



CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS

COMUNIDADE EVANGÉLICA LUTERANA "SÃO PAULO"
Recredenciado pela Portaria Ministerial nº 3.607 - D.O.U. nº 202 de 20/10/2005

RAFAEL FONSECA DUARTE

**UM PROTÓTIPO DE UM APLICATIVO PARA RECOMENDAÇÃO DE
FILMES A PARTIR DE AVALIAÇÕES REALIZADAS PELOS
USUÁRIOS**

Palmas

2014

RAFAEL FONSECA DUARTE

**UM PROTÓTIPO DE UM APLICATIVO PARA RECOMENDAÇÃO DE
FILMES A PARTIR DE AVALIAÇÕES REALIZADAS PELOS
USUÁRIOS**

Trabalho de Conclusão de Curso (TCC)
elaborado e apresentado como requisito
parcial para obtenção do título de
bacharel em Sistemas de Informação pelo
Centro Universitário Luterano de Palmas
(CEULP/ULBRA).

Professor Mestre Edeilson Milhomem.

Palmas

2014

RAFAEL FONSECA DUARTE

**UM PROTÓTIPO DE UM APLICATIVO PARA RECOMENDAÇÃO DE
FILMES A PARTIR DE AVALIAÇÕES REALIZADAS PELOS
USUÁRIOS**

Trabalho de Conclusão de Curso (TCC)
elaborado e apresentado como requisito
parcial para obtenção do título de
bacharel em Sistemas de Informação pelo
Centro Universitário Luterano de Palmas
(CEULP/ULBRA).
Professor Mestre Edeilson Milhomem.

Aprovada em junho de 2014.

BANCA EXAMINADORA

Prof. M. Sc. Edeilson Milhomem
Centro Universitário Luterano de Palmas

Prof. M. Sc. Jackson Gomes
Centro Universitário Luterano de Palmas

Prof. M. Sc. Parcilene Fernandes de Brito
Centro Universitário Luterano de Palmas

**Palmas
2014**

DEDICATÓRIA

Dedico este trabalho primeiramente a Deus, por me conceder o dom da vida; aos meus pais, Raimundo e Lucia, pelo esforço, amor e dedicação em todos os momentos; e a todos aqueles que me apoiaram.

AGRADECIMENTOS

Muitos me ajudaram a concluir este trabalho, por isso aqui deixo os meus sinceros agradecimentos:

Em primeiro lugar agradecer a Deus pelo dom da vida e por me dar forças e sabedoria nesta fase difícil.

Agradecer aos meus pais, Raimundo e Lucia, por toda confiança, amor, paciência e compreensão durante minha caminhada. Agradeço a eles por terem me mostrado o caminho certo a ser seguido.

A minha namorada Karoline, por estar ao meu lado me incentivando e apoiando com todo seu amor.

Ao meu orientador Edeilson Milhomem, pela paciência e confiança depositados em minha pessoa. Agradeço a ele por todos os seus puxões de orelha, pois graças a eles conclui o trabalho. Agradeço pela sua disposição diante de todas as minhas dificuldades enfrentadas no trabalho.

Agradeço a todos meus amigos pelas conversas de incentivo para que fosse possível concluir este trabalho. Quero agradecer em especial a Cristiane Cunha, por diversas vezes tirar um pouco do seu tempo para me auxiliar em correções e conselhos.

E todos os que contribuíram para a realização deste trabalho.

RESUMO

Os Sistemas de Recomendação utilizam as informações dos usuários para realizar recomendação de itens ou serviços de acordo com as suas preferências, desta forma os usuários encontram algo de seu interesse com mais facilidade. Para realizar uma recomendação personalizada, os Sistemas de Recomendação utilizam técnicas de filtragens, onde é realizado um processamento com as informações dos usuários e, a partir disso, é gerada uma recomendação personalizada. Para compreender como funciona o Sistema de Recomendação, neste trabalho serão apresentados os conceitos de Sistema de Recomendação, técnicas de recomendação, como: Filtragem Baseada em Conteúdo e Filtragem Colaborativa, e a API *Apache Mahout Taste*. O objetivo deste trabalho é apresentar os conceitos e as etapas realizadas no desenvolvimento de um protótipo de recomendação utilizando a Filtragem Colaborativa, que irá gerar recomendações de filmes de acordo com as preferências dos usuários.

PALAVRAS-CHAVE: Sistema de Recomendação, Filtragem Colaborativa, Taxonomia, API Apache Mahout Taste.

LISTA DE FIGURAS

Figura 1 Recomendação Não-Personalizada.	14
Figura 2 Recomendação Efêmera.	15
Figura 3 Recomendação Persistente.	16
Figura 4 Taxonomia dos Sistemas de Recomendação. Traduzida de SCHAFER <i>etal.</i> (2001).	17
Figura 5 - Técnicas de Recomendação em um Sistema de Recomendação (GODOY NETO, 2009, p.48).	22
Figura 6 API Apache Mahout Taste.	32
Figura 7 Arquitetura do Sistema de Recomendação	37
Figura 8 Modelo Relacional da Base Movielens Adaptada	42
Figura 9 Base de Teste.	43
Figura 10 Resultados Gerados Pelo Protótipo	44

LISTA DE TABELAS

Tabela 1 Similaridades entre usuários utilizando Pearson.....	30
Tabela 2 Modelo da Base de Dados	35
Tabela 3 Modelo da Base de Dados Perfil dos Usuários.....	36
Tabela 4 Similaridade dos Usuários	45
Tabela 5 Avaliações dos Usuários.....	45

LISTA DE QUADROS

Quadro 1 Exemplo de Recomendação Baseado em FC.....	26
---	----

LISTAGEM

Listagem 1 Classe Responsável em Gerar Recomendação Colaborativa	39
Listagem 2 Interface DataModel	39
Listagem 3 Interface UserSimilarity	40
Listagem 4 Interface UserNeighborhood	40
Listagem 5 Interface Recommender.....	41

LISTA DE ABREVIATURAS

SR – Sistema de Recomendação

FBC – Filtragem Baseada em Conteúdo

FC – Filtragem Colaborativa

IDE – Integrated Development Environment

SUMÁRIO

1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO	14
2.1.	Sistemas de Recomendação	14
2.1.1.	Taxonomia dos Sistemas de Recomendação	17
2.2.	Técnicas de Recomendação	21
2.3.	Filtragem Baseada em Conteúdo	22
2.4.	Filtragem Colaborativa.....	24
2.4.1.	Técnica de Filtragem Colaborativa	26
2.4.2.	Exemplo de Filtragem Colaborativa.....	29
3	MATERIAIS E MÉTODOS	32
3.1.	Materiais	32
3.1.1.	API Apache Mahout Taste	32
3.2.	Metodologia.....	34
4	RESULTADOS E DISCUSSÃO	37
4.1.	Arquitetura da Ferramenta.....	37
4.2.	Implementação	38
4.3.	EXPERIMENTOS	41
4.3.1.	Cenário	41
4.3.2.	Resultados	43
6	CONSIDERAÇÕES FINAIS	48

1 INTRODUÇÃO

Um Sistema de Recomendação utiliza dados do usuário para realizar uma recomendação que esteja de acordo com a sua preferência. Com o crescimento acelerado das informações contidas na WEB, devido aos avanços da tecnologia e principalmente da internet, têm se tornado cada vez mais diversificadas as formas de encontrar itens ou serviços, por exemplo, através de pesquisas em *engines* de busca, ou através de recomendações. As recomendações podem ajudar os usuários a estabelecer filtros de informações relevantes através das suas preferências (MILIOLI, 2007).

Um Sistema de Recomendação tem como objetivo facilitar aos usuários o acesso às informações que mais lhes interessam. No mercado, uma recomendação pode fazer toda a diferença para conquistar ou perder um cliente. Por exemplo, diante da seguinte situação hipotética: os usuários que possuem cadastro no sistema de uma empresa constantemente recebem e-mails de promoções. Os usuários ficarão satisfeitos se a promoção enviada pela empresa no seu e-mail contiver apenas itens ou serviços que atendam às suas preferências ou o mais próximo possível disso, de acordo com o seu perfil. Mas, caso o usuário comece a receber várias recomendações que não estejam de acordo com as suas preferências, o usuário ficará insatisfeito em receber estas recomendações.

Mas, caso o usuário receba várias recomendações no e-mail que não estejam de acordo com as suas preferências, trará aos usuários insatisfação, podendo levá-los a cancelar suas contas com a empresa.

Para realizar recomendações em ambientes computacionais é necessário utilizar técnicas de recomendação. Cada uma destas técnicas possui suas particularidades e podem ser aplicadas em diferentes contextos. Existem três tipos de técnica de recomendação, são elas: Recomendação Baseada em Conteúdo, Recomendação Colaborativa e Recomendação Híbrida.

A técnica de Recomendação Baseada em Conteúdo parte da premissa que, se um usuário consumiu um item ou utiliza algum serviço, então ele continuará consumindo itens ou serviços semelhantes, ou seja, baseia-se nas características dos produtos ou serviços já utilizados (Torres, 2004). Já na técnica de Recomendação Colaborativa, as recomendações dependem de análise humana, dos itens ou serviços. Esta análise gera um conjunto de avaliações, isto é, as notas fornecidas pelos usuários sobre a satisfação de um item ou serviço. A partir destas avaliações, é possível aplicar este método de recomendação. Além destas, ainda há a técnica Recomendação Híbrida, que engloba as características de ambas as técnicas anteriores, isto é, faz uma (fusão) entre elas para proferir as recomendações.

A utilização de técnicas de recomendação em comércio eletrônico e de grande importância. GrefLinden (2007) foi quem programou o primeiro Sistema de Recomendação para empresa Amazon, segundo Marshal (2006) o Sistema de Recomendação da Amazon é responsável por 35% das vendas de produtos da empresa. Outra empresa que mostra esta situação é a Netflix, uma das maiores lojas virtuais de filmes dos Estados Unidos, que no ano de 2006 ofereceu um prêmio de um milhão de dólares para o grupo ou pesquisador que conseguisse uma melhoria de pelo menos 10% no Sistema de Recomendação da empresa (NETFLIX 2006). Somente em 2009 um grupo composto com sete membros ganhou este prêmio.

