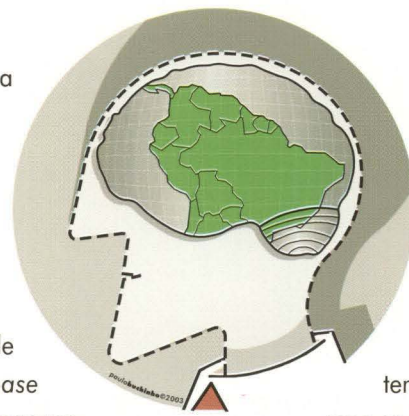
Micro-segmentação de clientes com base em dados de consumo: Modelo RM-Similis

por Luís Cavique

Em marketing relacional, pretende associar-se ao cliente «certo» o produto «certo» com a mensagem apropriada. Neste estudo, é apresentado um modelo que responde a questões, tais como: encontrar os clientes mais lucrativos, determinar os clientes que irão responder a uma campanha, ou prever qual o próximo produto de um cliente específico. Apresenta-se um novo critério de segmentação de clientes, designado "denominador RM", inspirado na combinação de critérios referidos na literatura, que, para além da sua simplicidade, apresenta a vantagem de permitir criar um maior número de estratégias de comunicação. Para descobrir a melhor combinação de produtos de um determinado cliente, é utilizada a análise do cabaz de compras. Para o efeito, descreve-se um algoritmo eficiente desenvolvido pelo autor, o algoritmo Similis. O modelo RM-Similis, que pretende encontrar o cliente «certo» e o produto «certo», resulta da associação do referido algoritmo com o novo critério de segmentação.

Palavras-chave: *Marketing Relacional,* Database Marketing, RFM, Cross-Selling, *Cabaz de Compras*



Tal como o e-mail tem revitalizado a arte na escrita de cartas, as compras *online* devolveram algumas das vantagens que os clientes tinham nas antigas mercearias locais, onde eram conhecidas as suas preferências e os produtos eram entregues em casa. A evolução dos sistemas e tecnologias de informação e comunicação, *Database Marketing* e CRM *(customer relationship management)*, têm permitido a criação de uma nova forma de marketing. O marketing relacional ou individualizado vai mais além do que o marketing segmentado, permitindo atingir o cliente individualmente através de micro-campanhas planeadas e de carácter permanente. O marketing relacional tem como objectivo aumentar o consumo de cada cliente existente *(share of wallet)* e de conquistar novos clientes, recorrendo normalmente à recomendação de clientes defensores do produto/serviço.

Neste artigo, para além de desenvolver o tema da micro-segmentação de clientes, iremos enquadrar o tema a montante e a jusante. A montante, os dados de uma aplicação cliente e, a jusante, as estratégias de comunicação. É apresentado um modelo que responde a questões essenciais, tais como: quais são os clientes mais lucrativos, que clientes irão responder a uma campanha, ou ainda prever qual o próximo produto de um determinado cliente.

Na segunda parte abordamos a estrutura essencial de uma *database marketing*. A terceira e quarta partes são dedicadas aos algoritmos: na terceira parte é apresentado o novo critério de segmentação RM para segmentar clientes, e na quarta parte é referido o inovador algoritmo Similis uti-

Luís Cavique
Doutor em Engenharia de Sistemas (Instituto Superior Técnico). Professor Adjunto da Escola Superior de Comunicação Social do Instituto Politécnico de Lisboa. Investigador do Instituto Superior Técnico da Universidade Técnica de Lisboa, Lisboa, Portugal.
E-mails: lcavique@escs.ipl.pt e pcp1184@alfa.ist.utl.pt

Recebido em Abril de 2003 e aceite em Julho de 2003.

lizado na determinação de cabazes de compras (Cavique, 2002). À associação dos dois dá-se o nome de modelo RM-Similis. Na quinta parte, são exemplificadas algumas estratégias de comunicação com o cliente. Finalmente, na última parte são apresentadas as conclusões.

