

May social behavior reveal preferences on different contexts? Recommending movie titles based on tweets

Frank Helbert Borsato

*Coordenação de Ciência da Computação
Universidade Tecnológica Federal do Paraná
Campo Mourão - Paraná, Brasil
frankhelbert@utfpr.edu.br*

Ivanilton Polato

*Coordenação de Ciência da Computação
Universidade Tecnológica Federal do Paraná
Campo Mourão - Paraná, Brasil
ipolato@utfpr.edu.br*

Abstract—Recommendation systems allow users to deal with information overload by generating personalized recommendations that guide them through the universe of available options. These systems have been successfully used for over a decade in various application domains. However, most recommendation techniques do not consider context, generating recommendations which do not consider user's daily generated information. The increasing use of social networks and microblogs has created a valuable source to extract information that might help improve the results of recommender systems. Specifically, microblogging messages may help on contextualized movie recommendation. In this paper, a hybrid recommender is proposed merging a collaborative filtering recommender with a content based recommender. The results confirm the usefulness of the proposed recommender, where the user's exchanged messages define a bias to insert context in the recommendations.

Keywords—Microblogging, recomendação contextualizada, filtragem colaborativa, recomendador por conteúdo, recomendador híbrido

I. INTRODUÇÃO

Sistemas de recomendação despontaram como importantes meios pelos quais as pessoas descobrem informações, produtos e serviços [8]. Esses sistemas complementam os resultados disponíveis pelos sistemas de recuperação de informações, baseados em consultas, oferecendo a possibilidade de descoberta de novos itens de maneira que o usuário interaja mais proativamente, baseado nas preferências manifestadas em oportunidades anteriores. Um efeito desses sistemas é o de guiar os usuários por um espaço de escolhas substancialmente superior a sua capacidade de explorá-las [5].

Esse processo de guiar o usuário individualiza a resposta do sistema, melhorando a usabilidade à medida em que traz o que é considerado útil e interessante. É isso que separa os sistemas de recomendação dos sistemas simples de busca, que inicialmente devolviam os itens que casavam os termos procurados. Entretanto, a distinção entre esses sistemas se tornou bem mais obscura depois do desenvolvimento do Google¹.

Basicamente, existem duas abordagens no contexto de sistemas de recomendação: filtragem colaborativa [8], e baseado em conteúdo [6]. Na primeira abordagem, as recomendações são geradas a partir dos dados de usuários com interesses similares, na qual o interesse é obtido por avaliações que representam quanto um indivíduo gosta de um determinado item. Analisando as avaliações conhecidas, o sistema fornece qual seria a afinidade para um item ainda não avaliado. A similaridade entre os usuários é geralmente obtida pela correlação entre as avaliações dos itens feitas pelos usuários em questão. Já na abordagem baseada em conteúdo, são utilizados metadados que descrevem os itens para determinar sua similaridade. Gerar recomendações nessa abordagem é uma questão de identificar os itens similares aos que o usuário já manifestou interesse no passado (comprou, gostou, adicionou à lista de desejos).

Ambas as abordagens possuem suas fraquezas e méritos. Muitos pesquisadores já as combinaram de diferentes formas, chegando a sistemas chamados de híbridos, em busca de um melhor desempenho [5], [4]. Algo considerado crucial para a escolha da técnica utilizada é a natureza dos dados disponíveis. Esses dados podem ser notas dadas por outros usuários, tags associadas, metadados dos itens e podem ainda não estar diretamente disponíveis. Nesse último caso, os dados devem ser minerados em busca de informações relevantes.

Minerar dados do Twitter² mostra-se bem promissor, sendo utilizado em muitas situações, tanto para alimentar recomendadores, como para fazer previsões ou estimativas de dados da vida cotidiana [13], [3]. O Twitter é um *microblog* com uma grande base de utilizadores, consistindo em dezenas de milhões de usuários que participam ativamente na criação e propagação de conteúdo. As mensagens postadas e os sentimentos extraídos delas podem ser utilizados para prever o resultado de pesquisas [13] e até estimar o lucro de filmes [3]. Filmes são um tópico especialmente popular entre os usuários das redes sociais, caracterizado por um grande número de pessoas discutindo e uma grande

¹<http://www.google.com>

²<http://www.twitter.com>

variedade de opiniões [10].

