



## ANÁLISE RFV DO CLIENTE POR ALGORITMOS GENÉTICOS NA OTIMIZAÇÃO DE ESTRATÉGIAS DE MARKETING

RFV CUSTOMER ANALYSIS BY GENETIC ALGORITHMS IN OPTIMIZATION  
OF MARKETING STRATEGIES

**Anderson Guimarães de Pinho**

### RESUMO

O presente artigo objetivará a evolução de regras de decisão por Algoritmos Genéticos as quais classifiquem corretamente futuros clientes evasivos para a empresa. Em estratégias de marketing, é de grande dúvida para a empresa quais clientes abordar numa campanha, ou quais clientes apresentam maiores chances de evasão. Para responderem a isto, muitos pesquisadores têm recorrido a informações de recência, frequência e valor do cliente, na mineração de conhecimento valioso o qual possa ser utilizado. Neste artigo, utilizaremos Algoritmos Genéticos como um diferencial competitivo na explicitação deste conhecimento. Diferencial competitivo, pois a classificação de um comportamento futuro do cliente ocorre na forma de regras do tipo "If Then Else". Isto nos permitirá uma simples integração dos resultados obtidos com processos empresariais, de fácil entendimento para o usuário. Trata-se de uma nova forma de se alinhar estratégias de marketing com inteligência de mercado na área de conhecimento chamada database marketing.

**Palavras-chave:** Análise RFV. Algoritmos genéticos. Data mining. Previsão a churn. Computação evolucionária.

### ABSTRACT:

This article aim to discover rules of decision by Genetic Algorithms which correctly classify evasive future customers for a company. In marketing strategies, it is a big question for the enterprises which customers should they focus, or which customers are more likely to churn. To respond this, many researchers have used the information of recency, frequency and value of customers when decided to mining valuable knowledge which can be used. In this article we use Genetic Algorithms as a competitive differential in the explanation of knowledge. Competitive, because the classification of a customer's future behavior occurs in the form of rules such as "If Then Else." This will allow us to a simple integration of the results of business processes, easy to understand for the user. This is a new way to align marketing strategies with market intelligence in the area of knowledge called database marketing.

**Keywords:** RFV analysis. Genetic algorithms. Data mining. Forecasting the churn. Evolutionary computation.

## INTRODUÇÃO

RFV (ou RFM em algumas literaturas) entende-se como recência, frequência e valor monetário do cliente. *Recência* como uma medida de quanto tempo se passou desde a última transação com a empresa. *Frequência* como uma medida de quão freqüente um cliente efetua transações. E *Valor Monetário* como o gasto médio feito por transação.

Estratégias baseadas em RFV buscam métricas ou regras para avaliar *o comportamento e valor do cliente* para a empresa. Perguntas como “quais clientes devem ser impactados por uma ação de marketing” ou “quais clientes são mais valiosos para a empresa em termos de contribuição financeira passada e futura” são encontradas freqüentemente por pesquisadores na gestão do relacionamento com o cliente (Customer Relationship Management ou CRM).

Nestes casos, a análise de RFV pode conter informação valiosa para a empresa na resposta a estes questionamentos. Toda esta informação necessária para análise, encontra-se em histórico transacional de vendas a clientes disponível nos bancos de dados de grandes empresas.

São dos mais diversos, os estudos envolvendo RFV. Num primeiro exemplo, Peter et AL (2005) apresentou um modelo estocástico estimar o Valor Financeiro do Tempo de Vida do Cliente (Customer Life Time Value ou CLTV), utilizando como variáveis explicativas RFV em empresas com vínculo não contratual. Em um segundo estudo, Colombo et al (1999) introduz um simples modelo estocástico baseado em RFV para responder *a quais clientes uma firma deve focar para fazer uma oferta de produto*. Ambos estudos levam em comum o mesmo princípio motivacional: medidas comportamentais de clientes são indicadores chaves para predizer comportamento futuro.

Sobre o problema de Colombo, sabemos que as empresas podem maximizar o retorno de campanhas e minimizar custos de marketing se souberem quais clientes endereçar uma ação de venda. Estes clientes podem ser assim considerados de maior valor para a empresa, pois seu comportamento passado indica uma intenção positiva de manutenção do relacionamento.

Por outro lado, clientes menos valiosos seriam aqueles que não apresentam uma intenção de recompra futura. Conseqüentemente, apresentariam baixas chances de resposta a uma ação de venda marketing, seja ela de *cross-selling* ou *up-selling* (Berry, 115).

