

Técnica de Recomendação Baseada em Metadados para Repositórios Digitais Voltados ao Ensino

Marcus Filipi Rosso Casagrande, Gustavo Kozima, Roberto Willrich

Programa de Pós-Graduação em Ciência da Computação – Departamento de
Informática e Estatística – Universidade Federal de Santa Catarina (UFSC)
Caixa Postal 476 – 88.040-900 – Florianópolis – SC – Brasil

{marcusrc, kozima, willrich}@inf.ufsc.br

Abstract. *This paper proposes a simple and scalable recommender technique applied to digital repositories aimed on the educational content. The proposed technique is based on the user's profiles construction by observing the metadata values of the accessed content. This proposed technique is valuated in a digital repository used for literature teaching.*

Resumo. *Este artigo propõe uma técnica de recomendação simples e escalável aplicada a repositórios digitais voltados ao ensino. A técnica proposta é baseada na construção de perfis de usuário via observação dos valores de metadados dos conteúdos acessados. A técnica proposta é avaliada em um repositório digital utilizado no ensino de literatura.*

1 Introdução

Repositórios digitais (RD) são sistemas, geralmente Web, que oferecem recursos para organizar coleções digitais e para construir um catálogo dos metadados, que facilitam a organização e a descoberta destas coleções. Em um contexto educacional, as coleções oferecidas pelos RDs são compostas por conteúdos, ou objetos, de aprendizagem. A organização de conteúdos em RDs é baseada em metadados, que são dados sobre os dados. Existem algumas iniciativas de padrões de metadados, sendo que os principais são Dublin Core [Dublin Core 2013] e IEEE LOM [IEEE 2002].

A Biblioteca Digital de Literatura Brasileira (BD-LB) (www.literaturabrasileira.ufsc.br) é um exemplo de RD voltado ao ensino. Este repositório disponibiliza uma grande coleção de obras literárias e outros conteúdos relacionados ao ensino de literatura, além de manter um banco de dados sobre autores brasileiros. A BD-LB vem sendo utilizado em diversas atividades de ensino no Curso de Literatura da Universidade Federal de Santa Catarina.

Apesar dos RDs oferecerem meios para a organização dos conteúdos, com o aumento das coleções, o usuário se depara com o problema clássico de sobrecarga de informação. Portanto, é importante adicionar funcionalidades aos RDs que permitam a localização mais eficiente de conteúdos. Uma alternativa para isto são os sistemas de recomendação, que apoiam os usuários na localização de conteúdos através da apresentação de uma lista de conteúdos possivelmente interessantes ao usuário.

Este artigo propõe uma técnica simples e eficiente de recomendação de conteúdos em RDs baseada em metadados descritores dos conteúdos e agrupamento de usuários. Nesta técnica, o perfil do usuário é construído implicitamente observando os valores dos metadados dos conteúdos acessados. A redução da complexidade da técnica proposta é

devido a utilização de valores dos metadados das obras acessadas, não exigindo técnicas complexas de análise destes conteúdos. O agrupamento de usuários com perfis similares também é outra técnica visando a redução do tempo de processamento. A técnica de recomendação proposta foi implementada na BD-LB, com a qual foram realizados testes de precisão e de tempo de processamento.

Este artigo está organizado na forma que segue. Na seção 2 são revisados conceitos gerais sobre sistemas de recomendação. A seção 3 apresenta os trabalhos relacionados. A seção 4 apresenta a técnica de recomendação proposta. Na sequência, a seção 5 descreve a implantação e apresenta os resultados dos testes realizados. Finalmente, as conclusões e trabalhos futuros são apresentados na seção 6.

2 Sistemas de Recomendação para Repositórios Digitais

Um sistema típico de recomendação procura identificar dentro de um conjunto de itens (conteúdos, produtos ou serviços) aqueles que possam ser de interesse de um determinado usuário. Estes sistemas são úteis quando o universo de escolha é muito grande e é possível realizar uma estimativa das preferências dos usuários. Sistemas de Recomendação já vem sendo usados a algum tempo no contexto educacional, como em ([Ferro 2011] e [Sibaldo 2007]). O objetivo do uso de sistemas de recomendação neste contexto é de buscar e filtrar informações relevantes a um aluno ou professor em particular a partir de uma coleção geral de conteúdos educacionais.