Muitos pesquisadores já exploraram o tema, mesmo antes do uso massivo das redes sociais. Fisk [8] utiliza a filtragem colaborativa para implementar um recomendador de filmes. Basu *et al.* [4] utiliza os dados extraídos do IMDb³ para melhorar um recomendador híbrido de filmes que apresenta um desempenho superior se comparado às iniciativas que se utilizam somente da filtragem colaborativa.

Entretanto, a maioria das técnicas de recomendação não considera o contexto, o que leva a resultados estáticos, que somente mudam após um longo período de interação dos usuários [15]. Dessa forma, as recomendações geradas são independentes de fatores comuns do dia a dia, como o dia de trabalho, clima, a região onde o usuário vive e os eventos locais. Sendo assim, ainda há espaço para melhorias, tanto que em 2010 houve o Challenge on Context-aware Movie Recommendation⁴, evento cujo objetivo foi o de impulsionar a pesquisa em recomendadores dependentes de contexto. Os trabalhos utilizaram bases de dados anônimas extraídas de comunidades de filmes, gerando assim um ambiente padronizado que possibilitou a comparação dos métodos [15], [9]. Nota-se, entretanto, que o contexto foi limitado a comunidades de filmes, ambiente bem diferente do encontrado em um sistema genérico de *microblogging*, como o Twitter.

Mas até que ponto as informações postadas em um sistema de *microblogging* refletem as opiniões e preferências de uma pessoa? Na busca de um caminho para responder a essa pergunta considera-se o problema de recomendar títulos de filmes a um usuário baseado em seus tweets recentes. Nesse contexto, duas questões fundamentais são abordadas: (QP1) Os termos encontrados nos tweets de um indivíduo podem ser relacionados aos seus filmes prediletos através da análise de conteúdo (sinopse do filme)? (QP2) Os métodos de recomendação atuais, especificamente os baseados em conteúdo, podem ser utilizados quando os contextos dos itens são extremamente divergentes (conteúdo dos tweets x sinopse do filme)? Em seguida, descrevemos os métodos e estudos desenvolvidos para responder a essas questões. Especificamente, desenvolvemos um sistema de recomendação híbrido de filmes dependente de contexto, que se utiliza dos tweets dos usuários para melhorar os resultados do recomendador. Aplicado a dados reais minerados do Twitter, o sistema apresenta um bom desempenho.

II. SISTEMAS DE RECOMENDAÇÃO

Basicamente, sistemas de recomendação fazem sugestões personalizadas e tornaram-se aplicações fundamentais em comércio eletrônico, guiando o usuário por um subconjunto do universo de dados, e ainda atendendo efetivamente suas necessidades e preferências [5]. Há mais de uma década

os sistemas de recomendação fazem parte de grandes sítios de comércio eletrônico, como a Amazon.com, CDNow e MovieFinder [17].

Os recomendadores podem ser classificados de acordo com os dados que utilizam e como os processam. Especificamente, temos os dados da base, que são aqueles que o sistema possui para fazer as recomendações; os dados de entrada, específicos ao usuário para o qual o sistema gerará a resposta; e o algoritmo, responsável por combiná-los. Dessa forma, podemos distinguir cinco técnicas de recomendação diferentes [5], das quais consideraremos somente duas: colaborativa e baseada em conteúdo.

Na técnica colaborativa, as recomendações são resultado da comparação das avaliações ou recomendações de itens atribuídas por todos os outros usuários da base. Reconhecidas as incidências comuns, possibilita-se a “previsão” de uma recomendação ainda não conhecida. Nesses sistemas, os dados de um usuário são representados como um vetor de itens contendo suas avaliações. Sendo assim, o desempenho do sistema melhora com o tempo, conforme o usuário interage, avaliando novos itens e consequentemente incrementando o vetor que representa suas preferências. Entretanto, o tempo em geral não é considerado, embora já tenha sido mostrado que os interesses das pessoas mudam com sua passagem [2]. Isso pode ser implementado utilizando um fator multiplicativo sobre as avaliações.