Bases de dados de clientes

Na caracterização de clientes, são geralmente utilizados dados relacionados com o indivíduo, que incluem dados sócio-demográficos, geográficos ou de estilo de vida. Para obter este tipo de dados são requeridos dispendiosos inquéritos. Por outro lado, os dados de consumo existem na contabilidade das empresas, na maior parte das vezes basta associar os produtos comprados a um determinado cliente. Neste trabalho, o modelo apresentado utiliza exclusivamente dados de consumo, disponíveis em qualquer sistema de informação empresarial, evitando as demoras e os custos dos dados resultantes de inquéritos.

Com o apoio das Bases de Dados ou Aplicações para Clientes, pretende construir-se relações lucrativas e duradouras, ao comunicar com o cliente certo, utilizando o produto certo com a mensagem certa - emitida no momento certo e através do canal certo (Cavique e Themido, 2002). Deste objectivo, distinguem-se claramente três tipos de conjuntos de dados: os dados do cliente, os dados da compra do cliente e os dados relativos à comunicação com o cliente (Brito, 2000).

Para responder a esta solicitação, existe uma estrutura de dados, como se apresenta em seguida:

- Dados do Cliente: nome, contactos/canal, contacto/canal preferencial, datas mágicas, dados demográficos, sócio-gráficos e psico-gráficos;
- Compras do Cliente: data, produto, quantidade, preço, forma de pagamento, vendedor, descontos, devoluções, reclamações, ofertas, pontos;
- Comunicações com o Cliente: data-hora, origem, destino, assunto, conteúdo da mensagem, canal.

As Bases de Dados de Clientes, para além de registarem os dados do cliente e das suas compras como qualquer sistema de contabilidade, integram o registo das comunicações que darão suporte a programas de fidelização.

Para o sucesso das aplicações de cliente, a qualidade dos dados e a agregação dos mesmos são dois factores críticos. Ao utilizar dados de diversas fontes, a garantia da qualidade dos dados é uma tarefa de extrema importância no que diz respeito à uniformização dos formatos dos dados e à remoção de dados duplicados nos atributos críticos da aplicação. Este tipo de tarefas foram durante muito tempo negligenciadas, estando actualmente a tomar a sua devida importância nas aplicações de clientes. Por outro lado, os dados devem estar devidamente consolidados e estáveis por forma a facilitar as pesquisas. Este tipo de conceito desenvolvido para ambientes de *Data Warehouse* é também necessário nas Bases de Dados de Clientes, permitindo a agregação dos dados e a capacidade de absorver periodicamente novos dados.

Cliente certo: Critério RM

Por forma a aumentar as baixas taxas de resposta, em *Database Marketing* é usual recorrer à técnica de segmentação RFM, onde R (qualidade de ser recente) é dada pela última data da visita à loja, F representa a frequência de compras na loja, e M o valor monetário global do cliente.

Existem várias formas de calcular os valores para RFM, e neste estudo vamos utilizar a especificação de Hughes (2000). Se escolhermos para cada atributo RFM cinco classes, com os números de 1 a 5, obtemos 125 classificações diferentes. Assim, o cliente 555 é um cliente muito recente, muito frequente e com um alto volume de compras, enquanto que um cliente 111 é pouco recente, pouco frequente e com baixo volume de compras. Existe uma possibilidade de hierarquizar as classificações, seguindo a valorização que é comum dar aos números inteiros. O critério de segmentação começa por ordenar a tabela de clientes de forma crescente pela data da última compra, num segundo passo classifica os primeiros n/5 clientes com o número 1, os segundos n/5 clientes com o número 2 e assim sucessivamente até ao número 5. O processo repete-se para os atributos da frequência e valor monetário. Da concatenação dos três atributos resulta a classificação RFM de cada cliente. Os clientes a seleccionar para cada campanha - ou micro-campanha - são os enésimos primeiros clientes com maior valor de RFM, permitindo uma única estratégia.

Como forma de ultrapassar as referidas desvantagens, apresentamos um novo critério de segmentação de clientes inspirado nas vantagens dos critérios RFM e CLV. A medida mais importante para o critério RFM é a relativa a ser ou não recente, e, por outro lado, a medida do CLV é o valor monetário. O critério proposto RM cruza as variáveis relativas a ser ou não recente e o valor monetário, permitindo definir quatro estratégias diferentes de comunicação com o cliente.

