



**Tulio Jorge de Alcântara Neves de Souza Anibolet**

**Boosting para  
Sistemas de Recomendação**

**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. Ruy Luiz Milidiú

Rio de Janeiro

Julho de 2008



**Tulio Jorge de Alcântara Neves de Souza Anibolet**

**Boosting para  
Sistemas de Recomendação**

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(a) Setorial do Centro Técnico Científico – PUC-Rio

Rio de Janeiro, 30 de julho de 2008

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

### **Tulio Jorge de Alcântara Neves de Souza Anibolet**

Recebeu seu título de Bacharel em Informática pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) em 2004. Desde 2007, integra, na mesma instituição, a equipe de pesquisade Sistemas de Recomendação, do Laboratório de Engenharia de Algoritmos e Redes Neurais (LEARN). Possui 4 anos de experiência profissional atuando como analista de sistemas e trabalha no Banco Nacional de Desenvolvimento Econômico e Social (BNDES).

#### Ficha Catalográfica

Anibolet, Tulio Jorge de Alcântara Neves de Souza

*Boosting* para Sistemas de Recomendação / Tulio Jorge de Alcântara Neves de Souza Anibolet; orientador: Ruy Luiz Milidiú. – Rio de Janeiro: PUC, Departamento de Informática, 2008.

60 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. Sistemas de Recomendação. 5. *Boosting*. I. Milidiú, Ruy Luiz. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

Às minhas queridas tias Márcia e Iara.

## Agradecimentos

O maior agradecimento deste trabalho vai para as minhas tias pelo carinho, dedicação e apoio.

Agradeço também ao meu orientador, Prof. Ruy Luiz Milidiú, pelo incentivo, liderança e sugestão do tema deste trabalho.

O agradecimento seguinte vai para os professores Ivan Mathias Filho e Sérgio Colcher, pela confiança em mim depositada ao me recomendarem para o mestrado e para os professores Daniel Schwabe e Marcus Vinícius Soledade Poggi de Aragão, por aceitarem participar da banca examinadora.

Não posso me esquecer dos meus amigos do LEARN e todos os outros conquistados na PUC-Rio, aos quais agradeço por consolidarem um ambiente repleto de idéias criativas: Roberto Cavalcante, Diogo Mendonça, Leonardo Majowka, Cícero Nogueira, Júlio Duarte, Sérgio Ciglione, Andrew Diniz, Bruno Siqueira, Rodnei Couto, Márcio Aguiar e Eduardo Fonseca.

Agradeço, ainda, aos meus grandes amigos, Allan Miranda, Shirley Arruda e José Carlos Seraphim Jr. pelos conselhos nas horas adequadas, energia e estímulo nas horas em que mais precisei.

## Resumo

Anibolete, Tulio Jorge de Alcântara Neves Souza; Milidiú, Ruy Luiz. ***Boosting para Sistemas de Recomendação***. Rio de Janeiro, 2008. 60p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Com a quantidade de informação e sua disponibilidade facilitada pelo uso da Internet, diversas opções são oferecidas às pessoas e estas, normalmente, possuem pouca ou quase nenhuma experiência para decidir dentre as alternativas existentes. Neste âmbito, os Sistemas de Recomendação surgem para organizar e recomendar automaticamente, através de Aprendizado de Máquina, itens interessantes aos usuários. Um dos grandes desafios deste tipo de sistema é realizar o casamento correto entre o que está sendo recomendado e aqueles que estão recebendo a recomendação. Este trabalho aborda um Sistema de Recomendação baseado em Filtragem Colaborativa, técnica cuja essência está na troca de experiências entre usuários com interesses comuns. Na Filtragem Colaborativa, os usuários pontuam cada item experimentado de forma a indicar sua relevância, permitindo que outros do mesmo grupo se beneficiem destas pontuações. Nosso objetivo é utilizar um algoritmo de *Boosting* para otimizar a performance dois Sistemas de Recomendação. Para isto, utilizamos uma base de dados de anúncios com fins de validação e uma base de dados de filmes com fins de teste. Após adaptações nas estratégias convencionais de *Boosting*, alcançamos melhorias de até 3% sobre a performance do algoritmo original.

## Palavras-chave

*Boosting*, Filtragem Colaborativa, Sistemas de Recomendação, Aprendizado de Máquina

## Abstract

Anibolet, Tulio Jorge de Alcântara Neves de Souza; Milidiú, Ruy Luiz (Advisor). ***Boosting for Recommendation Systems***. Rio de Janeiro, 2008. 60p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

With the amount of information and its easy availability on the Internet, many options are offered to the people and they, normally, have little or almost no experience to decide between the existing alternatives. In this scene, the Recommendation Systems appear to organize and recommend automatically, through Machine Learning, the interesting items. One of the great recommendation challenges is to match correctly what is being recommended and who are receiving the recommendation. This work presents a Recommendation System based on Collaborative Filtering, technique whose essence is the exchange of experiences between users with common interests. In Collaborative Filtering, users rate each experimented item indicating its relevance allowing the use of ratings by other users of the same group. Our objective is to implement a Boosting algorithm in order to optimize a Recommendation System performance. For this, we use a database of advertisements with validation purposes and a database of movies with testing purposes. After adaptations in the conventional Boosting strategies, improvements of 3% were reached over the original algorithm.