O ponto forte das técnicas colaborativas é que elas podem ser aplicadas a qualquer tipo de objeto, funcionando muito bem para aplicações complexas como recomendação de filmes e músicas. Nessa abordagem, não é necessário criar uma representação para o objeto do mundo real, basta apenas criar um mecanismo para identificá-lo. Em contraste, nas técnicas baseadas em conteúdo, o objeto deve ser definido baseado em suas características (metadados). O que o recomendador faz é encontrar objetos com características semelhantes aos que o usuário já tenha manifestado interesse no passado. Novamente, o resultado do recomendador melhora progressivamente conforme novos dados são adicionados.

Embora a técnica colaborativa seja uma boa escolha no contexto de recomendação de filmes, seu desempenho depende de o quanto as avaliações do usuário se sobrepõem a base já existente. Se o espaço de avaliações for esparsos demais ou se o conjunto de itens mudar rapidamente, a técnica colaborativa não será a melhor escolha [5].

Neste trabalho utilizamos um conjunto selecionado de itens (título de filmes) e uma base de tamanho considerável, disponível no sítio do GroupLens⁵, para garantir uma sobreposição suficiente das avaliações. Os usuários foram convidados a avaliar todos os seus itens conhecidos, evitando-se assim dois problemas conhecidos da técnica: o problema do novo usuário, que com poucos itens avaliados

³<http://www.imdb.com>

⁴<http://www.dai-labor.de/camra2010>

⁵<http://www.grouplens.org>

acaba com uma recomendação pobre; e o problema do novo item, o qual não foi avaliado por ninguém e dessa forma não seria recomendado [12].

Utilizando a técnica baseada em conteúdo, não se evidencia o problema encontrado com a abordagem colaborativa com relação a um novo item ainda sem avaliações. Como a técnica se utiliza dos metadados, um novo item é tão elegível quanto qualquer outro da base. Entretanto, essa técnica pode se mostrar ineficiente, pois, ao contrário das pessoas, possui dificuldade em discernir conteúdo de alta e baixa qualidade [5], [7]. Por exemplo, um título de comédia pode ser recomendado a alguém que assiste somente a filmes de suspense, baseado na informação dos atores ou do diretor presentes no filme.

Para lidar com estes problemas os sistemas híbridos combinam duas ou mais técnicas de recomendação, atenuando os pontos fracos de cada uma das técnicas individuais [5]. Em geral, a técnica colaborativa é combinada com alguma outra. Neste trabalho em especial, vamos voltar nossa atenção para a técnica híbrida ajustada (com pesos). Nessa técnica, a avaliação computada pelo recomendador é obtida a partir dos resultados de todos os recomendadores presentes no sistema. A avaliação resultante para um item é resultado da soma das avaliações intermediárias ajustadas pelo peso de cada recomendador constituinte.

A. Filmes x Tweets

No experimento, a filtragem colaborativa é utilizada para determinar a avaliação de cada filme que o usuário ainda não tenha assistido. Considera-se que todos os usuários possuem uma quantidade suficiente de avaliações iniciais e a base é suficientemente grande, eliminando-se assim grande parte dos problemas encontrados na abordagem da técnica colaborativa. Isso é obtido através dos dados utilizados, visto que o foco não está ligado ao tratamento das limitações da filtragem colaborativa sozinha, e sim ao seu contexto.

O resultado da filtragem é combinado à análise de conteúdo, ponderado com pesos específicos. A ideia é que a hibridização forneça um viés ao resultado da filtragem colaborativa para refletir os últimos eventos da vida cotidiana. Acredita-se que isso irá gerar recomendações localizadas, tornando o resultado dinâmico, além de permitir a exploração de um espaço maior de itens.