Atualmente, existem várias empresas que utilizam serviços de sistemas de recomendação para recomendar filmes, tais como Netflix, Crackle e outros. As recomendações são baseadas em informações implícitas, como o número de vezes que o usuário assiste a um filme, e também são baseadas em informações explícitas, ou seja, o usuário realiza uma avaliação sobre o filme assistido ou apenas visualizado.

O trabalho foi dividido nas seguintes seções.

- Sistemas de Recomendação, Seção 2.1: Nesta seção é realizado um estudo sobre SR, desde como utilizar um SR até sua representação através de uma taxonomia;

- Técnicas de Recomendação, Seção 2.2: Nesta seção são apresentadas as técnicas de recomendação mais utilizadas em SR. Com isto é possível um entendimento sobre como cada técnica é utilizada em um SR;
- Filtragem Baseada em Conteúdo, Seção 2.3: Nesta seção a apresentação de como é realizado uma recomendação utilizando a FBC, considerando as vantagens e desvantagens que esta técnica possui;
- Filtragem Colaborativa, Seção 2.4: Nesta seção foi realizado um estudo detalhado sobre a FC, desde a explicação contextual até a apresentação de exemplos.

2 REFERENCIAL TEÓRICO

Nesta seção são apresentados os principais conceitos referentes ao projeto proposto neste trabalho, são eles: Sistemas de Recomendação (Seção 2.1), Técnicas de Recomendação (Seção 2.2), Recomendação Baseada em Conteúdo (Seção 2.3), Técnica de Recomendação Baseada em Conteúdo (Seção 2.3.1), Recomendação Colaborativa (Seção 2.4), Técnica de Recomendação Colaborativa (Seção 2.4.1).

2.1. Sistemas de Recomendação

Sistema de Recomendação (SR) tem sido utilizado com o intuito de minimizar o esforço de busca, provendo a utilização de um filtro de informação para apresentar serviços ou produtos, como CDs, livros, filmes, restaurantes e outros (SHAFFER, KONSTAN e RIEDL, 2001). Segundo Torres (2004), SR sugerem os produtos ou serviços que provavelmente serão de interesses dos usuários, uma vez que as recomendações podem ser baseadas em informações contidas em seus perfis. Para Torres (2004), os Sistemas de Recomendação podem possuir algumas estratégias ou grau de personalização para realizar recomendação, como: não personalizada, efêmera e persistente.

Os SR não personalizados são aqueles em que a mesma recomendação é realizada para todos os usuários. Por exemplo, tem-se o contexto da recomendação de filmes, em que todos os filmes em lançamento são recomendados aos usuários do sistema. Como pode ser observado, na Figura 1 é apresentada uma exemplificação de como é realizada uma recomendação não personalizada.



Figura 1 Recomendação Não-Personalizada.

A Figura 1 representa um cenário hipotético para recomendações de filmes que foram lançados no ano de 2013. Neste cenário, o SR realiza uma seleção dos filmes que foram recém-lançados no mercado, gerando assim uma lista de filmes que podem ser recomendados para os clientes. Esta lista com várias recomendações de filmes recém-lançados é então enviada para todos os e-mails dos usuários cadastrados no sistema da loja virtual.

O SR efêmera é quando o sistema realiza uma indicação apenas com base nas informações do usuário no instante que o indivíduo está acessando o sistema (TORRES, 2004). Por exemplo, quando o usuário acessa uma loja virtual e realizar uma busca específica utilizando a ferramenta *engines* ou apenas adicionando algo no carrinho de compra, o sistema utiliza estas informações para gerar recomendações para o usuário. A Figura 2 exemplifica uma situação de recomendação efêmera.



Figura 2 Recomendação Efêmera.

Como pode ser observado na Figura 2, para receber recomendações, primeiro o usuário deve alimentar o sistema, ou seja, fornecer alguma palavra-chave para ferramenta de buscar ou adicionar algo de seu interesse no carrinho de compra. A recomendação efêmera utiliza vários métodos, e dois deles são apresentados pela Figura 2, outros métodos seria o tempo de permanência em uma página da loja virtual e o movimento do mouse na página. Como se pode observar, é utilizada a palavra-chave, ou seja, palavras ou textos digitados pelos usuários para

realizar a busca de um determinado filme e o carrinho de compra, que são filmes já selecionados pelo usuário para uma possível compra.

Ao utilizar a palavra-chave através de pesquisas em *engines* de busca para realizar uma recomendação efêmera, o sistema retornará as buscas para o usuário a partir de dados fornecidos ao sistema no momento em que estiver realizando alguma ação. Já para as recomendações realizadas pelo carrinho de compra, o sistema leva em conta as informações de filmes já adicionados no carrinho de compra para realizar uma nova recomendação para o usuário.

O SR persistente ou personalizado é quando o sistema utiliza as informações armazenadas no perfil de preferência dos usuários, considerando os históricos de acessos, para gerar as recomendações (TORRES, 2004). Um exemplo deste tipo de recomendação é: em uma loja virtual quando se realiza o cadastro de um usuário e preenchem-se quais são as preferências do cliente no que tange aos gêneros de filmes. O sistema pode utilizar estas informações para recomendar filmes ou promoções considerando as preferências dos usuários. A Figura 3 exemplifica uma situação de recomendação persistente ou personalizada.

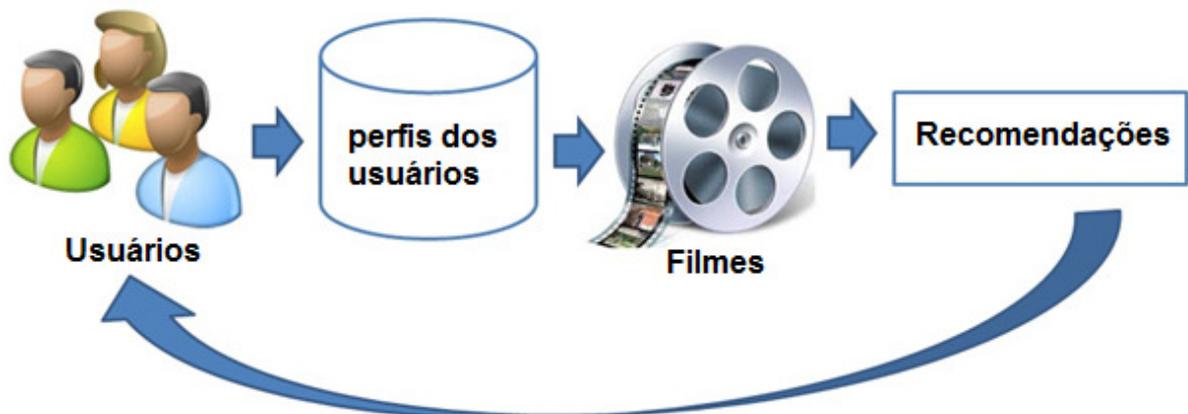


Figura 3 Recomendação Persistente.

Observando a Figura 3, percebe-se que os usuários cadastrados no sistema têm armazenado em seu perfil as suas preferências, estas pode ser desde: avaliações, gêneros de filmes, idade, profissão. Com os dados dos usuários armazenados no sistema de uma loja virtual, são realizadas as comparações utilizando técnicas de recomendação, ou seja, Filtragem Colaborativa, Baseada em Conteúdo, para descobrir a similaridades entre os filmes já comprados, assistidos ou

apenas visualizados, em relação aos novos filmes disponíveis no sistema. A partir da análise da similaridade dos perfis dos usuários, é possível listar/recomendar filmes que sejam das preferências dos usuários ou o mais próximo possível.

A projeção e o desenvolvimento de um SR podem seguir uma determinada classificação taxonômica, isto é, uma determinada organização, que necessita de alguns elementos. Esta organização é apresentada na seção 2.1.1.

2.1.1. Taxonomia dos Sistemas de Recomendação

Para que um SR possa ser projetado e desenvolvido, é necessário seguir algumas etapas, conforme pode ser observado na Figura 4: **entrada**, importante para que se tenham dados de entrada parâmetros para recomendação; **métodos e técnicas de recomendação**, essenciais para que as recomendações possam ser geradas; e **saída de dados**, que são os resultados provenientes do SR; além de outros aspectos do sistema, tal como o nível de personalização (SCHAFER ET al., 2001, tradução nossa),.



Figura 4 Taxonomia dos Sistemas de Recomendação. Traduzida de SCHAFER et al. (2001).

A Figura 4 ilustra a taxonomia do sistema de recomendação, e como cada etapa no processo de recomendação é classificada. Observando a Figura 4, o sistema SR necessita coletar entradas no sistema, estas entradas podem ser obtidas através de usuários e das comunidades de usuários.

Entradas obtidas pelos usuários do sistema geram vantagens para realizar recomendações, afinal quanto mais informações os usuários disponibilizam no sistema, maior a oportunidade de realizar uma recomendação mais próxima das preferências dos usuários, ou seja, mais próximo do seu gosto. As coletas de dados dos usuários podem ser realizadas através da navegação implícita, navegação explícita, avaliações e histórico do usuário (TORRES, 2004).

Na navegação implícita, as informações dos usuários são coletadas automaticamente pelo sistema quando o usuário utiliza o mesmo. A coleta é realizada de forma discreta, sem que o usuário perceba que dados estão sendo coletados para realizar uma recomendação. Os dados podem ser obtidos através de movimento do mouse, tempo de permanência em uma página de determinado produto, páginas visitadas no sistema, entre outras formas (TORRES, 2004).

Na navegação explícita, o usuário informa ao sistema quais as suas preferências através de formulários de cadastro no sistema, ou de avaliação de produtos etc. A coleta dos dados é realizada a partir do fornecimento explícito de determinados dados fornecidos pelo usuário ao sistema conscientemente, assim fornecendo subsídios aos sistemas de recomendação para realizar recomendações através de avaliações de itens ou serviços, comentários, e entre outras formas. Desta forma tornando possível realizar recomendações de itens ou serviços para os usuários (MOBASCHER e ANAND 2013).

Entradas obtidas através de avaliações dos usuários pela navegação explícita ocorrem quando o usuário avalia um item ou serviço que já tenha conhecimento ou que já tenha adquirido. A avaliação é realizada em uma escala numérica, como avaliar um filme com nota de 1 a 5. As informações sobre avaliações realizadas pelos usuários são utilizadas como fonte de informação para mensurar as semelhanças entre os perfis dos usuários, desta forma utilizando as avaliações e as semelhanças dos usuários, será recomendado algo que venha a ser do interesse

deles. Suponha que o usuário João avalie um filme de comédia com nota 4, Maria possui em seu perfil o gênero de comédia como preferência, logo, o sistema poderá indicar para Maria o filme avaliado por João.

