

CENTRO UNIVERSITÁRIO FEEVALE

SIDINEI PEREIRA GONCHOROSKI

UTILIZAÇÃO DE TÉCNICAS DE KDD
EM UM CALL CENTER ATIVO
(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo, abril de 2007

SIDINEI PEREIRA GONCHOROSKI

sidinei.pereira@gmail.com

UTILIZAÇÃO DE TÉCNICAS DE KDD

EM UM CALL CENTER ATIVO

(Título Provisório)

Centro Universitário Feevale
Instituto de Ciências Exatas e Tecnológicas
Curso de Ciência da Computação
Anteprojeto de Trabalho de Conclusão

Professor orientador: Juliano Varella de Carvalho

Novo Hamburgo, abril de 2007.

RESUMO

Para divulgação e aumento de seus índices de vendas, o mercado explora os meios de comunicação. Para acompanhar o ritmo dos negócios, as empresas encontram, através dos *Call Centers* Ativos que possuem estrutura para realizar atendimentos pelo telefone e campanhas de *telemarketing*, as soluções mais simples e acessíveis para atingir seus objetivos. Com o armazenamento de dados de suas operações, puderam criar listas de clientes que são enviadas aos *Call Centers* para realizar as campanhas. No entanto, as empresas do ramo de atendimento possuem certas limitações e dificuldades de explorar e aproveitar o conhecimento nas bases de dados, já que podem não conhecer todas as informações sobre o melhor perfil de vendas ou ter dificuldade em criar uma estratégia para aumentar o desempenho do *telemarketing*. Através do processo de descoberta de conhecimento em banco de dados (KDD) é possível identificar regras e padrões válidos aplicando as técnicas e algoritmos da tecnologia de mineração de dados. Sendo assim, esse trabalho tem como objetivo estudar e utilizar, através da ferramenta *Weka*, técnicas e algoritmos de descoberta de conhecimento, além de realizar o tratamento dos dados e visualização do resultado, proporcionando ao *Call Center* Ativo, traçar estratégias e aumentar suas vendas diminuindo a quantidade de contatos telefônicos.

Palavras-chave: Descoberta de Conhecimento. Técnicas de Mineração de Dados. *Weka*. Algoritmos de Mineração de Dados. *Call Center* Ativo.

SUMÁRIO

MOTIVAÇÃO	4
OBJETIVOS	8
METODOLOGIA	9
CRONOGRAMA	10
BIBLIOGRAFIA	11

MOTIVAÇÃO

O mundo dos negócios sempre explorou os meios de comunicação para aumentar seus índices de vendas e divulgar seus produtos a um público maior. Dentre os meios de comunicação existentes, o telefone continua sendo uma opção para se comunicar com as pessoas. As empresas cada vez mais preocupadas em acompanhar o ritmo constante da concorrência e dos desafios do mercado utilizam o telefone para comercializar seus produtos e aumentar seu público. Um *Call Center* Ativo é um tipo de empresa dotada de toda a estrutura física, humana e tecnológica capaz de realizar atendimentos através do telefone e promover campanhas de *Telemarketing*¹. É através dos *Call Centers* Ativos que as empresas encontram as soluções mais simples e acessíveis de atingir seus objetivos e explorar um mercado mais abrangente.

Vários setores da sociedade passaram, a partir do crescimento da computação e dos bancos de dados (BD), a armazenar suas operações e produções. Estes dados formaram grandes bases que não possuem um tratamento específico e que não disponibilizam aos profissionais, muitas vezes, informações e conhecimento que possam ser utilizados em seu trabalho de modo eficiente. Através desses dados as empresas puderam criar grandes listas de *prospects*², chamadas de *mailing*. O trabalho do *Call Center* Ativo entra em cena no momento que essas empresas decidem realizar uma campanha de *Telemarketing* para recuperar clientes, aumentar vendas ou adquirir novos clientes.

Ao acompanhar as empresas do ramo é possível notar que existem certas limitações e dificuldades de explorar e aproveitar o conhecimento nas bases de dados geradas e atualizadas dentro das empresas de atendimento. Os operadores realizam os contatos corrigindo os dados dos clientes e preenchendo as propostas com os dados ainda não conhecidos.

¹ “A utilização planejada de recursos de telecomunicações e informática como forma de se obter lucro direto ou indireto, através da satisfação do mercado consumidor de qualquer bem ou serviço.” (DANTAS, 1994, p. 47).