III. COLETA DE DADOS E METODOLOGIA

O estudo contou com a participação de 15 voluntários, que tinham perfil ativo no Twitter. O conjunto de tweets dos participantes foi coletado manualmente, acessando o perfil de cada um deles diretamente no navegador. As coletas foram realizadas entre os dias 02 e 18/05/2012. De cada voluntário foi coletado um mínimo de 300 tweets retroativamente, contados a partir do último tweet publicado no dia anterior à data da coleta.

Os dados coletados do Twitter podem conter ruídos que poderiam atrapalhar substancialmente no desenvolvimento dos experimentos. O conjunto de tweets originalmente coletados dos voluntários sofreu um processo de verificação para remoção desse tipo de conteúdo. Assim, foram removidas todas as URLs e menções encontradas. Também foram desconsiderados tweets que continham duas ou menos palavras, ou continham apenas URLs. Esse processo justifica a coleta de uma quantidade de tweets maior do que o realmente utilizado, uma vez que não se pode saber com antecedência a quantidade de ruído encontrado nos tweets de um usuário. Depois da coleta, para cada voluntário, foram selecionados grupos com os 20 e os 50 tweets mais recentes.

A coleta de dados para a filtragem colaborativa foi feita com os voluntários preenchendo uma planilha contendo 118 filmes selecionados. Para cada filme que um voluntário já havia assistido, uma nota de avaliação foi preenchida, com escala entre 1 e 5, sendo 1 o pior valor da escala, no qual o filme foi considerado ruim e 5 o maior valor, aplicado para os filmes considerados excelentes. O preenchimento dos dados se deu maneira individual de forma que um voluntário não fosse influenciado por outro, sendo vedado a eles o acesso aos dados preenchidos por outros voluntários.

O conjunto de filmes que recebeu avaliações dos voluntários foi escolhido incluindo-se filmes considerados clássicos, com grande fama e número de expectadores, mas também filmes com apenas boas avaliações de expectadores. A lista inclui filmes nacionais e internacionais, inclusive filmes em outras línguas além do português e do inglês. Para obter resultados qualitativos na filtragem colaborativa foi utilizada uma base de dados que o GroupLens disponibiliza online. As bases são compostas por coletas realizadas no sítio MovieLens⁶, mantido por pesquisadores do grupo. A base original escolhida contém 10 milhões de avaliações feitas por 72 mil usuários em um grupo de aproximadamente 10 mil filmes. Do conjunto de 118 filmes escolhidos para nosso estudo, 106 tinham avaliações na base do MovieLens. Os outros 12 filmes são mais recentes e ainda não foram inseridos na base do grupo. Assim, a base original do MovieLens foi reduzida para acomodar apenas as avaliações dos 106 filmes escolhidos que possuíam correlação. Com essa operação, a base passou de 10 milhões para aproximadamente 488 mil avaliações.

Finalmente, para que a recomendação por conteúdo fosse gerada, foi necessário coletar as sinopses de todos os 118 filmes incluídos no estudo. Para a coleta das sinopses foi adotada a língua portuguesa como padrão. As sinopses foram retiradas do sítio da Wikipedia. Quando as sinopses não estavam disponíveis em português na Wikipedia, foram adotadas alternativas como o acesso à sítios especializados em filmes e, em última instância, a tradução da sinopse do inglês para o português. Ao fim desse processo de coleta,

⁶<http://www.movielens.org>

as sinopses coletadas foram armazenadas em arquivos texto separados.