Entradas obtidas pela navegação implícita podem ser feitas através do histórico do usuário. O sistema utiliza esta informação para realizar comparações com outros itens ou serviços do sistema, assim obtendo dados de forma implícita e recomendando algo para o usuário. Por exemplo, o usuário tem armazenado no seu histórico de compras filmes do gênero comédia. O sistema está oferecendo para os usuários uma promoção dos filmes do gênero comédia, desta forma o sistema realizará a comparação do histórico de compra dos usuários com o gênero dos filmes em promoção. Realizando esta comparação dos dados será feita a recomendação dos filmes aos usuários.

Os dados de entrada do sistema também são obtidos através das comunidades. Estes dados de entrada da comunidade são realizados através de todos os usuários que utilizam o sistema. As coletas de dados realizadas pela entrada da comunidade podem ser realizadas através da popularidade de itens externos à comunidade, avaliações e comentários textuais. A entrada de dados da comunidade pode ser realizada através de popularidade de itens externos, que podem ser, por exemplo, referentes às listas de itens ou serviços mais vendidos, ou lançados no ano corrente.

Entradas de dados da comunidade, são as avaliações realizadas por diversos usuários do sistema, e com estas avaliações obtém uma média de avaliação da comunidade. As avaliações realizadas pelos usuários das comunidades também ocorrem em uma escala numérica; por exemplo: usuários da comunidade avaliam filmes com notas de 1 a 5. Com as avaliações realizadas pelos usuários da comunidade em determinados itens ou serviços, são realizadas as recomendações aos usuários do sistema que participa da comunidade avaliada. Outra forma de obter dados das comunidades são comentários textuais, estes comentários são opiniões que os usuários têm sobre itens ou serviços já utilizados. Desta forma disponibilizando as opiniões para outros usuários executarem decisões sobre os itens ou serviços da comunidade, que venham a ser de interesse dos usuários.

Obter os dados de entrada para o SR é importante, pois a partir desses dados são geradas as saídas, que podem oferecer algo aos usuários de acordo com as suas preferências. Segundo Torres (2004), as saídas geralmente ocorrem através de recomendação, avaliações ou revisões.

As saídas de dados utilizando avaliações da comunidade são realizadas por todos os usuários do sistema. Estas avaliações são realizadas em uma escala numérica, por exemplo, de 0 a 5. Os itens ou serviços que obtiverem o maior índice de avaliações são recomendados para os usuários do sistema. Por exemplo, 80% dos usuários do sistema avaliam o filme X do gênero comédia com notas entre 4 e 5. Logo, o sistema irá realizar uma recomendação do filme X, para todos os usuários do sistema, devido o valor obtido nas avaliações dos usuários.

São utilizadas também as revisões dos usuários como saídas de dados para o SR. As revisões utilizam as opiniões que os usuários podem informar em um item ou serviço, as quais permitem que os usuários tenham livre expressão sobre a influência de um item ou serviço. Por exemplo, o usuário compra um filme, mas percebe que o áudio da gravação ficou com algum problema. Desta forma, o usuário pode relatar esta informação para os outros usuários do sistema.

Após apresentar a forma que as informações são obtidas pelo sistema e como às saídas são realizadas, é necessário também saber como utilizar as informações obtidas para gerar uma recomendação. Os métodos de recomendações mais utilizados são: seleção manual, baseado em atributo, correlação item a item e correlação usuário a usuário. O método de recomendação de seleção manual é considerado por Torres (2004) como a forma mais simples, na qual os editores e especialistas que conhecem os usuários propõem as recomendações.

O método de recomendação baseada em atributo utiliza os atributos que os itens ou serviços possuem para realizar a recomendação, segundo Torres (2004). Por exemplo: o usuário assiste filmes do gênero comédia; assim, quando um filme do gênero comédia for lançado, o sistema possivelmente fará uma recomendação deste novo filme para o usuário, pois o atributo gênero comédia é semelhante ao histórico de filmes assistidos por ele.

A correlação item a item como método de recomendação tenta identificar item ou serviço que de alguma forma tenha relação com item ou serviço que o usuário tenha mostrado interesse no momento que utilizou o sistema. Exemplo da utilização deste método é a venda casada, em que um usuário compra um filme de comédia e são oferecidas para o usuário, logo após a compra do filme, outras edições ou outros filmes de comédia. A correlação usuário a usuário no método de recomendação, leva em consideração usuários que têm preferências similares. Por exemplo, João tem como preferência no seu perfil filmes do gênero comédia; Maria também tem o gênero comédia como preferência no seu perfil. Maria faz a locação de um filme de comédia que João ainda não locou. Com base nisso, o sistema fará a recomendação do filme locado por Maria para João.

A utilização do SR é baseada em técnicas de recomendação, as quais contemplam os principais objetivos do SR, ou seja, consiste em realizar recomendação de itens ou serviços. Portanto, para desenvolver um SR é necessário realizar um estudo sobre as técnicas de recomendação. A seção 2.2 apresenta as Técnicas de Recomendação.

2.2. Técnicas de Recomendação

A definição dos perfis dos usuários é essencial para que as recomendações possam ser geradas, mas, além disso, é necessário aplicar algumas técnicas de recomendação para viabilizá-las, são elas: Filtragem Baseada em Conteúdo e Filtragem Colaborativa. A Filtragem Baseada em Conteúdo parte da premissa que, se um usuário consumiu um item ou utiliza algum serviço, então ele continuará consumindo um item ou serviço semelhante, ou seja, baseiam-se nas características dos itens ou serviços já utilizados (TORRES, 2004). Na Filtragem Colaborativa, as recomendações dependem de análise humana, dos itens ou serviços. Esta análise gera um conjunto de avaliações, isto é, as notas fornecidas pelos usuários sobre a satisfação de um item ou serviço. Com base nas avaliações realizadas pelos usuários são realizadas as recomendações.

A Figura 5 apresenta como ocorre o funcionamento do SR utilizando as técnicas de filtragem baseada em conteúdo e filtragem colaborativa.

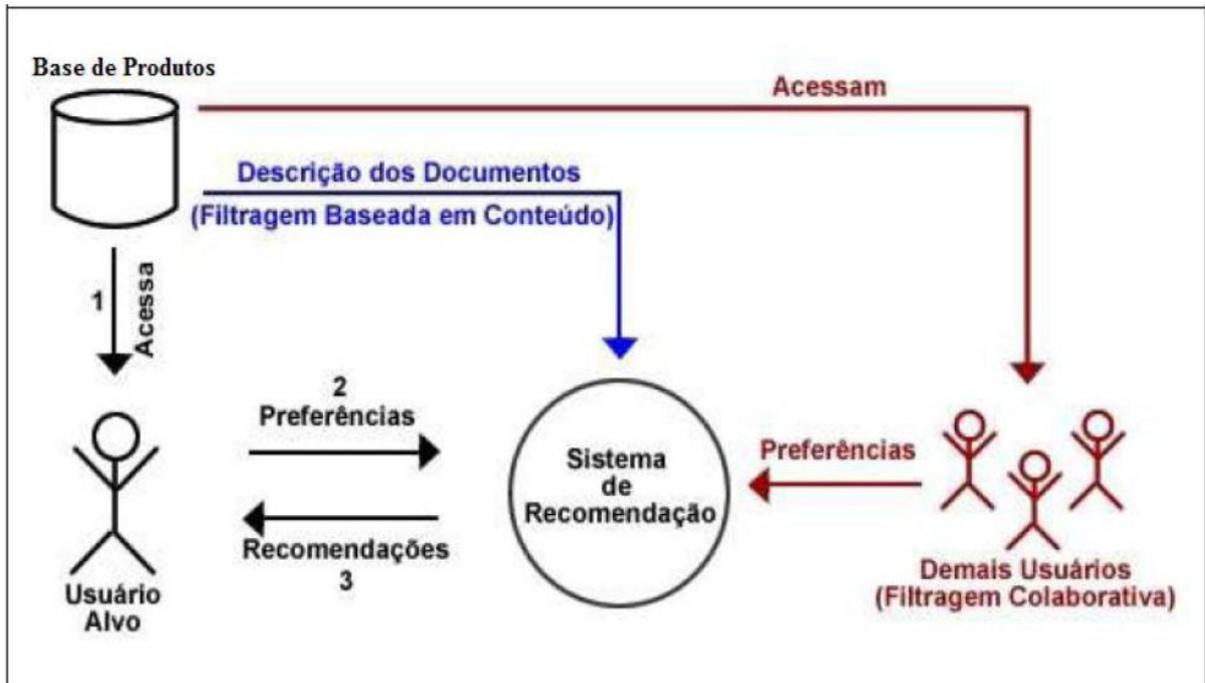


Figura 5 - Técnicas de Recomendação em um Sistema de Recomendação (GODOY NETO, 2009, p.48).

A Figura 5 demonstra o modelo genérico de um SR utilizando a técnica de recomendação colaborativa e baseada em conteúdo. A utilização da técnica de filtragem colaborativa leva em consideração as preferências do usuário e dos demais usuários do sistema, desta forma realiza uma comparação entre o perfil do usuário alvo e os perfis dos demais usuários assim verificam quais perfis dos usuários são mais similares com o do usuário alvo.

Já a utilização da filtragem baseada em conteúdo necessita apenas do perfil do usuário e a base de produtos, pois assim é realizada uma comparação com a preferência contida no perfil do usuário com o conteúdo dos produtos, desta forma é selecionado o produto que possa ser de interesse do usuário e recomendando ao usuário alvo. A seguir na seção 2.3 será abordada a técnica de Filtragem Baseada em Conteúdo.

2.3. Filtragem Baseada em Conteúdo

A Filtragem Baseada em Conteúdo (FBC) é responsável por realizar uma análise sobre o conteúdo dos itens e do perfil dos usuários (MILANI e CAZELLA, 2007). A Filtragem Baseada em Conteúdo compara a semelhança do conteúdo dos itens para realizar uma recomendação, desta forma leva em conta os itens adquiridos pelo

usuário anteriormente e a preferência no perfil, recomendando assim, itens que tenham a mesma similaridade de conteúdo (TORRES, 2004).

