



Pós-Graduação em Ciência da Computação

**“Aprendizagem Ativa em Sistemas de Filtragem
Colaborativa”**

Por

Igor Azevedo Sampaio

Dissertação de Mestrado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE, FEVEREIRO/2006



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

IGOR AZEVEDO SAMPAIO

“Aprendizagem Ativa em Sistemas de Filtragem
Colaborativa”

*ESTE TRABALHO FOI APRESENTADO À PÓS-GRADUAÇÃO EM
CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE INFORMÁTICA DA
UNIVERSIDADE FEDERAL DE PERNAMBUCO COMO REQUISITO
PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA
COMPUTAÇÃO.*

ORIENTADOR: Geber Lisboa Ramalho
CO-ORIENTADOR: Ricardo Bastos Cavalcante Prudêncio

RECIFE, FEVEREIRO/2006

Sampaio, Igor Azevedo.

Aprendizagem ativa em sistemas de filtragem colaborativa / Igor Azevedo Sampaio. - Recife: O autor, 2006.

vii, 75 folhas: il. fig., tab.

Dissertação (mestrado) Universidade Federal de Pernambuco.CIN.Ciência da Computação,2006. Inclui bibliografia.

006.31

CDD (22.ed.)

CIN2006-022

Agradecimentos

Em primeiro lugar agradeço a Deus por sua graça e pelas bênçãos derramadas constantemente sobre mim de maneira maravilhosa. Por ter me cercado de pessoas especiais e indispensáveis.

Agradeço a minha família, meus pais e meus irmãos, pelo amor e por fornecer o suporte emocional imprescindível em todos os momentos de minha vida.

A Emília, por ser minha companheira fiel e maior incentivadora. Por me apoiar com seu carinho e cuidado e por estar ao meu lado mesmo nas horas mais difíceis. Emília, este trabalho é dedicado a você, ao lado de quem espero passar os meus dias.

Aos professores Geber Ramalho e Ricardo Prudêncio pela paciência e dedicação na orientação deste trabalho.

Resumo

Nos dias de hoje, a quantidade de informação disponível é muito maior do que nossa capacidade de tratá-la. Vemos-nos diante de centenas de canais de televisão, dezenas de filmes para ver e milhares de produtos nas lojas de comércio eletrônico. Quando precisamos tomar uma decisão e não conhecemos todas as alternativas possíveis, uma abordagem bastante freqüente é buscar a recomendação de outras pessoas.

Na década de 1990 surgiram sistemas computacionais capazes de automatizar o processo de recomendações. Em geral os *Sistemas de Recomendação*, como ficaram conhecidos, coletam indicadores das preferências dos usuários para fornecer-lhes uma visão personalizada da informação. Uma abordagem amplamente empregada nos *Sistemas de Recomendação* é a *Filtragem Colaborativa (FC)*, em que a produção das sugestões é feita com base na similaridade entre usuários. Assim, para prever a relevância que um item i terá para um usuário alvo u , o sistema se baseia nas opiniões dos usuários com preferências similares às de u sobre i .

Um problema freqüente nos *Sistemas de Recomendação* diz respeito à chegada de um novo usuário. Nessa situação, o sistema não conhece nada a respeito das preferências dele e também não é capaz de gerar-lhe recomendações. Nos sistemas que utilizam FC isto também ocorre, pois a similaridade entre os usuários é computada com base nos itens que eles avaliaram em comum. Para amenizar esse problema, uma solução é que haja uma etapa inicial na utilização do sistema em que sejam apresentados alguns itens para o usuário novato avaliar. No entanto isso precisa ser feito de maneira eficiente, para que o sistema adquira o máximo de informação com um mínimo de esforço do usuário.

O paradigma de aprendizagem em que o algoritmo controla os exemplos utilizados no treinamento para otimizar o processo é chamado de *aprendizagem ativa*. A aplicação dessa técnica para melhorar o processo de aquisição das preferências do usuário em sistemas de FC tem sido alvo de vários estudos. Em um deles foi proposto o método *ActiveCP* que combinava a controvérsia e da popularidade de um item para determinar a ordem em que seriam apresentados para serem avaliados pelo usuário. O método apresentou bons resultados experimentais.

Neste trabalho, é investigada a utilização de uma nova medida de controvérsia capaz de resolver várias das restrições presentes na metodologia originalmente proposta no *ActiveCP*. É também apresentada uma nova metodologia, mais simples, com uma melhor aplicabilidade prática e que mantém os ganhos de informação na aquisição das preferências dos usuários obtidos pelo método original. Finalmente, a nova metodologia é avaliada em uma base de usuários com avaliações de filmes que simula a base de dados de um sistema em início de operação.

Palavras-chave: sistemas de recomendação, filtragem colaborativa, problema do usuário novo, aprendizagem ativa.

Abstract

Nowadays, the amount of available information is much greater than our ability to manage it. There are hundreds of TV channels, dozens of movies on theatres and thousands of products in on-line stores that we can choose from. When we need to take a decision without being able to know all the possible alternatives, a common approach is to rely on recommendations of other people.

In the 1990s a new category of computer systems has appeared to automatize the recommendation process. Usually the *Recommender Systems*, as they came to get known, acquire some indications of user preferences for providing them with a personalized view of the information. One technique that has been widely used in *Recommender Systems* is the *Collaborative Filtering (CF)*, which uses the similarities among users to make the recommendations. So, for discovering the relevance of a given item i to a target user u , the system relies upon the opinions of other users that have preferences similar to those of the target user u .

However a problem that occurs frequently on *Recommender Systems* is about the arrival of a new user. In that situation, since the newcomer has not rated any information item, the system has no information about his interests. Furthermore the system is not able to generate any recommendation for that user. That problem also occurs on CF-based systems, once the calculation of the similarities among users is based on the items the users have rated in common. A possible solution for diminishing the new user problem is to include an initial preference-acquisition step, where the system presents some items for the new user to rate. However, that must be made in a very efficient way for the system to acquire the most information with a minimum user effort.

The machine learning technique where the algorithm controls the presentation order of training examples for optimizing the learning process is called *active learning*. The application of such technique for improving the acquisition process of user preferences in CF-based systems has been the aim of several studies. In one of the previous work it has been proposed a method called *ActiveCP* that manages to combine the controversy and popularity of a given information item for deciding the order the items should be presented for receiving user rating. The method has obtained good experimental results.

In this work, the use of a new controversy measure is investigated. That measure is capable of solving several restrictions originally present in the proposed methodology for *ActiveCP*. Furthermore, a new methodology is presented; witch is simpler and has a better applicability for real situations. It keeps the good recommendation results that the original method has obtained. Finally, the new methodology was evaluated in a database of user ratings for movies that emulates the database of a recommender system at startup phase.

Keywords: recommender systems, collaborative filtering, new user problem, active learning.

Índice

1. Capítulo 1	1
1.1. O Método ActiveCP	4
1.2. Objetivos	5
1.3. Trabalho Realizado	6
1.4. Organização da Dissertação	7
2. Capítulo 2	9
2.1. Introdução	10
2.2. Tarefas de um sistema de recomendação	12
2.3. Considerações Finais	17
3. Capítulo 3	18
3.1. Introdução	19
3.2. Sistemas de Recomendação Não-Personalizada	19
3.3. Sistemas de Recomendação Personalizada	20
3.4. Filtragem Baseada em Conteúdo	21
3.5. Filtragem Colaborativa	22
3.5.1. Algoritmos de Filtragem Colaborativa	24
3.5.2. Implementação de FC Com <i>K-Nearest Neighbors</i>	25
3.6. Filtragem Híbrida	27
3.7. Avaliando a Qualidade das Recomendações	28
3.7.1. Área Sob Curva ROC	29
3.7.2. Breese	32
3.7.3. Outras métricas	33
3.8. Considerações Finais	33
4. Capítulo 4	35
4.1. Introdução	36
4.2. Aprendizagem Ativa	37
4.2.1. Utilizando Aprendizagem Ativa no Problema da Aquisição de Perfil do Usuário	38
4.3. ActiveCP	40
4.3.1. Controvérsia	41
4.3.2. Popularidade	42
4.3.3. Combinando os Critérios da Controvérsia e da Popularidade	42
4.3.4. Metodologia dos Experimentos do <i>ActiveCP</i>	43
4.3.5. Problemas na Metodologia do <i>ActiveCP</i>	45
4.4. Considerações Finais	48
5. Capítulo 5	49

5.1.	Introdução	50
5.2.	Nova medida de controvérsia	50
5.3.	Nova Metodologia de Experimentos	51
5.4.	Análise Experimental.....	53
5.4.1.	Base de dados	53
5.4.2.	Métodos avaliados.....	55
5.4.3.	Métricas de Avaliação dos Resultados.....	57
5.4.4.	Descrição dos experimentos.....	57
5.4.5.	Resultados Obtidos.....	57
5.4.6.	O Desvio Como Estratégia de Seleção “Combinada”	63
5.5.	Considerações Finais	66
6.	Capítulo 6.....	68
6.1.	Conclusões	69
6.2.	Contribuições.....	69
6.3.	Limitações e Trabalhos Futuros.....	70
	Referências.....	72

Lista de Figuras

Figura 2.1 - Modelo do processo de recomendação.	11
Figura 2.2 - Opiniões dos usuários do site Americanas sobre um item da loja.	13
Figura 2.3 - Lista dos CDs mais vendidos da seção <i>Jazz</i> do site Submarino.....	14
Figura 2.4 - Exemplo de uso da Amazon.com em que o sistema sugere outros produtos relevantes para o usuário interessado no disco <i>Kind Of Blue</i>	15
Figura 2.5 - Recomendação gerada pelo sistema do IMDB a partir do filme <i>As Good As It Gets</i>	16
Figura 3.1 - Exemplo de uma matriz de avaliações de itens por usuários. A célula contendo "?" indica que o sistema precisa prever a relevância do Item 4 para a usuária Paula.	23
Figura 3.2 - Curvas da distribuição dos valores para elementos relevantes (b) e não relevantes (a). O limiar t determina a precisão (taxa de acerto) e o ruído (taxa de erro) do sistema.	29
Figura 3.3 - Exemplo de uma curva ROC. Os pontos t indicados na curva representam diferentes limiares usados para separar itens relevantes de não relevantes para um sistema com sinal de saída entre 1 e 5. Por exemplo, se o limiar usado for de $t=3$ a cobertura será igual a 0,85 enquanto o ruído será igual a 0,6.	30
Figura 3.4 - Exemplos de curvas ROC da cobertura vs. ruído. A área A sob a curva indica a separação entre os elementos relevantes dos não relevantes.....	31
Figura 4.1 - Modelo de coleta “passiva” de avaliações em um sistema de recomendação.	39
Figura 4.2 - Modelo de coleta “ativa” de avaliações em um sistema de recomendação.	39
Figura 4.3 - Metodologia dos experimentos utilizada no ActiveCP.	44
Figura 4.4 - Restrições da metodologia do <i>ActiveCP</i>	47
Figura 5.1- Distribuição da quantidade de itens avaliados por usuários.....	54
Figura 5.2 - Distribuição da quantidade de avaliações recebidas por filmes.....	55
Figura 5.3 - Gráficos da performance das predições usando perfil do usuário formado pelos itens selecionados com os critérios isolados.	60
Figura 5.4 - Gráficos da performance das predições usando perfil do usuário formado pelos itens selecionados com os critérios combinados.....	62
Figura 5.5 - Comparação entre Desvio isoladamente e o Desvio combinado com a Popularidade.....	65

Lista de Tabelas

Tabela 5.1 - Diferenças entre as metodologia do <i>ActiveCP</i> e a nova metodologia proposta.....	52
Tabela 5.2 - Valores médios da medida ROC para as metodologias de seleção isoladas (apenas controvérsia e apenas popularidade).	58
Tabela 5.3 - Valores médios da medida Breese para as metodologias de seleção isoladas (apenas controvérsia e apenas popularidade).....	58
Tabela 5.4 - Valores médios da medida ROC para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas).	61
Tabela 5.5 - Valores médios da medida Breese para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas).	61
Tabela 5.6 - Valores médios da medida ROC para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas) e o desvio.....	64
Tabela 5.7 - Valores médios da medida Breese para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas) e o desvio.....	64



Capítulo 1

Introdução

Nos anos 1990, juntamente com a popularização da *World Wide Web* e da Internet como um todo, ocorreu uma verdadeira explosão da quantidade de informação disponível [Queiroz 03]. Nos nossos dias, as pessoas têm acesso a dezenas ou centenas de canais de TV, milhares de filmes, milhões de CDs e livros e bilhões de documentos on-line. Na infinidade de possibilidades oferecidas pela *Web* há ainda uma quantidade imensa de foros de discussão e comunidades para se participar, além de uma infinidade de lojas virtuais e itens à venda. No entanto, a capacidade das pessoas de avaliar a qualidade e a relevância de uma informação permanece a mesma.