Com os dados coletados teve início a fase de experimentação. A primeira fase envolveu a geração das recomendações usando filtragem colaborativa. Foram utilizadas as bibliotecas do Apache Mahout⁷, mais especificamente o RecommenderJob, desenvolvido para gerar as recomendações usando algoritmos disponíveis para execução no arcabouço Apache Hadoop⁸. Para obtenção dos resultados, foi utilizado um aglomerado com 15 nós executando o Hadoop. Devido ao grande número de avaliações envolvidas nas bases (mais de 488 mil), a execução em uma máquina simples poderia demorar muito, e em alguns casos gerar erros de execução por falta de memória disponível. Para cada filme que os voluntários não assistiram, o RecommenderJob gerou recomendações com notas que variam de 1 a 5. Para o cálculo de erro nos experimentos, o Mahout também gerou recomendações para os 33% dos filmes avaliados pelos voluntários que foram removidos aleatoriamente e intencionalmente para a geração do RMSE (Root Mean Squared Error), ou em português erro quadrático médio. O RecommenderJob foi executado usando dois algoritmos para o cálculo das recomendações: coocorrência e similaridade de cossenos [14]. O algoritmo de coocorrência trabalha com o conceito de coexistência de itens. Assim, são calculados os números de vezes que cada par de itens aparecem juntos no conjunto de dados. No caso do algoritmo de similaridade de cossenos, o cosseno dos ângulos dos vetores contendo avaliações dos usuários é utilizado como medida de similaridade. As recomendações para os voluntários foram obtidas usando os dois algoritmos escolhidos.

Na segunda fase foram geradas recomendações baseadas em conteúdos [1]. Para obtenção dos resultados, ao conjunto de arquivos de sinopses dos filmes foi adicionado um arquivo contendo os tweets dos voluntários. O algoritmo de similaridade foi executado para cada arquivo de tweets coletado de forma que os tweets de um voluntário não interferissem nas recomendações de outro voluntário. O algoritmo calcula a similaridade dos conteúdos dos arquivos de entrada usando um analisador ajustado para a língua portuguesa, passando por um processo de remoção de stopwords para a geração do vetor de termos. Depois, para cada vetor é calculado o *tf/idf* dos termos. Assim, a similaridade entre dois conteúdos é medida pela similaridade associada aos vetores de termos gerados para cada arquivo de entrada. Logo, o algoritmo consegue verificar quais sinopses de filmes são mais similares ao arquivo contendo os tweets de um usuário.

Após a obtenção dos resultados iniciais, foram calculados os erros para cada um dos conjuntos de dados. Em seguida os dados foram usados para a geração das recomendações híbridas e obtenção dos resultados finais para as análises,

conforme descrito nas seções seguintes.

IV. RESULTADOS

O sistema proposto é um recomendador híbrido ajustado, formado por um recomendador colaborativo (RC) e por um recomendador baseado em conteúdo (RBC). Dependendo do ajuste (peso) utilizado, o resultado de um dos recomendadores se torna mais evidente. Apresentamos uma série de ensaios considerando diferentes pesos para cada um dos sistemas. O peso é representado pela letra p , sendo o resultado da recomendação híbrida para um determinado item e usuário dado pela fórmula:

$$R_H(i, u) = (1 - p) \cdot R_{RC}(i, u) + p \cdot R_{RBC}(i, u)$$

onde R_{RC} é a avaliação obtida para o item i e usuário u pelo recomendador colaborativo e R_{RBC} para o recomendador por conteúdo. Além da mudança no peso, são utilizados dois algoritmos para calcular a similaridade entre os itens da base. Para determinar o desempenho do sistema, determina-se o erro de cada um dos recomendadores em separado. O erro é determinado utilizando-se o RMSE entre os valores previstos e os valores reais avaliados. A intuição por trás do RMSE é que recomendações muito ruins (baixa qualidade) contribuem quadraticamente no valor do erro. Nesse caso, quanto menor o valor do RMSE, melhor.