² Pessoa que é alvo do atendimento e candidata a adquirir o produto ou serviço oferecido.

As estratégias geradas a partir da experiência dos responsáveis podem falhar, neste caso é preciso realizar uma nova análise até que se descubra, por exemplo, que uma característica não considerada importante antes é o diferencial para o sucesso da venda em questão. Quando não existem indicações sobre quais são as características que classificam um cliente como potencial, realizar a análise pode acarretar perda de tempo e mais demora a se realizar uma venda, já que os clientes serão atendidos aleatoriamente.

A descoberta de conhecimento em banco de dados, do termo em inglês *Knowledge Database Discovery* (KDD) representa, “O processo não-trivial de identificar válidos, novos, potencialmente utilizáveis e por fim, padrões compreensíveis dentro dos dados.” (tradução nossa) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996, p. 6). Estas características significam:

Dados: Conjunto de informações que são armazenados em Sistemas Gerenciadores de Banco de Dados (SGBD), em vários registros, com os mesmos atributos e que representam um tipo de coleção. Ex: Todas as vendas de um produto.

Padrão: Conforme (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), padrão é o grupo de itens que tem maior incidência em um conjunto de dados. Por exemplo, analisando um grupo de pessoas para descobrir o tipo de cliente que adquire capitalizações, tendo posse dos valores de rendimento e da quantidade de crediários que cada um possui, chega-se a conclusão que a segunda não influencia na aquisição e somente os que têm rendimentos superiores a certo limiar acabam as adquirindo, independente dos crediários.

Processo: Se trata do processo envolvido na descoberta do conhecimento, que passa pela preparação dos dados, garimpagem e análise. As etapas do processo de KDD, de acordo com (FELDENS, apud CARVALHO; SAMPAIO; MONGIOVI, 1999), são: seleção, pré-processamento, transformação, garimpagem, análise e assimilação.

Válido: São aqueles padrões considerados válidos e interessantes ao objetivo traçado.

Novo: Representa todo conhecimento adquirido e que não estava previsto ou não poderia ser deduzido através de hipótese. Um padrão gerado por hipótese não é considerado novo já que poderia ser comprovado através, por exemplo, de estatística. Um destes exemplos é a Análise Exploratória de Dados (AED) do termo em inglês *Exploratory data analysis*, que é uma técnica da área de estatística lançada por (TUKEY, 1977) que visa obter o máximo de informações ocultas entre os dados (HOAGLIN; MOSTELLER; TUKEY, 1992).

Potencialmente Utilizável: Alguns dos padrões encontrados podem acabar não sendo úteis. Para que a descoberta de conhecimento seja relevante, é preciso que o resultado não represente algo totalmente sem sentido para o negócio. Padrões que sejam muito amplos ou que tenham pouca variação em relação a outros acabam não tendo muita utilidade.

Compreensível: É poder criar padrões que possam ser entendidos pelos seres humanos e acrescentem conhecimento útil para a tomada de decisões.

A tarefa de descobrir conhecimento não é simples. Os dados recolhidos e armazenados não são preparados de forma que a qualquer momento sejam analisados para que mostrem ao usuário os relacionamentos e padrões. Os dados acabam vindo de várias fontes e acaba necessitando de um tratamento, um pré-processamento para se definir informações importantes e corrigir possíveis imperfeições, já que tiveram origem em outros locais não sendo fruto de técnicas de KDD, conforme WIEDERHOLD (1996). Estes dados podem não estar completos.

A mineração de dados é uma área interdisciplinar que agrupa estatística, BD e inteligência artificial munida de vários algoritmos (FREITAS, 1998) e pode ser definida como:

O termo *data mining* é normalmente utilizado pela comunidade de estatísticos, analistas de dados e os MIS (*Management Information Systems*). [...] a visão de que KDD é todo processo de descoberta de conhecimento útil em dados enquanto *data mining* se refere à aplicação de algoritmos para extração de padrões em dados sem os passos adicionais do processo de KDD. (tradução nossa) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996, p. 3-4).