Muitas técnicas de recomendação se baseiam na construção de um perfil de usuário, mantendo dados que permitam estimar as preferências e interesses de um determinado usuário. Tal perfil pode ser construído com a ajuda explícita dos usuários. A desvantagem desta construção explícita é a necessidade do usuário se dar ao trabalho de preencher as informações requisitadas, sendo que alguns trabalhos apontam que o usuário tende a não dar este tipo de informação [Carroll 1987]. A construção implícita do perfil do usuário é estimada a partir da interação deste com o sistema (baseado nos conteúdos acessados, critérios de consultas e/ou histórico de navegação).

Em geral, os sistemas de recomendação voltados para RDs utilizam as técnicas clássicas de Filtragem Colaborativa (FC) e/ou Filtragem Baseada em Conteúdo (FBC). A FBC se baseia na construção de um perfil de usuário que mantém a lista de conteúdos já acessados por ele e as características destes conteúdos. O perfil é então utilizado para identificar conteúdos relevantes ao usuário comparando o seu perfil com os descritores (ou o próprio conteúdo) de cada conteúdo. Uma das desvantagens da FCB se dá pelo fechamento da recomendação: características dos conteúdos usados para a recomendação podem acabar por gerar recomendações com um escopo muito limitado. Por exemplo, se o usuário acessa apenas conteúdos do tipo de recurso educacional texto, talvez não sejam recomendados conteúdos do tipo software. Outro aspecto crítico é a escalabilidade, caso o número de conteúdos a serem avaliados sejam muito grande.

Nas técnicas de FC, os conteúdos a serem recomendados a um usuário são determinados com base no conjunto de conteúdos já avaliados ou acessados pelos “vizinhos próximos” deste usuário. Vizinhos próximos são usuários com perfis similares ao usuário foco da recomendação. Técnicas de FC possuem dois problemas conhecidos: esparsidade e escalabilidade ([Sarwar 2002], [Park 2012]). O problema da esparsidade se deve ao fato de que alguns conteúdos acessados por um usuário podem ter tido pouco (ou até nenhum) acesso por outros usuários do sistema, dificultando a recomendação de

conteúdos similares a tais conteúdos impopulares. Conteúdos mais populares tendem a ser mais recomendados. A escalabilidade está relacionada à complexidade dos algoritmos de FC. Geralmente a quantidade de informações a serem consideradas é tão grande que pode tornar o tempo de resposta um problema.

Uma das principais técnicas utilizadas para melhorar a escalabilidade em sistemas de recomendação é o agrupamento (também chamado de clusterização) de usuários ou conteúdos. O objetivo é reduzir a quantidade de dados sobre o qual a FC será realizada e com isto reduzir o tempo de processamento.

3 Trabalhos Relacionados

Trabalhos como [Herlocker 2012] e [Cazella 2008] propõem técnicas de recomendação baseadas em FC para RDs, mas ambas as propostas se baseiam na avaliação explícita do usuário dos conteúdos. Esta é justamente uma das deficiências destas propostas, pois os usuários geralmente tendem a não oferecer este tipo de informação. Além disso, elas também sofrem dos problemas de esparsidade e escalabilidade associadas a FC.

Uma das formas de tratar o problema da escalabilidade é a redução do espaço de busca via o agrupamento de usuários [Sarwar 2002], ou de conteúdos [Li 2003]. Geralmente estes trabalhos utilizam a técnica *k-means* [Jain 1988]. Nela é definida inicialmente uma quantidade *k* de clusters a serem criados e são escolhidos aleatoriamente *k* dados (perfis ou conteúdos) do conjunto a ser comparado e os coloca como centroide de cada cluster. Em seguida, são distribuídos os demais dados nos clusters. Um dado é alocado a um cluster quando o dado é mais similar com o centroide do respectivo cluster. A cada ciclo, os centroides de cada cluster são atualizados.