Encontrar informação relevante nesse contexto, freqüentemente envolve a enfadonha análise de grandes quantidades de dados. A natureza repetitiva e tediosa de tal tarefa favoreceu o surgimento de sistemas computacionais que automatizam ou facilitam esse processo de seleção. As tecnologias surgidas para tratar o problema de sobrecarga de informação podem ser geralmente divididas em duas categorias principais: *recuperação de informação* (RI) e *filtragem de informação* (FI) [Belkins 92].

A *recuperação de informação* baseia-se em que o usuário se engaje ativamente em uma atividade de busca pela informação de que necessita. Essa atividade geralmente ocorre na forma de uma consulta formulada pelo próprio usuário em uma linguagem entendida pelo sistema. Há um grande conjunto de técnicas avançadas que realizam busca e extração de informações não estruturadas em vários tipos de documentos, notadamente texto e documentos em HTML [Baeza-Yates 99]. A RI identifica a parte relevante do conteúdo dos itens presentes em sua base de informações e classifica a sua importância para a busca feita pelo usuário. Para realizar essa classificação, existe hoje uma grande variedade de heurísticas disponíveis. Exemplo de sistemas de RI de grande importância são os engenhos de busca na Internet como o Google¹.

As técnicas de RI são bastante úteis quando o usuário consegue descrever o que procura de maneira compatível com a descrição do conteúdo dos itens de informação presentes da base. Os sistemas de RI não distinguem seus usuários, nem identificam seus interesses individuais. Desta forma, também não são capazes de conhecer os interesses de

¹ www.google.com

longo termo de cada usuário, mas apenas o interesse imediato do momento em que uma busca é realizada.

Os sistemas de *filtragem de informação*, por outro lado, geralmente atuam baseados nas necessidades persistentes dos usuários. Para isto, eles costumam manter um perfil com os interesses estáveis do usuário (e.g. manter-se atualizado sobre determinado tópico) para usá-lo na busca de informação. De acordo com esses interesses o sistema aplica filtros sobre novos itens de informação do sistema e realiza a ação apropriada. Essas ações podem ser de trazer o item ao conhecimento do usuário ou simplesmente ignorar a informação [Herlocker 00]. Nesse tipo de sistema, os usuários tendem a se engajar de maneira mais passiva na tarefa de busca por informação e a terem interesses que variam lentamente com o tempo.

FI e RI, embora atuem na solução do mesmo problema, o fazem de maneira diferente e se aplicam em contextos distintos. Uma categoria de sistemas de informação surgida também para tratar o problema da sobrecarga de informação são os *sistemas de recomendação*. Esse tipo de sistema freqüentemente utiliza-se de técnicas de FI e RI para automatizar ou dar suporte computacional ao processo de geração e fornecimento de recomendações.

Há ainda sistemas de recomendação que são capazes de identificar as necessidades individuais de informação de cada usuário e são, por isso, chamados de *sistemas de recomendação personalizada*. Uma abordagem de geração de recomendações que tem sido amplamente empregada nesse tipo de sistema é a *Filtragem Colaborativa (FC)* [Herlocker 04]. Ela funciona estabelecendo relações entre usuários, detectando similaridade entre eles para automatizar o processo no qual, pessoas com preferências semelhantes auxiliam umas às outras na busca por informação. Por exemplo, para gerar recomendação de um dado filme i para um usuário u utilizando essa abordagem, o sistema teria de identificar que outros usuários demonstraram gostos similares aos de u e utilizar as opiniões deles para prever a relevância de i .

Para gerar a predição em sistemas de FC um dos algoritmos mais amplamente utilizados tem sido o kNN (*k-nearest neighbors*) [Herlocker 00]. Assim, para um dado item que se deseja sugerir a um usuário alvo, tem-se um kNN onde os exemplos são os usuários que avaliaram o item em questão, e a classificação de cada exemplo é a avaliação para o

item feita pelo usuário que o exemplo representa. Para fazer a predição de um item, o sistema procura os k usuários mais próximos do usuário alvo, dentre aqueles que avaliaram o item em questão.

Antes que um sistema de recomendação personalizada seja capaz de gerar boas recomendações, ele precisa conhecer as preferências e gostos do usuário. Esses indicadores de preferências são obtidos frequentemente através de notas atribuídas pelos usuários aos itens do sistema. No caso em que o kNN é utilizado, essas notas são utilizadas pelo algoritmo para cálculo de similaridade entre usuários. A quantidade de itens que um usuário precisa avaliar antes que o sistema comece a fornecer-lhe boas recomendações pode ser grande, o que pode gerar desestímulo para um usuário novato. Uma vizinhança calculada para um usuário com apenas dois filmes, por exemplo, pode não ser muito confiável para gerar predições. Esse problema tem ficado conhecido como “problema do usuário novo” e várias abordagens de aprendizagem ativa têm sido propostas para resolvê-lo [Yu 04], [Boutilier 03], [Rashid 02].

1.1. O Método ActiveCP

Aprendizagem ativa em um algoritmo de aprendizagem supervisionada é o processo pelo qual o algoritmo tem algum controle sobre quais exemplos serão utilizados em seu treinamento. A aplicação dessa técnica em geral consiste em apresentar uma seqüência de itens para serem avaliados pelo usuário novato. Os itens selecionados são aqueles que o sistema julga que, uma vez avaliados, trarão o maior ganho de informação sobre as preferências do usuário. O objetivo é obter do usuário um número suficiente (ou menor possível) de avaliações para que o sistema de recomendação obtenha um bom desempenho de precisão.

Em sua dissertação de mestrado [Teixeira 02a], Teixeira propôs um método de aprendizagem ativa para sistemas de Filtragem Colaborativa em que a ordenação dos itens apresentados ao usuário era feita utilizando os conceitos de controvérsia e popularidade de um item. O conceito de popularidade foi definido como sendo simplesmente a quantidade de avaliações que um item havia recebido. Para medir a controvérsia, foi utilizada a variância de todas as notas recebidas por um item. O método consistia basicamente em

gerar duas listas ordenadas dos itens. Na primeira eles eram ordenados pela popularidade e na outra pela controvérsia. Em seguida, as duas listas eram combinadas gerando a ordenação final em que os itens seriam apresentados ao usuário para que ele lhes atribuísse nota. Um algoritmo de FC utilizando kNN utilizava as notas atribuídas pelo usuário para gerar recomendações de novos filmes.

O método *ActiveCP* foi testado experimentalmente em uma base de dados consistindo de avaliações de filmes feita por usuários reais. Os resultados dos experimentos mostraram que a utilização do método representava um ganho de informação considerável quando comparado com a seleção aleatória de itens e também com a seleção feita utilizando apenas a controvérsia ou a popularidade.

No entanto, o *ActiveCP* tal como proposto apresenta algumas limitações decorrentes do fato de a medida de controvérsia utilizada (a variância) não refletir em seu resultado a quantidade de avaliações recebidas por um item. Em seu trabalho, Teixeira decidiu fixar a quantidade de avaliações utilizadas no cálculo da variância para todos os itens. Essa abordagem apresentou êxito nos testes realizados em uma base de dados relativamente grande, com muitas avaliações. No entanto, não foi realizado um estudo mais aprofundado do impacto dessa abordagem em uma base com poucas avaliações, onde a quantidade de avaliações de um item pudesse ser pequena.

1.2. Objetivos

Nesta dissertação, será analisado o problema da seleção de itens na aplicação da aprendizagem ativa em sistemas de filtragem colaborativa. Em especial, investigamos os problemas do método *ActiveCP* [Teixeira 02b] que utiliza os conceitos de controvérsia e popularidade para selecionar itens. Em resumo, os principais objetivos desta dissertação são:

1. Investigar os problemas na metodologia proposta pelo *ActiveCP*, em grande parte ocasionados pela medida que foi utilizada para medir a controvérsia de um item, no caso, a variância das notas atribuídas ao filme;

2. Investigar o impacto da possível perda de informação ocasionada pelos problemas do *ActiveCP* em uma base com dados escassa, que simula a situação de um sistema de recomendação em início de operação;
3. Apresentar uma nova medida de controvérsia, que chamaremos de desvio, que resolva os problemas inerentes à medida da variância;
4. Propor uma nova metodologia de experimentos utilizando a nova medida de controvérsia, que mantenha ou melhore o ganho de desempenho obtido com a abordagem anterior e que seja mais diretamente aplicável na prática.

Analisaremos detalhadamente o método de aprendizagem ativa para sistemas de filtragem colaborativa chamado *ActiveCP*. Veremos como ele combina os critérios de controvérsia e popularidade para selecionar os itens mais informativos das preferências de um usuário. Discutiremos como a variância, medida de controvérsia utilizada, pode dificultar a utilização do método em uma situação prática.

Introduziremos uma nova medida de controvérsia e apresentaremos uma nova metodologia de experimentos que não herdamos as limitações da que foi utilizada no *ActiveCP*. Por último, apresentaremos os resultados de experimentos executados com a nova metodologia em uma base de dados de um sistema de recomendação de filmes.

1.3. Trabalho Realizado

Neste trabalho, foi investigado o impacto dos problemas ocasionados pela utilização da variância como medida de controvérsia no *ActiveCP*. Uma nova medida de controvérsia, que chamamos de desvio, foi proposta com o objetivo de resolver os problemas demonstrados pela variância. A metodologia de experimentos também foi modificada para tirar proveito das vantagens da medida do desvio e para tornar-se mais simples de ser empregada na prática.

Os experimentos foram realizados em uma base pequena, com 300 usuários e 394 filmes com uma média de 54 avaliações por filme. Nessa base foram executados testados experimentalmente os seguintes métodos de seleção:

- Aleatória, para ser utilizada como base na comparação do desempenho das outras metodologias;

- Controvérsia dada pela variância isolada com quantidade fixa de avaliações por filme;
- Controvérsia dada pela variância isolada utilizando todas as avaliações de um filme;
- Controvérsia dada pelo desvio;
- Popularidade;
- Seleção combinando as medidas de controvérsia listadas anteriormente com a popularidade.

A quantidade de filmes selecionados para serem avaliados pelos usuários (i.e. tamanho do perfil do usuário) foi variado entre 2 e 10 avaliações.

Ao final, os resultados dos experimentos nos permitiram confirmar a existência de uma limitação na utilização da variância como medida de controvérsia. Além disso, o desvio mostrou-se muito útil ao ser utilizado na seleção de itens para aplicações de aprendizagem ativa em sistemas de FC com kNN.

1.4. Organização da Dissertação

O restante da dissertação está organizado em cinco capítulos, conforme descrição abaixo:

- **Capítulo 2 – Sistemas de Recomendação:** Apresenta uma visão geral dos sistemas de recomendação, suas tarefas e sua utilização.
- **Capítulo 3 – Técnicas de Sistemas de Recomendação:** Faz uma breve análise das duas categorias mais amplas de sistemas de recomendação: os sistemas personalizados e os não-personalizados. Em seguida, descreve as principais técnicas empregadas nesses sistemas.
- **Capítulo 4 – O método *ActiveCP*:** Descreve o problema da aquisição do perfil do usuário e a metodologia de aprendizagem ativa utilizada para tratá-lo chamada *ActiveCP*.
- **Capítulo 5 – Propostas:** Descreve a nova medida de controvérsia juntamente com uma nova metodologia experimental propostas. Os resultados experimentais utilizando a nova metodologia são apresentados.

- **Capítulo 6 – Conclusões:** Sumariza as principais conclusões do trabalho juntamente com possíveis trabalhos futuros.



Capítulo 2

Sistemas de Recomendação

2.1. Introdução

Receber e fornecer recomendações é um ato de comunicação bastante comum no nosso dia-a-dia. Frequentemente procuramos saber a opinião de colegas sobre filmes em cartaz, consultamos a crítica especializada em revistas e jornais, ou pedimos a sugestão do vendedor da loja de discos sobre de que álbuns um apreciador de MPB provavelmente iria gostar.

Esses exemplos nos ajudam a entender o conceito de recomendação. Em geral, um indivíduo se encontra diante de uma decisão a ser feita em meio a um universo de alternativas. Esse universo é geralmente enorme, tornando até mesmo impossível para o indivíduo conhecer ou avaliar satisfatoriamente cada uma das possibilidades.

Uma recomendação pode se basear nas preferências de quem a faz e pode ser dirigida a um indivíduo específico, no caso de uma sugestão feita a um amigo, ou para um público mais geral, no caso de uma crítica escrita por um especialista. Para a pessoa que recebe a recomendação, ela funciona como um filtro ou uma visão particular de um universo de possibilidades geralmente inacessível. Ela pode levar em consideração também a preferência de quem está à procura de sugestões e não apenas de quem a faz. Por exemplo, ao sugerir discos de música brasileira a alguém, certamente será relevante saber quais estilos ou artistas tal pessoa prefere (e.g. gosta da Bossa Nova de João Gilberto, mas não gosta do Tropicalismo de Caetano Veloso). É possível até mesmo fazer recomendação baseada nas opiniões de outras pessoas. Alguém que não é admirador do gênero *Jazz* pode recomendar discos baseado no que seus amigos que apreciem tal estilo costumam ouvir. Ainda, a recomendação pode incluir explicações sobre como ela foi gerada para permitir que o seu recebedor a avalie.

