5CONTECSI June 04 - 06, 2008 São Paulo, Brazil

5th International Conference on Information Systems and Technology Management
5º Congresso Internacional de Gestão da Tecnologia e Sistema de Informação
De 04 a 06 de Junho de 2008 - São Paulo - Brasil

PS-1043

A DATA MINING METHODOLOGY TO SUPPORT CUSTOMER RELATIONSHIP MANAGEMENT

Clodis Boscarioli (West Paraná State University, Paraná, Brasil) –boscarioli@unioeste.br
Leandro Augusto da Silva (University of São Paulo, São Paulo, Brasil) – leandro.augusto@ieee.org
Renato José Sassi (University of São Paulo, São Paulo, Brasil) – sassi@lsi.usp.br
Emilio Del Moral Hernandez (University of São Paulo, São Paulo, Brasil) – emilio_del_moral@ieee.org

In the current organizational scenario, the tools that allow the Customer Relationship Management (CRM) are essentials. Data mining techniques can provide important information to a better iteration with customer. The purpose of this work is address a data mining methodology applied in the CRM, specifically in data clustering. This methodology shows the importance of attributes of the database in automatic generation of clusters, and the real relation between the disposed clusters, to a better employment of the knowledge to a personalized marketing.

Keywords: Data-mining, Customer Relationship Management, Artificial Neural Network, Self-Organizing Maps, Data Clustering.

UMA METODOLOGIA DE MINERAÇÃO DE DADOS PARA AUXÍLIO AO GERENCIAMENTO DO RELACIONAMENTO COM O CLIENTE

No cenário organizacional atual, faz-se necessário ter ferramentas que permitam o gerenciamento do relacionamento com o cliente (CRM). Técnicas de mineração de dados podem prover as informações necessárias para essa melhor interação com os clientes. O objetivo deste trabalho é descrever uma metodologia de mineração de dados aplicada a CRM, mais especificamente de agrupamento de dados, que enfatiza a análise da influência dos atributos de uma base de dados na formação automática de *clusters* (grupos), e a relação existente entre os grupos encontrados, para um melhor aproveitamento do conhecimento para o marketing personalizado.

Palavras-chave: Mineração de Dados, Gerenciamento do Relacionamento com o Cliente, Redes Neurais Artificiais, Mapas Auto-organizáveis, Agrupamento de Dados.

1. Introdução

No inicio das relações comerciais, utilizava-se a filosofia do relacionamento personalizado: Conhecia-se os clientes pelo nome, sabia-se onde moravam, onde trabalhavam, que tipo de produtos necessitavam, como queriam pagar suas contas e quanto tinham para gastar na compra. Intencionalmente ou não, dividia-se os clientes em grupos, de acordo com seu valor à empresa, dando atendimento prioritário aos de maior valor, e investindo naqueles com grande potencial em tornar-se bons clientes.

Com o tempo, a venda e o marketing personalizado deram lugar às estratégias baseadas em pesquisas e estatísticas de mercado. O cliente perdeu sua identidade e passou a ser apenas um consumidor que, mesmo sem participar das pesquisas mercadológicas, foi incluído nas estatísticas.

No entanto, os melhores clientes podem não estar, ou sequer desejarem constar nessas estatísticas. Com o surgimento de novas tecnologias e o fácil acesso aos recursos computacionais, houve novamente uma grande mudança: os conceitos de marketing de massa e personalizado fundiram-se, e deram lugar à Personalização de Massa, que corresponde à análise dos dados disponíveis para atingir o conhecimento dos vários tipos de clientes (Peppers & Rogers, 2000).

Emerge ai o conceito de Gerenciamento do Relacionamento dos clientes (CRM - *Customer Relationship Management*), que é a personalização do atendimento baseado em níveis de clientes. Não há a necessidade de um atendimento diferenciado individual. Cada cliente está relacionado a um grupo que possui necessidades de atendimento e intenções de compra comuns. Com o conhecimento de a qual agrupamento o cliente pertence, pode-se personalizar o atendimento de acordo com uma classe de requerimentos.

Neste contexto, um conjunto de ferramentas de software e metodologias para gestão do negócio foi desenvolvido e aplicado, sob a denominação de Sistemas de Inteligência de Negócios (BI - *Business Intelligence*), para análise de informações internas e externas à empresa que auxiliem na tomada de decisão. São exemplos dessas ferramentas: *Data Warehouse*, *Data Mining*, *Balanced Scored Cards*, *Dashboards*, relatórios Ad-hoc e OLAP (*On-line Analytical Processing*), podendo-se afirmar que a aplicação de *Data Mining* ainda é a menos encontrada em soluções de software gerenciais.

O objetivo deste trabalho é descrever uma metodologia de mineração de dados (*Data Mining*, DM) aplicada a CRM, que visa à formação de agrupamentos de dados (Data Clustering), dando ênfase na análise da influência dos atributos de uma base de dados na formação automática de *clusters* (grupos ou segmentos).

Uma vez que grupos com atributos comuns tendem a estar próximos, a compreensão da relação de um grupo com o restante da base de dados, principalmente com os grupos presentes em seu entorno, faz-se necessária, uma vez que essa relação é bastante importante às aplicações de marketing personalizado.

