



**Roberto Pereira Cavalcante**

**Filtragem Colaborativa aplicada a  
Publicidade Direcionada**

**Dissertação de Mestrado**

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

Orientador: Prof. Ruy Luiz Milidiú

Rio de Janeiro

Abril de 2008



**Roberto Pereira Cavalcante**

## **Filtragem Colaborativa aplicada a Publicidade Direcionada**

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. Ruy Luiz Milidiú**

Orientador  
Departamento de Informática – PUC-Rio

**Prof. Daniel Schwabe**

Departamento de Informática – PUC-Rio

**Prof. Marcus Vinicius Soledade Poggi de Aragão**

Departamento de Informática – PUC-Rio

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro Técnico Científico – PUC-Rio

Rio de Janeiro, 3 de abril de 2008

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

### **Roberto Pereira Cavalcante**

Recebeu seu título de Bacharel em Informática pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) em 2006. Sua experiência acadêmica inclui atuação como monitor na disciplina de Projeto e Análise de Algoritmos. Desde 2004, realiza atividades relacionadas com pesquisa e desenvolvimento no LEARN (Laboratório de Engenharia de Algoritmos e Redes Neurais) da PUC-Rio.

#### Ficha Catalográfica

Cavalcante, Roberto Pereira

Filtragem Colaborativa aplicada a Publicidade Direcionada / Roberto Pereira Cavalcante; orientador: Ruy Luiz Milidiú. – Rio de Janeiro: PUC, Departamento de Informática, 2008.

62 f. : il. (col.) ; 29,7 cm

1. Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Teses. 2. Filtragem Colaborativa. 3. Aprendizado de Máquina. 4. Publicidade Direcionada. 5. World Wide Web. I. Milidiú, Ruy Luiz. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Aos meus pais e a Ana Carolina.

## Agradecimentos

Em primeiro lugar, agradeço aos meus pais, Raimundo (*in memoriam*) e Luzia, pela constante dedicação, incentivo e apoio em todas as áreas de minha vida. Devo a eles tudo o que sou e o pouco que sei.

À minha namorada, Ana Carolina, por ser a pessoa mais especial que já participou da minha vida, pelo amor e carinho que ela me dedica a cada dia, e pela paciência e compreensão com minha ausência durante a produção deste trabalho. E como se não bastasse tudo isso, também pelo grande apoio emocional durante a confecção e pela ajuda na revisão deste texto.

Ao meu orientador, Prof. Ruy Luiz Milidiú, por não apenas exercer tão bem o papel de orientador acadêmico, mas também ocupar o papel de grande conselheiro de minha vida profissional, e pela amizade construída ao longo desses três anos e meio de convivência.

Aos Profs. Daniel Schwabe e Marcus Poggi de Aragão, por aceitarem o convite para participar da banca examinadora deste trabalho.

Aos antigos e atuais colegas do LEARN, por criarem um excelente ambiente de trabalho, permeado de problemas interessantes e discussões inteligentes: Alberto Sardinha, Patrick, Marco (Caveira), Pedro, Fred Liporace, Raúl, Cícero, Julio,

Fred Pessoa (Frodo), Fred Barros, Críston, Tulio, Diogo, Leonardo, Prof. Laber, Eduardo e Caio.

Aos professores e funcionários do Departamento de Informática, por comporem essa instituição de ensino de excelência acadêmica, onde pude crescer tanto pessoal e profissionalmente.

Aos amigos, por contribuírem para que esses dois anos fossem um pouco menos estressantes. Cito aqui alguns deles, sob o risco de deixar alguém injustamente fora da lista: Saymon, Adem, Vitor, Andrew, Bruno, Estela, Tati, Roberta, Fábio, Daniel Fleischman e Daniel Rocha.

Ao CNPq e à FAPERJ, pelo apoio financeiro durante o mestrado, sem o qual a conclusão deste trabalho seria inviável.

A todos vocês, muito obrigado!