Com a expansão do CRM, é introduzida uma nova medida para os clientes - o *customer lifetime value (CLV)*. Esta medida é calculada com base no lucro resultante do total de transações dos clientes durante o seu período de vida. Desta forma, o *CLV* devolve valor financeiro de um cliente médio, tornando-se difícil prever a atitude de cada cliente individualmente (Verhoef e Langerak, 2002; artigo publicado na *Revista Portuguesa e Brasileira de Gestão* nº 3/2002).

Como forma de ultrapassar as referidas desvantagens, apresentamos um novo critério de segmentação de clientes inspirado nas vantagens dos critérios RFM e *CLV*. A medida mais importante para o critério RFM é a relativa a ser ou não recente, e, por outro lado, a medida do *CLV* é o valor monetário. O critério proposto RM cruza as variáveis relativas a ser ou não recente e o valor monetário, permitindo definir quatro estratégias diferentes de comunicação com o cliente.

Para obter a classificação RM, basta ordenar pela data mais recente e afectar R como foi definido anteriormente. O processo repete-se para classificar o valor monetário M. O resultado é uma matriz RM, cujas células têm valores diferentes de clientes.

Depois de classificar os clientes com base na relação com ser ou não recente e na frequência, podemos realizar a análise segundo as duas variáveis em simultâneo, dividindo os clientes em quatro quadrantes por forma a adoptar estratégias diferentes para cada grupo, conforme está representado na Figura 1.

Os clientes R^+M^+ são mais recentes, mais lucrativos, compram as opções mais caras e têm menos custos no processo de venda. Este tipo de clientes, para além de serem os mais lucrativos, têm uma importância acrescida, recomendam a empresa a potenciais clientes, fazendo referência aos seus produtos/serviços que originam novas aquisições. A estratégia a utilizar, neste caso, é de premiar a fidelização, que não se esgota numa campanha, mas pelo contrário, tem um carácter permanente. Por outro lado, aos clientes R^-M^- não deve recair qualquer esforço, dado que são clientes muito pouco rentáveis ou mesmo prejudiciais à empresa.

Figura 1
Segmentação RM

	+ Recentes	– Recentes
+ Valor Monetário	Premiar	Reter
– Valor Monetário	Estimular	Esquecer

Para os clientes R^+M^- recentes e pouco frequentes que correspondem a novas aquisições da empresa, pretende-se estimular a compra por forma a migrá-los para R^+M^+. O mesmo objectivo de migração existe para os clientes R^-M^+ muito frequentes mas que não visitam a loja há muito tempo. Neste caso, devem ser implementadas estratégias de retenção.

No recente artigo de Verhoef e Langerak (2002), estes referem que as empresas não deveriam fazer incidir os esforços de marketing exclusivamente nos seus clientes mais lucrativos. O critério RM enquadra todos os clientes, indo excluir somente os clientes pouco recentes e pouco lucrativos.

Produto certo: análise do Cabaz de Compras

As estratégias de *cross-selling* são suportadas por algoritmos de análise do Cabaz de Compras. Quando se fala em Cabaz de Compras, compreende-se um conjunto de N produtos adquiridos por um cliente numa visita a um loja. Dado que cada cliente compra um conjunto diferente de produtos, em quantidades diferentes e em diferentes instantes de tempo, torna a determinação de Cabaz de Compras um problema complexo. Com base nos atributos escolhidos <cliente, produto>, iremos definir o problema do Cabaz de Compras como a procura dos grupos mais frequentes de N produtos que são comprados em conjunto.

Definido o Cabaz de Compras tipo com N produtos que

represente um conjunto grande de clientes, podemos partir para as estratégias de *cross-selling* e *up-selling*. Com a primeira estratégia, depois de determinado o Cabaz, vamos procurar todos os clientes que tenham adquirido N-1 produtos desse Cabaz e sugerir a compra do produto que falta, o próximo-produto.