No entanto, é importante dizer que não há garantias de que após um longo período de inatividade, um cliente dado no passado como baixa chance de recompra, virá a efetuar uma transação. Em casos afirmativos, dizemos que o evento “transação com a empresa” representa um processo sem memória, de difícil modelagem, onde a ocorrência depende somente de um instante de tempo imediatamente anterior ao ocorrido (que neste caso, encontra-se também no futuro).

Por estes motivos, empresas não se preocupam em investigar um comportamento de compra futuro tão distante, uma vez que a dinâmica de mercados mais longínquos pode não ter dependência ou correlação com o presente. Em outras palavras, as chances de um cliente se tornar de alto valor num futuro distante, dificilmente encontrariam explicações no comportamento presente.

Numa visão inversa ao problema de Colombo, poderíamos trabalhar ações de marketing específicas para clientes com menos chances de respostas. Tais ações teriam como objetivo a

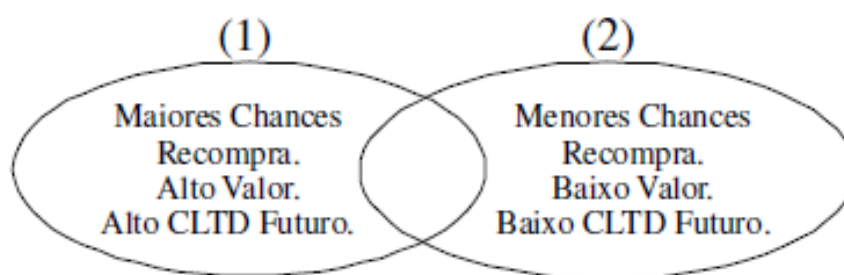
mudança comportamental em termos de recência, frequência e valor, afim de transformá-los em maior valor para a empresa.

Como por exemplo, suponha uma empresa administradora de investimentos na Bolsa de Valores de São Paulo (BOVESPA). Para clientes com baixa intenção de manutenção do relacionamento, poderiam ser oferecidos cursos e palestras sobre investimentos em ações. Tal ação teria como objetivo secundário, oferecer aos clientes ferramental intelectual, o suficiente para que estes possam continuar operando no mercado de ações pela empresa administradora. Conseqüentemente, o aumento de lucros pelo aumento do tempo de relacionamento.

Embora a discriminação de clientes mais e menos valiosos atenda a múltiplos objetivos, este último apresentado torna-se mais atraente, pois vai ao encontro com a retenção de clientes ativos na base de dados como conseqüência do aumento da duração do tempo de vida do cliente (Customer Life Time Duration ou CLTD). Sendo assim, esta será a principal motivação de nossos estudos nos próximos capítulos.

A figura a seguir mostra a distinção destes dois grupos acima discutidos:

**Figura 1:** Tipos de Clientes em Análise de RFV



**Fonte:** Própria Autoria

Para problemas mais simples como a *classificação* de dois grupos de clientes, muitas outras técnicas de menos complexidade (em comparação à modelagem estocástica) têm sido aplicadas. Kuman (2005, p.129-132) destaca o uso da Regressão Logística e Árvore de Decisão como solução a problemas envolvendo RFV. Tais técnicas apresentam suas características, as quais dividem pesquisadores e acadêmicos na sua aplicação.

Neste artigo, abordaremos a técnica de *Algoritmos Genéticos* (Michalewicz, 1999) para problemas de classificação de grupos. Veremos que esta técnica se apresentará como um diferencial competitivo, pois fornecerá uma solução de fácil entendimento e implementação em sistemas de informação através da evolução de regras de classificação. Desta forma, tecnologia, pessoas e processos numa empresa poderiam se alinhar de forma a contemplarem um novo conhecimento descoberto, aumentando lucros e competitividade no mercado

### **CUSTOMER LIFE TIME DURATION (CLTD) e RFV**

O paradigma dos problemas de RFV apresenta-se como o seguinte: clientes com baixa recência, alta frequência, e alto valor, apresentarão um alto CLTD e conseqüentemente estarão mais dispostos a manter um vínculo contínuo com a empresa, respondendo melhor a

campanhas de marketing. Contrariamente, clientes com alta recência, baixa frequência e baixo valor, são mais propensos à interrupção do vínculo empresarial, respondendo pior a campanhas, pois já sem encontram no fim do CLTD.