A KDD utiliza os algoritmos e técnicas de Mineração de Dados para adquirir informações importantes em bases que não estão visíveis ou identificáveis pelos usuários e que podem ser úteis na tomada de decisão. KDD também significa extrair informações de grandes bases de dados sem possuir hipóteses previamente criadas, conforme (CABENA et al, 1997). Através da Mineração de dados e seus algoritmos é possível descobrir regras de associação, classificação e *Clustering* (FAYYAD et al, 1996) (FREITAS, 1998). São exemplos de algoritmos para mineração de dados: *LargeKItemSets* (AGRAWAL; IMIELINSKI; SWAMI, 1993), *Apriori* (AGRAWAL; SRIKANT, 1994), *AprioriTid* (AGRAWAL et al, 1993), *Partition* (SAVASERE; OMIECINSKY; NAVATHE, 1995) e *Multiple Level* (ML-T2L1) (SRIKANT; AGRAWAL, 1996).

Com base nos conceitos citados, a motivação do trabalho é desenvolver uma ferramenta que facilite o tratamento e pré-processamento dos dados, que utilize a ferramenta

*Weka*³, *open source e freeware* para processar a base de dados com alguns dos algoritmos de mineração e que voltado para um *Call Center* ativo, consiga descobrir padrões e regras que possam auxiliar na seleção dos *prospects* das listas de *mailing* que serão trabalhados, a fim de ter um índice de vendas melhor, com um número menor de tentativas com os clientes. Desta maneira, espera-se melhorar o desempenho e aproveitar eficientemente os recursos e informações disponíveis, mas não utilizadas até o momento, por não existir uma ferramenta especializada para tal.

³ WEKA 3: Data Mining Software in Java. Nova Zelândia. Universidade de Waikato, 2007. Apresenta todas as características do projeto e do software Weka. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/index.html>>. Acesso em: 28 mar. 2007.

OBJETIVOS

Objetivo geral

Explorar uma ferramenta de descoberta de conhecimento para auxiliar o departamento de análise de dados e *mailing* de um *Call Center* ativo, a fim de realizar uma seleção mais eficiente dos *prospects*, aumentar as vendas e reduzir a quantidade de contatos telefônicos com os clientes.

Objetivos específicos

- Descrever o conceito de KDD e de mineração de dados, apresentando fundamentação teórica sobre o assunto;
- Estudar e detalhar algoritmos de mineração de dados;
- Estudar a ferramenta *Weka*, como ela utiliza os algoritmos e apresenta os resultados obtidos;
- Definir e modelar a estrutura escolhida para trabalhar com os dados, na própria base de dados ou externamente.
- Desenvolver um *software* para auxiliar na alimentação dos dados, onde serão aplicadas técnicas de mineração;
- Desenvolver um *software* para visualização dos dados resultantes da aplicação das técnicas de mineração de dados;
- Realizar uma bateria de testes e fazer uma discussão sobre as melhores técnicas usadas, características das mesmas, seus dados e resultados;
- Comparar o desempenho das listas trabalhadas após a descoberta de conhecimento na base de dados com as listas sendo trabalhadas sem a aplicação das técnicas propostas;
- Apresentar o relatório final com os resultados referentes à solução proposta.

METODOLOGIA

As atividades a serem realizadas para a conclusão deste trabalho estão organizadas da seguinte maneira:

1. **Elaboração do anteprojeto** e a definição do escopo do trabalho com auxílio do professor orientador;
2. **Pesquisa na bibliografia existente** em livros, artigos, revistas e páginas da internet para o desenvolvimento da fundamentação teórica para o trabalho;
3. **Análise e estudo da Ferramenta *Weka***, como trabalham os algoritmos, testes iniciais e como são apresentados os resultados;
4. **Redação do Trabalho de Conclusão I;**
5. **Revisão e Entrega do Trabalho de Conclusão I;**
6. **Definição e modelagem da estrutura escolhida** para o tratamento dos dados;
7. **Desenvolvimento do software** para o tratamento dos dados;
8. **Desenvolvimento da solução** visualização das análises do *software Weka*;
9. **Realização de testes e comparativos** de desempenho entre as atividades antes e após a utilização de KDD com mineração de dados;
10. **Conclusão.** Desenvolvimento do relatório apresentando os resultados;
11. **Redação do Trabalho de Conclusão II;**
12. **Revisão e Entrega do Trabalho de Conclusão II;**
13. **Banca.** Defesa da monografia perante a banca avaliadora.