Este artigo está organizado da seguinte forma:

A Seção 2 traz conceitos de CRM, necessários à contextualização junto ao objetivo proposto. Na Seção 3 uma visão geral sobre Mineração de Dados é dada, com ênfase na tarefa de agrupamento de dados e nas técnicas utilizadas. A relação de CRM e Mineração de Dados é brevemente apresentada na Seção 4. A

metodologia utilizada e seus resultados são apresentados na Seção 5. Por fim, a Seção 6 discute os resultados obtidos, as conclusões e os trabalhos futuros dessa pesquisa.

2. CRM (*Customer Relationship Management*)

Segundo Bretzke (2000), CRM é a combinação da filosofia do marketing de relacionamento com a tecnologia da informação, que provê os recursos de informática e telecomunicações, integrando os canais de relacionamento como o *call center*, a Internet, a força de vendas, e toda a empresa, de uma forma singular, que permite gerenciar o relacionamento com o cliente, agregando valor a cada relação.

CRM é uma estratégia que, a partir de um conjunto de processos, proporcionam à empresa meios para atender e conhecer os requerimentos dos clientes em tempo real, passando essas informações a todos os departamentos da empresa, para oferecer tratamento diferenciado ao cliente em qualquer setor. Para (Peppers & Rogers, 2000), um dos benefícios do CRM é a aprimorada habilidade de ver e analisar as atividades dos clientes.

CRM, contudo, requer sério planejamento, além do comprometimento de todas as pessoas da organização. De acordo com (Thearling et. al, 1999), há vários fatores que vêm aumentando a complexidade do relacionamento com os clientes, como tempo reduzido para uma campanha de marketing, aumento dos custos com marketing, fluxo de novos produtos oferecidos e nichos concorrentes.

Oliveira (2000) divide o processo de implantação de CRM em três etapas:

i) Conhecimento do cliente por meio de análises e pesquisas que revelem seu perfil e intenções de compra.

ii) Planejamento de campanhas de marketing e interação com o cliente, envolvendo análise dos dados coletados na primeira etapa, com definição de estratégias de atendimento.

iii) Efetivação das ações de marketing e venda.

De acordo com Tschohl (1996), um dos maiores problemas na implantação do CRM nas empresas é a insistência dos gerentes em considerar o relacionamento com o cliente apenas como uma estratégia de marketing. O marketing é um dos aspectos importantes na adoção desta nova filosofia, mas seu grande diferencial está no acompanhamento do ciclo de vida de um cliente, desde a sua primeira visita até sua última compra, existindo assim, a preocupação com todas as interações do consumidor com a empresa.

O CRM é dividido em três tipos: o CRM analítico, o operacional e o colaborativo.

O CRM analítico consiste em acompanhar as transações dos clientes para descobrir informações que ajudem a diferenciá-los em níveis de importância à empresa, além de realizar um acompanhamento das atividades de cada cliente.

O CRM operacional – que de acordo com (Peppers & Rogers, 2000), a maioria das empresas está focada – engloba ferramentas, informatizadas ou não, que melhoram a eficiência do relacionamento entre o cliente e a empresa. É nesse tipo de CRM que se prevê a integração de todos os produtos de tecnologia para proporcionar o melhor atendimento ao cliente. Estão entre os produtos de

CRM operacional as aplicações de automação da força de vendas, personalização de produtos, automação de marketing, reservatório central de conhecimento e dos sistemas de comércio eletrônico.

O CRM colaborativo é a aplicação da tecnologia de informação (TI) que permite a automação e a integração entre todos os pontos de contato do cliente com a empresa.

O foco deste trabalho está no CRM analítico, e na primeira etapa descrita por (Oliveira, 2000) para um processo de CRM.

3. Mineração de Dados

Mineração de Dados (DM - *Data Mining*) é um termo genérico que identifica o processo de transformar dados armazenados em conhecimento, expresso em termos de formalismos de representação como regras e relações entre dados (Rezende, 2005). O objetivo deste processo é descobrir, de forma automática ou semi-automática, o conhecimento que está implícito nas grandes quantidades de dados armazenados em bancos de dados.

Para que a mineração de dados possa encontrar padrões e tendências desejadas, ela deve realizar algumas tarefas, como:

Classificação: Tarefa que mapeia os dados de entrada em um número finito de classes, onde cada exemplo pertence a uma única classe, entre um conjunto pré-definido de classes. O objetivo de um algoritmo de classificação é encontrar algum relacionamento entre os atributos e uma classe, de modo que o processo de classificação possa usar esse relacionamento para predizer a classe de um exemplo desconhecido.

Estimativa: O objetivo da tarefa de estimativa é determinar um valor mais provável para um índice diante dos dados do passado ou de dados de outros índices semelhantes sobre os quais se tem conhecimento (Paini, 2003). Um exemplo deste tipo de tarefa é estimar o valor de um empréstimo que pode ser concedido a uma empresa, onde, a partir da análise das informações obtidas, estima-se o valor máximo que pode ser oferecido.

Regras de Associação: Uma regra de associação caracteriza o quanto a presença de um conjunto de itens nos registros de uma base de dados implica na presença de um outro conjunto distinto de itens nos mesmos registros. Desse modo, o objetivo das regras de associação é encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados.

Agrupamento de Dados (*Data Clustering*): É o processo de agrupamento de dados, tal que objetos dentro de um grupo tenham alta similaridade, quando comparados a outros objetos deste, e alta dissimilaridade a objetos de outros grupos. Por ser considerada uma das tarefas mais usuais do processo de DM para descobrir grupos e identificar distribuições e padrões interessantes que estão subjacentes aos dados (Halkidi et al., 2001), e por ser o foco deste trabalho, segue melhor detalhado.