Os erros são obtidos para cada usuário, removendo aleatoriamente 33% de suas avaliações, que recebem recomendações do sistema. Em seguida são comparadas as recomendações geradas pelo sistema com as avaliações originais do usuário. Considere o vetor A_u contendo a avaliação para todos os itens da base para os quais o usuário u tenha visto. Considere também o vetor A'_u como o vetor com os itens escolhidos para determinar o erro. Considerando R_u como a saída do sistema, o erro é dado pela fórmula:

$$RMSE_u(A'_u, R_u) = \sqrt{\sum_{i=1}^N \frac{(a_i - r_i)^2}{N}}$$

onde A'_u é representado pelo vetor $[a_1, a_2, \dots, a_n]$ e R_u pelo vetor $[r_1, r_2, \dots, r_n]$, onde N representa a quantidade de itens considerados para a determinação do erro. Para determinar o erro total, faz-se a média do erro obtido para cada usuário. O cálculo do erro foi repetido, mas desta vez foram removidos aleatoriamente, para cada usuário, 5 dos filmes com as melhores avaliações (Top5). Isso foi feito com base na aplicação de sistemas semelhantes, que raramente exibem mais de 5 resultados para o usuário (Tabela I).

Utilizar o RMSE como métrica é uma forma simplista de avaliar o sistema, já que descarta uma das premissas desse trabalho: o contexto. Determinar o erro é uma métrica quantitativa, e de fato válida, mas é necessário algo mais qualitativo para avaliar o desempenho em termos da utilidade do recomendador.

⁷<http://mahout.apache.org>

⁸<http://hadoop.apache.org>

Table I
RMSE DOS RECOMENDADORES

Recomendador colaborativo			
Algoritmo de Similaridade	Coocorrência	Cosseno	
RMSE (33%)	1.30708	1.28879	
RMSE (Top 5)	1.06794	1.04655	
Recomendador baseado em conteúdo			
Número de posts	20	50	
RMSE (33%)	2.13865	2.02330	
RMSE (Top 5)	2.45670	2.32623	
Recomendador híbrido proposto			
	$p = 0.5$	$p = 0.1$	$p = 0.05$
RMSE (33%) - 20 posts	1.47049	1.26858	1.27467
RMSE (Top 5) - 20 posts	1.72006	1.17208	1.10847
RMSE (33%) - 50 posts	1.41861	1.26777	1.27511
RMSE (Top 5) - 50 posts	1.66312	1.16296	1.10411

Por estarmos especialmente interessados no contexto, utilizamos como métrica de utilidade do sistema a cobertura de catálogo (*catalog coverage*) [11]. Para isso, vamos considerar que cada usuário recebe somente as 5 melhores recomendações. Para cada recomendador, obtivemos o número de elementos no conjunto formado pela união de todos os itens recomendados a todos os usuários. A cobertura é expressa como a porcentagem em relação aos itens do universo, no caso, os títulos de filmes considerados. A interpretação intuitiva para a cobertura é que, quanto maior seu valor, mais os usuários são expostos a diferentes itens.

Table II
COBERTURA DE CATÁLOGO

Recomendador colaborativo (cosseno)			
18.64%			
No. de posts	Recomendador baseado em conteúdo		
20	22.04%		
50	13.56%		
Recomendador híbrido proposto			
	$p = 0.5$	$p = 0.1$	$p = 0.05$
20	24.57%	23.72%	21.18%
50	16.10%	22.03%	20.33%

A última métrica utilizada é a *reversal rate* [16]. A taxa de reversão é uma medida de quão comumente o sistema comete grandes erros, que fazem o usuário perder a confiança no resultado. Para isso, consideramos os casos em que o sistema erra em mais de 2.8 na previsão. Esse valor foi escolhido pois representa um grande erro (70%) entre a previsão e a avaliação do usuário. A taxa de reversão é obtida dividindo-se a quantidade de casos onde o erro é superior a 70% pelo total de previsões realizadas.

V. DISCUSSÃO

O recomendador colaborativo possui um bom desempenho, isso foi confirmado pelos resultados presentes na Tabela I. O erro médio na estimativa da avaliação dos usuários foi de aproximadamente 1.29, em uma escala de 0 a 4. O recomendador baseado em conteúdo não se saiu

Table III
TAXA DE REVERSÃO

Recomendador colaborativo (cosseno)			
5.50%			
No. de posts	Recomendador baseado em conteúdo		
20	13.80%		
50	10.73%		
Recomendador híbrido proposto			
	$p = 0.5$	$p = 0.1$	$p = 0.05$
20	1.00%	3.50%	4.00%
50	2.50%	3.50%	4.50%

tão bem, com um erro considerado grande. Assim, não poderia atuar sozinho no cenário considerado, sob pena de desacreditar sua utilidade. Já o recomendador híbrido proposto se saiu muito bem, especialmente com p igual a 0.1.

