

Trabalho Prático 0 – Revisão de Programação e Tipos Abstratos de dados

Valor: 5 pontos

Data de entrega: 01/09/2014

Um sistema de recomendação tem como objetivo selecionar um conjunto de objetos que podem ser de interesse de um dado usuário do sistema. Por exemplo, o Netflix, um serviço web para exibição de filmes, tem um sistema capaz de recomendar um conjunto de filmes dadas as preferências (características) de um usuário.

Uma das formas mais simples de recomendação, chamada de recomendação dos itens mais populares, recomenda os mesmos itens a todos os usuários do sistema, e por isso não é personalizada. Nesse caso, para cada item na base de dados, conta-se quantas vezes ele foi acessado, considerando todos os usuários do sistema. Por exemplo, no caso do Netflix, computa-se a lista dos filmes mais assistidos, e os n filmes mais vistos são recomendados para todos os usuários do sistema.

Um segundo tipo de recomendação, que leva em conta o perfil dos usuários, é a recomendação chamada de filtragem colaborativa. Nesse tipo de sistema, são recomendados para um usuário u os filmes que foram assistidos por usuários parecidos com ele, e uma função de similaridade é utilizada para determinar a semelhança entre dois usuários quaisquer. Essa função de similaridade pode ser definida de diversas formas. Neste trabalho utilizaremos uma função simples, chamada de similaridade de Jaccard.

Dados dois usuários u_1 e u_2 , a similaridade de Jaccard é definida como:

$$J(u_1, u_2) = \frac{\text{items}(u_1) \cap \text{items}(u_2)}{\text{items}(u_1) \cup \text{items}(u_2)}$$

Em outras palavras, no caso de recomendação de filmes, a similaridade de Jaccard considera a intersecção dos filmes já vistos por dois usuários dividida pelo número total de filmes já vistos pelos dois.

Neste trabalho, implementaremos estas duas formas de recomendação para o cenário de filmes. Dado um usuário u , seu sistema deve recomendar sempre 3 filmes. Para isso, você deve criar um tipo abstrato de dados do tipo *Item*, que guarda o número de identificação do filme, o nome, número de identificação no IMDB e o ano. Além disso, você também precisará de um TAD usuário, que guarda o identificador do usuário e a lista de filmes que ele já assistiu. Se necessário, você pode assumir que o número máximo de usuários do sistema é 2.110 e o número máximo de filmes é 500.

Para simplificar o problema, a recomendação dos itens mais populares será a mesma para todos os usuários, independente dele já ter assistido o filme (embora em um sistema real itens repetidos nunca sejam recomendados). No caso de vários filmes com a mesma popularidade (ou seja, já assistidos pelo mesmo número de pessoas), você deve recomendar os filmes mais recentes (de acordo com o ano de lançamento).

No caso da recomendação por filtragem colaborativa, caso vários usuários apresentem o mesmo valor de similaridade com o usuário para o qual se quer recomendar, deve-se selecionar o usuário com o menor número identificador. Caso a lista de filmes desse usuário selecionado seja maior que 3, da mesma forma que na recomendação por popularidade, deve-se recomendar os 3 filmes mais recentes.

Bases de dados

Duas bases de dados com informações sobre os filmes a serem recomendados e seus usuários serão disponibilizadas.

A primeira, relacionada aos filmes, traz a lista de informações sobre cada um dos n filmes que o usuário pode ter assistido. Para cada filme, as seguintes informações são disponibilizadas, sempre separadas por tab:

movie_id	título	imdb_id	ano
1	Toy story	0114709	1995

A segunda base, relacionada aos usuários, traz o identificador do usuário seguido da lista de filmes que ele assistiu. Cada filme é representado por um número binário, onde 1 significa que o usuário já assistiu o filme e 0 que ele não assistiu. Um exemplo desse arquivo é dado abaixo, onde n representa o número de filmes na base de dados.

user_id	m_1	m_2	m_3	...	m_n
12	0	1	1		0

Arquivos de entrada

Seu programa deve receber como entrada um arquivo *input.txt*. A primeira linha do arquivo deve conter o nome da base de dados de filmes seguida do nome da base de dados de usuários (separadas por tab). As linhas seguintes devem conter o identificador do usuário para o qual se deseja fazer a recomendação.

Exemplo do arquivo input.txt

```
movies.txt    users.txt
```

```
321
543
12
....
```

O arquivo de saída, *output.txt*, deve conter a primeira linha idêntica ao arquivo de entrada. Para as linhas seguintes, deve-se colocar o identificador do usuário seguido de “:” e do título dos 3 filmes recomendados, separados por tab. Primeiro você deve listar o nome dos filmes recomendados de acordo com a abordagem do filme mais popular, e depois da filtragem colaborativa.

Exemplo do arquivo output.txt

```
movies.txt    users.txt
```

Most popular

```
Ice Age      Divergent      Noah
```

Personalizada

```
321: Ice Age      Toy Story 2      Despicable me
543: Matrix 2     X-Men            Batman Returns
12: Noah          Thor             Lord of the Rings
....
```

A saída será comparada com a de uma implementação padrão dentro do Prático.

Comentários Gerais:

1. Comece a fazer este trabalho logo, enquanto o problema está fresco na memória e o prazo para terminá-lo está tão longe quanto jamais poderá estar.
2. Clareza, indentação e comentários no programa também serão avaliados.
3. O trabalho é individual.
4. A submissão será feita pelo Prático (`aeds.dcc.ufmg.br`)
5. Trabalhos copiados, comprados, doados, etc. serão penalizados conforme anunciado.
6. Penalização por atraso: $(2^d - 1)$ pontos, onde d é o número de dias de atraso.