Para utilizar a FBC é preciso saber se um item possui alguma similaridade com outro item. Assim, é preciso realizar a comparação entre o conteúdo do item com o conteúdo dos itens já adquiridos anteriormente pelo usuário ou com os dados armazenados no perfil do usuário. No perfil do usuário são armazenadas as palavras-chave, com estas palavras armazenadas é possível realizar uma correlação com itens ou serviços que venham a ser de preferência do usuário. Os itens já consumidos pelo usuário também possuem conteúdo sobre a descrição e as suas palavras-chave.

Um exemplo que pode ser aplicado na FBC é a seguinte situação: uma loja virtual de filmes disponibiliza vários filmes diariamente sobre os mais variados gêneros, como drama, comédia e terror. A loja virtual disponibiliza todos os filmes da base de dados para os usuários. Na loja virtual é realizado o cadastro dos usuários e são armazenados nos perfis do usuário quais gêneros são de sua preferência. O usuário X tem armazenado em seu perfil o gênero drama e terror, como preferência, além das suas preferências são salvas no perfil histórico de pesquisas conforme filmes locados. Como a quantidade de filmes disponível para os usuários em uma loja virtual é muito grande, o usuário X perde muito tempo procurando um filme que atenda a sua preferência.

Com os dados fornecidos no ato do cadastro e com os dados de locações anteriores do usuário, é possível realizar recomendação para o usuário X a partir dos históricos salvo no perfil do usuário, realizando uma comparação com o conteúdo contido nos filmes, assim tornando mais ágil a busca por um filme que satisfaça o interesse do usuário. Por exemplo, o histórico do filme X é semelhante ao conteúdo do filme Z e Y, com esta comparação é possível recomendar filmes aos usuários.

A vantagem de se usar um sistema de Filtragem Baseada em Conteúdo é que todos os filmes armazenados na base de dados da loja virtual possui algum tipo de conteúdo, seja ele: comentários, descrição, gênero ou sinopse. Estes dados armazenados são utilizados para realizar similaridade com os dados contidos no

perfil do usuário. Assim o usuário obtém uma recomendação mais próxima do que deseja, pois são utilizadas apenas informações do usuário alvo (BORGES, 2011).

Uma desvantagem de se utilizar a FBC é que, caso o usuário queira receber informação de um novo gênero de filme que não esteja como preferência no seu perfil, ou que ainda não tenha consultado algum filme daquele gênero que deseja receber recomendação, ficará impossibilitado de receber a recomendação, pois a Filtragem Baseada em Conteúdo necessita de palavra-chave, comentários, descrição, sinopse e outras informações para verificar a similaridade (BORGES, 2011).

A seção 2.4 apresenta a Filtragem Colaborativa, descrevendo funcionamento da Filtragem Colaborativa e algumas técnicas.

2.4. Filtragem Colaborativa

A Filtragem Colaborativa (FC) tem sido umas das técnicas mais utilizadas para recomendar algo aos usuários (SCHAFER ET AL, 2001). A FC utiliza a análise humana no processo de avaliar um item ou serviço e, ao contrário da FBC, não analisa o conteúdo dos itens ou serviços para recomendar algo. A FC realiza análise entre preferências comuns em um grupo de pessoas, ou seja, analisa os perfis de outros usuários para recomendar algo a um usuário alvo. Segundo Sampaio (2007), “o processo de sugestão está apoiado inteiramente na similaridade entre os usuários de sistema e não na similaridade dos itens de informação”.

Um exemplo em que pode ser aplicada a FC no contexto de filmes: quando uma pessoa assiste a um filme que tenha gostado, provavelmente irá fazer a recomendação para outras pessoas assistirem o mesmo. Em um sistema de FC para a recomendação de filmes, o usuário teria que realizar avaliações dos filmes a partir da atribuição de notas. Esta avaliação pode ocorrer de forma numérica de modo que quanto maior o número, maior é a avaliação dada pelo usuário, ou também pode ocorrer apenas por opinião do usuário, ou seja, apenas dizer se gostou ou não do filme.

A Filtragem Colaborativa possui diversas vantagens e desvantagens. Uma das vantagens de se usar técnica de FC está na possibilidade de se fazer recomendação de itens ou serviços a partir da similaridade do perfil baseando nas avaliações dos usuários.

Uma desvantagem de se usar Filtragem Colaborativa está na inclusão de itens ou serviços na base de dados do sistema, pois se tratando de algo recém-cadastrado, não haverá avaliações, positivas ou não. Logo, não serão realizadas as recomendações sobre os itens ou serviços novos, até que alguns usuários comecem a realizar avaliações sobre eles (MILANI e CAZELLA, 2007).

O processo de recomendação utilizando a técnica FC pode ser dividido em três passos de forma generalizada, que são eles: representação dos dados de entrada, formação de vizinhança e geração de recomendação (QUEIROZ, 2003).

- **Representação dos dados de entrada:**na representação dos dados de entrada na técnica FC, os usuários expressam suas preferências avaliando itens ou serviços do sistema. As avaliações podem ser coletadas tanto a partir de indicações explícitas quanto implícitas. As avaliações revelam o interesse que os usuários demonstram sobre itens e serviços específicos e são armazenadas no perfil dos usuários.
- **Formação de Vizinhança:**para realizar uma recomendação, o sistema compara o perfil do usuário alvo com os perfis de outros usuários para gerar a similaridade de preferências entre os usuários.
- **Geração de Recomendação:**a partir do cálculo da vizinhança, é possível que o sistema realize recomendação para o usuário alvo, ou seja, o sistema sugere para o usuário alvo os itens ou serviços que podem ser do seu interesse.

Para exemplificar o processo de FC, o Quadro 1 apresenta um cenário em que alguns usuários realizam avaliações em filmes. As notas que compõem cada célula do quadro estão no intervalo de 1 a 10, em que a nota 1 indica a nota mínima sobre o item avaliado, e 10 a nota máxima. A célula que não contém valor, ou seja, não possui nota indica que o filme não foi avaliado pelo usuário e a célula que

possui o sinal de interrogação indica que o sistema deve prever qual a sua relevância do filme5 para o usuário alvo da recomendação, neste caso a Maria.

#Usuários	Filme1	Filme2	Filme3	Filme 4	Filme5
Rafael	1	1	3	3	1
Karol	4	4	3	1	5
Paulo	10	2	4	1	10
Maria	9	1	5	1	?

Quadro 1 Exemplo de Recomendação Baseado em FC

Como pode ser observado no Quadro 1, ele apresenta as avaliações de preferência de todos os usuários. O próximo passo do sistema de FC deve ser realizar a identificação da vizinhança do usuário alvo, com base nos filmes que os usuários tenham avaliações em comum. No exemplo do Quadro 1, o usuário alvo é a Maria, e os seus vizinhos são Paulo, Karol e Rafael. Paulo possui filmes avaliados em comum com Maria, ou seja, filmes com as notas das avaliações semelhantes. Com essa análise a técnica de FC deve então considerar as notas do item e serviço dos usuários que tenha mais similaridade com a nota do usuário alvo para recomendação, análise que é apresentada a seguir.

Neste exemplo, o usuário que possui uma nota mais similar com Maria é Paulo pelo fato de Paulo ter avaliado os filmes 1, 2, 3 e 4 com as notas (10, 2, 4 e 1), e Maria ter avaliado os mesmos filmes com as notas (9, 1, 5, e 1). Com estas similaridades contidas nas avaliações de Paulo e Maria, a possibilidade é que o filme 5 seja recomendado para Maria.

Na seção 2.4.1 serão apresentadas as etapas utilizadas para realizar uma recomendação baseada em FC, que incluem, também, a definição matemática de cada uma destas etapas. Além disso, em cada uma destas etapas são utilizadas equações para realizar os cálculos para realizar uma recomendação com técnica FC.

2.4.1. Técnica de Filtragem Colaborativa

Para aplicar a técnica de FC, há diferentes abordagens e algoritmos que podem ser utilizadas para realizar este processo de forma eficiente. Estas abordagens e estes

algoritmos funcionam a partir de avaliações realizadas nos itens ou serviços. Estes valores são as avaliações que os usuários atribuem aos produtos ou serviços. As avaliações são realizadas em uma escala numérica, por exemplo, de 1 a 10.

A aplicação da FC, conforme já mencionado, é formada por dois passos: formação de vizinhança e calcular predição, ou seja, nota que o usuário daria a determinado item ou serviço. No primeiro passo ocorre a formação de vizinhança do usuário alvo. A vizinhança é formada por um conjunto de usuários do sistema. Para realizar a formação da vizinhança, é utilizadas as avaliações armazenadas no perfil dos usuários. Sobre estas avaliações é aplicado um algoritmo para ser calculada a similaridade entre as avaliações, para verificar quais avaliações estão mais próximas da avaliação do usuário alvo, ou seja, a partir deste algoritmo é possível determinar a vizinhança do usuário alvo. Assim pode-se definir vizinhança como usuários que têm perfis semelhantes, ou seja, possuem similaridades em avaliações realizadas em itens ou serviços.

A seguir, na Equação 1, é apresentada a fórmula do cálculo do coeficiente de Pearson, que é aplicada para calcular a similaridades que os vizinhos de um usuário têm sobre um determinado item ou serviço a partir das suas avaliações.

$$W_{a,u} = \frac{\sum_{i=1}^m (R_{a,i} - \bar{R}a) * (R_{u,i} - \bar{R}u)}{\sqrt{\sum_{i=1}^m (R_{a,i} - \bar{R}a)^2} * \sqrt{\sum_{i=1}^m (R_{u,i} - \bar{R}u)^2}} \quad (1)$$

A Equação 1 apresenta o coeficiente de Pearson, que é utilizado para calcular a similaridade entre os perfis dos usuários conforme *apud* TORRES (2004). Os perfis dos usuários são formados pelas avaliações que são dadas aos itens ou serviços. Na fórmula, $W_{a,u}$ é a correlação de um usuário alvo a em relação a outro usuário u . $R_{a,i}$ é a avaliação que o usuário a realiza sobre o item i ; $\bar{R}a$ é a média das avaliações do usuário ativo, a ; m é a quantidade de itens; $R_{u,i}$ é a avaliação que o usuário u atribuiu para o item i ; $\bar{R}u$ é a média das avaliações do usuário alvo u , conforme TORRES (2004).

