



Livia Fonseca Fracalanza

**Mineração de Dados voltada para Recomendação no Âmbito de
Marketing de Relacionamento**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para
obtenção do título de Mestre pelo Programa de Pós-
Graduação em Informática da PUC-Rio.

Orientador: Prof. Marco Antonio Casanova

Rio de Janeiro
Maio de 2009



Livia Fonseca Fracalanza

Mineração de Dados voltada para Recomendação no Âmbito de Marketing de Relacionamento

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Marco Antonio Casanova
Orientador
PUC-Rio

Prof Antonio Luz Furtado
PUC-Rio

Prof^a. Karin Koogan Breitman
PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico - PUC-Rio

Rio de Janeiro, 02 de abril de 2009.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Livia Fonseca Fracalanza

Bacharel em Informática com ênfase em Análise de Sistemas graduada pela Pontifícia Universidade Católica do Rio de Janeiro em julho de 2006. Atualmente trabalha gerenciando projetos e participa no desenvolvimento de softwares de *Business Intelligence*.

Ficha Catalográfica

Fracalanza, Livia Fonseca

Mineração de dados voltada para recomendação no âmbito de marketing de relacionamento / Livia Fonseca Fracalanza ; orientador: Marco Antonio Casanova. – 2009.

59 f. : il. (col.) ; 30 cm

Dissertação (Mestrado em Informática)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.

Inclui bibliografia

1. Informática – Teses. 2. Market basket analysis. 3. Mineração de dados. 4. Problema do clique máximo. I. Casanova, Marco Antonio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

Agradeço à minha família pelo apoio de sempre, por torcerem por mim e pelo amor incondicional.

Todos os meus amigos foram muito importantes nessa fase da minha vida, mas em especial gostaria de agradecer ao João Rafael Peixoto Leite, amigo desde graduação, pois sem seu incentivo, paciência e carinho, este trabalho não teria começado e hoje o título de mestre ainda seria um sonho.

Agradeço ao meu orientador, Marco Antonio Casanova, por entender o quão difícil é estar no mercado de trabalho louco e acelerado e ao mesmo tempo decidir fazer mestrado. Obrigada por ter acreditado na minha capacidade, pela paciência e pela orientação.

Por último, porém não menos importante, obrigada a todos os meus companheiros da Stone Age Tech pelo apoio moral e incentivo.

Resumo

Fracalanza, Livia Fonseca; Casanova, Marco Antonio; **Mineração de Dados voltada para Recomendação no Âmbito de Marketing de Relacionamento**. Rio de Janeiro, 2009. 59p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Cross-selling é uma estratégia de vendas de produtos baseada em uma análise das compras passadas de um cliente ou nas compras passadas de outros clientes com o mesmo perfil. O algoritmo mais conhecido para análise da cesta de compras de um cliente é conhecido por *market basket analysis*. Este trabalho aborda a descoberta de padrões sequenciais em grandes bases de dados e tem por objetivo apresentar um algoritmo eficiente que transforma o problema da cesta de compras em um problema de clique máximo. Primeiramente, os dados de entrada são transformados em um grafo e o problema da descoberta do *clique* máximo é resolvido revelando as relações mais recorrentes entre os itens em questão. Os experimentos apresentados na dissertação demonstram a eficiência do algoritmo em grandes volumes de dados.

Palavras-chave

1. Informática – Teses; 2. Market basket analysis, 3. mineração de dados 4. problema do clique máximo

Abstract

Fracalanza, Livia Fonseca; Casanova, Marco Antonio;. **Recommendation based on Data Mining for Relationship Marketing.** Rio de Janeiro, 2009. 59p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Cross-selling is a strategy to recommend products to customers based on their past purchases or the purchases of other customers with the same profile. The best known algorithm for the analysis of a client shopping basket is known in the literature as market basket analysis. This dissertation discusses the discovery of sequential patterns in large databases and aims at implementing an efficient algorithm that transforms the shopping cart problem into a maximum *clique* problem. First, input data is transformed into a graph and maximum cliques are detected to discover the most frequent relationship between the items on the transaction. The dissertation also includes experiments that evaluate the efficiency of the algorithm for large data volumes.

Keywords

1. information science – thesis; 2. Market basket analysis; 3. data mining; 4. maximum clique problem

Sumário

1 Introdução	10
1.1. Motivação	11
1.2. Objetivos	11
1.3. Trabalhos Relacionados	12
1.4. Organização do Trabalho	13
2 Mineração de Dados	15
2.1. Processo de Busca de Conhecimento	15
2.2. Tipos de Predição	17
2.2.1. Classificação	17
2.2.2. Segmentação	18
2.2.3. Regressão	18
2.2.4. Associação	18
2.3. Técnicas de Mineração de Dados	18
2.3.1. Algoritmo Genético	19
2.3.2. Redes Neurais	19
2.3.3. Árvores de Decisão	19
2.3.4. Regras de Associação	19
2.3.5. Análise de Vizinhaça	20
2.4. Aplicações	20
3 <i>Market Basket Analysis</i> - MBA	21
3.1. Marketing de Relacionamento	22
3.2. Regras de Associação	23
3.3. Algoritmos de MBA	24
3.3.1. Artificial Immune System	25
3.3.2. Set-oriented Mining	25
3.3.3. Algoritmo Apriori	26
3.4. Implementação baseada em Grafo	27
3.4.1. Conceitos de Grafos	27
3.4.2. Montando a Matriz de Adjacências	29

3.4.3. Buscando a <i>Clique</i> de Tamanho Máximo	29
4 Testes e Resultados	31
4.1. Descrição do Ambiente de Testes – Protótipo	31
4.2. Testes	34
5 Conclusão e Trabalhos Futuros	40
5.1. Conclusões	40
5.2. Trabalhos Futuros	41
6 Referências	43
7 APÊNDICE A – Código da Implementação	45

Lista de figuras

Figura 1 Etapas para descoberta de conhecimento	16
Figura 2 Iteratividade entre Funcionalidades, Técnicas e Algoritmos de Mineração de Dados	17
Figura 3 Exemplo de iteração do algoritmo AIS	25
Figura 4 Exemplo de iteração do algoritmo SETM	26
Figura 5 Exemplo de iteração do algoritmo Apriori	27
Figura 6 Exemplo de <i>clique</i> máxima	28
Figura 7 Protótipo: escolha do conector ODBC	32
Figura 8 Protótipo: Autenticação no ODBC	32
Figura 9 Protótipo: Escolha da base de dados	33
Figura 10 Protótipo: Inclusão de filtro na consulta	33
Figura 11 Protótipo: Exibição das informações relevantes	34