Para (Cormack, 1971), a tarefa de agrupamento de dados está pautada em duas idéias básicas, a coesão interna dos objetos (homogeneidade) e o isolamento externo (separação) entre os grupos. Esta mesma definição é adotada por (Hair et al., 1998) e (Lattin et. al., 2003).

Para (Everitt et al., 2001) a análise de agrupamentos é um termo genérico para uma ampla escala de métodos numéricos utilizados para examinar dados multivariados visando encontrar conjuntos de observações homogêneas. Dada uma amostra de *n* dados (ou indivíduos), cada um deles medido segundo *p* variáveis, o objetivo é procurar um esquema que os agrupe em *g* grupos. Com este agrupamento, é possível identificar relacionamentos úteis entre os dados, como similaridades e diferenças antes não reveladas.

É um processo de aprendizado não-supervisionado, uma vez que não há classes ou exemplos pré-definidos que evidencie que algum tipo de relação deva ser válida entre os dados, ou mesmo a presença de tutores do domínio para supervisionar o aprendizado. Tanto o número ótimo de grupos quanto as características particulares que revelam semelhanças (ou diferenças) devem ser determinados pelo próprio processo.

Para a tarefa de agrupamento de dados, este trabalho utiliza, de forma combinada, os algoritmos SOM e *K-means*, abaixo descritos.

Self-organizing Maps

Um Mapa Auto-organizável (SOM - *Self-Organizing Maps*) é uma arquitetura de rede neural, com aprendizado não-supervisionado, baseada em um mapa de neurônios, cujos pesos são adaptados para verificar entradas de vetores semelhantes em relação a um conjunto de treinamento (Kohonen, 2001). Sua principal característica é o mapeamento ordenado dos padrões de entrada de elevada dimensão em mapas de neurônios de saída com menor dimensão, comumente em duas dimensões. Com os dados dispostos no mapa em duas dimensões, esses se tornam mais simples de serem agrupados, bem como de serem visualizados, o que têm motivado a aplicação e aprimoramento deste algoritmo em aplicações de DM (Kaski & Kohonen 1996), (Vesanto, 1997), (Vesanto & Alhoniemi, 2000) e (Jin et al., 2004).

SOM é estruturado em duas camadas, de entrada e de saída. Os neurônios da camada de saída são comumente dispostos em um mapa de duas dimensões, preservando a relação de vizinhança dos dados de entrada.

O algoritmo de treinamento do SOM é um processo iterativo e também chamado de competitivo. Em cada passo do processo (ou época), uma tupla (registro de dados) é aleatoriamente escolhida da base de dados. É então calculada a distância, geralmente a euclidiana, entre a tupla e todos os vetores protótipos. A unidade com menor distância, chamada de *bmu* (*best-matching unit*) é o *u* com protótipo *m* mais próximo à tupla:

$$||\text{tupla} - m_{bmu}|| = \min_i\{||\text{ tupla}-m_i||\}$$

Em seguida, os vetores protótipos são atualizados. O *bmu* e sua vizinhança topológica são movidos para próximos a tupla no processo, como se fosse um “arraste”.

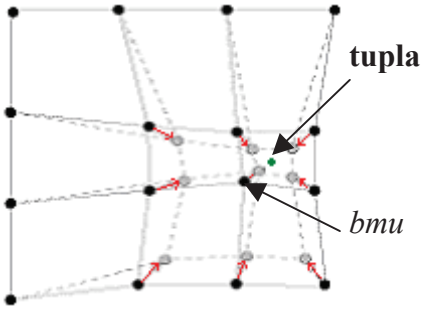

Figura 1 – Atualização do Neurônio Vencedor

Na Figura 1 são apresentadas a atualização do neurônio vencedor (*bmu*) – também chamado de vetor protótipo - e sua vizinhança em direção à tupla. Os círculos em preto e cinza correspondem às situações anterior e posterior à atualização, respectivamente. As linhas mostram a relação da vizinhança.

A regra para a atualização dos vetores protótipos da unidade *i* é:

$$mi(t+1) = mi(t) + \alpha\,(t)\, h_{bi}(t)\,[tupla - m_i(t)],$$

onde *t* é o tempo, α (t) é a taxa de aprendizado e h_{bi} (t) é o *kernel* da vizinhança centrado no neurônio vencedor. O *kernel* pode ser Gaussiano:

$$h_{bi}(t) = e^{-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}},$$

tal que r_b e r_i são as posições do neurônio *b* e *i* no mapa do SOM e σ(t) é o raio da vizinhança. Conforme a distância entre *bmu* e *i* e o tempo aumenta, $h_{bi} \rightarrow 0$. A taxa de aprendizado α(t) e o raio da vizinhança σ(t) diminuem monotomicamente com o tempo[1].

Algoritmo K-means

O Algoritmo K-*means* é baseado em centróide, ou seja, usa o centro geométrico de cada grupo para representá-lo. A Figura 2 ilustra esse conceito, onde os círculos maiores representam o centróide de três grupos existentes.

