

1 Introdução

1.1. Motivação

Qualquer empresa sabe, nos competitivos dias de hoje, onde reside sua verdadeira riqueza: em seus clientes. Não importa em que indústria estejam inseridas, em que setor operem. Os clientes são quem move o empreendimento, quem sustenta a operação e, em última análise, a própria razão de uma empresa existir.

Ao longo dos anos muitos pressupostos e práticas de negócios antes tomadas como ideais foram caindo por terra, dando lugar a novos conceitos utilizados mais e mais a cada dia. O mundo e os mercados em ritmo de rápida mudança obrigam as empresas a evoluírem sempre, se desejarem manter sua prosperidade. O foco deixou de estar no produto e passou para o mercado e o cliente. Deixou de ser doméstico para ser global. Produtos padronizados deram lugar a produtos personalizados, o marketing de massa se transformou em marketing direcionado, caminhando para um marketing individualizado para cada cliente [KOTL99].

Na busca por vantagem competitiva e melhor compreensão de seu grande tesouro - os clientes - as empresas atuais se voltam para a imensa quantidade de dados corporativos que possuem. A explosão da quantidade de dados disponíveis em meios magnéticos e a rápida evolução da tecnologia da informação, permitindo sua organização, armazenamento e utilização, criaram uma verdadeira fonte de conhecimento a ser explorada na batalha pela liderança e sobrevivência em mercados cada vez mais competitivos. Em sua maior parte tais dados não são coletados com a intenção inicial de serem analisados e utilizados em alguma modelagem preditiva; eles são simplesmente adquiridos num esforço de aumentar a eficiência das operações de algum negócio [BERR00]. Entretanto, uma vez coletada, essa grande massa de dados representa uma valiosa fonte de informação que pode ser utilizada na melhoria do processo de tomada de decisões empresariais em praticamente todos os níveis.

No entanto, transformar imensas quantidades de dados em retratos fiéis de clientes é uma tarefa muito árdua que exige arte e ciência, matemática e intuição. É nesse contexto que surge o conceito de mineração de dados. Minerar dados não é nada mais do que a combinação de grande poder computacional, algoritmos inteligentes, conhecimento do negócio e intuição humana aplicada a grandes massas de dados, objetivando a descoberta do conhecimento útil para o negócio mas oculto em milhares de terabytes de dados [KELL95].

É muito importante ter sempre em mente que mineração de dados não é somente aplicar algoritmos sofisticados como redes neurais [BISH96] [MOZE00] [ZHEN04], árvores de decisão [QUIN87] [BERR00], algoritmos genéticos [MICH96] [LOP99-1] [LOP99-2], sistemas neuro-fuzzy [CHUN00] [GONÇ01] [JANG97] [SOUZ99] ou máquinas de vetor de suporte [ARCH04] [BURG98] [ZHEN04] em qualquer tipo de dado e esperar a obtenção de modelos que são a solução ideal para todos os problemas de uma empresa. É essencial unir profundo conhecimento do negócio ao uso sábio e judicioso de métodos e algoritmos computacionais, através de um sistema que englobe e leve em consideração todas as peculiaridades do problema a ser solucionado. Só assim o já mencionado conhecimento será realmente encontrado em meio às grandes quantidades de dados disponíveis. Somente desta forma será gerada inteligência real sobre o negócio, inteligência que pode e deve servir de base no processo decisório.

Neste trabalho busca-se a solução de um problema de gerenciamento de relação com o cliente através da criação, estudo e otimização do processo de mineração de dados como um todo. A questão da perda de um cliente para um concorrente (*churn*) na indústria de telefonia celular [ARCH04] [AU03] [GEPP02] [MATT01] [MOZE00] [MOZE02] no Brasil [FERR03-1] [FERR03-2] [FERR04-1] [FERR04-2] [FERR05] será avaliada em busca de uma solução satisfatória. Um processo de mineração de dados cuidadoso, partindo do estudo do problema, passando pelo tratamento dos dados e culminando na busca e otimização do modelo mais adequado para a tarefa será a arma utilizada no combate ao *churn*. Visa-se também demonstrar o grande valor e vantagem competitiva que podem ser gerados com a utilização de tais processos em estratégias de retenção de clientes, através da análise da lucratividade do sistema de retenção de clientes a partir de mineração de dados desenvolvido.

1.2. Objetivos

O principal objetivo deste trabalho foi a elaboração de um sistema de retenção de clientes baseado em mineração de dados para a indústria de telefonia celular que chegasse o mais próximo possível de uma solução satisfatória para o problema do *churn* (perda de um cliente para uma operadora concorrente). Este problema em particular foi escolhido, principalmente devido à existência de dados reais, como forma de ilustrar o processo completo de mineração de dados direcionado para a resolução de uma questão específica em uma indústria qualquer. Espera-se que o leitor seja facilmente capaz de utilizar o conhecimento relatado e as etapas descritas no estudo para problemas de mineração de dados em outras indústrias.