Diversos trabalhos demonstram que sistemas de recomendação também oferecem vantagens no contexto educacional. [Primo et al. 2010] propõe um modelo de recomendação de conteúdos educacionais descritos por metadados. Tal modelo visa capturar as informações do usuário implicitamente, utilizando o padrão de metadados OBAA [Bez et al. 2009]. A proposta mostra uma arquitetura interessante, que inclui elementos de Web Semântica no perfil dos usuários e nos conteúdos disponíveis. Porém é um modelo teórico, sem formalizar o procedimento realizado para a recomendação.

Em [Ferro et al. 2011], os autores fazem uso de um sistema híbrido de recomendação de materiais didáticos, que combina recomendação pela avaliação de conteúdos feitas por outros usuários, recomendação de conteúdos semelhantes ao perfil do usuário, e recomendação de conteúdos mais populares. Em tal proposta, o perfil do usuário é modelado de forma explícita, onde ele deve informar suas áreas de interesse.

As propostas acima descritas utilizam diferentes técnicas de recomendação. Esta proposta se destaca pela junção de características que simplificam a complexidade de todo o processo de recomendação. Utilizando metadados específicos voltados ao contexto textual, a FC consegue facilmente identificar usuários similares. Além disso, se definiu uma técnica de agrupamento que reduz o tempo de resposta da recomendação.

4 Técnica de recomendação proposta

A técnica de recomendação proposta para RDs utiliza a FC combinada com uma técnica de agrupamento de usuários com perfis similares. Esta técnica é aplicável exclusivamente

em RDs, onde os conteúdos disponibilizados são descritos por um conjunto de metadados. A técnica proposta envolve quatro atividades descritas a seguir.

4.1 Perfil do Usuário

O modelo de perfil de usuário adotado é o mesmo adotado na nossa técnica de recuperação personalizada de informações descrita em [Furtado, 2009]. Formalmente, o perfil de um usuário u_i é definido por:

$$pu_i = (PE_i, CA_i) \quad (1)$$

onde Preferências Específicas (PE) são as preferências do usuário, e Conteúdos Acessados (CA) são os conteúdos acessados pelo mesmo usuário. As PE de um usuário são estimadas implicitamente com base na análise da frequência de ocorrência dos valores dos Metadados Observados (MO) dos conteúdos acessados pelo usuário. A seleção dos MO depende do domínio do repositório. Para tal, é necessário analisar que metadados são relevantes na determinação das preferências do usuário. A título de exemplo, na BD-LB, mantendo conteúdos na área de literatura, observou-se que o autor e gênero literário são elementos de metadados (adotados na BD-LB) que são relevantes para determinar se uma obra é relevante ou não para o usuário. Já para objetos de aprendizagem, metadados LOM como *Keyword*, *InteractivityType*, *LearningResourceType*, *ContextDifficulty* podem ser utilizados como MO. As preferências específicas de um usuário i (PE_i) é definida da seguinte forma:

$$PE_i = \{SV_j \mid j \in [1, J]\} \quad (2)$$

onde J denota o número de MO, SV_j é um conjunto de valores ponderados do metadado observado MO_j construído a partir dos conteúdos acessados pelo usuário u_i . Neste conjunto, cada valor de MO_j está associado a um peso de preferência (w) que pode variar entre 0 e 1. Assim, SV_j é representado como um conjunto de valores ponderados:

$$SV_j = \{vw_n \mid n \in [1, N_j]\} \quad (3)$$

onde N_j é o número de valores distintos do MO_j encontrados nos conteúdos acessados pelo usuário; vw_n é o valor ponderado (v_n, w_n), onde v_n é um valor do metadado MO_j e $w_n \in [0,1]$ representa o peso deste valor. Quanto maior for esse peso, maior será o interesse do usuário para este valor de elemento específico.