Para analisar o recomendador híbrido proposto em termos de utilidade são usadas duas métricas, cobertura de catálogo e taxa de reversão. A primeira nos diz quanto do universo está sendo recomendado aos usuários, enquanto a segunda nos informa sobre os erros graves cometidos pelos recomendadores, que poderiam abalar a confiança do usuário nas recomendações. Quanto maior a cobertura melhor, pois mais itens são exibidos e podem ser experimentados, levando a novas descobertas que em outras condições jamais seriam feitas. A cobertura também é um indicativo de utilização do contexto nas recomendações.

Considerando a Tabela II, podemos responder à nossa primeira questão de pesquisa (QP1) de maneira positiva. Os termos encontrados nos tweets de um indivíduo podem sim, ser relacionados aos seus filmes prediletos através da análise de conteúdo. Observamos que o sistema híbrido supera tanto o recomendador por conteúdo, quanto o recomendador colaborativo nesse quesito. Isso se deve às diferentes mensagens trocadas, no Twitter, pelos usuários voluntários da pesquisa que introduzem um viés ao resultado, responsável pelo aumento na variedade dos títulos apresentados. Obviamente, isso introduz um impacto no RMSE, observado na Tabela I. Mas seria o aumento no erro experienciado pelos usuários?

Para dirimir quaisquer dúvidas a esse respeito foi elaborada a Tabela III, que apresenta a taxa de reversão. Podemos perceber que o sistema híbrido proposto não comete mais erros graves que seus sistemas constituintes. Dessa forma, podemos concluir que, embora numericamente as previsões tenham uma pior qualidade, isso não afeta a confiabilidade nas estimativas do recomendador.

Quanto à segunda questão de pesquisa (QP2), considerando o recomendador baseado em conteúdo para relacionar tweets e títulos de filmes, podemos concluir que seu trabalho sozinho não é suficiente para refletir satisfatoriamente os gostos dos autores dos tweets. Isso pode ser concluído a partir dos resultados da Tabela I. Entretanto, isso não quer dizer que o recomendador está falhando.

De fato, a interpretação dos resultados com o recomendador híbrido levam a acreditar que a recomendação por conteúdo está desempenhando o papel esperado. Diretamente, pode-se perceber um aumento na diversidade dos resultados. Indiretamente, podemos refletir sobre o seguinte cenário: estamos nos aproximando do Natal, no Twitter, começam a ser trocadas mensagens relacionadas à data. O recomendador por conteúdo irá ranquear os filmes relacionados ao tema no topo da lista. Como a contribuição desse é pequena, o recomendador baseado em conteúdo introduzirá um viés ao resultado do recomendador colaborativo, podendo introduzir um ou mais títulos de filmes relacionados ao período natalino, ambos com previsões de avaliação muito boas.

No que tange a influência do número de tweets sobre os resultados, percebemos indicativos de que um número maior de mensagens melhora os resultados das recomendações (expressados pelo RMSE), entretanto, afeta negativamente a cobertura do sistema. Os dados sobre seu efeito na taxa de reversão são inconclusivos (Tabela III). Dessa forma, novos estudos devem ser conduzidos visando esclarecer seu efeito.

Podemos concluir então que o recomendador por conteúdo desempenha com sucesso o papel de contextualizar o resultado do recomendador híbrido baseado nas mensagens enviadas no Twitter.

