



Bruno de Figueiredo Melo e Souza

**Modelos de fatoração matricial para
recomendação de vídeos**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática da PUC-Rio

Orientador: Prof. Ruy Luiz Milidiú

Rio de Janeiro
Agosto de 2011



Bruno de Figueiredo Melo e Souza

**Modelos de fatoração matricial para
recomendação de vídeos**

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

Prof. Ruy Luiz Milidiú

Orientador

Departamento de Informática – PUC-Rio

Prof. Marco A. Casanova

Departamento de Informática – PUC-Rio

Prof. Daniel Schwabe

Departamento de Informática – PUC-Rio

Prof. José Engenio Leal

Coordenador Setorial do Centro Técnico Científico
PUC-Rio

Rio de Janeiro, 29 de agosto de 2011

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

Bruno de Figueiredo Melo e Souza

Graduou-se em Engenharia Elétrica com ênfase em Eletrônica e Computação na UFRJ – Universidade Federal do Rio de Janeiro – em 2006. Atualmente trabalha como Coordenador de Tecnologia na Globo.com liderando projetos de infra-estrutura para codificação e distribuição de vídeos para Internet.

Ficha Catalográfica

Souza, Bruno de Figueiredo Melo e

Modelos de fatoração matricial para recomendação de vídeos / Bruno de Figueiredo Melo e Souza ; orientador: Ruy Luiz Milidiú. – Rio de Janeiro : PUC-Rio, Departamento de Informática, 2011.

[], 67 f. : il. ; 30 cm

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

Inclui bibliografia

1. Informática – Teses. 2. Filtragem Colaborativa. 3. Sistemas de Recomendação. 4. Aprendizado de Máquina. 5. Modelos de Fatoração Latente. 6. Fatoração Matricial. I. Milidiú, Ruy Luiz. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Agradecimentos

À Globo.com que financiou por completo o curso de mestrado.

À minha família, pelo apoio incondicional a todas as minhas escolhas de vida.

Ao meu orientador Ruy Luiz Milidiú, por me guiar nas pesquisas e decisões ao longo do curso.

Resumo

Souza, Bruno de Figueiredo Melo e; Milidiú, Ruy Luiz . **Modelos de Fatoração Matricial para Recomendação de Vídeos**. Rio de Janeiro, 2011. 67p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A recomendação de itens a partir do *feedback* implícito dos usuários consiste em identificar padrões no interesse dos usuários por estes itens a partir de ações dos usuários, tais como cliques, interações ou o consumo de conteúdos específicos. Isso, de forma a prover sugestões personalizadas que se adéquem ao gosto destes usuários. Nesta dissertação, avaliamos a performance de alguns modelos de fatoração matricial otimizados para a tarefa de recomendação a partir de dados implícitos no consumo das ofertas de vídeos da Globo.com. Propusemos tratar estes dados de consumo como indicativos de intenção de um usuário em assistir um vídeo. Além disso, avaliamos como os vieses únicos dos usuários e vídeos, e sua variação temporal impactam o resultado das recomendações. Também sugerimos a utilização de um modelo de fatoração incremental otimizado para este problema, que escala linearmente com o tamanho da entrada, isto é, com os dados de visualizações e quantidade de variáveis latentes. Na tarefa de prever a intenção dos usuários em consumir um conteúdo novo, nosso melhor modelo de fatoração apresenta um *RMSE* de 0,0524 usando o viés de usuários e vídeos, assim como sua variação temporal.

Palavras-chave

Filtragem Colaborativa; Sistemas de Recomendação; Aprendizado de Máquina; Modelos de Fatoração Latente; Fatoração Matricial.

Abstract

Souza, Bruno de Figueiredo Melo e; Milidiú, Ruy Luiz (Advisor). **Matrix factorization models for video recommendation**. Rio de Janeiro, 2011. 67p. MSc. Dissertation – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Item recommendation from implicit feedback datasets consists of passively tracking different sorts of user behavior, such as purchase history, watching habits and browsing activities in order to improve customer experience through providing personalized recommendations that fits into users' taste. In this work we evaluate the performance of different matrix factorization models tailored for the recommendation task for the implicit feedback dataset extracted from Globo.com's video site's access logs. We propose treating the data as indication of a positive preference from a user regarding the video watched. Besides that we evaluated the impact of effects associated with either users or items, known as biases or intercepts, independent of any interactions and its time changing behavior throughout the life span of the data in the result of recommendations. We also suggest a scalable and incremental procedure, which scales linearly with the input data size. In trying to predict the intention of the users for consuming new videos our best factorization models achieves a *RMSE* of 0,0524 using user's and video's bias as well as its temporal dynamics.