Com o coeficiente de Pearson é possível obter o resultado das similaridades entre o usuário alvo para recomendação e os usuários com perfis similares, de modo

que só serão utilizados no cálculo de predição os usuários com perfis mais similares, ou seja, perfis de usuários que tenha uma proximidade em notas oferecidas em avaliações de itens ou serviços. O resultado obtido com o coeficiente de Pearson está em uma escala numérica entre -1 e 1 para verificar a similaridade, ou seja, os usuários que obtêm notas mais próximas de 1 são os mais similares. Após executar os passos anteriores para calcular as similaridades dos usuários, o próximo passo é realizar as predições para o usuário alvo, considerando os valores obtidos pela Equação 1.

Segundo Milioli (2007), “para geração de uma recomendação, faz-se necessário o cálculo da predição, suposta avaliação que o usuário daria para este item, caso o consumisse”. O cálculo da predição é realizado por meio de uma média ponderada das avaliações realizadas pelos vizinhos do usuário alvo. A Equação 2 apresenta como é calculada a predição.

$$P_{a,i} = \bar{R}_a + \frac{\sum_{u=1}^n (R_{u,i} - \bar{R}_u) * W_{a,u}}{\sum_{u=1}^n |W_{a,u}|} \quad (2)$$

De acordo com a Equação 2, a fórmula apresenta como calcular a predição, em que: $P_{a,i}$ é usado para calcular a predição do usuário alvo a com o item ou serviço i , \bar{R}_a é a média das avaliações do usuário alvo; \bar{R}_u é a média das avaliações dos vizinhos u ; $W_{a,u}$ é a similaridade do usuário alvo em relação ao usuário vizinho u ; n indica a quantidade de item ou serviço (MILIOLI, 2007).

A Equação 2 é utilizada para calcular como cada vizinho avaliaria um item ou serviço, utilizando a similaridade com o usuário alvo. Desta forma, deve ser calculada a predição para todos os usuários que são vizinhos do usuário alvo (TORRES, 2004). A utilização desta fórmula retorna a idéia de que o usuário alvo poderia avaliar determinados itens ou serviços com uma nota X, esta nota X seria retornada com base nas avaliações dos vizinhos do usuário alvo. Assim o SR pode determinar se o item ou serviço será recomendado ao usuário ou não de acordo com o valor encontrado no cálculo de predição.

Na seção 2.4.2 será apresentado um exemplo utilizando as equações da técnica de FC utilizando as avaliações contidas no Quadro 1. Desta forma

aparentará um passo a passo de como é a aplicação das equações com valores na técnica de FC para realizar uma recomendação.

2.4.2. Exemplo de Filtragem Colaborativa

Para aplicação do exemplo da FC simulando valores reais serão utilizados os valores contidos no Quadro 1, que são supostas avaliações realizadas por usuários. Observando o Quadro 1 pode perceber que Maria é o usuário alvo e tem avaliações similares às de Paulo. Para que seja comprovado que os usuários são similares, é necessário coletar as avaliações dos itens em comum dos dois usuários. A usuária Maria avaliou os filmes 1, 2, 3 e 4 com notas 9, 1, 5, e 1. Já o usuário Paulo avaliou os mesmos filmes com notas 10, 2, 4 e 1.

Para realizar o cálculo de similaridade entre os usuários é necessário aplicar a equação do coeficiente de correlação de Pearson, que utiliza as médias das avaliações de Maria (4) e Paulo (4,25). Na Equação 3 é apresentado o passo a passo do cálculo do coeficiente de correlação de Pearson, entre o usuário Maria (usuário alvo) e Paulo. Com o resultado do cálculo entre os dois usuários é possível obter o valor de similaridade dos usuários sobre itens ou serviços, o resultando será utilizado para calcular a predição entre os usuários.

$$\begin{aligned}
 W_{a,u} &= \frac{((9 - 4) * (10 - 4,25)) + ((1 - 4) * (2 - 4,25)) +}{\sqrt{(9 - 4)^2 + (1 - 4)^2 + (5 - 4)^2 + (1 - 4)^2} *} \\
 &\frac{((5 - 4) * (4 - 4,25)) + ((1 - 4) * (1 - 4,25))}{\sqrt{(10 - 4,25)^2 + (2 - 4,25)^2 + (4 - 4,25)^2 + (1 - 4,25)^2}} \\
 &= \frac{45}{\sqrt{44} * \sqrt{48,75}} = \frac{45}{6,63 * 6,98} = 0,97 \quad (3)
 \end{aligned}$$

O desenvolvimento da Equação 1, a fórmula apresenta como calcular o coeficiente de correlação de Pearson. Na Equação 3 os valores destacados em azul são as notas do usuário alvo, ou seja, do usuário Maria, e os valores em preto são os valores do usuário Paulo.

Repetindo-se o cálculo da Equação 3 com os valores do usuário alvo, ou seja, da usuária Maria com os valores de todos os outros usuários do Quadro 1, pode

comprovar que o usuário Paulo é mais similar a Maria. Na Tabela 1 são apresentados todos os usuários do Quadro 1 com os respectivos valores de similaridade com Maria.

Tabela 1 Similaridades entre usuários utilizando Pearson

Usuários	Similaridade
Rafael	-0,30
Karol	0,49
Paulo	0,97

De acordo com os dados apresentados na Tabela 1, pode-se obter o valor de 0,97 de similaridade entre o usuário Maria e Paulo. Desta forma pode-se afirmar que Paulo é bastante similar a Maria devido ao valor obtido na Equação 3 ser muito próximo a 1.

Após realizar o cálculo das similaridades dos usuários, o próximo passo para gerar uma recomendação é calcular a predição, ou seja, verificar uma suposta avaliação que o usuário Maria daria ao item 5, caso consumisse (RESNICK, 1994). Este cálculo de predição pode ser realizado através da equação de predição, tal como será apresentada na Equação 4.

$$\begin{aligned}
 P_{\text{maria},5} &= 4 + \frac{((1 - 2) * (-0,30)) + ((5 - 3) * 0,49) + ((10 - 4,25) * 0,97)}{0,30 + 0,49 + 0,97} \\
 &= 4 + \frac{6,85}{1,76} = 4 + 3,89 = 7,89(4)
 \end{aligned}$$

De acordo com os cálculos realizados na Equação 4, que faz a predição, caso Maria consumisse o item 5, a avaliação supostamente realizada por ela seria 7,89. Este resultado influencia diretamente em itens a serem recomendados a Maria, pois considerando em um intervalo de avaliação de notas entre 0 e 10, a nota 7,89 está próximo da nota máxima. Para gerar esta predição, como é possível observar na Equação 4, são utilizadas as avaliações dos usuários ao item 5, a média de avaliações dos usuários e o grau de similaridade dos usuários com o usuário alvo.

Como foi possível observar nas equações anteriores, foi apresentado os cálculos que a Filtragem Colaborativa realiza antes de realizar uma recomendação.

Estes cálculos serão realizados pelo protótipo para gerar recomendações aos usuários.

3 MATERIAIS E MÉTODOS

3.1. Materiais

Os materiais utilizados no desenvolvimento deste trabalho foram divididos em duas partes, que são: Software e Fontes Bibliográficas.

- **Software:** No desenvolvimento do protótipo de recomendação foi utilizado o ambiente de desenvolvimento Eclipse Luna, junto com a API Apache Mahout Taste, implementado com a linguagem de programação Java. Nas subseções abaixo serão apresentadas as tecnologias e ferramentas utilizadas neste trabalho.
- **Fontes Bibliográficas:** Os materiais utilizados para produção bibliográfica foram de fontes como: teses, artigos, dissertações, publicações científicas e textos técnicos sobre Sistema de Recomendação.

3.1.1. API Apache Mahout Taste

A utilização desta API é responsável em gerar recomendações. A API é um mecanismo de recomendação de código aberto desenvolvida em Java e fornece uma biblioteca Java utilizada na técnica de FC, classificação e clustering para realizar uma recomendação. O desenvolvimento desta API é um projeto realizado pela *Apache Software Foundation*. Na Figura 6 é apresentada uma estrutura da API.

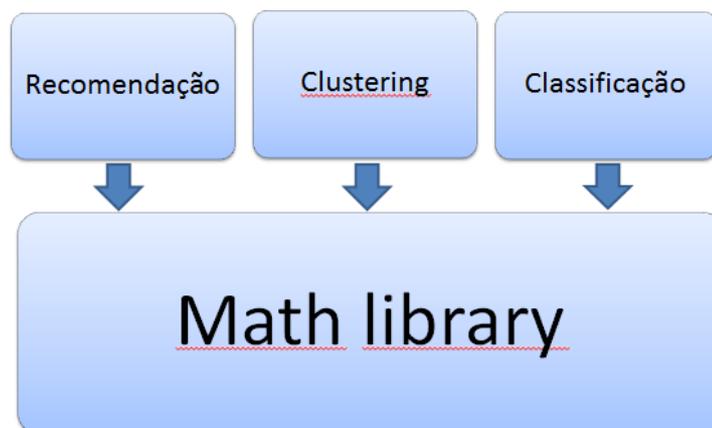


Figura 6 API Apache Mahout Taste.

Como pode ser observada na Figura 6, a utilização da API é dividida em módulos, ou seja, ela possui algoritmos específicos para cada problema voltado para os contextos, que são recomendação, clustering e classificação. Para este trabalho,

o módulo importante é o de recomendação, que utiliza a *Math Library*, uma biblioteca que define operações matemáticas que são utilizadas no módulo de Recomendação. Na *Math Library* estão algoritmos que realizam os cálculos vizinhança e predição, conforme os dados contidos na base de dados e, assim, podem-se obter resultados de uma possível recomendação aos usuários do sistema.

A versão da API utilizada neste trabalho é a versão 0.9 que pode ser obtida no próprio site da *Apache Mahout* Taste. Além da API é necessário instalar alguns pacotes, que são: *Mahout-core-0.9.jar*, *uncommons-math-1.0.2.jar*, *slf4-nop-1.6.1.jar*, *slf4j-api-1.6.1.jar*. ao projeto, clicar com o botão direito na pasta e selecionar a opção add JAR.

A biblioteca mahout-core 0.9 possui algumas classes já implementadas. Neste trabalho foram utilizadas as seguintes classes: *DataModel*, *UserSimilarity*, *UserNeighborhood*, *Recommender*, *RecommenderItem*.