Para ilustrar os conceitos apresentados, considere o repositório BD-LB e os metadados observados “autor” e “gênero literário”. Considere também três usuários ($u_1, u_2, e u_3$) com os seguintes PEs:

- $PE_1 = \{(Machado de Assis, 0.7), (José de Alencar, 0.3)\}, \{(Conto, 0.8), (Romance, 0.2)\}$
- $PE_2 = \{(Gregório de Matos, 0.4), (Machado de Assis, 0.6)\}, \{(Poesia, 0.5), (Conto, 0.5)\}$
- $PE_3 = \{(Machado de Assis, 0.3), (Aluizio Azevedo, 0.2), (José de Alencar, 0.3), (Gregório de Matos, 0.2)\}, \{(Conto, 0.3), (Romance, 0.2), (Poesia, 0.2), (Teatro, 0.3)\}$

Neste exemplo, as PEs do usuário u_1 foram geradas após ele ter acessado dez obras literárias, sete dos quais de Machado de Assis (peso 7/10) e três de José de Alencar (peso 3/10). Além disso, oito dessas obras foram do gênero Conto (peso 8/10) e dois do gênero Romance (peso 2/10).

4.2 Agrupamento de Usuários

Visando a redução do tempo de processamento do algoritmo de recomendação em RDs com grande número de usuários, foi definida uma técnica de agrupamento de usuários também baseada em valores de metadados. Nesta técnica, cada grupo reúne usuários com o mesmo interesse. Interesse este determinado pelos valores de metadados com maiores frequência de ocorrência.

O Algoritmo 1 especifica a técnica de agrupamento proposta. Este algoritmo é executado periodicamente (p.e. uma vez a cada 24h). Inicialmente o conjunto de grupo G é vazio. Em seguida, para cada perfil de usuário do RD é analisado o peso de cada um dos valores dos metadados observados. Caso o valor de um destes metadados atingir um limiar de agrupamento, o usuário será incluído no grupo identificado por este valor. Por exemplo, se o usuário realiza muitos acessos a obras de determinado autor ou palavra-chave, o usuário será incluído no grupo deste autor e no grupo desta palavra-chave.

```
Inicialização:
  G = {};
Para todos os usuários  $u_i$  do RD:
  Para todo  $SV_j \in PE_i$ 
    Para todo  $vw_n \in SV_j$ 
      Se  $w_n \geq \text{limiarAgrupamento}_j$ 
        Se grupo  $g_{j,vn}$  não existir, inclua este grupo em G
        Inclua o usuário  $u_i$  no grupo  $g_{j,vn}$ 
```

Algoritmo 1. Algoritmo de agrupamento adotado

O valor do limiar de agrupamento é específico do metadado observado. Seu valor depende principalmente da quantidade de valores distintos possíveis dos metadados observado. Por exemplo, os metadados observados da BD-LB são autores e gêneros literários. O número de autores possíveis atualmente é na ordem de milhares, enquanto os gêneros possíveis são 23 valores. Quando o elemento de metadado tem muitos valores possíveis, como autor, o limiar de agrupamento poderia ser menor, devido à possibilidade de maior dispersão dos usuários.

A título de exemplo, considere novamente os usuários u_1 e u_3 , com metadados Autor e Gênero usados para agrupamento, e um limiar comum de 0,4. Aplicando o algoritmo de agrupamento, u_1 será incluído em dois grupos: Autor Machado de Assis; e Gênero Conto. Já o usuário u_3 não será incluído em nenhum grupo. Neste caso pode-se recomendar conteúdos utilizando outras técnicas mais simples, como as obras mais populares, ou ainda, realizar a recomendação normal sem a etapa de agrupamento.

Um aspecto crítico na proposta é determinar que metadados observados serão considerados para o agrupamento. No contexto da BD-LB considerou-se autor e gênero literário como metadados de agrupamento, pois tais características são muito relevantes para levar um usuário a preferir determinada obra literária. Como contraexemplo, um metadado como idioma seria inapropriado para o agrupamento, pois se um usuário acessar muitas obras de determinado idioma, não quer dizer necessariamente que ele se interessaria por qualquer conteúdo escrito em tal idioma.