Quando se fala sobre CLTD, nem todas as relações cliente-empresa são iguais. Dependendo do tipo de serviço ou produto ofertado, clientes podem assumir um relacionamento contratual ou não contratual. Conforme se observa em Kuman (p.103), *casos contratuais* constituem a mais precisa observação do tempo de vida do cliente. Uma simples medida do tempo decorrido desde o início do relacionamento (ou início de uma janela de análise) até o fim do relacionamento (ou fim de uma janela de análise) pode ser obtida facilmente, determinando assim o CLTD. Desta forma, um cliente torna-se inativo quando não ocorre uma renovação de contrato. Neste caso, dizemos que ocorreu uma evasão ou “churn” de cliente.

Já em casos *não contratuais*, onde não há uma informação explícita sobre o fim de um relacionamento. Clientes neste mercado não têm barreiras que os impeçam de continuar ou interromper o relacionamento quando bem quiserem, sem alguma comunicação formal à empresa. O que nos proporcionará o ambiente ideal para aplicação de Algoritmos Genéticos.

Nestes ambientes não contratuais, a forma mais utilizada para cálculo do CLTD é emular uma regra de classificação de clientes ativos ou inativos em um tempo finito de relacionamento. Por exemplo, poderíamos definir uma regra baseada em RFV passado, para classificar clientes ativos e inativos após três meses de relacionamento, caracterizando *previamente* o fim ou manutenção do CLTD. Isto ofereceria parâmetros suficientes para que gestores de relacionamento ao cliente pudessem agir preventivamente na *retenção* destes classificados como futuros inativos, maximizando lucros da empresa pela permanência prolongada do status ativo. Conforme se observa em Karine (apud Reichheld & Sasser Jr., 1990), dependendo do setor de atuação, as empresas podem rentabilizar seus negócios em lucros de 25% a 85%, reduzindo em apenas 5% a perda de clientes.

Por estes motivos, ações focadas em grupos com *maiores chances de evasão* (menor chance de resposta a uma ação) passa a ser bastante atraente, pois vai ao encontro com a lucratividade futura da empresa e uma série de outros aspectos como satisfação e lealdade, sendo assim a estratégia defendida neste artigo.

## INTRODUÇÃO A TÉCNICA DE ALGORITMOS GENÉTICOS

Desenvolvido por John Holland na década de 60 e 70 [1], Algoritmos Genéticos (AGs) fornecem um mecanismo de busca adaptativa, inspirado na evolução natural de Darwin e reprodução genética humana, para resolução de problemas complexos de otimização. Fatores biológicos como seleção, reprodução, cruzamento e mutação de informação genética fornecem a estrutura necessária para resolução de problemas por AGs.

No mundo natural, restrições e incentivos de um ambiente em particular forçam diferentes espécies (e indivíduos dentro das espécies) a competirem e cruzarem para produção de filhos mais aptos. No mundo de AGs, a aptidão de várias potenciais soluções são comparadas, e as mais aptas terão mais chances de cruzarem entre si informação importante para o problema, produzindo soluções ainda mais aptas (Larose, p. 240).

Em AGs, várias soluções (ou indivíduos) são consideradas em paralelo a cada geração. Cada indivíduo possui os parâmetros necessários para resolução do problema, representados através de um *cromossoma* (ou *string* de caracteres), através do qual obtém-se um valor de

*aptidão* da solução. Cada cromossoma pode-se ser dividido em *genes*, que são pedaços ou blocos de DNA designados para codificarem uma determinada característica (exemplo: sexo).

A apresentação de uma determinada característica por um gene é dita como *alelo*, e a posição que ela ocupa no cromossoma como *locus*. AGs evoluem para soluções ótimas através de um processo adaptativo com o qual novos indivíduos são gerados, a partir dos operadores de *seleção*, *cruzamento* e *mutação* de antigos indivíduos. A seleção ocorre antes dos operadores genéticos de crossover e mutação.

Indivíduos são selecionados com base no seu valor de aptidão. Quanto maior a aptidão, maior é a probabilidade do indivíduo ser selecionado para cruzamento. O cruzamento, por sua vez, ocorre com a combinação de dois indivíduos selecionados, através da troca de partes do cromossoma de cada solução. E por último, o operador de mutação, quando da troca aleatória no gene sorteado de um cromossoma por um outro alelo.

Por diversos artigos e livros terem abordado esta técnica extensivamente, não entraremos em detalhe sobre operadores, técnicas e parâmetros de um GA, válida a exceção do problema abordada neste artigo sobre a evasão de clientes numa empresa. Para estudos mais detalhados sobre AGs, recomendamos a leitura de [1].