Com as classes e interfaces apresentadas acima, é possível realizar o desenvolvimento de implementação do protótipo de recomendação de filmes utilizando a técnica de Filtragem Colaborativa. Na próxima seção é apresentada a metodologia utilizada para o desenvolvimento deste trabalho.

3.2. Metodologia

O presente trabalho tem como objetivo apresentar um estudo sobre SR, especificando a técnica de FC, para que seja possível desenvolver um protótipo de recomendação de filmes. A técnica de FC definida para o desenvolvimento do protótipo utiliza as avaliações realizadas pelos usuários para recomendar filmes que sejam de interesses dos usuários. Com as avaliações são aplicados alguns cálculos como de vizinhança, ou seja, realizar um cálculo para verificar quais usuários possuem os perfis semelhantes e predição, que é importante para prever qual seria a nota que o usuário poderia avaliar um item ou serviço caso consuma se, com o resultado e recomendado aos usuários filmes mais próximos de suas preferências.

Assim, o desenvolvimento deste trabalho foi dividido em algumas etapas. A primeira etapa destinou-se a um estudo aprofundado sobre os conceitos fundamentais em um SR, tais como definição e aplicação. Com isto, foi possível entender o porquê da utilização das técnicas de recomendação e compreender como é o funcionamento de um SR.

Com este entendimento das técnicas de recomendações, realizou-se uma contextualização de algumas técnicas de recomendação, como a FBC e FC, que podem ser utilizadas em um SR. Para a FBC realizou-se apenas um rápido estudo sobre os conceitos e quando utilizá-los. Já na FC realizou-se um estudo aprofundado sobre os conceitos e as principais técnicas utilizadas para recomendar itens ou serviços aos usuários.

Ainda na FC, foram realizados testes manuais sobre as técnicas, que envolveram cálculos como coeficiente de correlação de Pearson e predição, pois realizando os cálculos de forma detalhada seria possível compreender o funcionamento da recomendação utilizando FC. Estas etapas realizadas foram primordiais para o andamento do trabalho, pois são os cálculos de correlação de Pearson e predição que realizam o processo de recomendar itens aos usuários de um SR. Após o entendimento de como são realizados os cálculos na FC para recomendar itens, pode-se definir qual é a melhor técnica que pode ser utilizada para o desenvolvimento do protótipo proposto neste trabalho, uma vez que o propósito é recomendar filmes baseando-se nas avaliações dadas pelos usuários.

A próxima etapa, também primordial para o desenvolvimento deste trabalho, foi à escolha da linguagem de programação, IDE e a API para auxiliar no desenvolvimento do protótipo. Neste projeto será utilizada a IDE NetBeans 8.0, utilizando a linguagem JAVA e a API Apache Mahout Taste para o desenvolvimento do projeto. Na seção 3.1 é realizado um detalhamento sobre a API.

Para que seja possível testar o protótipo desenvolvido foi utilizada a base de dados Movielens 100k, disponível em: <http://www.grouplens.org/node/73>. Esta base utiliza avaliações reais realizadas por usuários em diversos gêneros de filmes. A base de dados possui aproximadamente 65 mil avaliações realizadas por aproximadamente 943 usuários. Estas informações serão utilizadas para apresentar o funcionamento do protótipo. Os dados foram coletados através do site da Movielens (<http://www.movielens.umn.edu>). Na Tabela 2 é possível observar a estrutura da base de dados, na qual estão armazenadas as avaliações sobre os filmes.

Tabela 2 Modelo da Base de Dados

Usuário	Item	Avaliação
1	1	5
2	1	4
5	2	4
7	2	2

Como pode ser observada na Tabela 2, a base de dados possui os usuários, itens e avaliações. Cada usuário realiza avaliações sobre determinados itens. Estas avaliações podem ocorrer em uma escala numérica entre 0 e 5. Com os dados das avaliações, o id dos itens e o id dos usuários serão utilizados a API para realizar o cálculo de similaridade e também de predição. Com estas informações é possível gerar recomendação baseada apenas nas avaliações dos usuários, não levando em consideração nenhuma informação do perfil a não serem avaliações realizadas em alguns itens.

A base de dados também possui informações sobre o perfil dos usuários. Estas informações podem ser utilizadas para gerar uma recomendação personalizada para os usuários ou para gerar recomendação baseando em um

grupo de pessoas específico, ou seja, recomendar para uma média de idade. Na tabela 3 é possível observar a estrutura da base de dados, na qual estão armazenadas as informações do perfil dos usuários.

Tabela 3 Modelo da Base de Dados Perfil dos Usuários

Usuário	Idade	Sexo	Ocupação	Cep
1	24	M	Estudante	77000000
2	53	M	Executivo	77000003
5	23	F	Administrador	77000000
7	35	M	Advogado	77000012

Como pode ser observada na Tabela 3, a base de dados possui os usuários, idade do usuário, sexo e CEP. Para a realização dos testes manuais foi desenvolvido uma base de dados utilizando esta mesma estrutura. A base de testes manuais são compostas pelos 5 primeiros usuários da base do Movielens, contem 10 filmes e 30 avaliações. Na Seção 4, serão apresentados os resultados e discussão que o desenvolvimento do protótipo obteve.

4 RESULTADOS E DISCUSSÃO

4.1. Arquitetura da Ferramenta

Para gerar as recomendações, o protótipo desenvolvido parte do princípio já tenha a base de dados e os perfis dos usuários pré-definidos. Desta forma, há a necessidade de repassar algumas informações para o protótipo realizar recomendações, que são elas: usuário alvo, quantidade máxima de vizinhos e quantidade máxima de recomendação. A Figura 7, a seguir, apresenta a arquitetura do protótipo.



Figura 7 Arquitetura do Sistema de Recomendação

Como pode ser observada na Figura 7, a arquitetura do SR possui algumas etapas. A primeira etapa do sistema de recomendação é realizar o envio dos dados do usuário alvo, estas informações enviadas são: o ID do usuário, quantidade máxima de vizinhança e quantidade máxima de recomendação que deseja retornar ao usuário alvo. O SR recebe estas informações e repassa para próxima etapa, onde é realizado a Filtragem Colaborativa, a FC utiliza a API *Apache Mahout Taste* que auxilia o SR em gerar recomendação.

A próxima etapa do processo é apresentar os filmes recomendados conforme o retorno realizado pela API. Com estas recomendações retornadas, é passado para a próxima etapa que é retornar para o usuário alvo os filmes recomendado pela API. Os filmes retornados para o usuário alvo são as recomendações realizadas pelo protótipo.

Na seção 4.2 serão apresentadas as etapas utilizadas para realizar a implementação do protótipo, utilizando a Filtragem Colaborativa e a API *Apache Mahout Taste*.

4.2. Implementação

Nesta seção serão apresentados os passos utilizados no desenvolvimento do protótipo e suas funcionalidades.

O protótipo desenvolvido neste trabalho foi dividido em três etapas para que a recomendação possa ser gerada. A primeira etapa permite que seja informado ao protótipo qual o usuário alvo, ou seja, qual usuário irá receber a recomendação. A segunda etapa permite escolher a quantidade máxima de vizinhança do usuário alvo e a última etapa permite limitar a quantidade máxima de recomendação que o protótipo retornará para o usuário alvo. Através destes três passos, é possível que o sistema gere uma recomendação.

Para gerar as recomendações foi criada a classe `ItemRecommend`, esta é a classe principal utilizada do protótipo. Esta classe é utilizada para analisar os dados dos usuários, calcular a similaridade entre o usuário alvo e seus vizinhos e como saída gerar recomendação para o usuário alvo. A Listagem 1 apresenta a classe `ItemRecommend`, ou seja, a classe utilizada para recomendar, seguindo os padrões definidos pela API *Apache Mahout Taste*.

```

22
23 public class ItemRecommend {
24     public static void main(String[] args){
25         try{
26             Scanner v = new Scanner(System.in);
27             int qdtvizinhanca,UsuarioAlvo,qtdmaxrecomendacao;
28
29             DataModel dm = new FileDataModel(new File("data/teste.csv"));
30
31             UserSimilarity usuariosSemelhantes = new PearsonCorrelationSimilarity(dm);
32
33             System.out.println("Informe a Quantidade Máxima de vizinhança !");
34             qdtvizinhanca = v.nextInt();
35
36             UserNeighborhood vizinhanca = new NearestUserNeighborhood(qdtvizinhanca, usuariosSemelhantes, dm);
37
38             Recommender recomendacao = new GenericUserBasedRecommender(dm,vizinhanca, usuariosSemelhantes);
39
40             System.out.println("Informe o Usuário Alvo Para Recomendação !");
41             UsuarioAlvo = v.nextInt();
42
43             System.out.println("Informe o Quantidade Máxima de Recomendação !");
44             qtdmaxrecomendacao = v.nextInt();
45
46             List<RecommendedItem> recomendacoes = recomendacao.recommend(UsuarioAlvo, qtdmaxrecomendacao);
47             System.out.println("FILMES RECOMENDADO PARA O USUÁRIO.");
48             for (RecommendedItem recommendedItem : recomendacoes) {
49                 System.out.println(recommendedItem.getItemID());
50             }
51         } catch (IOException e){
52             System.out.println("Ocorreu Erro");
53             e.printStackTrace();
54         } catch (TasteException e) {
55             System.out.println("erro na API TASTE");
56             e.printStackTrace();
57         }
58     }
59 }

```

Listagem 1 Classe Responsável em Gerar Recomendação Colaborativa

Como pode ser observado na Listagem 1, à classe *ItemRecommend* utiliza 4 Interfaces, que são elas: *DataModel* na linha 29, *UserSimilarity* na linha 31, *UserNeighborhood* na linha 36 e *Recommender* na linha 38, que utilizam métodos para que sejam geradas as recomendações. As interfaces, métodos e classes serão detalhados a seguir:

- Interface *DataModel*: A implementação desta interface é realizada pela classe *FileDataModel*, que utiliza um arquivo CSV. TXT, .Date outros, como fonte de dados, ou seja, como banco de dados. A Listagem 2 apresenta a interface *DataModel*.

```

28
29     DataModel dm = new FileDataModel(new File("data/teste.csv"));
30

```

Listagem 2 Interface DataModel

Como pode ser observado na Listagem 2 é necessário informar para interface *DataModel* qual o nome do arquivo CSV, como pode ser observado. Como a classe *FileDataModel* utiliza um arquivo CSV, os dados contidos neste

arquivo devem estar na seguinte ordem *IdUsuário*, *IdFilme* e *AvaliacaoFilme*. E os dados devem ser separados por vírgula.