Keywords

Collaborative Filtering; Recommender Systems; Machine Learning; Latent Factor Models; Matrix Factorization.

Sumário

| | | |
|-----|---|----|
| 1 | Introdução | 11 |
| 2 | Sistemas de Recomendação | 15 |
| 2.1 | O Processo de Recomendação | 16 |
| 2.2 | O Problema de Recomendação | 16 |
| 2.3 | Estratégias de Recomendação | 19 |
| | Filtragem de Conteúdo | 20 |
| | Filtragem Colaborativa | 21 |
| | Comparação das Estratégias | 24 |
| 3 | Modelos de Fatoração Matricial | 27 |
| 3.1 | Introdução | 27 |
| 3.2 | Modelo Matemático | 28 |
| 3.3 | Fatoração Matricial Incremental | 30 |
| 3.4 | O Fator de Regularização | 33 |
| 3.5 | Adicionando Vieses ao Modelo | 34 |
| 3.6 | Adicionando Dinâmicas Temporais ao Modelo | 36 |
| 3.7 | O algoritmo final | 38 |
| 4 | Experimentos com um acervo de vídeos | 40 |
| 4.1 | Conjunto de dados | 40 |
| 4.2 | Métrica de Avaliação | 44 |
| 4.3 | Experimentos | 45 |
| | Quantidade de Passos de Treino | 45 |
| | Número de Variáveis Latentes | 46 |
| | Taxa de Aprendizado | 47 |
| | Fator de Regularização | 48 |
| | Comparação dos Modelos | 49 |
| 4.4 | Recomendando Vídeos | 50 |
| 5 | Trabalhos Relacionados | 52 |
| | Filtragem Colaborativa | 52 |
| | Otimizações de Performance | 54 |
| 6 | Conclusão e Trabalho Futuro | 56 |
| 7 | Referencias Bibliográficas | 59 |

Lista de Figuras

| | |
|---|----|
| Figura 2.1 - Processo de recomendação | 16 |
| Figura 2.2 - Recomendação baseada em filtragem de conteúdo..... | 20 |
| Figura 2.3 - Filtragem colaborativa na Amazon.com..... | 22 |
| Figura 2.4 - Processo de formação de vizinhança | 23 |
| Figura 2.5 - Processo de filtragem colaborativa | 24 |
| Figura 3.1 - Fatoração matricial incremental | 31 |
| Figura 4.1 - Catálogo de vídeos do Auto Esporte | 40 |
| Figura 4.2 - Quantidade de passos de treino vs. RMSE | 45 |
| Figura 4.3 - Número de variáveis latentes vs. RMSE..... | 46 |
| Figura 4.4 - Variável latente 1 vs. variável latente 2..... | 47 |
| Figura 4.5 - Efeito da taxa de aprendizado | 47 |
| Figura 4.6 - Efeito da regularização | 48 |
| Figura 4.7 - Comparação dos modelos de fatoração | 49 |

Lista de Tabelas

| | |
|---|----|
| Tabela 2.1 - Matriz de avaliações para recomendação de filmes | 18 |
| Tabela 3.1 - Um exemplo de matriz usuário item..... | 27 |
| Tabela 4.1 - Matriz de visualizações de vídeos..... | 42 |
| Tabela 4.2 - Formatação dos dados de entrada | 43 |
| Tabela 4.3 - Quantidade de passos de treino vs. <i>RMSE</i> | 46 |
| Tabela 4.4 - Quantidade de variáveis latentes vs. <i>RMSE</i> | 46 |
| Tabela 4.5 - Efeito da taxa de aprendizado..... | 48 |
| Tabela 4.6 - Efeito da regularização..... | 48 |
| Tabela 4.7 - Comparação dos modelos de fatoração | 49 |
| Tabela 4.8 - Exemplo de recomendações para um usuário..... | 51 |

Lista de Algoritmos

| | |
|---|----|
| Algoritmo 1 - Versão básica | 33 |
| Algoritmo 2 - Versão final com viés e dinâmicas temporais | 39 |