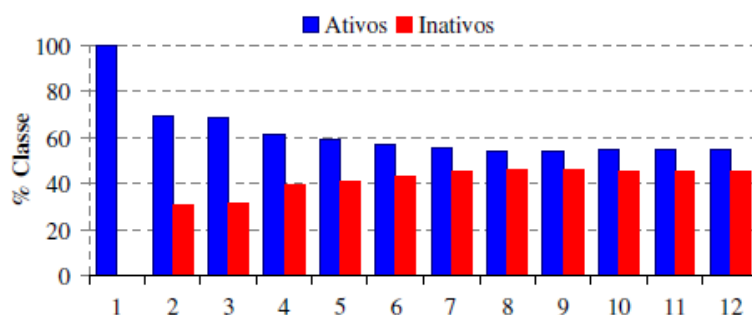
## **A EMPRESA E O PROBLEMA DE CHURN DE CLIENTES**

A empresa em estudo trata-se de uma administradora de investimentos com grande atuação na Bovespa (Bolsa de Valores do Estado de São Paulo). Clientes que optam por investir pelo sistema Home Broker executam ordens online, diversificando seus investimentos em até 4 categorias de investimentos: (1) compra e venda de ações; (2) cotas em fundos de investimentos; (3) bolsa de mercadorias e valores futuros (ou Bm&f); (4) e títulos do tesouro direto.

O problema da empresa apresenta-se da seguinte forma: clientes após a inclusão na base de dados e início das operações no sistema home broker, apresentam um decréscimo significativo na atividade até o 4<sup>o</sup> mês de relacionamento, identificado pela não intenção de continuar investindo. É fato para a empresa que após o 4<sup>o</sup> mês de relacionamento clientes que decidem por continuar suas operações na bolsa ou outro tipo de investimento o fazem continuamente ao longo de um horizonte de 12 meses ou mais.

Para que isto fique claro, apresentaremos o gráfico a seguir. Para tanto, separamos 12 safras mensais de entrada de cliente ao longo do ano de 2006, e verificou-se o status do cliente, mês a mês, por um período seguinte de 12 meses. Quando no mês de análise, após a inclusão no cadastro da empresa, não era verificada nenhuma operação no sistema home broker, o cliente era marcado como inativo. Contrariamente, recebia a marcação de ativo, caso viesse a efetuar alguma ordem de investimento no mês.

**Gráfico 1:** Curvas de Atividade e Inatividade



Fonte: Própria Autoria

Neste tipo de negócio, empresas gestoras de investimentos geralmente obtêm lucros através de uma taxa % sobre o valor movimento e/ou uma constante sobre cada ordem executada. Na empresa em análise, lucros provêm somente sobre uma valor constante para cada ordem executada.

Desta forma, podemos definir como variável de Valor, os lucros obtidos por quantidade de ordens executadas por cada cliente até um instante de tempo  $t$  dado pela função abaixo, onde  $t$  é uma medida mensal:

$$Valor_{Cliente}(t) = \sum_{i=1}^t Ordens_i * Const. \quad (1)$$

Esta última definição de valor seria um problema, pois pouco explicaria o potencial financeiro do cliente em questão, uma vez que o montante movimentado não é considerado no cálculo. Uma medida mais eficiente para Valor seria “a média de valor movimento por ordem executada até um instante de tempo  $t$ ”, e certamente seria mais discriminatória que a anterior.

$$Valor_{Cliente}(t) = \frac{\sum_{i=1}^t ValorMovimentado_i}{\sum_{i=1}^t Ordens_i} \quad (2)$$

Para Frequência, definimos como “a média mensal do número de ordens executadas até um instante de tempo  $t$ ”.

$$Frequência_{Cliente}(t) = \frac{\sum_{i=1}^t Ordens_i}{t} \quad (3)$$

E por último, Recência, como “o tempo de decorrido (em dias) até um instante de tempo  $t$ , desde a última ordem executada”.

Sendo assim, para nosso problema, consideraremos  $t = 3$  representando o terceiro mês de relacionamento do cliente. Com isto, buscaremos através da técnica de Algoritmos Genéticos descobrir regras que classifiquem futuros clientes inativos a partir do 4º mês de relacionamento. Desta forma, a empresa em questão poderá agir preventivamente através de ações de marketing de relacionamento, buscando a retenção destes clientes com maiores chances de evasão, chamados na literatura de RFV como de menor valor ou baixo CLTD futuro.

## **EVOLUÇÃO DE REGRAS DE DECISÃO POR ALGORITMOS GENÉTICOS**

A descoberta de conhecimento em grandes bancos de dados, ou data mining, tem inspirado muitos pesquisadores nos mais diversos campos da ciência. Uma dificuldade em processos de knowledge discover database (KDD), trata-se da extração do conhecimento correto, de fácil compreensão, e de grande utilidade para o usuário. Berry divide em 5 as responsabilidades atribuídas a mineração de dados em processos KDD: (1) classificação; (2) associação de regras; (3) perfilação de clientes; (4) clusterização; (5) estimação; e (6) predição.