- Interface *UserSimilarity*: A implementação desta interface é realizada pela classe *PearsonCorrelationSimilarity*, que é responsável em implementar um algoritmo de correlação, ou seja, implementa o algoritmo de correlação de Pearson. Este algoritmo é de grande importância para o sistema de recomendação, pois é responsável por realizar a identificação dos usuários que possuem correlações, ou seja, identificar os usuários que são mais semelhantes ao usuário alvo. A Listagem 3 apresenta a interface *UserSimilarity*.

```
30
31     UserSimilarity usuariosSemelhantes = new PearsonCorrelationSimilarity(dm);
32
```

Listagem 3 Interface UserSimilarity

Como pode ser observado na Listagem 3 para que a correlação seja realizada, é necessário especificar qual base de dados será utilizada, é passado por parâmetro à base de dados DM do tipo *DataModel*.

- Interface *UserNeighborhood*: A implementação desta interface é realizada pela classe *NearestNUserNeighborhood*, classe que é responsável por realizar a identificação da vizinhança do usuário alvo, ou seja, os usuários que possuem perfis semelhantes ao usuário alvo. A Listagem 4 apresenta a interface *UserNeighborhood*.

```
35
36     UserNeighborhood vizinhanca = new NearestNUserNeighborhood(qtdvizinhanca, usuariosSemelhantes, dm);
37
```

Listagem 4 Interface UserNeighborhood

Como pode ser observado na Listagem 4 para a utilização desta interface é necessário passar 3 parâmetros: *qtdvizinhanca*, *usuariosSemelhantes* e *DM*. No primeiro parâmetro é passado o número máximo de vizinhança que a recomendação terá como base para recomendar. No segundo parâmetro é passado o algoritmo que utilizará para realizar o cálculo de correlação entre os usuários (este cálculo foi definido na criação do objeto *UserSimilarity*). E o terceiro parâmetro é o modelo de dados do tipo *DataModel*.

- Interface *Recommender*: A implementação desta interface é realizada pela classe *GenericUserBasedRecommender*, que tem a responsabilidade de

gerar as recomendações para o usuário. A Listagem 5 apresenta a interface *Recommender*.

```
37
38 Recommender recomendacao = new GenericUserBasedRecommender(dm,vizinhanca, usuariosSemelhantes);
39
```

Listagem 5 Interface Recommender

Como pode ser observado na Listagem 5 para a utilização desta interface é necessário passar 3 parâmetros: *DM*, *Vizinhança* e *usuariosSemelhantes*.

- O método *recommend*: A implementação deste método ocorre através da interface *Recommend*. Este método tem como responsabilidade gerar a lista de recomendação para o usuário alvo. A Listagem 6 apresenta o método *recommend*.

```
45
46 List<RecommendedItem> recomendacoes = recomendacao.recommend(UsuarioAlvo, qtdmaxrecomendacao);
47
```

Listagem 6 Método Recommend

Como pode ser observado na Listagem 6 para geração desta lista de recomendação, é necessário repassar o *id* do usuário alvo e a quantidade máxima de recomendação que se deseja realizar para o usuário alvo. Na Listagem 6, é passado o usuário alvo e a quantidade máxima de recomendação.

Com as classes e interfaces apresentadas na Listagem 1, foi possível realizar todas as etapas para implementação da Filtragem Colaborativa no protótipo desenvolvido. Na seção 4.3 será apresentado os testes realizados sobre o protótipo desenvolvido utilizando as classes e interfaces da Listagem 1.

4.3. EXPERIMENTOS

O objetivo dos testes realizados é apresentar os resultados obtidos, ou seja, verificar as recomendações retornadas pelo protótipo de recomendação utilizando a Filtragem Colaborativa.

4.3.1. Cenário

Para aplicação de teste sobre o protótipo desenvolvido foi utilizada a base de dados Movielens100k. A base de dados do Movielens100k utiliza dados reais. Estes dados

foram adquiridos através de avaliações realizadas por diversos usuários em algum site parceiro do GrupMovielens. Esta base de dados possui aproximadamente 65 mil avaliações realizadas por aproximadamente 943 usuários. A Figura 8, a seguir, apresenta o modelo relacional da base de dados do Movielens.

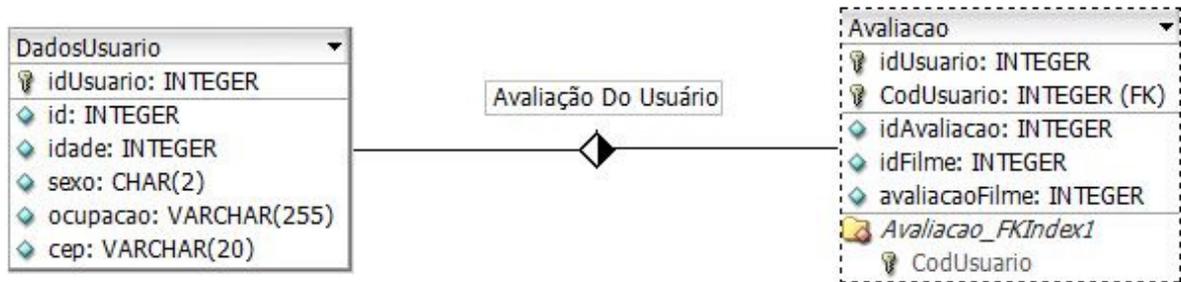


Figura 8 Modelo Relacional da Base Movielens Adaptada

A Figura 8 apresenta o modelo relacional da base de dados Movielens utilizada no protótipo. Como pode ser observada a base de dados possui duas tabelas se relacionando DadosUsuario e Avaliacao. A tabela DadosUsuários está armazenada as informações do perfil do usuário: *id*, *Idade*, *sexo*, *ocupação* e *CEP*; já na tabela Avaliacao estão armazenadas as avaliações realizadas pelos usuários sobre algum filme, a tabela possui as seguintes informações: *IdAvaliação*, *IdFilme*, *avaliacaoFilme* e como chave estrangeira o *CodUsuario*.

Para compreender os testes realizados e a realização de testes manuais com o protótipo de recomendação, os dados da base Movielens foram reduzidos. Nos testes foram utilizados apenas os 5 primeiros usuários da base de dados e apenas 30 avaliações com notas aleatórias em 10 filmes foram simuladas. A Figura 9, a seguir, apresenta a base de dados reduzida utilizada pra teste.

Dados dos Usuários				
idusuario	idade	sexo	ocupacao	cep
1	24	M	other	85711
2	53	F	technician	94043
3	23	M	technician	32067
4	33	M	writer	43537
5	24	F	other	15213

Avaliações dos Usuários		
idusuario	idfilmes	avaliacaofilme
1	8	2
1	10	3
1	2	2
1	4	1
1	3	5
1	1	4
2	7	3
2	2	3
2	5	4
2	4	4
2	1	2
2	9	4
3	4	4
3	3	1
3	6	3

Cont... →

3	8	4
3	10	2
3	2	3
4	3	3
4	4	5
4	10	2
4	7	3
4	2	5
4	1	2
5	3	3
5	5	1
5	2	1
5	6	3
5	1	4
5	9	4

Figura 9 Base de Teste

A redução na base de dados foi realizada em decorrência da necessidade de realizar os testes manualmente para comparar os resultados obtidos com os resultados do protótipo. Estes testes envolvem cálculos da Filtragem Colaborativa, portando seria inviável realizar esses cálculos manualmente utilizando uma base de dados volumosa.

4.3.2. Resultados

O protótipo de recomendação desenvolvido no trabalho utiliza a Técnica de Filtragem Colaborativa, por isso foi necessário realizar 3 passos para gerar recomendações. O primeiro passo foi definir a quantidade máxima de vizinhança, que na Figura 10 utiliza-se 2 como vizinhança máxima. A definição da vizinhança é importante para que os cálculos realizados para gerar recomendação sejam feitos apenas com a quantidade máxima de vizinhança definida, ou seja, os usuários que são mais similares com o usuário alvo de acordo com o limite da vizinhança. O próximo passo foi informar quem é o usuário alvo, ou seja, qual usuário recebe as recomendações do protótipo, na Figura 10 utiliza-se o usuário 3 como usuário alvo. O terceiro passo foi informar ao sistema qual a quantidade máxima de recomendação, no exemplo da Figura 10 o sistema deve fazer no máximo 10 recomendações para o usuário alvo. A Figura 10 apresenta os resultados das recomendações geradas pelo protótipo.

```

<terminated> ItemRecommend [Java Application] C:\Program Files\Java\jre8\bin\jav
Informe a Quantidade Máxima de vizinhança !
2
Informe o Usuário Alvo Para Recomendação !
3
Informe o Quantidade Máxima de Recomendação !
10
FILMES RECOMENDADO PARA O USUÁRIO.
7
1

```

Figura 10 Resultados Gerados Pelo Protótipo

Como pode ser observado na Figura 10, os id 's dos filmes recomendados para o- usuário- 3 foram os filmes 3 e 7. Com os resultados retornados pelo protótipo, serão realizados alguns passos para apresentar como o protótipo obteve os resultados apresentados na Figura 10. Os passos foram divididos em etapas, apresentadas a seguir:

A primeira etapa realizada pelo protótipo é o cálculo de similaridade para cada usuário da base de dados que esteja dentro do limite da vizinhança máxima do usuário alvo (usuário com $id=3$). Este cálculo tem como objetivo retornar os usuários que possuem mais similaridade com o usuário alvo. A equação utilizada para calcular a vizinhança entre os usuários foi a Correlação de Pearson. A Equação 5, a seguir, apresenta o cálculo de similaridade entre o usuário alvo e o usuário com $id=2$.

$$w_{u,u} = \frac{((3 - 3,5) * (3 - 3,5)) + ((4 - 3,5) * (4 - 3,5))}{\sqrt{(3 - 3,5)^2 + (4 - 3,5)^2} * \sqrt{(3 - 3,5)^2 + (4 - 3,5)^2}} = \frac{0,5}{0,7 * 0,7} = 1 \quad (5)$$

Como pode ser observado na Equação 5, o cálculo de similaridade realizado com o usuário 3 e 2, obteve a similaridade de 1. Este é o valor máximo da escala de recomendação, que é entre 1 e -1. A Tabela 4 apresenta todos os resultados obtidos pela equação de Correlação de Pearson, aplicada sobre os usuários da base de dados, que são as vizinhanças do usuário alvo.