A Figura 2.1 ilustra o modelo geral para o processo de recomendação apresentado por Terveen et al., com algumas adaptações [Terveen 01]. Ela será útil para entendermos o funcionamento e a utilização dos sistemas computacionais de recomendação.

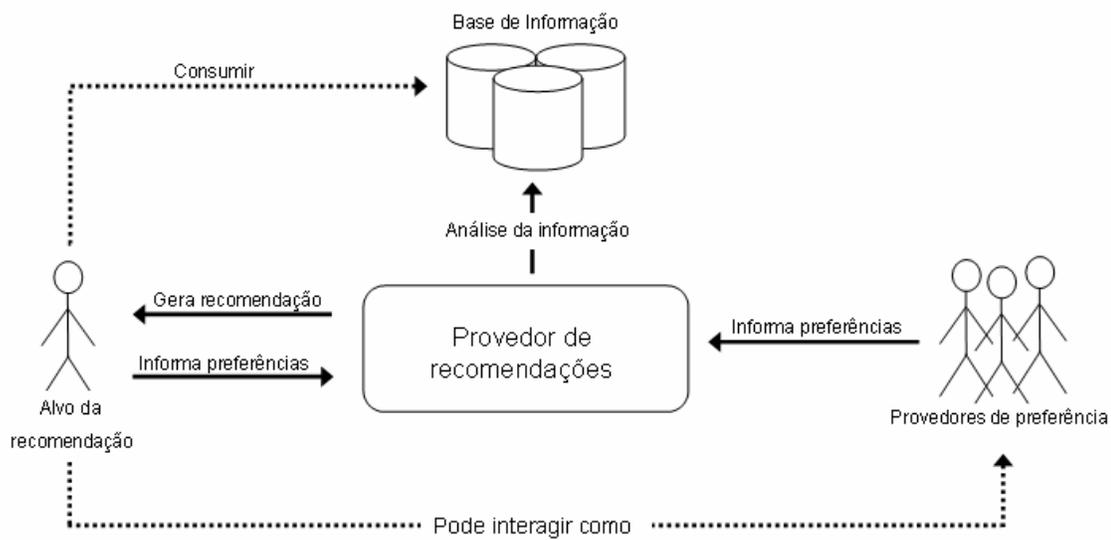


Figura 2.1 - Modelo do processo de recomendação.

Os elementos envolvidos no processo são:

- **Alvo da recomendação:** principal ator do processo e destinatário da recomendação. Ele pode apenas receber passivamente as sugestões ou pode fornecer também suas opiniões sobre os itens da base de informação.
- **Base de informação:** onde se encontra todo o universo de escolhas e onde o *alvo da recomendação* deseja encontrar informação de seu interesse.
- **Provedor de recomendações:** agente que seleciona e recomenda os itens da *base de informação* que têm maior relevância para *alvo da recomendação*. Para identificar quais são esses itens, o provedor pode se basear nas preferências informadas pelo próprio *alvo da recomendação*, nas preferências de outras pessoas e/ou no conteúdo dos itens.
- **Provedores de preferência:** todo processo de recomendação está baseado em indicadores humanos de preferência. Assim, os *provedores de preferência* são tipicamente um conjunto de pessoas em cujas opiniões o *provedor de recomendações* pode se basear para fazer as sugestões.

São chamados de **sistemas de recomendação** os sistemas de informação que automatizam ou dão suporte computacional a alguma das partes do processo descrito acima. Eles podem assumir o papel do *provedor de recomendações*, fornecendo recomendações aos usuários baseado em seus interesses e possivelmente levando em consideração também as opiniões de outros usuários. Nesse caso os sistemas de recomendação empregam técnicas de FI e de inteligência computacional na realização das tarefas de sugestão.

Diferentemente, os sistemas de recomendação podem apenas dar suporte à criação e ao compartilhamento de recomendações e opiniões entre as pessoas [Terveen 01]. Nesse caso, o objetivo do sistema é facilitar o processo de comunicação das sugestões entre a própria comunidade de usuários do sistema.

2.2. Tarefas de um sistema de recomendação

Para que possamos categorizar com clareza os sistemas de recomendação é importante identificar as tarefas frequentemente realizadas por esses sistemas. As tarefas de um sistema dizem respeito à sua comunicação com o usuário e a como os objetivos do usuário ao utilizá-lo são satisfeitos.

Em [Herlocker 00] e [Bezerra 04], são discutidas algumas tarefas frequentemente desempenhadas pelos sistemas de recomendação. A seguir estão descritas as que julgamos mais relevantes para este trabalho juntamente com exemplos de seu emprego em sistemas reais.

a) Mostrar as opiniões dos usuários do sistema acerca de um item de informação

Sistemas de informação on-line como *Americanas*² e *Mercado Livre*³ possibilitam que usuários forneçam suas opiniões a respeito de produtos de consumo e negociadores de leilão. Essas opiniões tornam-se visíveis a outros usuários do sistema, permitindo-lhes embasar a sua decisão nas opiniões de outras pessoas da comunidade.

² www.americanas.com.br

³ www.mercadolivre.com.br

A indicação de preferência é geralmente feita através de uma nota que o usuário atribui a um item podendo também anexar uma descrição textual de seu ponto de vista (Figura 2.2). Alguns sistemas permitem ao usuário indicar a relevância até mesmo dos comentários feitos pelos outros usuários sobre determinado item, permitindo ao sistema exibi-los ordenadamente.

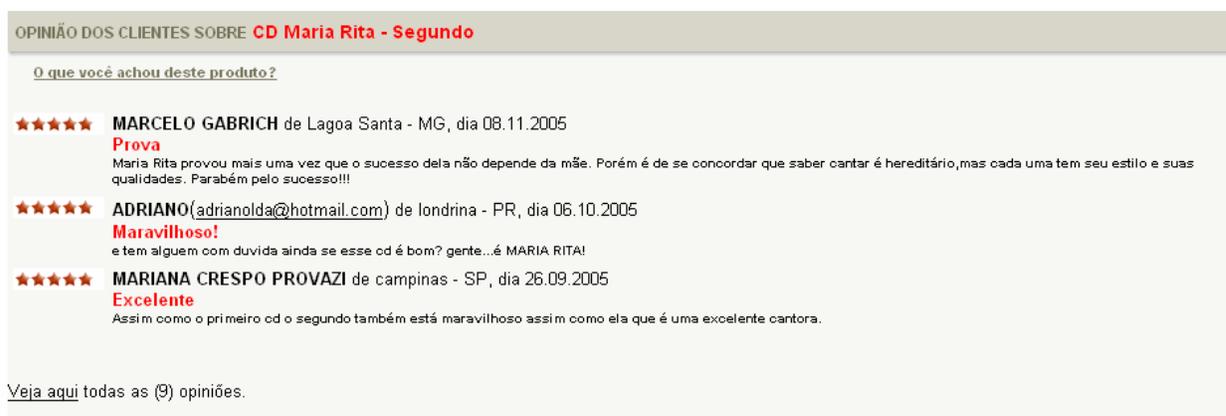


Figura 2.2 - Opiniões dos usuários do site Americanas sobre um item da loja.

b) Mostrar os itens mais comumente acessados da base

Essa é uma das tarefas mais comumente presentes em sistemas de recomendação. É comum em sites de comércio eletrônico haver uma lista de itens mais vendidos ou mais populares. Essa característica apóia-se no fato de que as pessoas frequentemente procuram saber a opinião da maioria no momento de tomar uma decisão.

Para esse tipo de sugestão não é necessário que os usuários criem o hábito de emitir suas opiniões sobre os itens explicitamente, pois as recomendações são geradas automaticamente pelo sistema baseado no uso comum que os usuários fazem dele.

The screenshot shows the Submarino website interface. At the top, there's a navigation bar with categories: LIVROS, CDs, DVDs, ELETRÔNICOS, CINE&FOTO, ELETRODOMÉSTICOS, ELETROPORTÁTEIS, UD, BRINQUEDOS, GAMES, INFORMÁTICA, TELEFONIA, ESPORTE&LAZER, BEBÊS, BELEZA&SAÚDE, PERFUMARIA, MODA, JÓIAS&RELÓGIOS, PESHOP, CAMA,MESA&BANHO, INST. MUSICAIS, FERRAMENTAS&AUTOMOTIVO. A search bar is located on the right with the text 'BUSCA: Palavra-chave'. Below the navigation bar, there's a date 'Terça, 15/11/2005'. The main content area is titled 'MAIS VENDIDOS' and features four CD listings:

- Billie Holiday BILLIE HOLIDAY**: Esta é mais uma coletânea fantástica da diva Billie Holiday, uma das maiores vozes do jazz de todos os tempos. São 12 faixas com o melhor de seu estilo, entre elas, "You've Changed" e "Lover Man", ...
- More Music From Ray RAY CHARLES**: *More Music From Ray* é a continuação da trilha sonora do aclamado filme sobre a vida do inesquecível Ray Charles, um dos maiores músicos de todos os tempos, vencedor de 12 prêmios Grammy. O f...
- GNT - New Jazz VARIOS**: Michael Bublé, Jamie Cullum, Diana Krall e Gwyneth Herbert estão entre os grandes ícones do jazz presentes *GNT - New Jazz*, a coletânea fantástica com o melhor do estilo. São 14 faixas compon...
- Genius Loves Company CD + DVD RAY CHARLES**: *Genius Loves Company* é último legado do genial músico Ray Charles, um disco de duetos, que foi gravado no próprio estúdio do músico e conta com a presença de doze artistas consagrados

Figura 2.3 - Lista dos CDs mais vendidos da seção *Jazz* do site Submarino

A Figura 2.3 mostra a lista de discos mais vendidos na seção *Jazz* do site de comércio eletrônico Submarino⁴.

c) Mostrar uma lista de itens relevantes para um usuário com base no contexto visualizado pelo usuário em dado momento.

Atualmente vários sites de comércio eletrônico são capazes de identificar os itens nos quais o usuário demonstra interesse ao longo de uma sessão de uso do sistema e, com base nisto, selecionar outros itens da base que podem ser relevantes para o usuário. Geralmente os itens que o sistema assume como sendo do interesse do usuário são os que os usuários procuram, visualizam ou compram.

A Figura 2.4 mostra um exemplo de uso da Amazon⁵ em que o sistema fornece recomendações de outros discos, DVD's e livros que podem interessar a um usuário que está procurando pelo álbum de *Kind Of Blue* do Miles Davis.

⁴ www.submarino.com.br

⁵ www.amazon.com

MUSIC INFORMATION
Explore this album
buying info
listen to samples
Amazon.com
Articles
editorial reviews
customer reviews

RATE THIS ALBUM
(Sign in to rate this item.)

Kind of Blue [ORIGINAL RECORDING REISSUED] [ORIGINAL RECORDING REMASTERED]
Miles Davis

List Price: \$44.98
Price: **\$10.99** and eligible for **FREE Super Saver Shipping** on orders over \$25. [See details.](#)
You Save: **\$0.99 (8%)**

Availability: Usually ships within 24 hours. Ships from and sold by Amazon.com. **Want it delivered Tomorrow?** Order it in the next 0 hours and 0 minutes, and choose **One-Day Shipping** at checkout. [See details.](#)

103 used & new from \$5.77
Club price: **\$8.49**

Share your own customer images ▶ [See more product details](#)

Customers who bought this title also bought:

- ◆ [A Love Supreme \[Original recording reissued\] \[Original recording remastered\]](#) ~ John Coltrane ([Rate it](#))
- ◆ [Time Out \[Original recording reissued\] \[Original recording remastered\]](#) ~ Dave Brubeck ([Rate it](#))
- ◆ [Birth of the Cool \[Original recording reissued\] \[Original recording remastered\]](#) ~ Miles Davis ([Rate it](#))
- ◆ [Thelonious Monk with John Coltrane](#) ~ Thelonious Monk with John Coltrane ([Rate it](#))
- ◆ [Sketches of Spain \[Original recording remastered\]](#) ~ Miles Davis, Gil Evans ([Rate it](#))
- ◆ [The Ultimate Blue Train \[Enhanced\] \[Original recording remastered\]](#) ~ John Coltrane ([Rate it](#))
- ◆ ['Round About Midnight \[Original recording remastered\]](#) ~ Miles Davis ([Rate it](#))
- ◆ [Saxophone Colossus \[Original recording reissued\]](#) ~ Sonny Rollins ([Rate it](#))

▶ [Explore Similar Items:](#) 20 in Music, 10 in Books, and 18 in DVD

Visit the **Books Store**
Kind of Blue: The Making of the Miles Davis Masterpiece by Jimmy Cobb (Foreword), Ashley Kahn

Figura 2.4 - Exemplo de uso da Amazon.com em que o sistema sugere outros produtos relevantes para o usuário interessado no disco *Kind Of Blue*.

Para descobrir em que outros itens da base de informações o usuário pode estar interessado, os sistemas de recomendação podem utilizar duas abordagens. Uma primeira maneira é relacionar os itens do sistema, isto é, identificar a similaridade entre itens. Assim, o sistema pode sugerir itens parecidos com o que o usuário buscou e está visualizando em dado momento. Por exemplo, para um usuário que está comprando uma câmera digital, o sistema pode sugerir que ele adquira também um cartão de memória.