Para a determinação do Cabaz de Compras, numa das primeiras aproximações destaca-se o programa *Collaborative Filtering* disponibilizado pela NetPerception, que determina para cada cliente a sua 'alma gémea'. Entende-se por 'alma gémea' um outro cliente que tenha os mesmos gostos, isto é, que tenha um Cabaz de Compras idêntico. Este programa está instalado em centenas de empresas, tendo, contudo, a desvantagem de estarmos a comparar simplesmente dois indivíduos, não tendo a visão do conjunto. Por exemplo, se o indivíduo X comprou dez livros sobre Internet e um livro de Culinária, e o indivíduo Y comprou oito livros sobre Internet, tal que Compras(Y) $\subseteq$ Compras(X), o programa irá sugerir a compra do livro de Culinária ao indivíduo Y.

Uma forma de representar o Cabaz de Compras é através de regras do tipo: "se <condição> então <acção>". Esta abordagem da análise do Cabaz de Compras foi inicialmente utilizada na descoberta de novos padrões, tendo sido popularizada pela descoberta do padrão que referia que "às 5ª Feiras nos supermercados, fraldas e cervejas são compradas em conjunto". Com base neste padrão, podemos gerar as seguintes regras: {fraldas}=>{cerveja} e {cerveja}=>{fraldas} (Berry e Linoff, 1997). Num dos pacotes de *Data Mining* mais conhecidos, o "Enterprise Miner" do SAS Institute, a análise do Cabaz de Compras está implementada utilizando este tipo de gerador de regras, que recorre ao algoritmo Apriori (Agrawal *et al.*, 1996).

Dada a elevada complexidade do algoritmo anterior, foi recentemente desenvolvido um algoritmo de mais baixa complexidade temporal. O algoritmo Similis (Cavique, 2002) que permite resolver um maior número de problemas reais. O algoritmo Similis estende a aplicação da análise do Cabaz de Compras até às centenas de produtos, permitindo a sua aplicação em ambientes como hipermercados ou livrarias virtuais.

Em seguida, apresentamos o procedimento para o algoritmo Similis, que se divide em dois passos:

Para validação do algoritmo Similis, foram utilizados dois problemas reais. O primeiro relativo às Pousadas de Portugal com 43 produtos, e o segundo referente aos ultra-congelados da Nestlé com 158 produtos. O algoritmo Similis apresenta um bom desempenho tanto a nível da qualidade de resultados, com um índice de qualidade médio sempre superior a 95%, com tempos computacionais reduzidos.

- Passo 1 - Transformação do problema: gerar um grafo ponderado G(V,A), onde cada vértice corresponde a um produto, e onde o peso de cada aresta representa a semelhança do par de produtos;
- Passo 2 - Pesquisa da solução: encontrar no grafo G(V,A) o sub-grafo completo (ou clique) de maior peso de dimensão k, que corresponde ao Cabaz de Compras mais frequente.

Para validação do algoritmo Similis, foram utilizados dois problemas reais. O primeiro relativo às Pousadas de Portugal com 43 produtos, e o segundo referente aos ultra-congelados da Nestlé com 158 produtos. O algoritmo Similis apresenta um bom desempenho tanto a nível da qualidade de resultados, com um índice de qualidade médio sempre superior a 95%, com tempos computacionais reduzidos. Foram ainda testadas várias medidas de semelhança por forma a validar o atributo mais sensível do algoritmo Similis, que é a noção de peso do clique, não se tendo encontrado diferenças significativas.

Comunicação individualizada

Depois de garantir a qualidade dos dados do cliente, e de escolher o melhor conjunto de parâmetros para o modelo RM-Similis, segue-se a estratégia de comunicação individualizada.

O conceito de fidelização de clientes tem sido amplamente divulgado, mas menos claro tem sido a forma de o operacionalizar. Em Bogmann (2000), são referidas diferentes abordagens nos programas de fidelização. Pretendemos neste trabalho distinguir três níveis de programas de fidelização: um nível mais elementar de satisfação do cliente; um segundo nível que recompensa o cliente com benefícios financeiros; e um nível mais alto de envolvimento do cliente.