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Mar/07	Abr/07	Mai/07	Jun/07
1	■			
2	■	■	■	■
3		■		
4			■	■
5				■

Trabalho de Conclusão II

Etapa	Meses				
	Ago/07	Set/07	Out/07	Nov/07	Dez/07
6	■				
7		■	■	■	
8		■	■	■	
9			■	■	
10				■	
11		■	■	■	
12				■	
13					■

BIBLIOGRAFIA

- AGRAWAL, Rakesh; IMIELINSKI, Tomasz; SWAMI, Arun. Mining Association Rules between Sets of Items in Large Databases. In: SIGMOD 5/93. Washington, USA. 1993. p. 207-216.
- AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. Fast Algorithms for Mining Association Rules. In: 20th VLDB Conference. Santiago, Chile. 1994. p. 487-498.
- CABENA, Peter; et al, **Discovering Data Mining from Concept to Implementation**. New Jersey-USA: Prentice Hall PTR. 1997. 193 p.
- CARVALHO, Juliano V.; SAMPAIO Marcus C.; MONGIOVI, Giuseppe. Utilização de Técnicas de “Data Mining” para o Reconhecimento de Caracteres Manuscritos. In: Simpósio Brasileiro de Banco de Dados. Florianópolis, Brasil. 1999. 15 p.
- CASTELLIANO, Tânia; FERREIRA, Norberto. **Telemarketing 100%**. Rio de Janeiro: Record. 1998. 204 p.
- DANTAS, Edmundo Brandão. **Telemarketing: A chamada para o futuro**. 2. ed. São Paulo: Atlas. 1994. 206 p.
- ELDER VI, John F; PREGIBON, Daryl. A Statistical Perpectiv on Knowledge Discovery in Databases. In: **ADVANCES in knowledge discovery and data mining**. Cambridge, Mass:AAAI/MIT Press. 1996. p. 83-113.
- FAYYAD, Usama M. (Org.) et al. **ADVANCES in knowledge discovery and data mining**. 1. ed. Califórnia: American Association for Artificial Inteligence. 1996. p. 1-120.
- FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery: An overview. In: **ADVANCES in knowledge discovery and data mining**. FAYYAD et al. G. Cambridge-Mass:AAAI/MIT Press. 1996. p. 1-27.
- FRAWLEY, William J.; PIATETSKY-SHAPIRO, Gregory; MATHEUS, Christopher J. Knowledge Discovery in Databases: An overview. In: **AI Magazine**. American Association for Artificial Inteligence. Menlo Park, CA, USA, 1992. p. 57 – 70.
- FREITAS, Alex A. Data Mining, In: XIII Simpósio Brasileiro de Banco de Dados. Maringá, Brasil. 1998. 104 p.
- HARRISON, Thomas H. **Intranet Data Warehouse: Ferramentas e técnicas para a utilização do Data Warehouse na intranet**. São Paulo: Berkeley. 1998. 358 p.
- HOAGLIN, David C.; MOSTELLER, Frederick; TUKEY, John W. Análise exploratória de dados: técnicas robustas: um guia. Lisboa: John Wiley, 1992. 446 p.

- NEMATI, Hamid R.; BARKO, Christopher D (editor). **Organizational data mining: leveraging enterprise data resources for optimal performance**. Hershey, PA: Idea Group. 2003. 388 p.
- PALMIERI, Antônio; BUAIRIDE, Ana M. **TeleMarketing**. São Paulo: SEBRAE-SP. 1997. 44 p.
- PRODANOV, Cleber C. **Manual de Metodologia Científica**. 3. ed. Novo Hamburgo: Editora Feevale. 2006. 77 p.
- SAVASERE, Ashok; OMIECINSKY, Edward; NAVATHE, Shamkant. An Efficient Algorithm for Mining Association Rules in Large Databases. In: 21st VLDB Conference. Zurich, Suíça. 1995. p. 432-444.
- SRIKANT, Ramakrishnan; AGRAWAL, Rakesh. Mining Quantitative Association Rules in Large Relational Tables, In: SIGMOD 6/96. Montreal, Canadá. 1996. p. 1-12
- TUKEY, John W. **Exploratory Data Analysis**. Reading: Addison-Wesley. 1977. 688 p.
- WEKA 3: Data Mining Software in Java. Nova Zelândia. Universidade de Waikato, 2007. Apresenta todas as características do projeto e do software Weka. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/index.html>>. Acesso em: 28 mar. 2007.
- WIEDERHOLD, Gio. On the barriers and Future of knowledge discovery. In: **ADVANCES in knowledge discovery and data mining**. Cambridge, Mass:AAAI/MIT Press, 1996. p. VI-XI.
- WITTEN, Ian H.; FRANK, Eibe. **Data Mining: Practical machine learning tools and techniques**. 2. ed. San Francisco: Morgan Kaufmann. 2005. 525 p.