Assim, apesar do sistema proposto focar em um problema específico, ele possui características e estágios comuns a diversos outros sistemas de mineração de dados que poderiam vir a ser desenvolvidos com foco em outras questões.

O desenvolvimento do sistema iniciou-se com o estudo abrangente do problema de retenção de clientes na indústria de telefonia celular brasileira, avaliando-se cuidadosamente o ambiente altamente competitivo no qual esta indústria se insere, suas causas e suas conseqüências [MATT01].

Com base no conhecimento da questão do *churn* e suas peculiaridades, principalmente no que diz respeito aos dados disponíveis e ao tipo de análise a ser realizada para a obtenção de um melhor resultado, foi realizado um extenso estudo de tratamento de dados [PYLE99] [MOZE00] [YAN04] e métodos de seleção de variáveis [BLUM97] [YI97] [CONT02] indicados para a otimização do problema de classificação de padrões, no qual os clientes de uma operadora eram classificados como possíveis abandonos ou não.

Para a tarefa de classificação, diversos modelos foram testados buscando-se o que provesse a classificação mais adequada para o problema [MOZE02]. Levou-se em consideração a importância e o impacto que cada tipo de erro de classificação (falso positivo ou falso negativo) [YAN01] poderia vir a ter quando integrado a um sistema de retenção de clientes. Entre os modelos avaliados destacam-se: classificadores bayesianos [DUDA00], redes neurais [BISH96] [MOZE00] [ZHEN04], árvores de decisão [QUIN87] [BERR00], algoritmos genéticos [MICH96] [LOP99-1] [LOP99-2], sistemas neuro-fuzzy [CHUN00]

[GONÇ01] [JANG97] [SOUZ99] e máquinas de vetor de suporte [ARCH04] [ZHEN04].

Finalmente, para concluir sobre a eficácia do sistema de retenção proposto, foi realizada uma análise da lucratividade [MOZE00] que uma operadora teria ao empregá-lo como parte de sua estratégia de retenção de cliente.

1.3. Descrição do Trabalho

Esta dissertação divide-se nas seguintes etapas: estudo da aplicação de mineração de dados a problemas de gerenciamento de relação com clientes (*Customer Relationship Management*); estudo do negócio e da estrutura da indústria de telefonia celular no Brasil, com foco na compreensão dos detalhes pertinentes ao problema de abandono de clientes para uma operadora concorrente (*churn*); aplicação de métodos de tratamento de dados, entre eles normalização, limpeza e *oversampling*; escolha e utilização de métodos de seleção de variáveis como o Método do Least Squares Estimator (LSE), o Método do Single-Input Effectiveness (SIE) e o modelo ANFIS; estudo, implementação e otimização de diversos modelos de classificação de padrões, entre os quais classificadores bayesianos, redes neurais, algoritmos genéticos, sistemas híbridos neuro-fuzzy e máquinas de vetores de suporte; aplicação de todos os métodos e modelos estudados a uma base de dados real de uma operadora de telefonia celular brasileira; avaliação de todos os resultados obtidos sob o prisma do problema que se busca solucionar, buscando os resultados mais interessantes; estudo da lucratividade do sistema de retenção proposto através da aplicação sobre uma base de dados real.

Em todos os testes e simulações realizados, foram utilizadas as ferramentas SAS System, SAS Enterprise Miner, SAS Enterprise Guide, MATLAB 6.1 e ferramentas criadas pelo laboratório ICA (Inteligência Computacional Aplicada) da PUC-Rio: Rule Evolver [LOPE99-1] [LOPE99-2] e o sistema Neuro-Fuzzy Hierárquico BSP [GONÇ01] [SOUZ99].

1.4. Organização da Dissertação

Esta dissertação é composta por mais 6 capítulos além deste:

O capítulo 2 apresenta o conceito da descoberta de conhecimento em bases de dados, discutindo pontos importantes e dando foco também a aplicações deste conceito em ações e estudos de CRM (*Customer Relationship Management*),

No capítulo 3 são apresentadas brevemente as técnicas e métodos utilizados no sistema de retenção de clientes a ser criado, desde a etapa de preparação dos dados até os modelos de classificação de padrões.

O capítulo 4 trata da indústria de telefonia celular no Brasil e discute em detalhes a questão do abandono de cliente (*churn*), investigando suas origens, causas, conseqüências e possíveis soluções.

No capítulo 5 é estruturado o sistema de retenção de clientes proposto, unindo o conhecimento do negócio e do problema com o ferramental computacional na busca de resultados de lucro para uma operadora.

O capítulo 6 corresponde ao estudo de caso da dissertação, com aplicação do sistema desenvolvido sobre uma base de dados real e a avaliação de seu desempenho.

O capítulo 7 apresenta as conclusões e sugestões para trabalhos futuros.