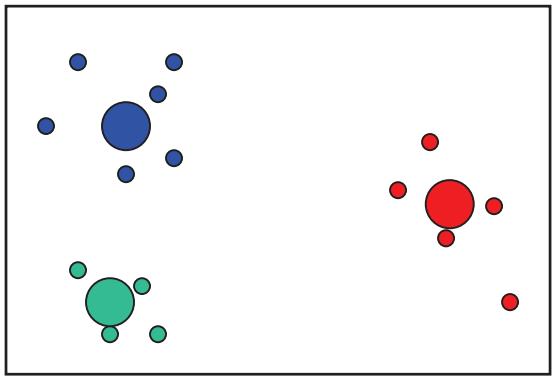

Figura 2 – Representação de grupos por centróides

[1] Para maiores detalhes sobre SOM, veja (Kohonen, 2001).

K-means é o mais conhecido dos métodos de agrupamento por particionamento, e de acordo com (Berkhin, 2002), é a ferramenta de agrupamento mais utilizada tanto em aplicações científicas quanto na indústria.

Este algoritmo, que pode ser resumido, em pseudocódigo, conforme o Quadro 1, estabelece um centróide inicial para os grupos, e vai ajustando-o aos itens de dados em um processo interativo, por meio de uma medida de distância.

```
Entrada: Conjunto de dados e um valor para k.
Saída: Conjunto de dados agrupado.
Selecione k pontos, aleatoriamente, como centróides iniciais.
repita
        Atribua cada ponto ao centróide mais próximo;
        Recalcule o centróide para cada grupo;
Até Estabilizar
```

Quadro 1: O algoritmo de agrupamentos K-means.

4. A Inter-relação entre CRM e Mineração de Dados

Uma vez que existem diferenças entre os clientes, é preciso entendê-las para otimizar as relações da empresa com os mesmos. Do conhecimento gerado pela mineração de dados advém uma maior capacidade de interação com os clientes.

Gestão das relações com clientes (CRM) é um processo que gerencia as interações entre a empresa e seus clientes.

Segundo (Thearling et al., 1999), para que um sistema de CRM seja bem-sucedido, faz-se necessário, em primeira instância, identificar os segmentos de mercado existentes nas bases de dados ou identificar os clientes potenciais, para então construir e executar campanhas que favoravelmente impactem sobre o comportamento desses indivíduos.

Identificar segmentos de mercado requer dados significativos sobre os potenciais clientes e seus hábitos de consumo. Quanto mais informação existir melhor é, pela ótica da análise. Contudo, maior é também a necessidade de uma boa ferramenta analítica. Os algoritmos de DM têm papel fundamental na busca de padrões de comportamentos de compra e similaridade entre perfis dos clientes.

Pode-se afirmar que a mineração de dados auxilia os tomadores de decisão no estabelecimento, com maior precisão, do alvo de campanhas publicitárias, bem como a melhor entender seus clientes, uma vez que se consegue extrair informações que antes não existiam nas bases de dados transacionais.

Contudo, a forma com que o resultado de um processo de mineração de dados impacta em uma empresa está muito mais voltada aos processos de negócio que nos processos dos algoritmos empregados para tal fim.

5. Metodologia

O agrupamento de dados deste trabalho é realizado de acordo com a Figura 3. Essa metodologia é proposta por (Vesanto & Alhoniemi, 2000).

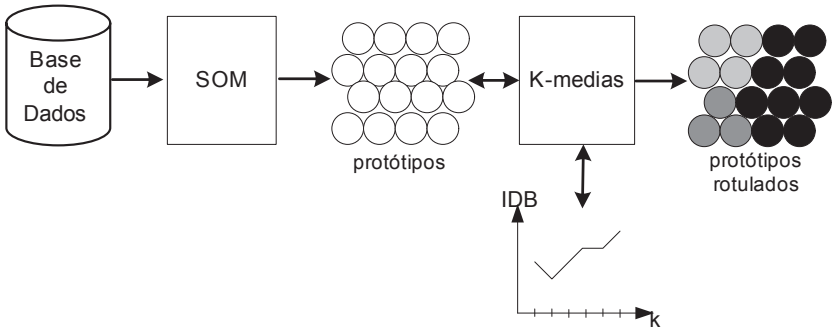

Figura 3: Esquema para segmentação do mapa SOM por K-means.

Um mapa SOM é gerado a partir da base de dados de entrada e, com base nos vetores protótipos (neurônios vencedores), é feita a segmentação pelo *K-means*, gerando como saída um mapa rotulado, onde é possível visualizar os grupos encontrados. Para a decisão sobre a qualidade do processo, um algoritmo de validação de agrupamentos, o Índice *Davies-Bouldin* (IDB) (Davies & Bouldin, 1979) é utilizado.

A base de dados *The Insurance Company*, definida em (Putten & Someren, 2000), foi escolhida para realização de experimentos por ser uma base real da área de *marketing* composta por dados de consumo e dados sócio-demográficos de 5.822 consumidores com 86 atributos numéricos. Esta base de dados está disponível no UCI (*Repository of Machine Learning Databases*), um repositório de bases de dados de acesso público, da Universidade da Califórnia, em Irvine (Newman, et al., 1998).

Os 5.822 registros contêm informações sobre 10 tipos de consumidores descritos a seguir:

1- consumidores bem-sucedidos;
2- produtores de hortifrutigranjeiros;
3- consumidores de classe média;
4- consumidores solteiros;
5- consumidores satisfeitos com sua condição social;
6- consumidores da terceira idade que gostam de viajar;
7- consumidores aposentados;
8- famílias sem crianças;
9- famílias conservadoras;
10- fazendeiros.