Observaremos que AGs podem revelar conhecimento de extrema simplicidade na solução de problemas do tipo classificação, através da extração de regras de grande banco de dados. Regras do tipo IF ... THEN, onde a parte IF se refere a um conjunto de atributos preditores ou independentes, e THEN a um atributo dependente, ou seja, a classe de predição (Santos et al, 1999). Desta forma, quando um conjunto de características antecedentes for verdadeiro, poderemos afirmar com uma certa chance de acerto, que uma classe de interesse específica é conseqüente.

Para sermos mais específicos, voltemos ao problema da empresa em estudo. Dado um padrão nos três primeiros meses de relacionamento em termos de RFV do cliente (antecedentes), poderemos classificá-lo como um futuro cliente inativo ou não, a partir do 4º mês, interrompendo assim o CLTD (conseqüente). Neste sentido, AGs extrairão conhecimento o suficiente para responder a este tipo de problema de classificação.

## **MODELAGEM POR ALGORITMOS GENÉTICOS**

### **REPRESENTAÇÃO**

Como mencionado, faz-se necessária a representação de uma solução por um string de caracteres ou cromossoma. Marco [14] detalha como até 6 as formas de se representar um cromossoma dependendo do tipo do problema. São elas a binária, real, lista, vetor, inteiro, e mista. A escolha da representação é importante, pois em alguns casos podem levar a problemas de convergência prematura do algoritmo, inabilidade de operar na presença de restrições não triviais e/ou inabilidade de operar localmente ao ótimo global (Michalewicz, p.97).

Neste artigo não trataremos as vantagens e desvantagens de uma forma de representação a outra, no problema em destaque. Simplesmente ficaremos sujeitos a forma de representação do software aqui utilizado (Evolver 4.0 for Excel) na forma de números reais e/ou inteiros dependendo do usuário.

Em mercados como o de investimento, o comportamento em termos de RFV tem grande correlação com outras variáveis externas sócio-econômicas. Desta forma, é de grande preocupação que o modelo aqui objetivado seja o menos dependente de comportamentos exógenos à empresa. Para ilustrarmos nosso raciocínio e utilizando um mercado hipotético,

poderíamos dizer que um comportamento em termos de recência do cliente superior a 20 dias sem operar na bolsa seria típico de um futuro cliente evasivo, mas que em outra época com menor instabilidade econômica, isto seria esperado do cliente. Desta forma, optou-se por trabalhar com decis de valores de recência, frequência, e valor, como seria em modelos clássicos de RFV (Kuman, p.119), ao invés dos valores como apresentados no capítulo 3.

O banco de dados utilizado apresenta uma amostra de 14.799 clientes (linhas). As variáveis de RFV (colunas) apresentam o seguinte domínio: R (dias) pertence ao Dom {0; 90}; F (média de ordens executadas mês) ao Dom {0,333; 2,86 x 10A}; e V (média de valor executado por ordem) ao Dom {3,485 x 10B ; 4,085 x 10C}. As siglas A, B e C representam potências as quais não serão divulgadas, pois serve para manter o sigilo da empresa divulgadora dos dados. Sendo assim, cada uma das variáveis foi codificada no intervalo de 1 a 10, conforme os decis de suas distribuições de frequência.

**Figura 2 - Representação dos Decis de RFV**

R		F		V	
1	Mais	1	Maior	1	Maior
2	Recente	2	Frequência	2	Valor
3		3		3	
4		4		4	
5		5		5	
6		6		6	
7		7		7	
8		8		8	
9	Menos	9	Menor	9	Menor
10	Recente	10	Frequência	10	Valor

**Fonte:** Própria Autoria

Como estratégia de representação de problemas de mineração envolvendo AGs, decidimos por representar simultaneamente 4+1 regras potenciais para classificação de clientes evasivos.