A outra abordagem consiste em sugerir os itens que foram consumidos por pessoas para as quais o tópico de interesse do usuário também foi relevante. Para isto é preciso fazer um cruzamento das informações de compras e/ou navegação dos consumidores com o item ou o produto visualizado pelo usuário alvo.

d) Mostrar ao usuário uma lista de itens interessantes levando em consideração o seu perfil

Neste caso, o sistema de recomendação é capaz de manter um perfil contendo os interesses demonstrados pelo usuário no decorrer dos usos que faz do sistema. Com essa característica o sistema é capaz de fornecer um alto grau de personalização nas

recomendações feitas, fornecendo ao usuário uma visão bastante particular da base de informação.

O IMDB⁶ (*Internet Movie Database*) é um sistema de recomendação que possui grande banco de dados de filmes e programas de TV e que também faz sugestões para seus usuários. O sistema armazena no perfil dos usuários os filmes assistidos por eles e as notas opcionalmente atribuídas aos itens. Com essas informações o sistema sugere itens semelhantes a um determinado item indicado pelo usuário. A Figura 2.5 mostra a lista de filmes sugeridos a partir do filme *As Good As It Gets*.

The screenshot shows the IMDb website interface. At the top, there is a navigation bar with links like 'NOW PLAYING', 'MOVIE / TV NEWS', 'MY MOVIES', 'DVD / VIDEO', 'IMDb TV', 'MESSAGE BOARDS', 'SHOWTIMES & TICKETS', and 'GAME BASE'. Below this is a search bar with the text 'Search the IMDb' and a 'go' button. The main content area is titled 'Recommendations for As Good As It Gets (1997)'. Below the title is a link 'How do these recommendations work?'. A table lists suggested movies with their respective user ratings. The table has five columns: 'Suggested by the database', 'Look up in IMDb', 'Showtimes (US only)', 'Available @Amazon', and 'User Rating'. The movies listed include 'Sideways (2004)', 'Used People (1992)', 'American Beauty (1999)', 'Todo sobre mi madre (1999)', 'Dogma (1999)', 'Cruel Intentions (1999)', 'Chasing Amy (1997)', 'Terms of Endearment (1983)', 'Terminator 2: Judgment Day (1991)', and 'Coming out (1989/1)'. A tip at the bottom of the table says: 'Tip: if you want to see if a movie is showing in a cinema near you, click the film roll. (USA only)'. On the left side of the page, there is a sidebar with navigation links like 'Overview', 'main details', 'combined details', 'full cast and crew', 'company credits', 'Awards & Reviews', 'user comments', 'external reviews', 'newsgroup reviews', 'awards & nominations', 'user ratings', 'recommendations', 'Plot & Quotes', 'plot summary', and 'nint kewnrns'.

Suggested by the database	Look up in IMDb	Showtimes (US only)	Available @Amazon	User Rating
Sideways (2004)	IMDb		DVD VHS	8.0
Used People (1992)	IMDb		VHS	5.5
American Beauty (1999)	IMDb		DVD VHS	8.5
Todo sobre mi madre (1999)	IMDb		DVD VHS	7.9
Dogma (1999)	IMDb		DVD VHS	7.3
Cruel Intentions (1999)	IMDb		DVD VHS	6.6
Chasing Amy (1997)	IMDb		DVD VHS	7.6
Terms of Endearment (1983)	IMDb		DVD VHS	7.3
Terminator 2: Judgment Day (1991)	IMDb		DVD VHS	8.2
Coming out (1989/1)	IMDb		DVD VHS	7.2

Tip: if you want to see if a movie is showing in a cinema near you, click the film roll. (USA only)

Figura 2.5 - Recomendação gerada pelo sistema do IMDB a partir do filme *As Good As It Gets*.

⁶ www.imdb.com

2.3. Considerações Finais

Neste capítulo, vimos como o processo de recomendação pode ser útil para auxiliar na busca por informação relevante. Também foi introduzida a categoria de sistemas computacionais que automatiza alguma ou várias das etapas do processo de recomendação. Eles têm sido chamados de sistemas de recomendação e 4 das principais tarefas geralmente desempenhadas por eles foram discutidas.

As tarefas **(a)**, **(b)** e **(c)** são mais simples e consistem em aplicações diretas de técnicas de filtragem de informação. Já para realizar a tarefa **(d)**, além do problema da filtragem da informação, há também o problema da aquisição de um perfil com os indicadores de preferência do usuário que seja eficiente do ponto de vista da qualidade das sugestões e computacionalmente tratável. Além disso, é importante que seu processo de montagem seja tão transparente para o usuário quanto possível.

Em sistemas que empregam essa tecnologia geralmente estão também os maiores problemas e áreas de investigação no campo da recomendação. É também na investigação de aspectos relacionados a esse tipo sistema que este trabalho dedicará a maior parte de sua atenção.



Capítulo 3

Técnicas para Sistemas de Recomendação

3.1. Introdução

Neste capítulo, discutiremos as diferentes tecnologias empregadas nos sistemas de recomendação. Começaremos fazendo uma breve análise das duas categorias mais amplas de sistemas de recomendação: os sistemas personalizados e os não-personalizados.

Em seguida, entraremos em maiores detalhes sobre os principais paradigmas e tecnologias utilizadas nos sistemas de recomendação personalizada. Analisaremos também as métricas mais utilizadas na avaliação dessa categoria de sistemas.

3.2. Sistemas de Recomendação Não-Personalizada

São sistemas cujo papel principal é o de mediador entre geradores e consumidores de recomendações, fazendo apenas a coleta e a distribuição das recomendações. As tarefas que eles realizam geralmente estão incluídas nas três primeiras listadas na seção 2.2.

O sistema Tapestry [Goldberg 92] é frequentemente citado como o primeiro sistema de recomendação conhecido. O Tapestry foi desenvolvido no Xerox Palo Alto Research Center com o objetivo de prover suporte a recomendações de mensagens eletrônicas em grupos de notícias ou listas de e-mails. O sistema permite que os usuários avaliem o conteúdo de uma mensagem como “bom” ou “ruim” e também que adicionem comentários expressando sua opinião. As avaliações e comentários são armazenados junto com as mensagens e ficam acessíveis aos filtros aplicados pelo outros usuários do sistema. Assim, um usuário do sistema poderia filtrar as mensagens com base não apenas no conteúdo delas, mas também nas opiniões das outras pessoas. Por exemplo, poderia-se aplicar um filtro do tipo “mostre-me as mensagens que Igor avaliou como boas”. Além de indicadores explícitos de opinião dos usuários, o Tapestry permite que os filtros levem em consideração também indicadores implícitos da opinião dos outros usuários. Isso possibilitava que um assinante de uma lista de discussão criasse um filtro para mostrar-lhe as mensagens que foram respondidas por pessoas que ele sabe que respondem às mensagens mais interessantes do fórum.

Os sistemas de recomendação não personalizada, embora se apoiem no processo de recomendação, não são capazes de identificar as preferências e necessidades de longo termo de seus usuários nem de gerar essas recomendações automaticamente com base nessas necessidades. Grande parte das vezes, o que eles fazem é fornecer o suporte para que as pessoas produzam e compartilhem recomendações sobre itens de interesse comum. Muitos sites de comércio eletrônico atuais incluem várias formas desse tipo de recomendação, como a possibilidade de os compradores darem uma nota e escreverem uma análise sobre um produto ou ainda a exibição da opinião de um especialista.

3.3. Sistemas de Recomendação Personalizada

Os sistemas capazes de aprender e identificar as necessidades individuais de um usuário e de gerar recomendações adequadas a essas necessidades são chamados de **Sistemas de Recomendação Personalizada**.

Ao coletar as preferências dos usuários, os Sistemas de Recomendação Personalizada as armazenam em uma representação interna chamada de **perfil**. Tradicionalmente, esses sistemas empregam técnicas de filtragem de informação e aprendizagem de máquina para gerar recomendações apropriadas aos interesses do usuário a partir da representação de seu perfil. No entanto, outras técnicas não inseridas no domínio da FI, como Redes Neurais, Redes Bayesianas e Regras de Associação, também são utilizadas na filtragem dos itens de informação [Schaffer 01].

O processo de filtragem de informação envolve a realização de duas atividades: encontrar a informação de que o usuário necessita (*filtering in*) e/ou eliminar a informação desnecessária (*filtering out*). Malone et al. [Malone 87] identificaram três formas como essas atividades podem ser realizadas:

- **Filtragem Cognitiva ou Baseada em Conteúdo:** seleciona a informação baseada no conteúdo dos elementos filtrados. Um filtro de mensagens eletrônicas que envie para a lixeira mensagens que contenham palavras indesejadas é um exemplo de filtragem cognitiva.
- **Filtragem Social ou Colaborativa:** é baseada no relacionamento entre as pessoas e nos seus julgamentos subjetivos a respeito da informação a ser filtrada. Um

exemplo simples de filtragem social é a seleção de mensagens eletrônicas com base na relação entre remetente e destinatários de uma mensagem.

- **Filtragem Econômica:** baseia-se na relação custo/benefício em produzir ou consumir um item de informação. Um filtro de mensagens eletrônicas utilizando essa abordagem poderia, por exemplo, estimar o custo de uma mensagem com base no tamanho de seu texto. Ou ainda, poderia levar em consideração o custo de produção de uma mensagem. Caso uma mensagem tivesse sido enviada para um grande número de destinatários, o seu custo de produção por usuário seria considerado pequeno e, conseqüentemente, uma baixa prioridade lhe seria atribuída.

Os sistemas de recomendação personalizada utilizam em sua maior parte, técnicas de duas categorias de filtragem de informação: a filtragem baseada em conteúdo e a filtragem colaborativa, ou ainda uma combinação das duas [Cohen 98].

3.4. Filtragem Baseada em Conteúdo

Os sistemas de recomendação que utilizam filtragem baseada em conteúdo fazem a sugestão de itens que sejam semelhantes aos que o usuário demonstrou interesse no passado. A aquisição das preferências neste tipo de filtragem apóia-se fortemente na descrição do conteúdo dos itens de informação (e.g. um filme poderia ser descrito pelo conjunto de atributos: nome, diretor, atores, ano, gênero). Com base nas opiniões que o usuário demonstra sobre itens da base e na descrição desses itens, o sistema deve ser capaz de criar um perfil que reflita os tipos de informação nos quais o usuário se interessa.

A maneira de representar o perfil do usuário depende das técnicas de aprendizagem de máquina e de recuperação de informação utilizadas. Por exemplo, o perfil pode ser utilizado em um algoritmo de aprendizagem de máquina para derivar um classificador capaz de prever a relevância de um item a partir dos seus atributos [Pazzani 97], ou ainda podem ser inferidas similaridades entre itens com base no conhecimento do domínio do sistema, transformando a filtragem em uma aplicação de CBR (*case-based reasoning*) [Burke 99]. É comum também a utilização de protótipos que representem os interesses do

usuário. Esses protótipos são compostos geralmente de um vetor de palavras e pesos associados à relevância da palavra [Cotter 00], [Krukwich 96], [Balabanovic 97].

Filtragem baseada em conteúdo está fora do escopo deste trabalho. Focaremos e entraremos em mais detalhes sobre as técnicas e os problemas de filtragem colaborativa.

3.5. Filtragem Colaborativa

A técnica de filtragem colaborativa (FC) baseia-se no fato de que as melhores recomendações para um indivíduo podem ser aquelas feitas por pessoas que possuem preferências similares às dele. Assim, o que os sistemas de FC fazem é identificar usuários similares ao alvo da recomendação e sugerir itens que esses “vizinhos” avaliaram como relevantes e que ainda não tenha sido consumido pelo usuário alvo. O processo de sugestão está apoiado inteiramente na similaridade entre os usuários de sistema e não na similaridade dos itens de informação.

O processo da filtragem colaborativa pode ser generalizado em três passos [Queiroz 03]:

- **Representação dos dados de entrada:** o usuário expressa suas preferências avaliando itens do sistema. As avaliações, tanto as positivas quanto as negativas, revelam o interesse do usuário em itens específicos e são armazenadas em seu perfil. Essas avaliações podem ser coletadas tanto a partir de indicações explícitas de preferência por parte do usuário quanto de maneira implícita (e.g. um sistema de comércio eletrônico assume que o usuário gosta dos itens que comprou).
- **Formação de vizinhança:** para fazer a recomendação o sistema compara o perfil do usuário alvo com os perfis de outros usuários do sistema para encontrar a similaridade entre eles. Existem várias formas de calcular esse conjunto de usuários similares, ao qual chamaremos de **vizinhança**.
- **Geração da recomendação:** finalmente, com base nas avaliações feitas pelos componentes da vizinhança aos itens de informação, o sistema gera a recomendação para o usuário alvo. Em outras palavras, o sistema sugere ao usuário alvo os itens que seus vizinhos mais gostaram. Novamente, a mecânica

utilizada para gerar as recomendações pode variar de acordo com o método de FC utilizado.