Pretende-se descobrir, a partir da mineração de dados, quais são os potenciais consumidores de seguros.

Após a rotulação do mapa SOM, formaram-se 13 (treze) grupos, cada um identificado pelo esquema de cores, como ilustrado à direita, na Figura 4. Esse mapa de similaridades por cores, indica quais consumidores foram agrupados em quais grupos. O mapa representa em duas dimensões, a distribuição e a topologia

dos dados. Existe também a relação entre grupos vizinhos no mapa, onde sua localização é feita de acordo com características (atributos) que esses grupos têm em comum.

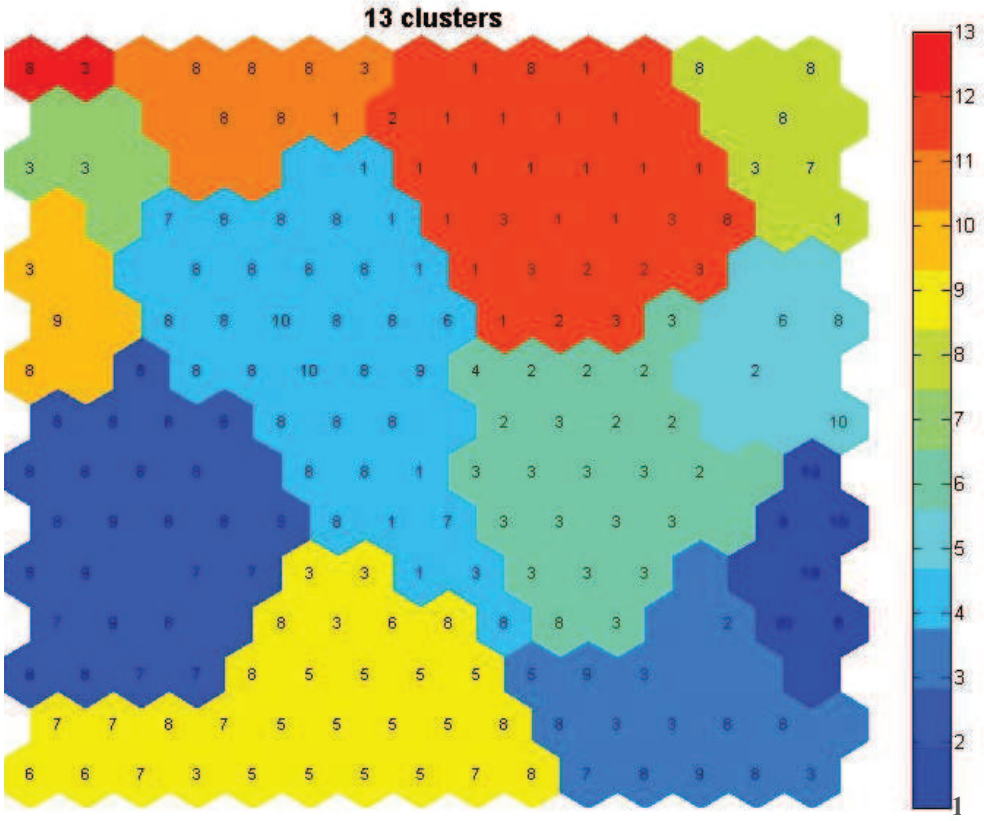

Figura 4: Mapa SOM gerado, rotulado pelas classes definidas a priori.

A base de dados está separada nos 10 grupos acima descritos. A classificação da base foi feita *a priori*, de acordo com classes definidas pela companhia de seguros (Putten & Someren, 2000). Entretanto, essa classificação a priori não considera todos os atributos da base de dados.

Desta forma, os dados da base são classificados pelos 13 novos rótulos e análises para verificar os atributos que fazem a diferença entre grupos foram realizadas. Para cada grupo são calculadas a média e a diferença desta para cada um dos registros que pertencem a tal grupo, como mostrado no Quadro 2.

```
Para cada grupo, faça
        MediaGrupo = media(registros ε grupo)
        Para cada registro, faça
                DifRegistro =  (MediaGrupo – registro)^2
        Fim para
Fim para
```

Quadro 2: Esquema para análise de variabilidade

Por fim, é feita a soma da variância por grupo e sua ordenação. Para fins de análise, são selecionadas as 10 menores variâncias, as quais devem ter as variáveis principais de cada grupo.

Para ilustrar os experimentos aqui realizados, uma análise com base nos Grupos 1, 4 e 9 é apresentada (Observe a Figura 4). O gráfico para cada um dos

grupos com os rótulos das variáveis que têm as 10 menores variâncias está ilustrado, respectivamente, nas Figuras 5, 6 e 7. Em cada figura, a identificação numérica em cada par (posição, valor) indica o atributo na base de dados.