Tabela 4 Similaridade dos Usuários

Id dos Usuários	Cálculo de Correlação de Pearson (similaridade dos usuários)
2	1
4	0,9
5	-0,5
1	-0,9

Como pode ser observado na Tabela 4, é possível verificar quais usuários possuem mais similaridade com o usuário 3. Com os valores da Tabela 4 é possível dizer que para realizar a recomendação dos filmes é utilizado apenas os usuários com *Id* 2 e 4, pois são usuários que ficaram mais próximo da escala de recomendação, ou seja, possui mais similaridade com o usuário alvo, dentro da vizinhança definida no protótipo.

Estas similaridades encontradas pelo cálculo obtiveram estes valores devido às avaliações realizadas pelos usuários nos filmes semelhantes aos que o usuário alvo também tenha avaliado. A Tabela 5 apresenta as avaliações realizadas pelos usuários.

Tabela 5 Avaliações dos Usuários

Usuários	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9	F 10
1	4	2	5	1				2		3
2	2	3		4	4		3		4	
3		3	1	4		3		4		2
4	2	5	3	5			3			2
5	4	1	3		1	3			4	

Como pode ser observado na Tabela 5, o usuário 1 tem mais filmes em comum avaliado com o usuário alvo, do que os usuários 2,4 e 5. Mesmo com mais filmes em comum avaliado, o usuário 1 foi o que teve o grau de similaridade -0,9, ou seja, o mais distante do usuário alvo como apresentado na Tabela 4. Este grau de similaridade ocorre devido ao fato de que nos filmes avaliados em comum as notas dos dois usuários foram bastante distintas, como pode ser observado na Tabela 5. Já o usuário 2, possui apenas 2 filmes avaliados em comum com o usuário alvo, e os filmes em comum obtiveram notas iguais, devido isto o seu grau de similaridade

calculado foi 1, como apresentado na Tabela 4 foi o maior comparando com os dos outros usuários.

O usuário 4 possui 4 filmes avaliados em comum com o usuário alvo, mas devido os valores das avaliações não serem iguais, o grau de similaridade foi 0,9, mas ainda é um usuário bastante similar com o usuário alvo. Já o usuário 5 possui 3 filmes avaliados em comum com o usuário alvo, mas as suas notas estão distintas, devido a esta diferença, sua similaridade teve o resultado -0,5. Após a realização dos cálculos de similaridade e o entendimento de como os resultados foram obtidos, o próximo passo é calcular a predição sobre os filmes recomendados.

A próxima etapa realizada pelo protótipo apresenta a validação dos testes realizados pelo cálculo de predição, ou seja, calcular qual nota o usuário alvo avaliaria os filmes que lhe foram recomendados. Para a realização do cálculo de predição foram utilizados os filmes que já foram avaliados pelos dois usuários mais semelhantes ao usuário alvo e ainda não avaliados pelo usuário alvo. Realizando um levantamento na base de dados, verificou-se que os filmes avaliados pelos usuários semelhantes e não avaliados pelo usuário alvo foram os filmes 1 e 7. Desta forma, o próximo passo é calcular a predição. A Equação 6, a seguir, apresenta qual a predição do usuário alvo sobre o filme 7.

$$P_{3,7} = 2,83 + \frac{((3-4,33)*0,9)+((3-3,5)*1)}{0,9+1} = 2,83 + \frac{-1,69}{1,9} = 2,83 - 0,8 = 2 \quad (6)$$

Como pode ser observado na Equação 6, o resultado da predição sobre o filme 7 resultou no valor de 2. A Equação 7, a seguir, apresenta qual a predição do usuário alvo sobre o filme 1.

$$P_{3,1} = 2,83 + \frac{((3-4,33)*0,9)+((3-3,5)*1)}{0,9+1} = 2,83 + \frac{-1,697}{1,9} = 2,83 - 0,8 = 2 \quad (7)$$

Como pode ser observado na Equação 7, o resultado da predição sobre o filme 1 resultou no valor de 2. Os cálculos de similaridades e predições realizados de forma manuais apresentaram os mesmos filmes recomendados pelo protótipo.

Esta seção teve como objetivo apresentar os resultados dos cálculos que são realizados para gerar recomendações. As equações exemplificadas acima

apresentam como são realizados os cálculos de similaridade e predição, cálculos que são realizados pela Filtragem Colaborativa. Os resultados apresentados pelo protótipo, conseqüentemente obteve os mesmos resultados apresentados pelos cálculos manuais, ou seja, gerou as mesmas recomendações. Com estes resultados encontrados é possível observar que as recomendações do protótipo obtiveram os mesmos resultados dos cálculos manuais.

6 CONSIDERAÇÕES FINAIS

Neste trabalho foram abordados os conceitos sobre Sistema de Recomendação, Técnicas de Recomendação, Filtragem Baseada em Conteúdo e Filtragem Colaborativa. Os estudos realizados sobre estes conceitos foram necessários para realizar a definição e o desenvolvimento do protótipo de recomendação, utilizando a Filtragem Colaborativa.

Para realizar o desenvolvimento do protótipo, foi necessário a implementação somente da Filtragem Colaborativa. A técnica de FC implementada utiliza o auxílio da API *Apache Mahout Taste*, que tornou possível a realização de um desenvolvimento rápido e prático. Mas para usar a API, foi necessário um estudo aprofundado sobre as funcionalidades oferecidas. No estudo feito sobre a API foram encontradas dificuldades em conseguir documentações atualizadas, ou até mesmo documentação que estava de acordo com a versão da biblioteca.

Para validação do protótipo desenvolvido, utilizou-se a base do MovieLens. Mas para realizar os testes utilizou-se de uma base de dados reduzida, devido os cálculos que seriam feitos manualmente, pois se torna inviável fazer os cálculos manuais em base grande. Após as realizações dos testes sobre diversas formas foi comprovada a funcionalidade do protótipo desenvolvido. Foi possível observar também que quanto maior for a base de dados de um Sistema de Recomendação, maior poderá ser a quantidade de recomendações retornadas pelos SR.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- BORGES, Deise Miranda. **Estudo de Técnicas de Recomendações Automáticas de Produtos**. 2010. 85p. Relatório de Estágio (Graduação em Sistemas de Informação) – Centro Universitário Luterano de Palmas, Palmas.
- BORGES, Deise Miranda, OLIVEIRA, Fernando Luiz. **Análise e comparação dos sistemas de recomendação de produtos existentes em três empresas de comércio eletrônico (CE): Saraiva, Submarino e Amazon.2010** Centro Universitário Luterano do Brasil (CEULP/ULBRA), Palmas .Disponível em: <<http://sistemas-humanocomputacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Mir2010.pdf>>. Acesso em 29 de Agosto de 2013.
- CAZELLA, Sílvio C.; REATEGUI, Eliseo B. **Sistemas de Recomendação**. São Leopoldo: XXV Congresso da Sociedade Brasileira de Computação, 2005.
- CAZELLA, S. C. **Aplicando a Relevância da Opinião de Usuários em Sistema de Recomendação para Pesquisadores**. Tese, disponível em: <http://www.inf.unisinos.br/~cazella/papers/VersaoFinal2006TeseSilvioCazellahomologacao.pdf>, Porto Alegre, 2006.
- FILHO, V. M. **Sistema Inteligente de Recomendação para Comércio Eletrônico**. TCC, disponível em: <http://tcc.dsc.upe.br/20062/TCCValmirFinal.pdf>. Recife, 2006.
- INGERSOLL, Grant. **Introdução ao Apache Mahout**. 2009. Disponível em: <<http://www.ibm.com/developerworks/br/java/library/j-mahout/>>. Acesso em: 15 de Junho de 2014.
- LOPES, Giseli Rabello. **Sistema de Recomendação para Bibliotecas Digitais sob a Perspectiva da Web Semântica**. 2007. 69p. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre. Disponível em: <http://www.lume.ufrgs.br/handle/10183/10747>. Acesso em: 19 de out de 2013.
- MACÁRIO FILHO, Valmir. **E-Recommendier: Sistemas Inteligente de Recomendação para Comércio Eletrônico**. 2006. 58p. Trabalho de Conclusão de Curso (Graduação

em Engenharia da Computação) – Escola Politécnica de Pernambuco, Recife.
Disponível em: http://dsc.upe.br/~tcc/20062/TCC_Valmir_Final.pdf. Acesso em: 16 de dezembro de 2009.

REATEGUI, Elisio Berni; CAZELLA, Sílvio César. *Sistemas de Recomendação*, 2005. Disponível em: <http://scholar.googleusercontent.com/scholar?q=cache:FXRXKpeLaC8J:scholar.google.com/+recomenda%C3%A7%C3%A3o+de+itens+na+internet&hl=pt-BR&as_sdt=0,5>. Acesso em: 07 Nov 2013.

SAMPAIO, Igor Azevedo. *Aprendizagem Ativa em Sistemas de Filtragem Colaborativa*, 2007.

Disponível em: <<http://www.cin.ufpe.br/~rbcp/dissertacoes/dissertacaoIAS.pdf> >. Acesso em: 09 Nov. 2013.

SCHAFER, J. B; KONSTAN, J A.; RIEDL, J. E-Commerce Recommendation Applications. GroupLens Research Project, Department of Computer Science and Engineering University of Minnesota, 2001.

SCHAFER, J. Ben; KONSTAN, Joseph A.; RIEDL, J John. E-Commerce Recommendation Applications: **Data Mining and Knowledge Discovery**. 2001.

Disponível em: <<http://www.springerlink.com/content/r24285574675qu7v/fulltext.pdf>>. Acesso em: 07 out. 2011.

SILVA, Edeilson Milhomemda. **SWEETS**: um Sistema de Recomendação de 85 Especialistas aplicado a Redes Sociais. 2009. 145p. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Pernambuco, Recife.

TORRES, R. *Personalização na Internet: como descobrir os hábitos de consumo de seus clientes, fidelizá-los e aumentar o lucro de seu negócio*. São Paulo: Novatec, 2004.

[Queiroz 03] Queiroz, S. R. M. 2003. *Group Recommendation Strategies Based On Collaborative Filtering*. Dissertação de Mestrado, Universidade Federal de Pernambuco.