A. Limitações e trabalhos futuros

Para o ajuste da ponderação de cada técnica foi utilizado um valor fixo global e imutável (p). Ajustar a ponderação de cada técnica de recomendação para cada indivíduo específico, otimizando-o de forma que o erro calculado para o indivíduo em questão seja reduzido, como realizado em [7], deve diminuir o erro observado e deve moldar o desempenho de acordo com os dados disponíveis para cada indivíduo considerado. O ajuste do peso somente pode ser feito se os usuários assistirem aos filmes recomendados e fizerem sua avaliação. Após várias iterações, que podem levar de várias semanas a meses, acredita-se que haveria uma melhora no desempenho através da redução no erro da predição. Entretanto, novos experimentos devem ser conduzidos para confirmar essa hipótese. Aplicar novos filtros aos tweets pode também ser uma forma de melhorar o desempenho. Muitos trabalhos utilizam postagens em comunidades específicas de filmes, assim, a grande maioria das mensagens trocadas estão contextualizadas [10]. Uma nova abordagem é avaliar o efeito de uma filtragem prévia sobre a cobertura.

REFERENCES

- [1] S. Alag. *Collective intelligence in action*. Manning, 2008.
- [2] J. Allan. Incremental relevance feedback for information filtering. In *Research and Development in Information Retrieval*, pages 270–278, 1996.
- [3] S. Asur and B. A. Huberman. Predicting the future with social media. In J. X. Huang, I. King, V. V. Raghavan, and S. Rueger, editors, *Web Intelligence*, pages 492–499. IEEE, 2010.
- [4] C. Basu, H. Hirsh, and W. Cohen. Recommendation as classification: using social and content-based information in recommendation. *AAAI '98/IAAI '98*, pages 714–720, Menlo Park, CA, USA, 1998. American Association for Artificial Intelligence.
- [5] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12:331–370, 2002.
- [6] S. T. Chelcea, G. Gallais, and B. Trousse. A personalized recommender system for travel information. In A.-M. Dery-Pinna and A. Giboin, editors, *UbiMob*, ACM International Conference Proceeding Series, pages 143–150. ACM, 2004.
- [7] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. Combining content-based and collaborative filters in an online newspaper. In *In Proceedings of ACM SIGIR Workshop on Recommender Systems*, 1999.
- [8] D. Fisk. An application of social filtering to movie recommendation. In *Software Agents and Soft Computing: Towards Enhancing Machine Intelligence, Concepts and Applications*, pages 116–131, London, UK, UK, 1997. Springer-Verlag.
- [9] Z. Gantner, S. Rendle, and L. Schmidt-Thieme. Factorization models for context/time-aware movie recommendations. *CAMRa '10*, pages 14–19, New York, NY, USA, 2010. ACM.
- [10] S. Garcia Esparza, M. P. O'Mahony, and B. Smyth. On the real-time web as a source of recommendation knowledge. *RecSys '10*, pages 305–308, New York, NY, USA, 2010. ACM.
- [11] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22:5–53, 2004.
- [12] J. A. Konstan, J. Riedl, A. Borchers, and J. L. Herlocker. Recommender systems: A groupLens perspective. In *Recommender Systems. Papers from 1998 Workshop. Technical Report WS-98-08*, pages 60–64. AAAI Press, 1998.
- [13] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *ICWSM*. The AAAI Press, 2010.
- [14] S. Owen, R. Anil, T. Dunning, and E. Friedman. *Mahout in Action*. Manning Publications Co., Manning Publications Co. 20 Baldwin Road PO Box 261 Shelter Island, NY 11964, first edition, 2011.
- [15] A. Said, S. Berkovsky, and E. W. D. Luca. Putting things in context: Challenge on context-aware movie recommendation. In *Proceedings of the Workshop on Context-Aware Movie Recommendation*, pages 2–6, New York, NY, USA, 2010. ACM.
- [16] B. M. Sarwar, J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl. Using filtering agents to improve prediction quality in the groupLens research collaborative filtering system. *CSCW '98*, pages 345–354, New York, NY, USA, 1998. ACM.
- [17] J. B. Schafer, J. Konstan, and J. Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic Commerce*, pages 158–166, New York, NY, USA, 1999. ACM.