## Keywords

Boosting, Collaborative Filtering, Recommendation Systems, Machine Learning

## Sumário

1 Introdução	12
1.1. Motivação	12
1.1.1. Desafio da Recomendação de Anúncios na Internet	14
1.1.2. Desafio da Recomendação de Filmes	16
1.2. Trabalhos Relacionados	18
1.3. Objetivo	20
1.4. Organização do Trabalho	21
2 <i>Boosting</i>	22
2.1. Problema de Classificação	22
2.2. Validação Cruzada	23
2.3. AdaBoost	25
3 Sistemas de Recomendação e Filtragem Colaborativa	30
3.1. Recomendação Colaborativa	30
3.2. Fatoração de Matrizes	33
3.3. Adaptação do AdaBoost	34
4 Experimentos	37
4.1. Conjunto de Dados de Anúncios	37
4.1.1. Geração da Base Artificial	38
4.2. Conjunto de Dados de Filmes	39
4.2.1. Segmentação dos Dados	39
4.3. Metodologia e Métricas de Avaliação	41
4.4. Resultados do algoritmo MM para Anúncios	42
4.5. Resultados do algoritmo FM para Anúncios	43
4.6. Resultados do <i>Boosting</i> com algoritmo FM para Anúncios	45
4.7. Resultados do algoritmo MM para Filmes	49
4.8. Resultados do algoritmo FM para Filmes	50
4.9. Resultados do Boosting com algoritmo FM para Filmes	51

5 Considerações Finais	55
5.1. Contribuições	55
5.2. Trabalhos Futuros	56
6 Referências Bibliográficas	57

## Lista de figuras

Figura 1: Motivação para os Sistemas de Recomendação.	12
Figura 2: Recomendação colaborativa a partir da experiência de outrem.	13
Figura 3: Lucro obtido através do clique nos anúncios.	15
Figura 4: Logística de aluguel da Netflix.	16
Figura 5: Ponto de partida do trabalho.	20
Figura 6: Objetivo do trabalho.	20
Figura 7: Exemplo de validação cruzada.	24
Figura 8: Exemplo de <i>repeated holdout</i> .	24
Figura 9: Algoritmo AdaBoost.	27
Figura 10: Distribuição dos exemplos na classe “+” ou na classe “-“.	28
Figura 11: Executando a primeira rodada do AdaBoost.	28
Figura 12: Executando a segunda rodada do AdaBoost.	29
Figura 13: Executando a terceira e última rodada do AdaBoost.	29
Figura 14: Combinando os resultados.	29
Figura 15: Grupo de filmes correlatos.	31
Figura 16: Exemplificação da recomendação de filmes.	32
Figura 17: Algoritmo AdaBoost original para problemas regressivos.	35
Figura 18: Algoritmo AdaBoost adaptado para a recomendação.	36
Figura 19: Métricas de posição em função das épocas de treinamento.	44
Figura 20: Calibrando o AdaBoost.RS para anúncios.	46
Figura 21: Melhora de <b>3%</b> no RMSE em relação ao FM.	48
Figura 22: Melhora de 2,5% no EAMP e 0,8% na PP em relação ao FM.	48
Figura 23: Calibrando o AdaBoost.RS para filmes.	52
Figura 24: Melhora de 0,2% e 0,13% no RMSE com relação ao FM.	54
Figura 25: <i>Ranking</i> da competição Netflix.	55

## Lista de tabelas

Tabela 1: Conversão da Posição em CTR Relativo.	38
Tabela 2: <i>Clusters</i> Utilizados na Recomendação de Filmes.	40
Tabela 3: Resultados do Sistema de Referência.	42
Tabela 4: Resultados do FM.	43
Tabela 5: Calibração do Fator de Suavização para Anúncios.	45
Tabela 6: Resultados do <i>Boosting</i> aplicado ao FM.	47
Tabela 7: Resultados do MM sobre os <i>Clusters</i> .	49
Tabela 8: Resultados do FM sobre o <i>Cluster 1</i> .	50
Tabela 9: Resultados do FM sobre o <i>Cluster 2</i> .	50
Tabela 10: Calibração do Fator de Suavização para Filmes.	51
Tabela 11: Resultados do <i>Boosting</i> aplicado ao FM no <i>Cluster 1</i> .	53
Tabela 12: Resultados do <i>Boosting</i> aplicado ao FM no <i>Cluster 2</i> .	53