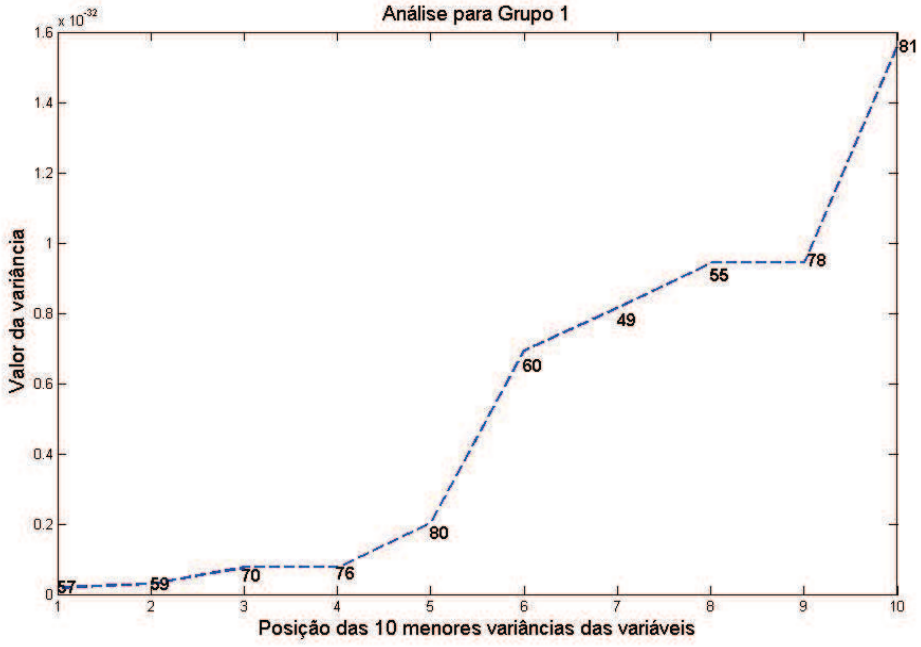

Figura 5: Gráfico com as 10 menores variâncias ordenadas para o Grupo 1.

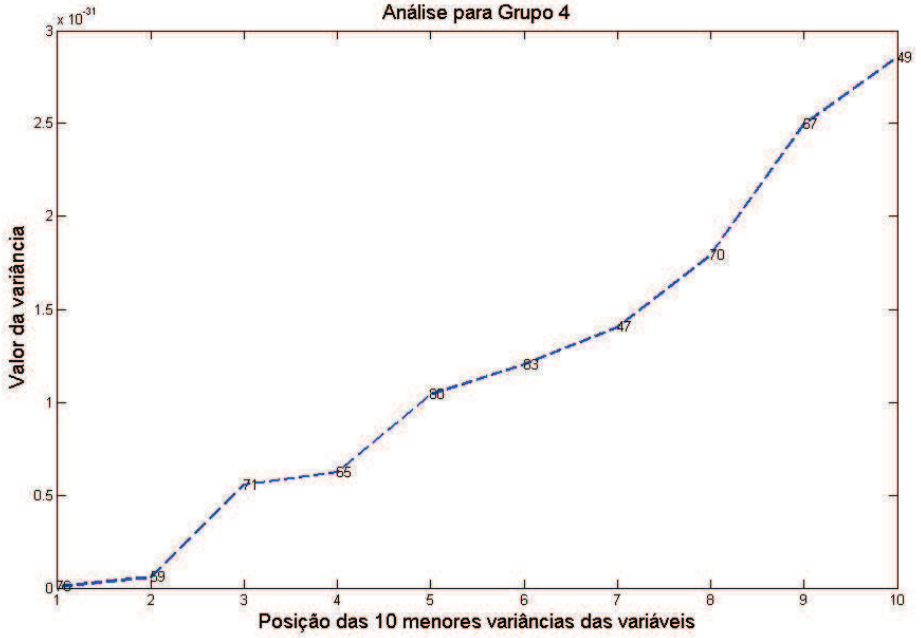

Figura 6: Gráfico com as 10 menores variâncias ordenadas para o Grupo 4.

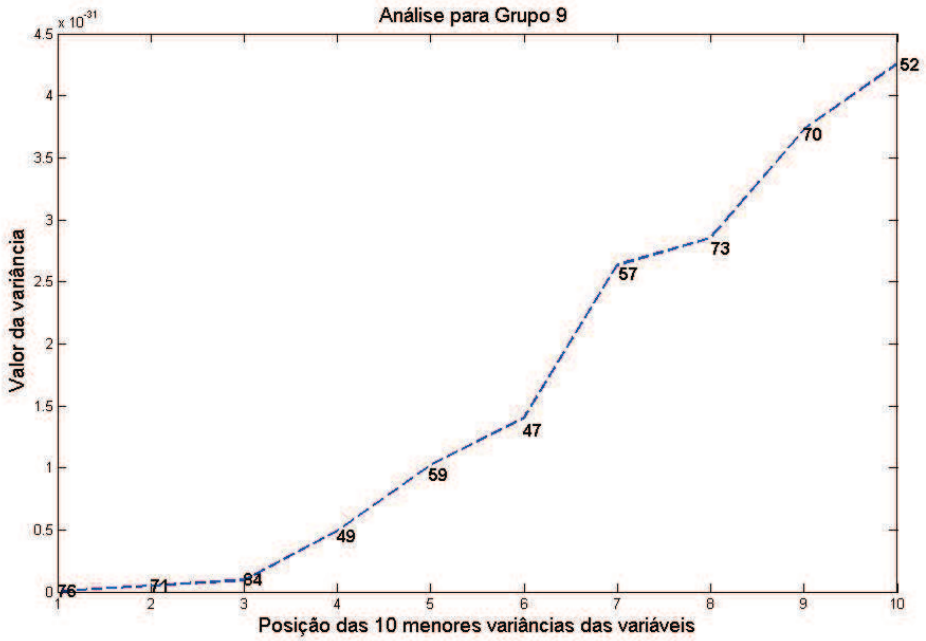

Figura 7: Gráfico com as 10 menores variâncias ordenadas para o Grupo 9.