O nível de satisfação tende essencialmente a minimizar os *gaps* entre a visão do fornecedor e as necessidades do cliente, desenvolvido no modelo de Zeithaml *et al.* (1992). Uma revisão do estado da arte encontra-se em Nascimento (2000).

O segundo nível de fidelização passa por recompensar a sua lealdade através da troca de pontos, cupões ou outros benefícios. Os programas de fidelização das companhias de aviação provaram ser um excelente método para adquirir e reter clientes fiéis. O desconto não deve ser usado, dado que pode ser copiado pela concorrência e provoca no cliente uma busca constante por um preço menor, indo no sentido inverso ao da fidelização. Note-se, contudo, que o número de consumidores com mais de um cartão-cliente tem vindo a aumentar. Para além dos benefícios na troca de pontos, é de referir outro tipo de benefícios, tais como os prémios de referência quando da aquisição de novos clientes. A maior parte dos programas de fidelização terminam neste nível: a empresa oferece um cartão-cliente e retribui o consumo com benefícios financeiros. Exemplos de programas de fidelização encontram-se em Brito (2000).

O terceiro nível de fidelização diz essencialmente respeito à comunicação personalizada. A personalização ou "customização" é alcançada através de uma estratégia de comunicação utilizando para o efeito o combinatório dos próximos-produtos e *newsletters*. Periodicamente, para cada cliente pode ser definido um nível máximo (saturação) de comunicações e programados "pontos de contacto" em datas-mágicas, com mensagens informativas e propostas de *cross-selling*. A comunicação deve ser gerida de forma diferenciada para cada grupo de clientes.

Utilizando o critério RM, num caso a um cliente $R^{+}M^{--}$ pode ser enviado um cartão de agradecimento com o cartão-cliente. Para um cliente $R^{+}M^{+}$ pode ser enviado um convite exclusivo à loja para apresentação de uma nova colecção. E, finalmente, para um cliente $R^{--}M^{+}$, um questionário de retenção ou pedido de opinião acerca de um produto recentemente adquirido.

Nesta secção, diferencia-se "satisfação" de "lealdade", indo ao encontro do pressuposto de Verhoef e Langerak (2002), que clientes satisfeitos nem sempre são leais. A satisfação é o nível mais elementar da fidelização, o envolvimento do cliente é atingido através de uma estratégia de comunicação adequada.

Conclusões

Este artigo divide-se em três grupos, que correspondem às formas de ver as Bases de Dados de Clientes: os dados, os modelos e a comunicação. Com a abordagem seguida, concretiza-se, de uma forma integrada, os objectivos definidos para uma Base de Dados de Clientes, onde se pretendem construir relações lucrativas e duradouras, ao comunicar com o cliente certo, utilizando o produto certo, com a mensagem certa, emitida no momento certo e através do canal certo. O cliente certo, associado à mensagem adequada, é dado pela segmentação RM e o produto certo - ou próximo-produto - pelo algoritmo Similis.

O procedimento geral a aplicar nas Aplicações de Clientes pode seguir a seguinte sequência: a) armazenar dados dos clientes; b) segmentar os clientes; e c) comunicar com os clientes. A re-alimentação do sistema é feita através das aquisições dos clientes, permitindo a actualização dos seus dados de compra. Todo o processo é mantido e gerido com base nos dados de consumo.

Neste artigo apresenta-se um novo critério de segmentação de clientes - o critério RM - e faz-se referência a um algoritmo recente para implementar estratégias de *cross-selling* - o algoritmo Similis. Apesar da sua simplicidade, o critério RM é de grande utilidade ao diferenciar quatro importantes estratégias de comunicação, mostrando-se, neste ponto de vista, superior ao critério RFM. O algoritmo Similis apresenta-se como um algoritmo muito eficiente para a determinação de cabazes de compras. Da associação do critério apresentado com o referido algoritmo, é desenvolvido o modelo RM-Similis.