Para ilustrar melhor esse processo, consideremos a forma mais simples de representação dos dados de entrada que é uma tabela em que as linhas representam os usuários, as colunas representam os itens e cada célula contém a avaliação dada pelo usuário ao item, conforme a Figura 3.1. As notas das avaliações encontram-se no intervalo de 1 a 5, onde 1 indica que o item tem relevância mínima para o usuário e 5 indicando relevância máxima. A célula contendo o sinal de interrogação indica que o sistema precisa prever a relevância do **Item 4** para a usuária **Paula**.

	Item 1	Item 2	Item 3	Item 4
Pablo	2	4		1
Bebeto		2	5	5
Paula	3	1	5	?

Figura 3.1 - Exemplo de uma matriz de avaliações de itens por usuários. A célula contendo "?" indica que o sistema precisa prever a relevância do Item 4 para a usuária Paula.

Com os indicadores de preferência armazenados para cada usuário, o sistema de FC precisa, no segundo passo do processo, identificar quem são os vizinhos do usuário alvo (no nosso exemplo, Paula) com base nos itens que eles tenham avaliado em comum. No exemplo da Figura 3.1, os vizinhos de Paula são os usuários Pablo e Bebeto, possuindo dois itens avaliados em comum com cada um. Pela análise das notas dadas aos itens avaliados em comum, percebemos que Paula possui maior afinidade de preferências com Bebeto, já que têm opiniões semelhantes quanto aos itens que avaliaram em comum; enquanto que com Pablo essas opiniões são mais discordantes. Com base nisto, é desejável que o sistema conclua que o Item 4 será relevante para Paula assim como foi para Bebeto e o sugira.

Nos sistemas de FC, não há necessidade de se representar o conteúdo dos itens, sendo suficiente apenas haver um identificador para cada item e para cada usuário. Toda a informação utilizada na geração das recomendações deriva diretamente do julgamento humano. Isso significa que características realmente importantes para a avaliação de qualidade são tacitamente levadas em consideração, mesmo que não sejam facilmente

representadas em computadores. Além disso, permitem que sejam recomendados itens cujo conteúdo seja muito difícil de analisar ou tratar automaticamente, como conteúdo multimídia de som e áudio.

3.5.1. Algoritmos de Filtragem Colaborativa

Conforme mencionado anteriormente, há várias abordagens e algoritmos sendo utilizados com sucesso no processo de filtragem colaborativa. Billsus e Pazzani [Billsus 98] apresentaram um modelo para FC em que empregavam Redes Neurais. Para cada usuário do sistema era treinada uma rede que serviria de modelo na geração de predições. Os exemplos de treinamento eram construídos com os itens que o usuário já tivesse avaliado. A entrada da rede eram as avaliações do item feitas por todos os outros usuários do sistema e a saída era a avaliação do usuário alvo. O problema dessa abordagem residia no fato de que o tamanho da entrada poderia ser muito grande dependendo da quantidade de usuários do sistema. Para tratar esse problema, Billsus e Pazzani utilizavam *singular value decomposition (SVD)* para reduzir a tabela de avaliações.

Lin, Alvarez and Ruiz [Lin 00] propuseram uma técnica de filtragem colaborativa baseada na mineração de regras de associação. De acordo com os autores, essa técnica consegue identificar regras que descrevem as relações de preferências entre os usuários e que são usadas na geração de predições. Cohen et al.[Cohen 98] também utilizou regras de associação entre os usuários, mas em conjunto com regras baseadas em informações sobre o conteúdo dos itens.

Em [Breese 98] é mostrada a utilização de algoritmos de aprendizado de Redes Bayesianas na FC. Nesse modelo, uma Rede Bayesiana é aprendida para cada usuário e utilizada na classificação dos itens como relevante ou não-relevantes.

Todos os algoritmos discutidos até aqui envolvem a construção de um modelo de predição. Embora muitas vezes sejam eficientes na geração de sugestões, nesses algoritmos existe o problema de construir o modelo. Essa construção pode ser custosa computacionalmente e, além disso, com o surgimento de novas avaliações, vai haver a necessidade de se gerar um novo modelo que incorpore essas novas informações.

Uma alternativa aos *algoritmos baseados em modelo* são os *algoritmos baseados em memória*. Esses algoritmos operam sobre a base de usuários, no momento da predição, para fazer a sugestão combinando as avaliações dos usuários semelhantes. Os principais algoritmos dessa categoria utilizam uma solução baseada em kNN (*k-nearest neighbors*).

Um dos primeiros sistemas a utilizar esse tipo de algoritmo em filtragem de informação foi o GroupLens [Resnick 94] de recomendação de mensagens de grupos de notícias. No momento em que uma predição precisa ser feita, o sistema utiliza uma medida de correlação para encontrar o N usuários mais similares ao usuário alvo. Uma vez que essa vizinhança é calculada, o sistema utiliza as notas dadas pelos vizinhos ao item que será sugerido para calcular uma nota que corresponde à relevância do item para o usuário alvo. É importante notar que a influência que um vizinho terá na predição da nota depende da similaridade dele com o usuário alvo. Vários outros sistemas de FC foram desenvolvidos posteriormente em cima da mesma idéia, como sistemas de recomendação de música [Shardanand 95], filmes [Hill 95] [Good 99] e programação de TV a cabo [Cotter 00]. Essa metodologia embora venha sendo utilizada desde os primeiros sistemas de filtragem colaborativa, ainda é uma das mais efetivas no que diz respeito à qualidade das recomendações [Herlocker 04].

3.5.2. Implementação de FC Com *K-Nearest Neighbors*

Como vimos na seção anterior, para que o sistema de FC com vizinhos mais próximos possa fazer uma predição, ele precisa calcular a similaridade entre os usuários. Essa similaridade é utilizada tanto para selecionar a vizinhança do usuário alvo quanto para ponderar a influência de cada usuário na predição do item.

Para a realização dessa tarefa, um grande número sistemas tem empregado a correlação de Pearson [Resnick 94], [Hill 95], [Billus 98], [Herlocker 00]. A similaridade $s_{a,u}$ entre os usuários a e u é dada pela correlação de Pearson de acordo com a seguinte fórmula:

$$s_{a,u} = \frac{\sum_{i=1}^m [(r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)]}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2 \sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}} \quad (\text{Eq. 3.1})$$

Na equação acima, $r_{a,i}$ e $r_{u,i}$ são, respectivamente, as avaliações de a e u para um item i e \bar{r}_a e \bar{r}_u são a média das avaliações de a e u . Note que todas as médias e somatórios da fórmula são calculados apenas entre os m itens avaliados por ambos os usuários.

Estudos comparativos mostram que essa medida de correlação está entre as que apresentam melhores resultados quando empregada na filtragem colaborativa com vizinhos mais próximos [Herlocker 00][Breese 98]. No entanto, um problema na utilização dessa medida diz respeito ao cardinal da interseção entre as avaliações feitas por dois usuários. Esse problema ocorre porque, quando esse cardinal é muito pequeno, o sistema pode julgar dois usuários como muito similares (ou muito diferentes) com base em pouca informação. Herlocker et al. [Herlocker 99] identificou a necessidade da aplicação de uma penalização ao cálculo da similaridade dado pela equação 3.1 quando a quantidade de itens em comum entre dois usuários fica abaixo de determinado limiar. Para a base de dados utilizada em seu trabalho, Herlocker et al. sugerem a seguinte medida de similaridade para dois usuários que possuem m itens avaliados em comum:

$$s'_{a,u} = \begin{cases} \frac{s_{a,u}}{50 - m} & , \text{ se } m < 50 \\ s_{a,u} & , \text{ se } m \geq 50 \end{cases} \quad (\text{Eq. 3.2})$$

A utilização dessa medida apresentou melhores resultados do que quando a penalização não é utilizada, e o valor de 50 utilizado na equação para o limiar foi encontrado empiricamente.

Após calcular a similaridade do usuário alvo com os outros usuários da base, o sistema precisa escolher quais serão selecionados para fazerem parte da vizinhança preditora, isto é, quais usuários terão suas avaliações levadas em consideração para prever a

relevância do item em questão. Nos primeiros sistemas de FC com vizinhos, todos os usuários que tivessem avaliado o item eram levados em consideração na geração da predição [Resnick 94]. Trabalhos posteriores sugeriram a adoção de um tamanho fixo de vizinhança, sendo selecionados aqueles usuários de maior similaridade. Em sua tese de doutorado [Herlocker 00], Herlocker confirmou que essa abordagem produzia melhores resultados e sugeriu a utilização de uma vizinhança de tamanho 40.

Uma vez calculadas as similaridades e selecionada a vizinhança, resta combinar as avaliações para gerar a predição. A maioria dos sistemas utiliza as similaridades dos vizinhos do usuário alvo para ponderar a influência que cada um terá na predição. Uma função de predição bastante utilizada é mostrada na equação 3.3 [Konstan 97], [Billsus 98], [Herlocker 00].

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n [(r_{u,i} - \bar{r}_u) w_{a,u}]}{\sum_{u=1}^n w_{a,u}} \quad (\text{Eq. 3.3})$$

Por essa fórmula, a predição $p_{a,i}$ do item i para o usuário a será a média de suas avaliações dos itens da base \bar{r}_a mais um desvio. A direção e intensidade desse desvio é calculada, na vizinhança de tamanho n , pela diferença entre a média \bar{r}_u das avaliações feitas por cada usuário u da vizinhança e sua avaliação $r_{u,i}$ para o item i , ponderada pela razão entre sua similaridade $w_{a,u}$ com o usuário alvo e a soma das similaridades do usuário alvo com todos os seus vizinhos.

Em seus estudos comparativos, Herlocker também chegou à conclusão de que a utilização da ponderação pela similaridade no cálculo da predição melhora a qualidade das recomendações [Herlocker 00].

3.6. Filtragem Híbrida

A aplicação da filtragem puramente baseada em conteúdo apresenta alguns problemas. Um dos principais está na dependência da representação do conteúdo dos itens.

Para alguns tipos de informação, o conteúdo não pode ser representado satisfatoriamente em forma de atributos. Em outros casos, embora essa representação seja possível, ela deixa de levar em consideração fatores importantes para o usuário na determinação da relevância dos itens. Por exemplo, na filtragem de filmes, características subjetivas como qualidade das atuações, trilha sonora ou desenvolvimento do enredo podem ser determinantes para o usuário. Outro problema freqüente nesse tipo de filtragem está na pouca diversidade das recomendações, uma vez que a tendência do sistema é sempre recomendar itens semelhantes aos itens que o usuário já avaliou no passado.

Por outro lado, a filtragem colaborativa também possui alguns problemas próprios. Um primeiro problema ocorre quando um novo item é adicionado à base de informação. Como os sistemas de FC baseiam-se na experiência dos usuários para gerar recomendações, o sistema será incapaz de recomendar o novo item até que tenha sido avaliado. Além disso, nesse tipo de filtragem, a qualidade das recomendações depende da quantidade de avaliações que seus usuários fornecem. Se o total de avaliações fornecidas pelos membros da comunidade for muito pequena comparada à quantidade de itens na base, um sistema puramente colaborativo poderá ter a precisão de suas recomendações comprometida.

Por causa desses problemas, inerentes à natureza das duas técnicas de filtragem discutidas, e com base na observação de que as duas abordagens possuem características complementares, começou-se a investigar a utilização de técnicas híbridas de filtragem. A idéia da incorporação de características das duas filosofias de filtragem é tirar proveito das vantagens de ambas, sem herdar as desvantagens de nenhuma. Um dos primeiros sistemas de filtragem híbrida, o FAB, surgiu ainda em 1997, na Stanford University e fazia filtragem de páginas da Internet. A partir daí surgiram muitos outros sistemas utilizando uma metodologia híbrida [Good 99] [Claypool 99] [Cotter 00]. É uma abordagem que se mostrou muito eficiente, sendo ainda muito utilizada atualmente [Bezerra 04].

3.7. Avaliando a Qualidade das Recomendações

A seguir serão discutidas as métricas que vem sendo utilizadas em análises estatísticas e na avaliação de desempenho dos sistemas de filtragem de recomendação. Além disso,

discutimos também as tarefas específicas envolvidas no processo de recomendação à que cada métrica melhor se aplica.

3.7.1. Área Sob Curva ROC

A curva “*relative operating characteristic*”, ou simplesmente curva ROC, foi introduzida na comunidade de RI por Swets [Swets 63]. A curva ROC é utilizada para medir o quanto um valor produzido por um sistema é capaz de distinguir os elementos relevantes dos não relevantes.

O sistema medido tem como saída uma variável de relevância associada a cada elemento. O conhecimento sobre a relevância dos elementos permite construir duas curvas de distribuição: uma para os valores obtidos para os elementos relevantes e outra para os elementos não relevantes. No momento de selecionar os elementos relevantes, o sistema usa um limiar t . Os elementos que ultrapassem esse limiar serão considerados relevantes e, portanto, selecionados. Caso contrário, esses elementos serão considerados irrelevantes e então rejeitados. Para cada valor escolhido para o limiar t é possível calcular a *cobertura* (proporção dos elementos relevantes que são selecionados) e o *ruído* (proporção dos elementos não relevantes que são selecionados). A Figura 3.2 ilustra as distribuições dos elementos relevantes e não relevantes e o acerto e erro para um limiar t escolhido.