O objetivo da técnica de agrupamento proposta é reduzir o espaço de busca da FC. Neste caso, o algoritmo de FC, ao explorar apenas os usuários de determinado grupo, tenderá a encontrar usuários com maior grau de similaridade entre si, sem a necessidade de explorar outros usuários que tenderão a ter pouca (ou nenhuma) relação com os

usuários do grupo pesquisado. Assim, pode-se beneficiar do custo reduzido de processamento sem prejudicar a qualidade da recomendação.

A técnica proposta pode ser facilmente estendida para que um grupo possa representar agrupamento de usuários considerando mais de um metadado de agrupamento. Por exemplo, um grupo pode ser identificado por usuários preferindo determinado autor e determinado idioma. Mas o grupo deve sempre ser identificado pelo menos por um valor de metadado de agrupamento.

4.3 Determinação dos vizinhos próximos

Nesta proposta, vizinhos próximos de um usuário u_i são os usuários que têm uma similaridade de perfil com o usuário u_i acima de certo valor, chamado aqui de limiar de similaridade. Tal limiar depende do domínio do repositório e dos metadados observados. No caso da BD-LB, após a realização de testes, definiu-se empiricamente o valor de 0,7 como limiar inicial. Os vizinhos próximos de um usuário u_i é definido por:

$$Vizinhos_i = \{v \mid similaridade_{perfil}(i, v) \geq limiar\} \quad (4)$$

onde v é um outro usuário que pertence a pelo menos um grupo que u_i faz parte.

A determinação da similaridade do perfil dos usuários é realizada em duas etapas. Na primeira é determinada a similaridade entre os conjuntos SVs (conjunto de valores ponderados dos metadados observados) contidos nos perfis de u_i e u_v . Para este cálculo foi adotado o coeficiente de Pearson, conforme Equação (4).

$$similaridade_{mo_i}(i, v) = \frac{\sum_{n=1}^N w_{in} \cdot w_{vn}}{\sqrt{\sum_{n=1}^N (w_{in})^2 \cdot \sum_{n=1}^N (w_{vn})^2}} \quad (4)$$

onde i representa o usuário foco da recomendação; v representa um vizinho próximo do grupo a ser comparado; N é o número de valores diferentes do elemento de metadado mo_j do conjunto SV_i do perfil de usuário i ; w_{in} é o peso de preferência do valor v_{in} do elemento de metadados mo_j ; w_{vn} é o peso de preferência do valor de v_{in} do elemento de metadados mo_j se presente no perfil do usuário v (senão, w_{vn} é zero).

A título de exemplo, considere a recomendação realizada para o usuário u_1 . Os usuários u_1 e u_2 terão seus perfis comparados, pois ambos estão contidos nos grupos “Machado de Assis” e “Conto”. Nesta etapa é calculada a similaridade dos metadados autor e gênero: $similaridade(mo_{autor})(1,2) = 0,76$; e $similaridade(mo_{gênero})(1,2) = 0,69$.

Em seguida, para determinação da similaridade geral, é feita a média simples entre as similaridades de cada metadado observado, conforme Equação (5).

$$similaridade_{perfil}(1,2) = \frac{\sum_{j=1}^J similaridade_{mo_j}}{J} \quad (5)$$

Para os usuários do exemplo, o cálculo de similaridade dos perfis de u_1 e u_2 resultaria em $similaridade_{perfil}(1,2) = 0,73$.

4.4 Construção da Recomendação

A atividade final tem por objetivo produzir a lista ordenada de conteúdos a serem recomendados a um determinado usuário. Esta lista é obtida a partir do conjunto de conteúdos acessados pelos vizinhos próximos do usuário foco da recomendação. Esta

técnica assume que a lista de recomendação deve ser constituída de conteúdos que não foram ainda acessados pelo usuário. Desta forma, estes conteúdos já acessados pelo usuário serão retirados da lista de conteúdos a recomendar.