Ainda no artigo de Verhoef e Langerak (2002), refere-se que a pirâmide de clientes, baseada na rentabilidade, nem sempre é um bom esquema de segmentação, esquecendo as necessidades específicas dos clientes. O modelo RM-Similis vai exactamente ao encontro deste pressuposto, segmentando os clientes pela sua relação com ser ou não recente e pela rentabilidade, e, em seguida, respondendo às necessidades de cada cliente propondo um próximo-produto.

Como conclusão, diremos que o modelo RM-Similis permite definir um conjunto de estratégias de comunicação para o cliente certo - encontrado segundo o critério RM -, como produto certo - encontrado com o algoritmo Similis -, com a mensagem certa - enviada numa data-mágica através do contacto/canal preferencial -, por forma a manter uma relação lucrativa e duradoura. ■

Referências bibliográficas

AGRAWAL, R.; MANNILA, H.; SRIKANT, R.; TOIVONEN, H. e VERKAMO, A. (1996), «Fast discovery of association rules» *in Advances in Knowledge and Data Mining*, U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth e R. Uthurusamy (eds.), MIT Press.

BERRY, M. e LINOFF, G. (1997), **Data Mining Techniques, for Marketing, Sales and Customer Support**, John Wiley and Sons.

BOGMANN, I. M. (2000), **Marketing de Relacionamento**, Nobel, São Paulo.

BRITO, C. M. (2000), «O Marketing Relacional» *in* **Os Horizontes do Marketing**, ed. Brito, C. M. e Lencastre P., Editorial Verbo, Lisboa, pp. 61-84.

CAVIQUE, L. (2002), ***Meta-heurísticas na Resolução do Problema da Clique Máxima e Aplicação na Determinação do Cabaz de Compras***, Dissertação de Doutoramento em Engenharia de Sistemas no Instituto Superior Técnico da Universidade Técnica de Lisboa.

CAVIQUE, L. e THEMIDO, I. (2002), «Estratégias de Comunicação em CRM», *in* **E-Portugal**, L. Valadares Tavares e M. J. Pereira (eds.), pp. 13-20.

HUGHES, A. M. (2000), **Database Marketing**, McGraw-Hill Companies.

NASCIMENTO, J. R. (2000), **Satisfação do Consumidor: o Caso da Televisão por Cabo em Portugal**, Principia, Cascais.

VERHOEF, P. C. e LANGERAK, F. (2002), «Onze preconceitos sobre a gestão das relações com o cliente», *Revista Portuguesa e Brasileira de Gestão* Nº 3/2002 (Out./Dez.), pp. 16-23.

ZEITHAML, V.; PARASURAMAN, A. e BERRY, L. (1992), «Problems and Strategies in Services Marketing» *in Managing Services Marketing*, J. Bateson, ed., The Dryden Press, Texas, pp. 48-67.

ESTAMOS HABILITADOS A RESPONDER

À SUA NECESSIDADE DE COMUNICAR

Somos um grupo de empresas, sediado em Lisboa, experiente e experimentado na produção de todo o tipo de informação destinada a ser veículada através do papel impresso.

Estamos preparados para concepcionar o seu suporte informativo desde o conteúdo até à sua imagem final.

Se pretende criar uma imagem editorial para a sua instituição ou alterar a sua actual, e veiculá-la para o exterior, consulte-nos.

Revistas institucionais, técnicas e científicas, Livros, Logotipos, Deplians, Brochuras, Catálogos, Cartazes, Folhetos, Newsletters institucionais e de empresa, e todos os mais variados suportes informativos são a nossa especialidade.

c3im
CONSULTORIA DE INFORMAÇÃO INOVAÇÃO, INTERNET E MULTIMÉDIA, IDA
oficina gutenberg
Concepção e coordenação de projectos, design, redacção e fotografia
Telefone e Fax: (351) 21 497 15 22
c3im@janelanaweb.com
www.c3im.pt

Gabinete Gráfico Imaginário, Lda.
Paginação electrónica e tratamento de imagem
Telefone e Fax: (351) 21 497 15 22
gabinetegráfico@netcabo.pt

ADVENTVS
A METANATIONAL MEDIA GROUP
Barcelona - Buenos Aires - Lisboa - Macao