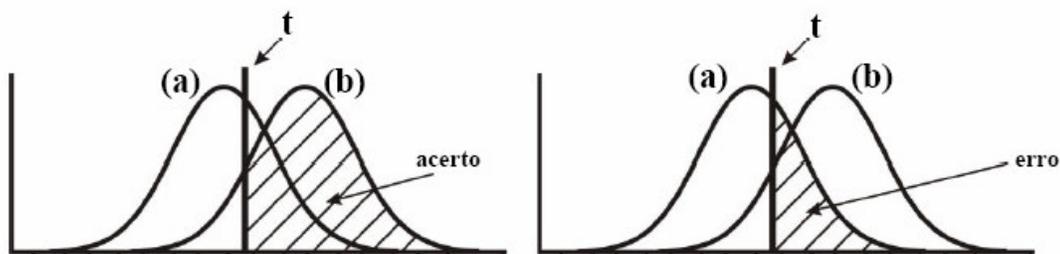


Figura 3.2 - Curvas da distribuição dos valores para elementos relevantes (b) e não relevantes (a). O limiar t determina a precisão (taxa de acerto) e o ruído (taxa de erro) do sistema.

A cobertura será a razão do número de acertos pelo número de elementos relevantes e o ruído a razão do número de erros pelo número de elementos não relevantes.

A utilização da cobertura e do ruído é bastante comum na avaliação de sistemas que tem como objetivo a seleção binária através de uma variável resultante separada por um

limiar. O problema é que os valores dessas grandezas dependem do limiar escolhido. A curva ROC por sua vez é capaz de avaliar a capacidade de o sistema separar os elementos relevantes dos não relevantes sem depender da escolha de um limiar, permitindo comparar dois sistemas mais facilmente. A curva ROC é a curva obtida quando desenhados em um plano cartesiano os valores da cobertura (ordenada) versus o ruído (abscissa) para diferentes valores do limiar t . A Figura 3.3 mostra um gráfico com uma curva ROC e alguns pontos para diferentes valores de t .

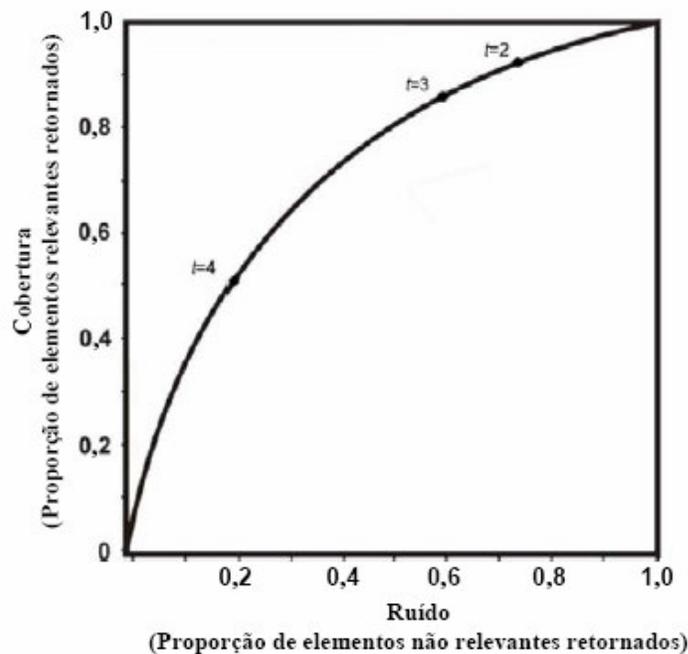


Figura 3.3 - Exemplo de uma curva ROC. Os pontos t indicados na curva representam diferentes limiares usados para separar itens relevantes de não relevantes para um sistema com sinal de saída entre 1 e 5. Por exemplo, se o limiar usado for de $t=3$ a cobertura será igual a 0,85 enquanto o ruído será igual a 0,6.

As curvas ROC são bastante úteis para observar a capacidade de um algoritmo de FI separar informação relevante da não relevante. Entretanto, comparar várias curvas ROC pode ser um processo trabalhoso e impreciso. A área sob uma curva ROC pode ser usada como um valor que expressa a capacidade do sistema discriminar os itens relevantes dos não relevantes. A Figura 3.4 mostra o exemplo de várias curvas ROC e a área A em cada situação.

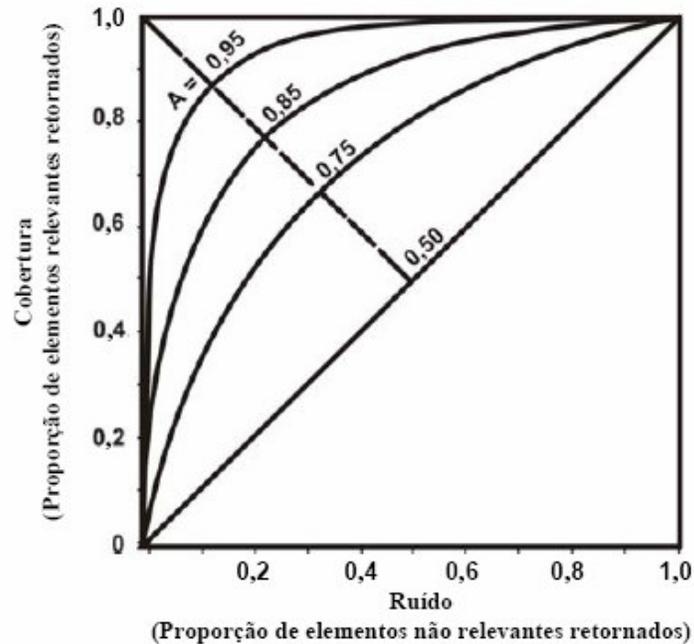


Figura 3.4 - Exemplos de curvas ROC da cobertura vs. ruído. A área A sob a curva indica a separação entre os elementos relevantes dos não relevantes.

Em [Hanley 82] é mostrado que a área sob a curva ROC representa a probabilidade de que, dados dois itens escolhidos aleatoriamente, sendo um relevante e outro não, o sistema distinguirá corretamente os dois elementos. É fácil concluir que, se a área sob a curva é de 0,5, então em 50% das vezes o sistema distinguiria corretamente os elementos e em 50% das vezes erroneamente, o que equivaleria a uma classificação puramente aleatória, ou seja, a um desempenho nada satisfatório. Valores menores que 0,5 indicam que o sistema está trocando os elementos relevantes pelos não relevantes, enquanto valores próximos de 1,0 indicam uma boa precisão na seleção dos elementos.

A área sob a curva ROC é uma métrica de seleção, ou seja, ela mede a capacidade do sistema selecionar os itens relevantes para o usuário. A área sob a curva ROC não leva em consideração a ordem gerada pela predição do sistema e também não mede a precisão do sistema na predição de cada item individualmente.

3.7.2. Breese

A métrica criada por Breese et al. [Breese 98] é usada para medir a utilidade de uma lista ordenada por um sistema de filtragem para o usuário. Em um sistema de filtragem de informação onde o resultado de uma filtragem é retornado como uma lista de itens ordenados pela relevância, é presumível que o usuário irá investigar apenas os primeiros elementos da lista na sua busca por itens que lhe interessam. A métrica Breese procura medir a utilidade de uma lista levando em consideração a tendência do usuário de observar apenas os primeiros itens da lista.

Para cada item de uma lista, sua utilidade será a diferença entre a avaliação dada pelo usuário e o valor médio d no intervalo de avaliação. Para a constante d geralmente se assume um valor no intervalo de avaliação que indica neutralidade em relação à relevância de um item. Na lista, cada item terá sucessivamente uma probabilidade menor de ser visto pelo usuário. A constante α , chamada de *half-life*, é a posição do item na ordem gerada pelo sistema que terá 50% de chance de ser observada. A utilidade de uma lista ordenada para o usuário a é então:

$$R_a = \sum_j \frac{\max(r_{a,j} - d, 0)}{2^{(j-1)/(\alpha-1)}} \quad (\text{Eq. 3.4})$$

Onde j é a posição do item na lista ordenada pelo sistema e $r_{a,j}$ é a avaliação feita pelo usuário ao item que ocupa a posição j . Breese utiliza um *half-life* α igual a 5 e acrescenta que a utilização de um valor igual a 10 apresentou pouca diferença nos resultados. Será assumido um *half-life* igual a 5 indicado na literatura como apropriado para avaliar as recomendações com a mesma base de dados utilizada nos experimentos deste trabalho [Breese 98]. O valor final da métrica para o conjunto de usuários testados é calculado como segue:

$$R = 100 \frac{\sum_a R_a}{\sum_a R_{\max}} \quad (\text{Eq. 3.5})$$

Onde R_{\max} é a utilidade máxima obtida quando todos os itens são ordenados de acordo com a avaliação do usuário, ou seja, quando a ordenação dos itens pelo sistema é igual à ordenação pelas avaliações do usuário. O valor que R_a pode assumir depende do número de itens retornados em uma lista e da distribuição das avaliações do usuário. Essa última transformação permite considerar os resultados independentemente desses fatores, pois o valor da métrica é normalizado através da divisão pelo valor máximo de utilidade de uma lista.

A métrica criada por Breese é apropriada para avaliar a capacidade do sistema de ordenar os itens de acordo com sua relevância para o usuário, também por dar mais ênfase à precisão dos primeiros itens na ordenação do sistema.

3.7.3. Outras métricas

Além das métricas apresentadas aqui existem outras métricas que poderiam ser utilizadas para medir a qualidade das recomendações. Métricas como precisão e cobertura, correlações de valores como Pearson ou de ordens como a Spearman e Kendall Tau e a *npdm* [Yao 95], são outras opções de uma vasta seleção de métricas já usadas para avaliar sistemas de filtragem de informação. Todas as métricas possuem vantagens e significados diferentes. Um estudo mais aprofundado das métricas encontra-se em [Herlocker 00].

3.8. Considerações Finais

Nesse capítulo nós analisamos os principais conceitos envolvidos na área de sistemas de recomendação, em especial nos sistemas de recomendação personalizada. Discutimos as duas principais técnicas de filtragem mais amplamente empregadas nesses sistemas: filtragem baseada em conteúdo e filtragem colaborativa; e ainda sobre como essas duas abordagens podem ser combinadas para superar limitações inerentes à utilização de uma delas isoladamente.

Sobre filtragem colaborativa (FC) aprofundamos mais a discussão, apresentando vários dos principais trabalhos na área, traçando um panorama geral dos algoritmos mais utilizados. Focamos nossa atenção na utilização de algoritmos de kNN, vimos como a

correlação de Pearson é utilizada para medir a similaridade entre os usuários e como essa similaridade é utilizada na geração da previsão de um item.

Por último, discutimos as métricas comumente utilizadas na área para medir a qualidade das recomendações produzidas. Apresentamos as métricas ROC e Breese que serão utilizadas neste trabalho, ressaltando as tarefas específicas dos sistemas de recomendação a cuja medição elas melhor se aplicam.

Do que vimos até agora, fica clara a importância do perfil do usuário no processo de geração de recomendações. No entanto, é freqüente que os sistemas de recomendação necessitem ter uma quantidade razoavelmente grande de avaliações no perfil antes de ser capaz de gerar boas recomendações para o usuário. Assim, torna-se importante a investigação de estratégias que possibilitem aprender sobre as preferências do usuário obrigando-o a fazer o menor número de avaliações possível. Uma forma de fazer isso é utilizando uma técnica de aprendizagem de máquina chamada *Aprendizagem Ativa*.

No próximo capítulo discutiremos o problema da aquisição do perfil do usuário e também uma metodologia de aprendizagem ativa capaz de tratá-lo com bastante êxito. O método chama-se *ActiveCP*, foi proposto por Teixeira et al. [Teixeira 02b] e no seu estudo reside o maior foco de interesse deste trabalho.



Capítulo 4

O Método ActiveCP

4.1. Introdução

A aprendizagem das preferências do usuário em um sistema de recomendação é parte primordial do processo de recomendação, pois quanto melhor o sistema conhecer os interesses do usuário, mais subsídio ele terá para gerar as sugestões. Atualmente existem duas abordagens principais utilizadas para aquisição do perfil do usuário em um sistema de recomendação: baseada em indicadores implícitos e baseada em indicadores explícitos [Claypool 01].

Na primeira delas, o sistema tenta obter informação sobre as preferências do usuário através do uso que ele faz do sistema. Em [Nichols 97] e [Oard 98] são discutidas várias formas de aquisição implícita de preferência. Por exemplo, eles sugerem que examinar um documento, imprimi-lo, salvar informações ou registrar no *bookmark* uma determinada página de um *website* são indicadores de que o usuário tem interesse na informação ali contida. A maior vantagem da aquisição implícita de preferências é que não há custo adicional para o usuário na utilização do sistema, ou seja, o usuário não precisa preocupar-se em fornecer suas opiniões diretamente. No entanto, nessa abordagem há o problema de identificar quais critérios implícitos são mais informativos das preferências dos usuários e também o de como medir o nível de relevância de cada critério na formação do perfil do usuário. Alguns trabalhos mostram que indicadores implícitos, embora geralmente menos precisos que indicadores explícitos, podem ser utilizados com sucesso na descoberta das preferências dos usuários [Morita 94], [Konstan 97] e [Claypool 01], mas ainda não são muito utilizados em sistemas de recomendação.

