

Marcos Masanobu Takahashi
Orientador: Prof. Roberto Hirata Jr.

INTRODUÇÃO

No comércio eletrônico as recomendações são amplamente utilizadas para vencer a concorrência que oferece produtos a preços semelhantes e dessa forma os clientes ficam cada vez menos no site, com média de 4 a 5 visualizações de página por visita.

Para que o cliente encontre o produto desejado no curto intervalo de tempo de acesso no site, são utilizados algoritmos de recomendação que abrangem dados como de compra, navegação e até mídias sociais e esses serão os dados utilizados no trabalho.



OBJETIVO

- Implementar dois algoritmos de recomendação: *Large-scale Parallel Collaborative Filtering* e *Fast Context-aware Recommendations with Factorization Machines*;
- compará-los em um contexto restrito do mundo real (e-commerce de Móveis e Utilidades Domésticas);
- comparar o retorno financeiro de cada algoritmo.

SIST. DE RECOMENDAÇÃO

Sistemas de Recomendação são comumente divididos em 3 tipos de algoritmos:

- **Content-based Filtering:** recomenda itens similares àqueles que o usuário já comprou ou interagiu através dos atributos dos mesmos.
- **Collaborative Filtering:** através dos dados de compra ou interação passadas, recomenda itens que pessoas com mesmo tipo de comportamento interagem.
- **Hybrid Recommender Systems:** sistemas híbridos que unem as abordagens Content-based e Collaborative Filtering.

Existe uma outra classe de sistemas de recomendação chamada Context-aware Recommender Systems, que utiliza também o contexto em que dado evento aconteceu.

REFERÊNCIAS

- [1] Robert Schreiber, Yunhong Zhou, Dennis Wilkinson e Rong Pan. Large-scale Parallel Collaborative Filtering for the Netflix Prize
- [2] Steffen Rendle. Fast Context-aware Recommendations with Factorization Machines

PARALLEL COL. FILTERING

Large-scale Parallel Collaborative Filtering [1] foi desenvolvido para a competição Netflix Prize e melhorou o desempenho do algoritmo que era utilizado pela Netflix (CineMatch) em 5.91%.

A abordagem tradicional dos algoritmos Collaborative Filtering é utilizar uma função f tal que:

$$f : User \times Item \rightarrow Rating$$

e utilizando os dados de ratings (nota) que se possui, estimar os ratings dos pares $(user, item)$ que não foram dados no problema.

Este algoritmo decompõe a matriz de ratings R em $U \subseteq \mathbb{R}^{n_u \times n_f}$ e $I \subseteq \mathbb{R}^{n_i \times n_f}$, onde U é a matriz de características do usuário, I a matriz de características do item, n_u é o número de usuários, n_i o número de itens e n_f o número de variáveis consideradas no modelo.

Com as matrizes decompostas, o rating $r_{j,k}$ dado pelo usuário j ao item k é:

$$r_{j,k} = \langle u_j, i_k \rangle$$

CONTEXT RECOM. WITH FM

O *Fast Context-aware Recommendations with Factorization Machines* [2] é um algoritmo que além de fazer recomendações muito boas, apresenta tempo de execução muito bom diferente dos algoritmos similares de Contexto.

O problema de recomendação utilizando Contexto se dá da seguinte forma:

$$y : U \times I \times C_3 \times C_4 \dots C_m \rightarrow \mathbb{R}$$

onde U são os usuários, I os itens e $C_3, C_4 \dots C_m$ os diferentes contextos.

Através de um mapeamento de dados da mesma forma que o exemplo da figura,

Recommender Data	Feature vector x	Target y
(A, TI, H, {C}, 5)	$x^{(1)}$ 1 0 0 1 0 0 0 0 0 1 0 0 1	5 $y^{(1)}$
(A, NH, S, {I}, 3)	$x^{(2)}$ 1 0 0 0 1 0 0 1 0 0 0 0 0	3 $y^{(2)}$
(A, SW, N, {B, C}, 1)	$x^{(3)}$ 1 0 0 0 0 0 1 0 0 1 0 0 0.5 0.5	1 $y^{(3)}$
(B, SW, N, {A, C}, 4)	$x^{(4)}$ 0 1 0 0 0 0 1 0 0 1 0 0.5 0 0.5	4 $y^{(4)}$
(B, ST, H, {I}, 5)	$x^{(5)}$ 0 1 0 0 0 0 0 1 0 0 1 0 0 0	5 $y^{(5)}$
(C, TI, S, {A}, 1)	$x^{(6)}$ 0 0 1 1 0 0 0 1 0 0 0 1 0 0	1 $y^{(6)}$
(C, SW, H, {A, B}, 5)	$x^{(7)}$ 0 0 1 0 0 1 0 0 0 1 0.5 0.5 0	5 $y^{(7)}$

tem-se os ratings aplicando a função:

$$y(x) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \hat{w}_{i,j} x_i x_j$$

TESTES E RESULTADOS

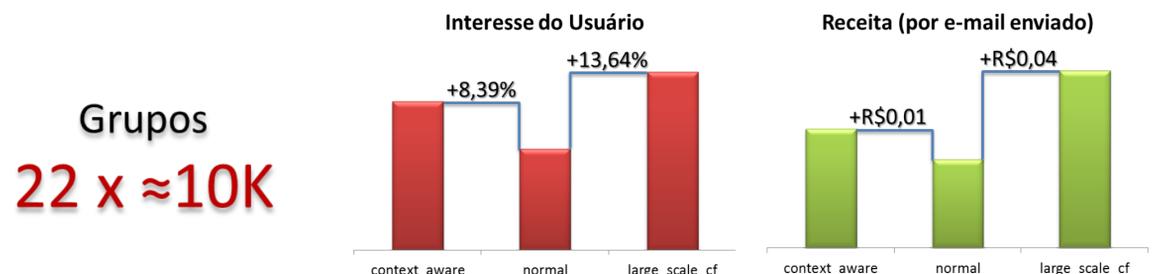
Todos os códigos e testes foram implementados em linguagem R e os dados utilizados para cada tipo de recomendação foram:

- **Large-scale Parallel Collaborative Filtering:** User (id_user), Item (id_item), Rating (# compras do item);
- **Fast Context-aware Recommendations with Factorization Machines:** User (id_user, sexo, idade, estado(geográfico)), Item (id_item, categoria, cor), Contexto (proximidade da compra com datas especiais, uso de cupom de desconto, recompra, likes em tags de Facebook, agrupamento de dados de navegação com abandono de carrinho).

A partir desses dados foram geradas recomendações que foram enviadas através de campanhas de E-mail Marketing, sendo que foram 5 dias de disparos de campanhas, com 20 grupos escolhidos aleatoriamente, 1 grupo selecionado através de Large-scale Parallel Collaborative Filtering e 1 grupo selecionado através de Fast Context-aware Recommendations with Factorization Machines.

Não houve sobreposição de pessoas dos mesmos algoritmos em disparos diferentes, ou seja, uma pessoa não recebeu mais de uma campanha de dado algoritmo.

A seguir, alguns números dos testes e os resultados obtidos:



TRABALHOS FUTUROS

- Identificação e Inclusão de novos contextos ao modelo Context-aware;
- Incorporação de dados de navegação e carrinho abandonado ao Large-scale Parallel Collaborative Filtering;
- Testes com intersecção de Recomendação com produtos dos dois modelos.

PÁGINA DO TRABALHO

A página do trabalho e os códigos utilizados podem ser encontrados em:



<https://linux.ime.usp.br/~marcost/mac0499/>