A lista de recomendação é ordenada com base no cálculo do peso de preferência definido por [Furtado, 2009]. Assim, os conteúdos que tiverem maior chance de serem interessantes para o usuário aparecem primeiro. Desta forma, a lista ordenada de conteúdos recomendados CR_i para um usuário u_i é definida como segue:

$$CR_i = (d_1, d_2, \dots, d_R) \text{ onde } \forall r \in [1, R-1], w_{pref}(u_i, d_r) \geq w_{pref}(u_i, d_{r+1}) \quad (6)$$

O peso de preferência w_{pref} de um conteúdo d_r a ser recomendado é estimado pela combinação dos pesos de preferência de cada um dos valores dos elementos de metadado observados deste conteúdo. Para exemplificar o cálculo do peso de preferência, considere novamente o usuário u_1 da BD-LB com seu perfil descrito na seção 4.1, e a lista de conteúdos a serem recomendados constantes na Tabela 1. Estes documentos foram supostamente acessados pelos vizinhos próximos de u_1 .

Tabela 1. Lista de Recomendação para o usuário u_1

Título	Autor	W_{autor}	W_{genero}	W_{pref}
Esaú e Jacó	Machado de Assis	0,7	0,2	0,45
A afilhada	Manuel de Paiva	0	0,8	0,4
Rio de Janeiro	José de Alencar	0,3	0	0,15

De acordo com as PEs de u_1 (apresentado na seção 4.1), o valor de preferência de cada conteúdo será calculado como a média entre o peso de preferência dos valores dos elementos de metadado autor e gênero. O resultado deste cálculo também é apresentado na Tabela 1. Finalmente, com base no peso de preferência calculado, a lista de conteúdo pode ser ordenada, conforme apresentado na Tabela 1.

5 Implementação e testes realizados

Esta seção apresenta uma visão geral da implementação da técnica de recomendação proposta na Biblioteca BD-LB e serão descritos os resultados dos testes realizados.

5.1 Implementação na BD-LB

A BD-LB contém uma vasta coleção de obras literárias de domínio público e informações sobre escritores brasileiros. Atualmente estão catalogadas 74.104 obras literárias (3.389 delas com textos integrais disponibilizados) de 17.836 autores. A plataforma é caracterizada pelo uso de código aberto, adotando-se LAMP (Linux, Apache, MySQL, PHP).

A BD-LB permite o cadastramento dos usuários e o sistema estima implicitamente o perfil do usuário como apresentado neste trabalho. A recomendação de obras é realizada na página principal da BD-LB após o usuário realizar a autenticação. Esta lista de recomendação é obtida usando a técnica proposta neste artigo.

A técnica de recomendação utiliza como limiar de agrupamento 0,4 tanto para o metadado autor quanto para o metadado gênero. Após o agrupamento, a FC usada filtra a lista de vizinhos próximos nos grupos a que pertencem o usuário foco da recomendação, onde foi adotado um limiar de similaridade de 0,7. O processo de agrupamento dos usuários (Seção 4.2) é realizado a cada 24 horas, durante o período de menor utilização

do servidor. O processo de determinação dos vizinhos próximos e determinação da lista de recomendação (Seções 4.3 e 4.4) são executados no momento em que o usuário cadastrado acesse página principal.

5.2 Testes Realizados

Foram realizados dois tipos de avaliação do sistema proposto. O primeiro visou avaliar a precisão da técnica de recomendação proposta. A segunda visou avaliar o impacto do agrupamento no desempenho do sistema.

5.2.1 Teste de Precisão

As principais métricas de validação de resultados de sistemas de recomendação são precisão e revocação. A técnica de precisão é medida pela razão entre o número de documentos relevantes recuperados e o número total de documentos recuperados pelo sistema. Já a revocação é definida pela relação entre o número de documentos relevantes recuperados e o número total de documentos relevantes contidos no sistema. Para avaliar a técnica de recomendação proposta, foi utilizada a métrica de precisão. A métrica de revocação não foi possível ser realizada devido ao grande número de obras disponíveis para que os usuários avaliassem quanto à relevância.