Por sua vez, na aquisição de preferências baseada em indicadores explícitos, o sistema não precisa inferir o grau de interesse do usuário por um item. Essa informação é fornecida de maneira clara e direta pelo próprio usuário, geralmente através de uma nota em uma escala quantitativa. Sistemas que utilizam indicadores explícitos geralmente disponibilizam, em sua interface, componentes que permitem a avaliação de cada item mostrado. Embora essa seja a abordagem mais utilizada pelos sistemas de recomendação, nela há a necessidade de que o usuário se engaje em uma atividade a mais no uso do sistema que é a de avaliar os itens de informação [Terveen 01]. Como frequentemente o número de avaliações necessárias para que o sistema forneça boas recomendações é

relativamente grande, pode-se gerar desinteresse por parte do usuário ao se requerer que ele forneça notas aos itens do sistema [Claypool 01].

Por esse motivo, ao se utilizar avaliações explícitas é importante que o sistema seja capaz de aprender o máximo possível sobre um usuário com cada avaliação fornecida por ele. Para atingir esse objetivo tem sido cada vez mais freqüente a utilização de estratégias de *aprendizagem ativa* em sistemas de recomendação. Aprendizagem ativa é o paradigma onde os algoritmos de aprendizagem possuem algum controle sobre as entradas a partir das quais é realizado o treinamento. A seguir, discutiremos rapidamente a aprendizagem ativa e sua utilização em sistemas de recomendação para, e em seguida, discutirmos em detalhes o método *ActiveCP* de aprendizagem ativa para sistemas de filtragem colaborativa proposto por Teixeira [Teixeira 02a]. É na investigação de alguns aspectos deixados em aberto no trabalho de Teixeira que reside o maior foco de interesse deste trabalho.

4.2. Aprendizagem Ativa

Na maioria dos algoritmos de aprendizagem indutiva é comum se assumir a disponibilidade de um grande conjunto de exemplos de treinamento previamente classificados. No entanto, em muitos problemas de aprendizagem o processo de classificação pode ser bastante caro e em alguns casos requerer trabalho manual. Nesses casos, é bastante útil reduzir a quantidade de exemplos de treinamento necessários para gerar o classificador sem, no entanto, comprometer sua precisão. O paradigma onde o algoritmo de aprendizagem tem algum controle sobre os exemplos que serão utilizados no treinamento é chamado de *aprendizagem ativa*. Os algoritmos de aprendizagem ativa se dividem em dois paradigmas principais: *membership queries* e *selective sampling*.

No paradigma de *membership queries* [Angluin 88], o algoritmo é capaz de construir exemplos artificiais e pedir sua classificação. Desta maneira o algoritmo constrói exemplos cuja classificação seja importante para que uma hipótese consistente seja construída rapidamente, ou seja, com a menor quantidade de exemplos possível. O maior problema dessa metodologia está na possibilidade de o exemplo construído não fazer sentido no domínio particular [Lindenbaum 99]. Por exemplo, em um problema de reconhecimento de caracteres, construir uma imagem arbitrária ou confusa.

A abordagem de *selective sampling*, por outro lado, assume que há um conjunto (potencialmente grande) de exemplos não classificados disponível. A tarefa do algoritmo então, consiste em selecionar os exemplos importantes na formulação de uma hipótese consistente e pedir sua classificação. Este procedimento pode ser descrito de forma iterativa em que, a cada iteração o algoritmo de *selective sampling* é chamado para selecionar um exemplo que deverá ser classificado pelo supervisor. Esse exemplo será então utilizado como exemplo de treinamento do algoritmo de aprendizagem. Esses passos são repetidos até alguma condição de parada seja atingida, em geral, quando o classificador apresenta um desempenho satisfatório.

Aprendizagem ativa tem sido utilizada para acelerar o aprendizado em diversos algoritmos de classificação, como Redes Neurais [Cohn 94] e kNN [Hasenjager 98], [Lindenbaum 99].

4.2.1. Utilizando Aprendizagem Ativa no Problema da Aquisição de Perfil do Usuário

Um sistema de recomendação personalizada, para ser capaz de fornecer sugestões de qualidade, precisa aprender as preferências de cada usuário através das avaliações de itens (implícitas ou explícitas) feitas por ele. Portanto, podemos caracterizar o problema da aquisição do perfil de um dado usuário como uma instância de um problema de aprendizagem.

Quando a avaliação dos itens é feita de maneira explícita, é delegada ao usuário a tarefa de indicar manualmente as suas opiniões sobre os itens. Nesse caso, é conveniente assumir que a classificação dos exemplos utilizados para aprendizagem é uma tarefa cara e, portanto, é desejável reduzir o número de exemplos de treinamento necessários para se atingir uma mesma qualidade de predições. Do que vimos na seção anterior, esse é um contexto potencialmente propício para a utilização de técnicas de aprendizagem ativa.

De fato, aprendizagem ativa tem sido empregada em sistemas de recomendação para acelerar o processo de aprendizagem do perfil do usuário. Sobretudo para usuários que ainda não tenham fornecido uma quantidade de avaliações suficiente para que o sistema

seja capaz de gerar boas recomendações. Esse problema é referenciado na literatura como “problema do usuário novo” (*new user problem*) [Yu 03], [Rashid 02].

Em sistemas de recomendação convencionais, a coleta de preferências é feita de maneira passiva, isto é, o sistema simplesmente espera o usuário fornecer as avaliações, como ilustra a Figura 4.1. Nesse modelo, os itens que serão avaliados e a ordem em que essas avaliações serão feitas pelo usuário são indefinidos. Ele simplesmente avalia os itens que vão surgindo ao longo da utilização do sistema.

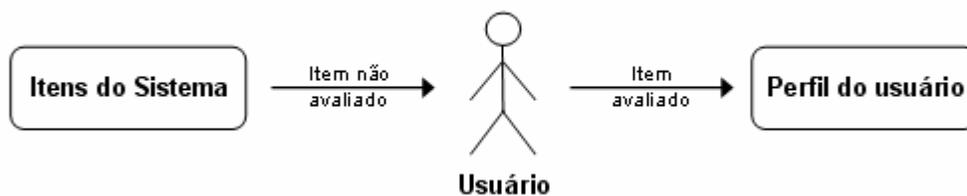


Figura 4.1 - Modelo de coleta “passiva” de avaliações em um sistema de recomendação.

Como alternativa a esse modelo, Pennock e Horvitz [Pennock 00] introduziram a ideia de utilizar aprendizagem ativa por *selective sampling* em uma etapa inicial de coleta de preferências. O objetivo dessa etapa seria o de possibilitar que o sistema estivesse apto a gerar boas recomendações para um indivíduo após consultar sua opinião a respeito de alguns itens. Esses itens precisam ser selecionados de maneira que, uma vez avaliados pelo usuário, permitam ao sistema discriminá-lo o melhor possível quanto às suas preferências. A Figura 4.2 ilustra esse processo.



Figura 4.2 - Modelo de coleta “ativa” de avaliações em um sistema de recomendação.

Em resumo, podemos dizer que a utilização de *selective sampling* em sistemas de filtragem tem os seguintes objetivos:

- i. Minimizar a quantidade de avaliações que um usuário precisa fornecer para que o sistema atinja uma dada qualidade de recomendações;
- ii. Maximizar a qualidade das recomendações para uma dada quantidade de avaliações feitas pelo usuário.

Trabalhos mais recentes têm explorado a utilização de aprendizagem ativa para tratar o problema do usuário novo e algumas metodologias têm sido propostas [Rashid 02], [Boutilier 03], [Yu 04]. Teixeira et al. [Teixeira 02b] apresentaram uma metodologia de aprendizagem ativa para acelerar a aquisição das preferências do usuário em sistemas de FC chamado *ActiveCP*. O *ActiveCP* obteve bons resultados experimentais utilizando um critério de seleção de exemplos baseado nos conceitos de controvérsia e popularidade e será detalhado na próxima seção.

4.3. ActiveCP

Em sistemas de FC com vizinhos mais próximos, quando um usuário u avalia um item i qualquer ele está, em última instância, modificando a forma com que ele está relacionado com os demais usuários do sistema que também avaliaram o mesmo item. Isso ocorre porque, como vimos anteriormente, a similaridade entre dois usuários é calculada com base nos itens que eles avaliaram em comum. Assim, quando outros usuários avaliam o item i com uma nota semelhante àquela dada pelo usuário u , eles tornam-se mais semelhantes a u . Inversamente, quando atribuem a i uma nota bastante diferente da nota dada por u , eles tornam-se menos similares a u .

Se considerarmos que para cada usuário existe uma vizinhança ótima, isto é, o conjunto de usuários com os quais o sistema de sugestão atinge a melhor qualidade de recomendações, seria necessário conhecer a avaliação de todos os usuários para todos os itens para determinar essa vizinhança. No entanto, como tipicamente cada usuário avalia apenas uma parcela dos itens do sistema, torna-se importante garantir que esses itens sejam aqueles capazes de determinar a melhor vizinhança possível.

Em sua dissertação de mestrado, Teixeira [Teixeira 02a] define o conceito de ganho de informação da avaliação de um item como sendo o quanto a avaliação de um item ajuda a discriminar um usuário dos demais membros da comunidade. Teixeira propôs um método de *selective sampling* para sistemas de filtragem colaborativa cujo critério de seleção baseia-se nos conceitos de controvérsia e popularidade de um item. Segundo esse critério, é possível selecionar os itens que, uma vez avaliados, proporcionarão o maior ganho de informação sobre um usuário.

4.3.1. Controvérsia

O critério da controvérsia baseia-se na intuição de que itens em que a maioria das pessoas concorda quanto à avaliação são, provavelmente, pouco discriminantes da vizinhança de um usuário. Isso porque, nessa situação, é estatisticamente provável que o usuário tenha uma opinião semelhante a dos outros membros da comunidade quanto ao item. Inversamente, um item para o qual os usuários tenham expressado opiniões bastante díspares é provavelmente mais informativo. De acordo com esse critério, o ganho de informação trazido pela avaliação de um item por um usuário, é maior quanto menos unânimes (ou mais controversas) forem as opiniões dos membros da comunidade a respeito da qualidade do item.

Em alguns algoritmos de aprendizagem ativa, a incerteza é utilizada na seleção de novos exemplos mais informativos [Lindebaum 99]. Um exemplo é selecionado pelo critério de seleção do algoritmo quando o grau de incerteza do classificador sobre sua classificação é alto. Esta abordagem permite que o classificador melhore seu desempenho à medida que são observados exemplos em regiões do domínio de exemplos onde a precisão do classificador é baixa. Sob esse aspecto, o critério da controvérsia apresentado por Teixeira é análogo à incerteza utilizada em aprendizagem ativa. Outros trabalhos [Rashid 02], [Boutilier 03], [Yu 04] sugerem a utilização do conceito de entropia das notas recebidas por um item em critérios de seleção para aprendizagem ativa em sistemas de FC, o que também um conceito bastante semelhante ao da controvérsia.

Para medir a controvérsia, Teixeira utiliza a *variância* da distribuição das notas dadas ao item. Assim, a controvérsia c_i de um item i é calculada segundo a seguinte equação:

$$c_i = \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_i)^2}{n} \quad (\text{Eq. 4.1})$$

Para um conjunto de n avaliações, $r_{u,i}$ é a avaliação do item i feita pelo usuário u e \bar{r}_i é a média de todas as avaliações feitas para o item i . A escolha da variância é bastante natural, já que essa é uma função frequentemente usada como medida de dispersão de uma distribuição, o que corresponde à idéia intuitiva de controvérsia apresentada por Teixeira. Possivelmente, outras medidas de dispersão também poderiam ser utilizadas para medir a controvérsia de um item.

4.3.2. Popularidade

Como vimos na seção 3.5.2, em um sistema de FC com vizinhos mais próximos, a vizinhança de um usuário é calculada com base nos itens que ele avaliou em comum com os outros usuários da comunidade. Assim, quando um usuário avalia um item popular, ou seja, um item que tenha sido avaliado por um grande número de pessoas, o sistema torna-se capaz de calcular a similaridade dele com um número grande de outros indivíduos. Consequentemente possibilita também estabelecer um relacionamento do usuário alvo com um conjunto mais significativo de usuários.

Em seu trabalho, Teixeira sugere que a popularidade é um critério relevante na seleção dos itens que serão apresentados ao usuário. Em outras palavras, ele sugere que a avaliação de um item popular por um usuário traz para o sistema um grande ganho de informação sobre a similaridade dele com outros usuários. A popularidade de um item é definida por Teixeira como sendo simplesmente a quantidade de avaliações que ele recebeu no sistema.