Agora, deseja-se saber quais são os atributos que os grupos têm em comum. A Figura 8 ilustra a interseção dos grupos em análise. Esses atributos influenciam na proximidade entre os grupos, sendo, portanto, uma informação bastante útil em aplicações CRM.

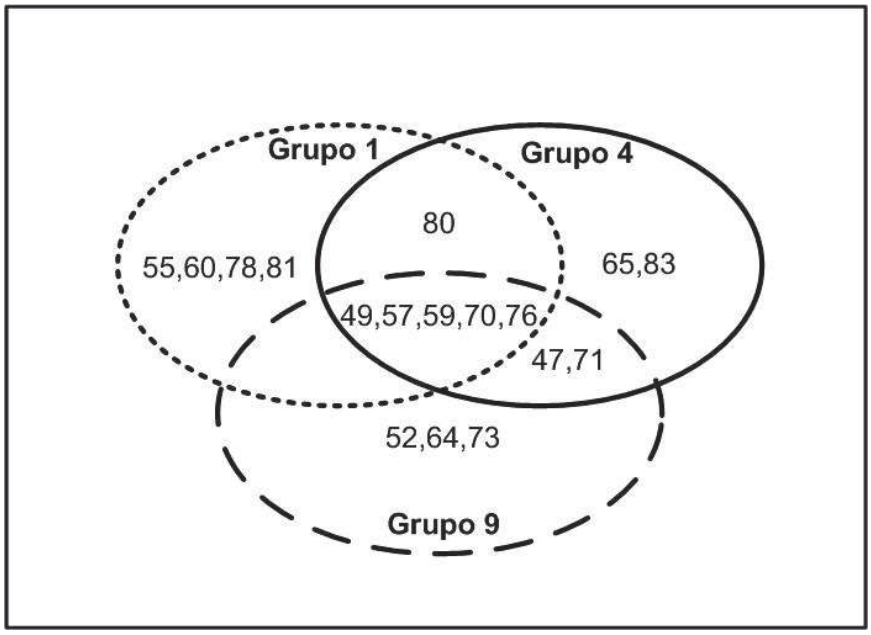

Figura 8: Atributos comuns entre os Grupos em análise.

No Quadro 3, uma descrição detalhada de cada um dos atributos comuns aos grupos em análise 1, 4 e 9 é dada. Pode-se observar que todos os atributos dizem respeito a questões de segurança.

49 - PMOTSCO Contribution motorcycle/scooter policies
57 - PGEZONG Contribution family accidents insurance policies
59 - PBRAND Contribution fire policies
70 - AMOTSCO Number of motorcycle/scooter policies
76 - ALEVEN Number of life insurances

Quadro 3: Descrição dos algoritmos comuns entre os Grupos 1, 4 e 9.

O número de registros para os três grupos analisados totaliza 2.358, o equivalente a 40,5% de toda a base de dados, podendo representar um bom nicho de mercado, o de pessoas fortemente preocupadas com segurança.

6. Conclusões

A classificação original da base de dados em 10 tipos de consumidores diferentes foi realizada considerando um número pequeno de atributos, como profissão, renda ou tamanho da família. A desconsideração dos demais atributos limitou o potencial de exploração da base de dados, impossibilitando uma análise mais profunda do perfil dos consumidores.

Desta forma, a rede SOM ao considerar todos os atributos dessa base de dados descobriu relações similaridades entre os 10 diferentes tipos de consumidores gerando novos grupos (segmentos), como visto na Figura 4. Por exemplo, para os grupos analisados a título de ilustração neste texto, observa-se que a maioria dos clientes é de famílias sem crianças, fazendeiros e consumidores aposentados e pessoas satisfeitas com sua condição social.

Este conhecimento revela um grande potencial de exploração da base que não se limita apenas na determinação de 10 tipos de consumidores diferentes, mas sim na combinação de todos os atributos e registros para determinar novos hábitos e perfis de consumo.

Este fato pode ser observado ao analisar o Quadro 3, que mostra os atributos responsáveis pela intersecção dos grupos 1, 4 e 9, demonstrando o perfil de consumidores preocupados com segurança.

Pode-se então utilizar a exploração dos resultados apresentados em marketing para direcionar de forma mais eficaz e eficiente o envio, por exemplo, de mala direta e de e-mail de marketing, direcionando mensagens apenas aos consumidores que se interessam ou que possam se interessar pelo produto ou serviço, auxiliando inclusive no desenvolvimento de técnicas adequadas ao público-alvo de telemarketing.

Pode-se, assim, enviar com maior grau de certeza de retorno, uma mala direta ou um e-mail de marketing sobre o interesse em adquirir seguro, aos consumidores pertencentes aos grupos 1, 4 e 9.

Os resultados obtidos pela mineração de dados também podem ser utilizados como fonte de informação para o CRM analítico, ajudando a transformar o conhecimento sobre o cliente em oportunidade, gerando valor à empresa e possível aumento de receita.

Não se pode esquecer que os outros 59,5% da base também poderá ser minerado abrindo novas possibilidades de extração e utilização do conhecimento gerado.