Quatro, pois acreditamos ser o suficiente para solução do problema, tendo em vista 4 cluster (ou perfis) de clientes pré identificados num outro instante, com a utilização de Redes Neurais Artificiais e Mapas de Kohonen (Haykin, p.483). E 1 regra adicional, para critério de desempate de classes, conforme decodificação a ser detalhada em 6.2. Sendo assim, para representar um cromossoma, utilizaríamos uma lista de números reais com 24 posições, cada qual assumindo valores de 1 a 10, o que nos dá um espaço de busca de 1024 possíveis soluções. A seguir:

**Figura 3: Representação do Cromossoma**

Regra 1					
R		F		V	
Mín	Máx	Mín	Máx	Mín	Máx
{ 1; 10}	{ 1; 10}	{ 1; 10}	{ 1; 10}	{ 1; 10}	{ 1; 10}
....					
Regra 5					
R		F		V	
Mín	Máx	Mín	Máx	Mín	Máx
{ 1; 10}	{ 1; 10}	{ 1; 10}	{ 1; 10}	{ 1; 10}	{ 1; 10}

**Fonte:** Própria Autoria



## DECODIFICAÇÃO

A decodificação de uma possível solução para classificar futuros clientes evasivos seguiria o seguinte raciocínio: dado um cromossoma  $i$ , se pela Regra1 um cliente apresenta-se a  $R_i$  entre  $\text{Min}(R1)$  e  $\text{Máx}(R1)$  e  $F_i$  entre  $\text{Min}(F1)$  e  $\text{Máx}(F1)$  e  $V_i$  entre  $\text{Min}(V1)$  e  $\text{Máx}(V1)$ , este seria classificado como cliente evasivo. Era feito o mesmo raciocínio para todas as outras regras ( $R2$  até  $R5$ ), e ao final classificava-se o cliente na classe com maior número de votos.

Exemplo: Seja um cliente avaliado pelo cromossoma  $i$ , sua avaliação recebeu três votos na categoria de evasivo dados pelas regras  $R1$ ,  $R3$ , e  $R4$ , e 2 votos para a categoria de retido pela regras  $R2$  e  $R5$ . Desta forma, classificou-se este cliente como futuro evasivo. Por isso utilizouse uma 5 regra adicional no cromossoma, para que não houvesse empate de classes.

## AVALIAÇÃO

A avaliação de um cromossoma  $i$ , pertencente a uma população na geração  $j$ , requer a leitura de toda uma base de dados de clientes. A forma mais usual de efetuar esta avaliação é utilizar a *acurácia* e *abrangência* de um cromossoma em toda a base de dados de clientes. Acurácia como o % de acerto dado pelo modelo na classe objetivada pelo problema (em nosso caso clientes evasivos), e abrangência como o % de cobertura da classe objetivada na base utilizada.

A modelagem de um problema que classifique corretamente futuros clientes como evasivos ou retidos, apresenta 4 possíveis ocorrência em virtude de seu histórico passado observado, conforme tabela abaixo.

**Tabela 1:** Espaços de Ocorrência de um Cromossoma

Classes		Classificada		
		Evasivo	Retido	Total
Observada	Evasivo	A	B	(A+B)
	Retido	C	D	(C+D)
	Total	(A+C)	(B+D)	(A+B+C+D)

Fonte: Própria Autoria

Onde A, B, C, e D são números inteiros, dados pelas clientes pertencentes a estas categorias. Defini-se então a acurácia de um cromossoma  $i$  como:

$$Ac_i = \frac{A_i}{(A_i + C_i)} \quad (4)$$

E abrangência como:

$$Ab_i = \frac{A_i}{(A_i + B_i)} \quad (5)$$

Suponham que a hipótese nula,  $H_0$ , de um modelo estatístico seja: assumir que todos os clientes são futuros clientes evasivos. Podemos definir dois tipos de erros encontrados em

testes de hipóteses estatísticos (Bussab, p.323) O primeiro, erro do tipo 1, a probabilidade de eu rejeitar  $H_0$  dado que ela é verdadeira, ou seja, a probabilidade de eu assumir que o cliente é futuro retido, dado que ele será um futuro evasivo. O segundo, erro do tipo 2, seria a probabilidade de eu aceitar  $H_0$ , dado que  $H_0$  é falsa. Podemos definir então ambos erros, para um cromossoma  $i$ , da seguinte forma:

$$ET1_i = \frac{B_i}{(A_i + B_i)} \quad (6)$$

$$ET2_i = \frac{C_i}{(A_i + C_i)} \quad (7)$$

Sendo assim podemos dizer que uma boa solução para o problema seria aquela que maximiza-se a  $A_i$  e  $D_i$ , e minimizassem os  $ET1_i$  e  $ET2_i$ . Reparem que  $A_i = (1 - ET2_i)$ , e  $D_i = (1 - ET1_i)$ . Reparem também que estamos lidando com múltiplos objetivos na avaliação de um cromossoma. Para problemas desta natureza, pode utilizar uma combinação de múltiplos objetivos numa única função  $f$ , de tal forma a maximizá-la ou minimizá-la, de acordo com suas características. Para o nosso problema, utilizou-se como objetivo a maximização da função abaixo.

$$f(A_i, D_i, ET1_i, ET2_i) = \frac{(A_i + D_i)}{(C_i + B_i)} \quad (8)$$

Percebam que a maximização da função acima atende aos quatro objetivos aquidetalhados: maximizar acurácia e abrangência, e minimizar erros do tipo 1 e 2.