Ao todo, participaram deste teste 20 usuários reais da BD-LB. Inicialmente os participantes acessaram 5 obras de sua preferência. Em seguida foi informado para cada usuário uma lista de 5 obras recomendadas. Os participantes então avaliaram cada obra com notas de 1 a 5 de acordo com seu nível de relevância para leitura. Para fins de cálculo da precisão, considerou-se que o conteúdo é relevante se a nota dada pelo usuário é superior ou igual a 3. Além disso, devido à baixa quantidade de usuários, adotou-se o limiar de similaridade de 0,5. O índice de precisão médio obtido neste experimento foi de 0,8 com desvio padrão de 0,21. Apesar do reduzido número de usuários, este valor demonstrou a boa qualidade dos conteúdos recomendados. Este índice tende a aumentar de acordo com o aumento do número de usuários. Para verificar isso, foi feito anteriormente um teste similar com apenas 6 usuários, onde a precisão média foi de 0,625 com desvio padrão de 0,14.

5.2.2 Tempo de Processamento

Esta seção apresenta os testes de tempo de processamento da técnica proposta, visando medir a escalabilidade do sistema proposto. Para a realização deste teste foram inseridos 10000 mil usuários fictícios no repositório e, para cada um deles, foi simulado o acesso a 10 obras escolhidas aleatoriamente dentre um conjunto de 1416 obras.

Para viabilizar a realização do teste, foram escolhidos aleatoriamente 100 destes 10000 usuários cujos tempos de processamento das recomendações seriam medidos. Para cada um destes usuários foram geradas duas listas de recomendação: uma sem o uso do agrupamento, e outra com o uso de agrupamento com limiar de agrupamento 0,4. Para estudo comparativo, foram medidos os tempos de processamento e também as listas de obras recomendadas. Para estes testes, foi utilizado um servidor não dedicado com processador Intel Quad-Core Xeon E5405 2GHZ, com 8GB de memória.

A Figura 1 apresenta os tempos de processamento médio das recomendações com e sem agrupamento. Os resultados mostraram que houve um ganho significativo de tempo de processamento com o uso da técnica de agrupamento. Em média, o tempo de

processamento da recomendação foi de 7,3s, com desvio padrão de 0,76s, enquanto a recomendação com agrupamento obteve uma média de 4,63s, com desvio padrão de 2,76s. O elevado desvio padrão com o uso do agrupamento se deve ao fato de que alguns usuários não foram incluídos em nenhum grupo (por não alcançarem frequência de acesso superior ao limiar de 0,4). Para tais casos, a FC sem agrupamento é realizada, sem prejudicar o tempo de processamento, como mostra a Figura 1.

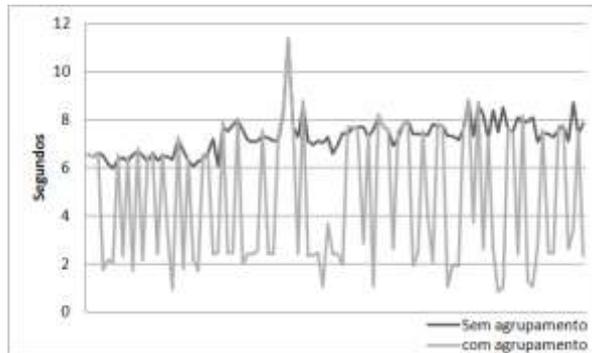


Figura 1. Tempos de processamento médio das recomendação

Ao todo, dos usuários que foram testados, 56 foram alocados em grupos, contra 44 que não foram. Analisando os resultados apenas dos que pertenciam a grupos, foi obtida uma média de tempo de recomendação de 2,24s, com desvio padrão de 0,6s. Este grande número de usuários não agrupados se deve ao maior número de autores. Como o teste selecionou aleatoriamente as obras, não há um comportamento natural dos usuários, de selecionarem menor número de autores e gêneros literários, além disso, não foi simulada a existência de obras mais populares, com um grande número de acessos.