Um dos objetivos de uma campanha de marketing é identificar que perspectivas têm maior probabilidade de responder a uma oferta especial – neste trabalho, a hipótese era identificar clientes que responderiam favoravelmente a uma oferta de seguro. As respostas a essas perguntas podem em muito ajudar na fidelização dos clientes, e aumentar a taxa de resposta à campanha, o que implicará em um acréscimo nas vendas e possivelmente, no retorno sobre o investimento.

Descobrir que tipos diferentes de consumidores podem apresentar hábitos e perfis idênticos ou semelhantes de consumo pode, também, auxiliar na alteração de produtos ou serviços já existentes e no desenvolvimento de novos produtos ou serviços.

Este artigo procurou fornecer uma visão geral sobre a aplicação de técnicas de mineração de dados à área de negócios, demonstrando que é possível reforçar e ajudar a redefinir o relacionamento com os clientes a partir da análise dos dados de uma organização.

Muito embora essas vantagens cada vez mais se imponham como fator de competitividade, ainda há uma lacuna entre a integração de técnicas de mineração de dados e os sistemas de apoio à decisão, sendo todo o processo de descoberta de novos conhecimentos realizado, ainda hoje, em separado. A real integração, em soluções de software, dessas duas disciplinas, apresenta-se como uma boa oportunidade para adquirir vantagens competitivas.

Como trabalhos futuros a essa pesquisa podem-se citar:

A verificação do desempenho da metodologia de análise de influência de atributos no agrupamento de dados em outras bases de dados reais, com características diferentes da utilizada e com a necessidade da produção de novos "*insights*" para auxílio à tomada de decisão.

Investigar requisitos de integração entre as soluções de mineração de dados e os sistemas de informação existentes, realizando, efetivamente, inteligência de negócios.

Referências Bibliográficas

BERKHIN, P. Survey of Clustering Data Mining Techniques. Relatório Técnico, Accrue Software, Accrue Software, San Jose, CA, 2002.

BRETZKE, M. Marketing de Relacionamento e Competição em Tempo Real com CRM. São Paulo: Atlas, 2000.

CORMACK, R. M. A Review of Classifications. Journal of the Royal Statistical Society, Series A, N° 134, pp. 321-353, 1971.

DAVIES, D. L.; BOULDIN, D. W. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-1, N°. 2, pp. 224-227, Abril, 1979.

EVERITT, B. S.; LANDAU, S.; MORVEN, L. Cluster Analysis. Hodder Arnold Publishers, 4ª Edição, Londres, 2001.

HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. On Clustering Validation Techniques. Journal of Intelligent Information Systems. Kluwer Academic Publishers. Vol 17, N° 2-3, pp. 107-145, Hingham, MA, USA, Julho, 2001.

HAIR, J. F.; TALHAM, R. L.; ANDERSON, R. E.; BLACK, W. C. Multivariate Data Analysis, Prentice Hall, 5ª Edição, New Jersey, 1998.

JIN, H.; SHUM, W. H.; LEUNG, K. S.; WONG, M. L. Expanding self-organizing map for data visualization and cluster analysis, Information Sciences, Nro. 163, pp. 157-173, 2004.

KASKI, S.; KOHONEN, T. Exploratory data analysis by the self-organizing map: structures of welfare and poverty in the world. In: Proceedings of the third International Conference on Neural Networks in the Capital Markets. World Scientific, Singapore, pp. 498-507, 1996.

KOHONEN, T. Self-Organizing Maps, Springer Series in Information Sciences, Vol. 30, Third Edition. Springer, Berlin, Heidelberg, New York, 2001.

LATTIN, J.; CARROLL, J. D.; GREEN, P. E. Analyzing Multivariate Data. Thomson Books, Duxbury Applied Series, Cole, 2003.

NEWMAN, D. J.; HETTICH, S.; BLAKE, C. L.; MERZ, C. J. UCI – Repository of Machine Learning Databases. 1998. Disponível na URL: http://archive.ics.uci.edu/ml. Data do último acesso: 07/04/2008.

OLIVEIRA, W. J. CRM e E-Business. Florianopólis: Visual Books, 2000.

PEPPERS D. & ROGERS M. CRM Series Marketing “1to1”, São Paulo: Makron Books, 2000.

PAINI, G. T. F. Uma Aplicação de Data Mining para Personalização de Sites e-Commerce. Monografia. Bacharelado em Informática. Universidade do Oeste de Paraná, Cascavel, Março, 2003.

PUTTEN, van der P.; SOMEREN, M. van. COIL Challenge 2000: The Insurance Company Case. Published by Sentient Machine Research, Amsterdam. Also a Leiden Institute of Advanced Computer Science Technical Report 2000-09. Junho, 2000.

REZENDE, S. O. Mineração de Dados. In: Anais do 15º Congresso da Sociedade Brasileira de Computação, pp. 397-433, São Leopoldo-RS, 2005.

THEARLING, K; BERSON, A.; SMITH, S. Building Data Mining Applications for CRM. Editora McGraw-Hill, 1999.

TSCHOHL, J. A satisfação do cliente – Como alcançar a excelência através do serviço ao cliente. São Paulo: Makron Books, 1996.

VESANTO, J. Data Mining Techniques Based on the Self-Organizing Map. Dissertação de Mestrado. Universidade de Helsing, Finlândia, 1997.

VESANTO, J.; ALHONIEMI, E. Clustering of the self-organizing map. IEEE Transaction on Neural Network, Vol.11, pp. 586-600, Maio, 2000.