## RESULTADOS OBTIDOS

Para evolução das espécies de cromossomas utilizou-se o software Evolver 4.0 for Excel.

Nesta etapa, clientes foram separados em dois grupos de análise. O primeiro consistia numa amostra equilibrada de 5000 clientes evasivos, e 5000 retidos, os quais foi utilizado para avaliação das regras evoluídas em todos os passos da modelagem. O segundo grupo foi usado como controle para avaliação do potencial de generalização das regras obtidas, e consistiu numa amostra de 1.085 clientes evasivos e 3.714 retidos.

Para os operadores genéticos, utilizamos o crossover uniforme para troca genética entre soluções, mutação como na forma clássica em AG, e elitismo para seleção do melhor indivíduo na próxima geração. Muitos testes foram feitos inicialmente para determinar as taxas de crossover e mutação – constantes em todo o processo de evolução –, bem como o número de geração e tamanho da população, mas nenhum resultado significativo foi obtido em termos de evolução. Observamos que a evolução tendia a privilegiar a classe de clientes retidos, minimizando assim  $A_i$  e  $D_i$ , dados pela cromossoma vencedor.

Por estes motivos, concluímos que era fundamental que existisse na primeira geração um cromossoma o qual fosse favorável a classe de clientes evasivos, visto o tamanho do espaço de

busca do problema. Este cromossoma de certa forma carregaria material genético importante na solução do problema.

Para tanto, a inclusão deste cromossoma “chave” na população inicial poderia buscar origem no resultado obtido de algum método estatístico (ou não) de classificação de padrões, ao exemplo de árvores de decisões. No entanto, optou-se pela forma mais simples de seleção deste cromossoma: inclui-se o cromossoma referente ao que seria a hipótese nula de um modelo estatístico ( $H_0$ ), ou seja, considerar a priori que todos os clientes são futuros evasivos.

Sendo assim, o cromossoma que representa esta hipótese nula trata-se de um string onde em um locus ímpar o valor seria igual a 1, e locus par, valor igual a 10. Desta forma, qualquer que fosse o cliente testado, atenderia sempre as 5 regras representadas no cromossoma, recebendo 5 votos para a classe de evasivos.

**Figura 4:** Representação do Cromossoma  $H_0$

Regra 1					
R		F		V	
Min	Máx	Min	Máx	Min	Máx
1	10	1	10	1	10
....					
Regra 5					
R		F		V	
Min	Máx	Min	Máx	Min	Máx
1	10	1	10	1	10

**Fonte:** Própria Autoria

A inclusão deste cromossoma na população demonstrou significativos avanços na direção de convergência do AG. No entanto, eram comuns o encontro e convergência de regras com grande abrangência e baixa acurácia, digamos,  $Ab = 90\%$  e  $Ac = 34\%$ .

Como trabalhávamos com amostras equilibradas, a obtenção do máximo global somente aconteceria quando  $Ab$  fosse o mais próximo possível de  $Ac$ . Regras as quais obtivessem estes resultados em comparação as demais, certamente deveriam ser favorecidas. Para tanto, inclui-se uma função de penalidade do tipo *Soft* definida pelo Evolver na fórmula  $100 * (\text{EXP}(\text{deviation}/100) - 1)$ , o qual penalizava soluções que não satisfizessem a seguinte condição:

$$|Ac_i - Ab_i| < 20\% \quad (9)$$

Resultados ao final de um ciclo de gerações igual a 500, com 1000 indivíduos cada, demonstraram-se bastante promissores. Parâmetros de crossover e mutação foram ajustados, respectivamente, para 70% e 5%. Cada ciclo foi repetido 10 vezes, mantendo-se sempre o melhor indivíduo do ciclo anterior, onde ao final observou-se a convergência ao que aparenta ser o ótimo global para o problema em questão. É importante dizer que somente no primeiro ciclo considerou-se o cromossoma  $H_0$  como possível solução. Um resumo para as estatísticas de  $Ac$ ,  $Ab$ , ET1, ET2, e  $f$ , obtidas pelo cromossoma vencedor ao final de cada ciclo, podem ser obtidas a seguir:

**Tabela 2:** Estatísticas de Desempenho e Avaliação do Cromossoma Vencedor – Amostra Avaliação

Ciclo	Ac	Ab	ET1	ET2	f
1	73,64%	70,96%	29,04%	26,36%	2,6738
2	78,43%	63,94%	36,06%	21,57%	2,7286
3	74,00%	75,82%	24,18%	26,00%	2,9355
4	73,13%	79,32%	20,68%	26,87%	3,0145
5	74,38%	77,34%	22,66%	25,62%	3,0568
6	74,80%	76,54%	23,46%	25,20%	3,0617
7	74,69%	77,36%	22,64%	25,31%	3,0933
8	74,78%	77,20%	22,80%	25,22%	3,0950
9	75,20%	76,42%	23,58%	24,80%	3,1000
10	75,20%	76,42%	23,58%	24,80%	3,1000

Fonte: Própria Autoria

**Tabela 3:** Representação do Cromossoma Vencedor

Regra	Recência		Frequência		Valor	
	Mín	Máx	Mín	Máx	Mín	Máx
1	1	10	1	10	1	10
2	4	10	1	10	1	10
3	5	10	6	10	7	10
4	7	10	1	10	1	10
5	1	5	9	10	1	10

Fonte: Própria Autoria

Para a amostra de controle, os resultados de generalização também foram satisfatórios com valor  $f$  ao final do 10º ciclo de 3,1988. A seguir, a representação do cromossoma vencedor:

## CONCLUSÕES E PRÓXIMOS PASSOS

A evolução de regras por algoritmos genéticos resultou no encontro de regras com alta acurácia e abrangência na solução do problema de evasão de clientes. Tais regras podem ser facilmente implementadas em sistemas inteligentes, bem como interpretadas por usuários envolvidos no processo de tomada de decisão de marketing. Uma vez objetivado reduzir as taxas de evasão de clientes, e aumentar a lucratividade futura da empresa pela permanência prolongada do status ativo do cliente, poderiam trabalhar ações específicas de marketing aos clientes classificados como futuros evasivos, reduzindo-se custos de alocação de marketing. Há de se reconhecer que a utilização de outras técnicas de inteligência computacional, ou estatística, poderiam produzir resultados melhores em termos de acurácia e abrangência, ficando este aspecto a ser investigado em passos futuros.

## REFERÊNCIA BIBLIOGRÁFICAS

[MICHALEWICZ, Zbigniew. **Genetic Algorithms + Data Structures = Evolution Programs**. 3rd rev. and extended ed. New York: Springer, 1999.

- COLOMBO, Richard; JIANG Weina. A Stochastic RFM Model. **Journal of Interactive Marketing**: Summer 1999.
- FADER, Peter; HARDIE, Bruce; LEE, Ka Lok. RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. **Journal of Marketing Research**: Vol. XLII (November 2005).
- PIERSMA, Nanda, et al. **Media Planning by Optimizing Contact Frequencies**. Econometric Institute Report EI 9856/A.
- COOPER, Lee, et al. Using Genetic Algorithms to Breed Competitive Marketing Strategies. **IEEE International Conference**: Volume 3, p.2367-p.2372, October 1998.
- REINARTZ, Werner. KUMAR, V. **Customer Relationship Management, A Database Approach**. John Wiley & Sons, June 2005.
- HUMBY, Clive; HUNT, Terry; PHILLIPS, Tim. **Scoring Points**. 2nd Ed. London: Kogan Page Limited, 2007.
- BARTH, Nelson Lerner. **Inadimplência: Construção de Modelos de Previsão**. São Paulo: Nobel Editora, 2004.
- BERRY, Michael; LINOFF, Gordon. **Data Mining Techniques for Marketing, Sales, and Customer Relationship Management**. John Wiley & Sons, Indianapolis, Indiana, 2004.
- LAROSE, Daniel; **Data Mining, Methods and Models**. John Wiley & Sons, New Jersey, Canada, 2006.
- SANTOS, Raul; Extração de Regras de Redes Neurais via Algoritmos Genéticos. **IV Congresso Brasileiro de Redes Neurais**, São José dos Campos, SP, Julho de 1999.
- HAYKIN, Simon. **Redes Neurais, Princípios e Práticas**. Paulo Matins Engel. 2ed. Porto Alegre: Bookman, 2001.
- BUSSAB, Wilton de O.; MORETTIN, Pedro A.. **Estatística Básica**. 5ed. São Paulo: Saraiva, 2006.
- AURÉLIO, Marco. **Notas de Aula do Curso Computação Evolucionária**. Pontifícia Universidade Católica, Rio de Janeiro, RJ, 2008.