6 Conclusões e trabalhos futuros

Este artigo propôs uma técnica de recomendação de conteúdos para repositórios digitais, considerando a existência de metadados que descrevem os conteúdos. O objetivo desta técnica é reduzir o esforço despendido pelo usuário na localização de conteúdos relevantes. Este benefício foi comprovado via uma avaliação empírica. A técnica proposta, ao contrário das existentes, analisa valores dos elementos de metadado dos conteúdos acessados pelos usuários a fim de estimar as suas preferências e formular a recomendação. O objetivo é determinar as preferências do usuário a partir dos dados que descrevem os conteúdos acessados anteriormente. O conjunto de elementos de metadado a ser observado é dependente do domínio do repositório, permitindo que metadados descritores de objetos de aprendizagem sejam considerados.

Como trabalhos futuros, pretende-se melhorar o algoritmo, incluindo o tratamento de casos especiais em que usuários não são alocados em nenhum grupo, além da implementação do agrupamento considerando mais de um metadado.

7 Agradecimentos

Este trabalho foi desenvolvido no contexto do projeto PRONEX Autores, obras e acervos literários catarinenses em meio digital, financiando pela FAPESC-CNPq. O autor Marcus Rosso Casagrande é financiado com uma bolsa de mestrado da CAPES.

8 Referências

- Bez, M. R., da Silva, J. M. C., Santos, E. R., Primo, T. e Bordignon, A. (2009) “Projeto obaa: Uma abordagem com objetos de aprendizagem interoperáveis baseados na web e na televisão digital”. *Informática na Educação: Teoria e Prática*, 12(1), p. 119–126.
- Carroll, J.M. e Rosson, M.B. (1987) “The paradox of the active user”. *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, p. 80-111.
- Cazella, S. C., Chagas, I. C., Barbosa, J. L., Reategui, E. B. (2008). “Um Modelo Para Recomendação de Artigos Acadêmicos Baseado em Filtragem Colaborativa Aplicado à Ambientes Móveis”. *Revista Novas Tecnologias na Educação* 7, p. 12-22.
- Dublin Core. (2013). Dublin Core Metadata Initiative. <http://dublincore.org>.
- Ferro, M.R.C., Nascimento Júnior, H.M., Paraguaçu, F., Costa, E.B., Monteiro, L.A.L. (2011) “Um Modelo de Sistema de Recomendação de Materiais Didáticos para Ambientes Virtuais de Aprendizagem”. *Simpósio Brasileiro de Informática na Educação*, p. 810-819.
- Herlocker, J., Jung, S., Webster, J. G. (2012) “Collaborative Filtering for Digital Libraries”. Technical Report Oregon State University. <http://hdl.handle.net/1957/28103>.
- IEEE LTSC (2004): “IEEE LTSC Working Group 12: Learning Object Metadata”. <http://ltsc.ieee.org/wg12/index.html>.
- Jain, A. K., Dubes, R. C. (1988). “Algorithms for Clustering Data”. Prentice Hall.
- Li, Q., Kim, B.M. (2003) “Clustering Approach for Hybrid Recommender System”. *IEEE /WIC International Conference on Web Intelligence (WI’03)*, p. 33-38.
- Furtado, C.A., Willrich, R, Fileto, R. Siqueira, F.L., Tazi, S. (2009) “Ordenação Personalizada na Recuperação de Informações em Bibliotecas Digitais.” In: *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*, 2009.
- Park D.H., Kim H.K., Choi I.Y., Kim J.K. (2012) “A literature review and classification of recommender systems research”. *Expert Systems with Applications*, 39 (11), p. 10059-10072.
- Primo, T.T., Vicari, R.M., Silva, J. M. C. (2010) “Rumo ao Uso de Metadados Educacionais em Sistemas de Recomendação”. *Simpósio Brasileiro de Informática na Educação*, pp. 4-8.
- Sarwar, B., Karypis, G., Konstan, J., Riedl, J. (2002) “Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering”. *Fifth International Conference on Computer and Information Technology*, p. 158–167.
- Sibaldo, M. A. A., Sales, T. B. M., Calado, I.A.A.R., Bittencourt, I.I., Costa, E.B. (2007) “Mobile GraW: Uma Aplicação para Dispositivos Móveis Baseada em Comunidades Virtuais de Aprendizagem Com Suporte A Recomendação”. *Simpósio Brasileiro de Informática na Educação*, p. 214-217.