and 6 and to all insurance product groups, especially product group 8. It is more likely to first acquire a checking account before subscribing to a loan.

Great cross-sell opportunities are present when the customer made a purchase in **product group 6** (i.e., checking accounts) over the last three purchases, as the repeat purchase probability is only 4.17%! Therefore, although checking accounts are not profitable in themselves, they might be valuable in inducing purchases in other, more profitable product groups. Checking accounts are rather basic products encouraging the purchase of credits (42.77%) and simple saving accounts (27.75%). To a lesser extent, they also lead to purchases of insurance policies. Finally, notice that the number of customers who opened a checking account and invest their money not needed on short term into product group 1 (27.75%) is much larger than those who put it in product group 2 (5.47%).

4.3 How to convert customers into real financial-services customers?

What cross-sell actions could convert solely bank customers into financial-services customers? Product groups 6, 5, 4 and 1 offer the best chances. If a customer recently opened a checking account, he might subscribe to another type of insurance (6.69%), a fire insurance (6.22%) or a car insurance (5.24%). Another possibility is cross-selling a car insurance policy to someone who recently took a loan (6.46%) or purchased within product group 1 (4.51%) or 4 (4.86%). The arrows in *Figure 1* illustrate paths to convert bank-only customers into financial-services customers. Only transitions of minimum 4% are considered. The percentages refer to the biggest insurance cross-sell opportunities from a given bank product group.

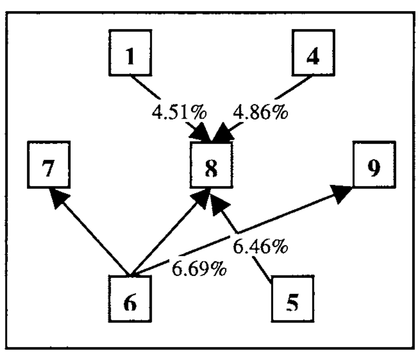

Figure 1: Paths to convert bank customers into financial-services customers.

How to persuade an insurance customer to also buy his bank products at the IFSP? In general, although it is more straight forward to cross-sell the remaining insurance product groups, the salesperson could help the IFSP transforming insurance-only customers into financial-services customers by promoting product groups 5 or 2. It is easier to cross-sell bank product groups 5 and 2 from product groups 7 or 9 than it is from product group 8. Finally, it might be preferred to promote product group 5 rather than product group 2, as the latter triggers almost no cross-product group purchases. See *Figure 2*.

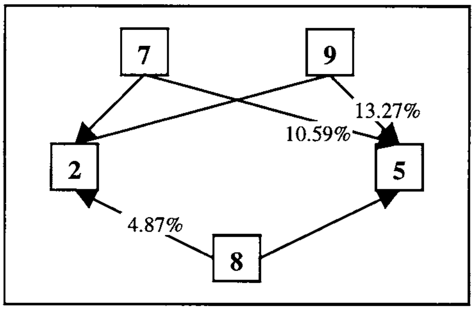

Figure 2: Paths to convert insurance customers into financial-services customers.

In conclusion, note that it is easier to transform an insurance customer into a financial-services customer than a bank customer.

5 Summary, limitations and directions for further research

With this paper we have tried to contribute to the existing research in three ways. *Methodologically*, we compared different techniques (Markov and MTD) for the description of sequences. We provide in a parsimonious model to estimate high-order Markov chains, and still being capable of interpreting results. *Methodologically for marketing*, we demonstrated that the parsimony of the MTD model is also welcome in casu of a random variable with a high number of possible values and not necessarily a high-order dependence. *Empirically for marketing*, we described the acquisition sequences of financial services and identified cross-sell opportunities. Unlike the studies of Kamakura et al. [7, 9], special attention is paid to possible paths to transform bank- or insurance–only customers into financial-services customers.

This study is not free of limitations. Firstly, we did not satisfy the discrete-time assumption of the discrete Markov and MTD model. The acquisition of financial services was not measured at constant discrete moments in time, because the latter would not give a total view on the sequence of purchases. Secondly, as Kamakura et al. [9] indicated, the use of single source data, in casu customer transaction information on the ownership of products at the IFSP, is not optimal for identifying cross-sell opportunities as it overlooks the possibility that the customer already owns the product at a competitor. Thirdly, using product groups rather than single products results in rather general cross-sell recommendations. However, in order to keep the number of states limited, some reduced picture of the total product assortment must be accepted.

Several paths are open for further research. Firstly, we overlooked the influence of customer heterogeneity on the cross-sell possibilities. Secondly, replicating the study at other companies, possibly in other countries/cultures, could enhance the external validity of the results.

References

[1] Feick, L., Latent Class Models for the Analysis of Behavioral Hierarchies. *Journal of Marketing Research*, **24** **(2)**, pp. 174-86, 1987.

[2] Hauser, J. R. & Urban, G.L., The value priority hypotheses for consumer budget plans. *Journal of Consumer Research*, **12** **(4)**, pp. 446-62, 1986.

[3] Pyatt, F. G., *Priority Patterns And The Demand For Household Durable Goods*, Cambridge University Press: Cambridge, 1964.

[4] Paroush, J., The Order of Acquisition of Consumer Durables. *Econometrica*, **33** **(1)**, p. 225-35, 1965.

[5] McFall, J., Priority Patterns and Consumer Behavior. *Journal of Marketing*, **33**, pp. 50-55, 1969.

[6] Stafford, E.F., Kasulis, J.J. & Lush, R.F., Consumer-behavior in accumulating household financial assets. *Journal of Business Research*, **10(4)**, pp. 397-417, 1982.

[7] Kamakura, W. A. & Ramaswami, S.N., Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing*, **8** **(4)**, pp. 329-49, 1991.

[8] Van den Poel, D. & Larivière, B., Customer attrition analysis for financial services using proportional hazard models. Forthcoming in *European Journal of Operational Research*, 2003.

[9] Kamakura, W. A., Wedel, M., de Rossa, F. & Mazzon, J.A., Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing*, **20** **(1)**, pp. 45-65, 2003.

[10] Raftery, A. E., A Model for High-order Markov Chains. *Journal of the Royal Statistical Society*, *B-series*, **47** **(3)**, pp. 528-39, 1985.

[11] Berchtold, A. & Raftery, A. E., The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. *Statistical Science*, **17** **(3)**, pp. 328-56, 2002.

[12] Pegram, G.G.S., An autoregressive model for multilag Markov chains. *Journal of Applied Probability*, **17** **(2)**, pp. 350-62, 1980.

[13] Schwarz, G., Estimating the Dimension of a Model. *Annals of Statistics*, **6** **(2)**, pp. 461-64, 1978.