## Resumo

Cavalcante, Roberto Pereira; Milidiú, Ruy Luiz. **Filtragem Colaborativa aplicada a Publicidade Direcionada**. Rio de Janeiro, 2008. 62p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O surgimento da World Wide Web representou uma nova oportunidade de publicidade, disponível para qualquer empresa: a possibilidade de exposição global para uma grande audiência a um custo extremamente pequeno. Como consequência disso, surgiu toda uma nova indústria oferecendo serviços relacionados à publicidade de busca, na qual uma empresa anunciante paga por uma posição de destaque em listas de anúncios. A fim de manter a credibilidade e a participação de mercado do serviço que os veicula – por exemplo, uma máquina de busca –, os anúncios devem ser exibidos apenas para os usuários que se interessem por eles, no que se chama de Publicidade Direcionada. Em virtude disso, surge a necessidade de se utilizar um sistema de recomendação que seja capaz de escolher que anúncios exibir para quais usuários. Nos sistemas de recomendação baseados em filtragem colaborativa, as preferências de outros usuários são utilizadas como atributos para um sistema de aprendizado, pois estas podem ser bastante detalhadas, gerando recomendações não só para os itens mais populares como também para nichos de itens. Neste trabalho, é desenvolvido um sistema de recomendação de anúncios que aplica Filtragem Colaborativa baseada em fatoração de matrizes ao problema de predição do *Click-Through Rate*, uma métrica em Publicidade Direcionada que expressa a relevância de um anúncio para os usuários que buscam por uma determinada palavra-chave. A fim de validar o método proposto de predição do *Click-Through Rate*, realizamos vários experimentos em um conjunto de dados sintéticos. Adicionalmente, o trabalho contribui para o projeto do LearnAds, um *framework* de recomendação de anúncios baseado em Aprendizado de Máquina.

## Palavras-chave

Filtragem Colaborativa, Aprendizado de Máquina, Publicidade Direcionada, World Wide Web

## Abstract

Cavalcante, Roberto Pereira; Milidiú, Ruy Luiz (Advisor). **Collaborative Filtering applied to Targeted Advertising**. Rio de Janeiro, 2008. 62p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The emergence of the World Wide Web represented a new advertising opportunity available to any company: the possibility of global exposure to a large audience at a very small cost. As a result, a whole new industry has emerged by offering services related to search advertising, in which an advertiser pays for a prominent position in lists of ads. In order to maintain the credibility and market share of the service that conveys them – for example, a search engine –, such ads must be displayed only to users who are interested in them, on what is called Targeted Advertising. Therefore, those services need to use a recommendation system that can choose which ads show to which users. Recommendation systems based on collaborative filtering use the preferences of other users as features to a learning system, since such preferences can be quite detailed, generating recommendations not only for the most popular items but also to item niches. In this work, we develop an ads recommendation system that applies Collaborative Filtering based on matrix factorization to the problem of predicting the Click-Through Rate, a Targeted Advertising metric that expresses the relevance of a particular ad for the users searching for a specific keyword. In order to validate the proposed method of Click-Through Rate prediction, we carry out several experiments on a synthetic data set. Additionally, the work contributes to the design of LearnAds, a framework for ads recommendation systems based on Machine Learning.

## Keywords

Collaborative Filtering, Machine Learning, Targeted Advertising, World Wide Web



# Sumário

1	Introdução	11
1.1.	Motivação	11
1.2.	Trabalhos Relacionados	14
1.3.	Objetivo	19
1.4.	Organização do Trabalho	20
2	Desenvolvimento da Publicidade de Busca	21
2.1.	Leilão de Primeiro Preço Generalizado	21
2.2.	Leilão de Segundo Preço Generalizado e o <i>Click-Through Rate</i>	24
3	Sistemas de Recomendação baseados em Filtragem Colaborativa	28
3.1.	Recomendação Colaborativa para Predição do CTR	31
3.2.	Fatoração de Matrizes	33
4	LearnAds: um <i>Framework</i> de Recomendação de Anúncios	36
4.1.	Requisitos Funcionais	36
4.2.	Casos de Uso	38
4.3.	Diagramas de Seqüência	40
4.4.	Diagramas de Classes	44
4.4.1.	Camada de Modelo	44
4.4.2.	Camada de Controle	46
5	Experimentos	48
5.1.	Conjunto de Dados	48
5.2.	Metodologia e Métricas de Avaliação	51
5.3.	Resultados	53
6	Considerações Finais	57
6.1.	Contribuições	57
6.2.	Trabalhos Futuros	58