4.3.3. Combinando os Critérios da Controvérsia e da Popularidade

Os experimentos preliminares executados por Teixeira utilizaram a controvérsia e a popularidade isoladamente. Os resultados desses experimentos, embora fornecessem

evidências de que esses critérios seriam relevantes para selecionar um item a ser avaliado pelo usuário, resultaram em ganho na qualidade das recomendações apenas quando determinadas quantidades de itens eram selecionados [Teixeira 02a]. Para tentar obter resultados melhores e mais consistentes, Teixeira tentou agregar em um critério único o ganho de informação das duas abordagens criando o método de seleção *ActiveCP*.

No *ActiveCP* são produzidas duas listas ordenadas com os itens do sistema. Uma lista com os itens ordenados em ordem crescente pela controvérsia e outra com os itens ordenados da mesma maneira pela popularidade. Com base nessas duas ordenações, são associados dois valores para cada item que indicam sua ordem de preferência dentre os itens do sistema segundo cada um dos critérios. Assim, para um item i na posição na posição p_i em uma lista com n itens, o valor v_i associado a esse item é calculado segundo a equação 4.2.

$$v_i = \frac{p_i}{1-n} - \frac{n}{1-n} \quad (\text{Eq. 4.2})$$

Essa equação mapeia a posição p_i de um item numa das listas ordenadas em um valor $v_i \in [0,1]$ onde ao primeiro elemento ($p_i = 1$) da lista será atribuído o valor $v_i = 1$ e ao último ($p_i = n$) o valor $v_i = 0$. Os valores vc_i e vp_i , correspondendo aos valores v_i de acordo com a controvérsia e a popularidade respectivamente, são depois combinados em um único valor CP_i segundo a seguinte fórmula:

$$CP_i = w_c * vc_i + w_p * vp_i \quad (\text{Eq. 4.3})$$

Onde w_c e w_p são os pesos dados à controvérsia e à popularidade, respectivamente. Como fica claro na equação 4.3, os valores dos pesos determinam o quanto cada critério influenciará no valor CP_i final, sendo sua soma sempre igual a 1.

4.3.4. Metodologia dos Experimentos do *ActiveCP*

Para testar a eficiência de sua metodologia, Teixeira executou uma série de experimentos utilizando uma base de dados formada por avaliações de usuários reais sobre filmes [McJones 97]. A base utilizada contém 72.916 usuários e 1.628 filmes, onde os

filmes são avaliados em uma escala de 1 a 5, cujos valores correspondem às seguintes avaliações: péssimo (1), ruim (2), regular (3), bom (4) e ótimo (5).

Para diminuir o tempo da execução dos experimentos, Teixeira restringiu seus testes a um subconjunto dessa base constituído de 10.000 usuários, aleatoriamente selecionados. Desse total, foram selecionados aleatoriamente 1.000 usuários que tivessem avaliado pelo menos 100 filmes. Para cada um desses usuários foram selecionadas exatamente 100 avaliações. A metodologia de seleção foi então, empregada no estudo de cada um desses 1.000 usuários.

A metodologia proposta consistia, para cada usuário selecionado, dividir os 100 itens avaliados em 5 conjuntos de 20 filmes cada. Os conjuntos eram utilizados para fazer validação cruzada de 5 iterações. Em cada uma das iterações, um dos cinco conjuntos é designado como conjunto de teste, ou seja, ele será utilizado para medir o desempenho do algoritmo de seleção. Dos outros 80 filmes, o algoritmo de seleção separa um subconjunto que considera mais importantes na determinação das preferências do usuário. As avaliações dos filmes selecionados irão formar o perfil do usuário que é utilizado na predição das avaliações dos itens do conjunto de teste utilizando o algoritmo de FC com vizinhos mais próximos descrito na Seção 3.5.2. Vários tamanhos de perfil foram testados nos experimentos realizados por Teixeira. A Figura 4.3 ilustra o funcionamento dessa metodologia.

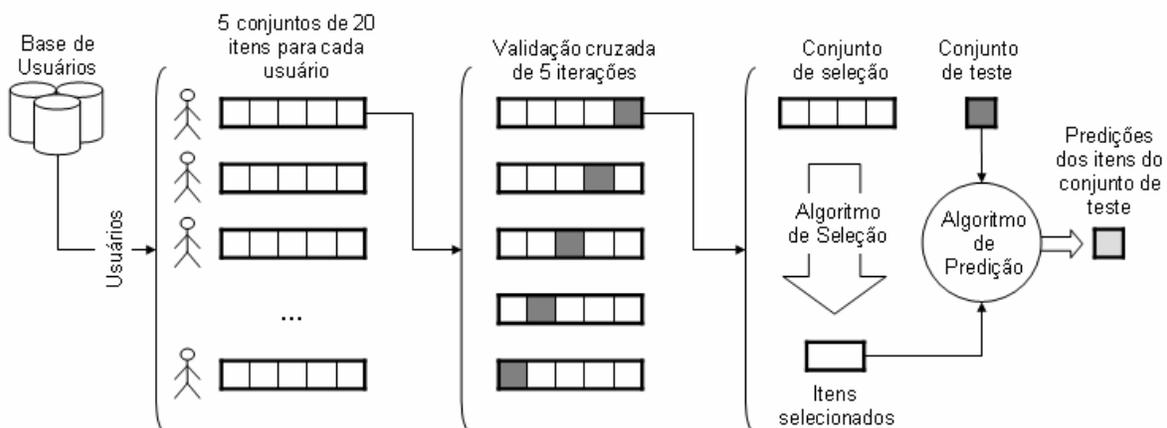


Figura 4.3 - Metodologia dos experimentos utilizada no ActiveCP.

Em seus experimentos, Teixeira comparou o critério *CP* com a seleção aleatória de itens por considerar esta última, a forma aproximada com que os usuários avaliam itens em um sistema de recomendação típico. Essa hipótese foi assumida com base na observação de que os usuários costumam fazer as avaliações de itens sem obedecer a nenhum critério ou seqüência específica.

Os testes do *ActiveCP* foram feitos para vários valores dos pesos w_c e w_p da equação 4.3, sendo que o melhor resultado nas predições foi obtido quando foi atribuído a ambos o valor de 0.5. Em outras palavra, o melhor resultado foi obtido quando a controvérsia e a popularidade tiveram a mesma influência na seleção dos itens. Para essa configuração de experimentos, Teixeira mostrou que o ganho na qualidade das predições em relação ao método de seleção aleatório é estatisticamente significativo; resultado que justifica a utilização do *ActiveCP* como método de *selective sampling* para FC com vizinhos mais próximos. Para medir a qualidade das recomendações foram utilizadas as métricas ROC e Breese apresentadas na Seção 3.7.

4.3.5. Problemas na Metodologia do *ActiveCP*

Embora tenha obtido bons resultados experimentais com sua metodologia de seleção, alguns problemas com a medida de controvérsia utilizada por Teixeira podem dificultar sua aplicação em um sistema real de filtragem colaborativa. O problema em se utilizar a variância como medida de controvérsia está relacionado com o fato de que, em um sistema de filtragem, o número de avaliações para os diferentes itens pode variar bastante. No entanto, a quantidade de notas recebidas por um item não influencia no valor de sua controvérsia quando a variância é utilizada. Por exemplo, um item que tenha sido avaliado por apenas dois usuários pode ter a mesma variância nas avaliações que um item que foi avaliado por uma grande quantidade de usuários. Nesse caso, um filme que possuísse apenas duas avaliações, uma com nota 1 e outra com nota 5, terá uma controvérsia alta, mas sua avaliação por um usuário traria pouca informação adicional para determinação da vizinhança e acarretaria em um ganho muito pequeno na precisão das recomendações [Rashid 02].

Sobre esse fato, Teixeira argumenta que a informação capturada pela variância corresponde apenas à “intensidade” da controvérsia das avaliações de um item e que seria importante o estudo de uma medida que refletisse também a “cobertura” da controvérsia, isto é, quantas avaliações participaram de seu cálculo. Ele chamou esse problema de “problema da cobertura versus intensidade da controvérsia”, mas não aprofundou a investigação de outras medidas controvérsia que pudessem solucioná-lo.

Para que esse problema não tivesse impacto em seus experimentos, a abordagem utilizada por Teixeira foi a de fixar a quantidade de avaliações que são utilizadas para calcular a controvérsia de um item [Teixeira 02b]. Dessa forma garantia-se que a cobertura no cálculo da controvérsia seria igual para todos os itens. Por esse motivo, nos experimentos do *ActiveCP* a controvérsia de um filme i qualquer era calculada sempre com base em exatamente 100 avaliações recebidas por i , independentemente de quantas outras avaliações o filme possuísse. Essas 100 avaliações eram escolhidas dentre as outras de maneira aleatória. Também a escolha dessa quantidade de avaliações também não é embasada, no trabalho de Teixeira, por uma análise empírica. Uma análise mais aprofundada dessa abordagem permite perceber que ela impõe algumas restrições ao método, fazendo com que informações potencialmente importantes sejam negligenciadas.

Uma restrição diz respeito ao fato de que o método despreza filmes que seriam propostos para compor o perfil do usuário por não terem uma quantidade mínima de avaliações. No caso específico das configurações utilizadas no *ActiveCP*, um filme precisava ter sido avaliado por pelo menos 100 usuários para poder ser considerado no momento da seleção. Assim, um filme extremamente controverso com 90 avaliações seria desconsiderado, enquanto que filmes potencialmente menos controversos que o primeiro seriam incluídos nos experimentos por possuírem mais de 100 avaliações. A Figura 4.4 ilustra essa situação em que os filmes f_{99} e f_{100} seriam desprezados pela metodologia por possuírem uma quantidade insuficiente de avaliações.

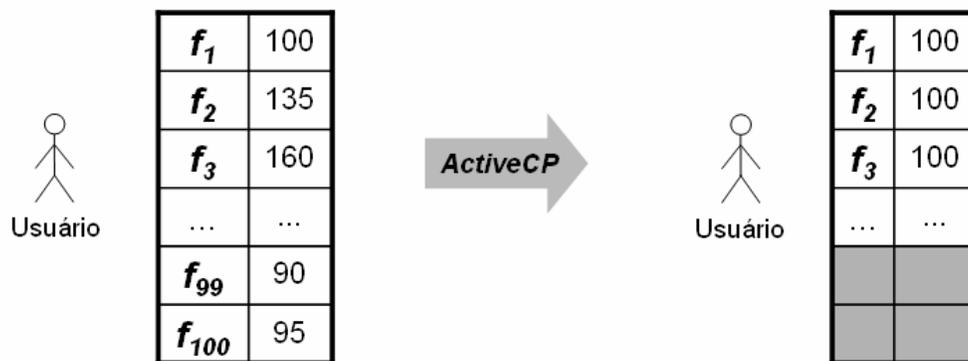


Figura 4.4 - Restrições da metodologia do *ActiveCP*.

Além disso, mesmo para os filmes incluídos nos experimentos, a metodologia deixa de levar em consideração a informação adicional contida nas avaliações que são descartadas do cálculo da controvérsia de um item. Como a solução adotada para resolver o problema “cobertura versus intensidade” foi de fixar o número de avaliações utilizadas no cálculo da controvérsia em 100, caso um filme possua, por exemplo, 130 avaliações a metodologia estará deixando de levar em consideração 30 delas. A Figura 4.4 ilustra essa situação para os filmes f_2 e f_3 .

Como consequência, pode ocorrer ainda uma redução da abrangência da metodologia experimental, uma vez que pode haver muitos usuários que não possuam a quantidade mínima necessária de filmes obedecendo a essa restrição. Um usuário como o da Figura 4.4, que tivesse avaliado 100 filmes onde apenas 98 deles tivessem mais de 100 avaliações não poderia ser incluído nos experimentos porque, como mencionamos na seção anterior, a metodologia utiliza exatamente 100 filmes de cada usuário para fazer a seleção e testar com validação cruzada.

As limitações metodológicas do *ActiveCP* podem tornar a sua aplicação menos eficaz em bases pequenas, ainda em formação; onde pode haver poucos filmes com a quantidade de avaliações suficiente para serem utilizados na aprendizagem ativa e onde o desprezo de avaliações no cálculo da controvérsia pode ser muito mais significativo.

4.4. Considerações Finais

Nesse capítulo vimos como a aprendizagem ativa pode ser utilizada para acelerar o processo de aquisição das preferências do usuário em um sistema de filtragem colaborativa. Em especial, discutimos o método *ActiveCP* proposto por Teixeira [Teixeira 02a]. Vimos como esse método utiliza os conceitos de controvérsia e popularidade de um item, combinando-os, para montar um critério de seleção capaz de:

- Minimizar a quantidade de avaliações que um usuário precisa fornecer para que o sistema atinja uma dada qualidade de recomendações;
- Maximizar a qualidade das recomendações para uma dada quantidade de avaliações feitas pelo usuário.

Em seguida analisamos a metodologia dos experimentos realizados para avaliação do *ActiveCP* e discutimos o problema da utilização da variância como medida de controvérsia, o qual foi chamado de problema da “cobertura versus intensidade”. A investigação de medidas de controvérsia mais adequadas à utilização em um método de *selective sampling* em sistemas de FC foi deixada em aberto por Teixeira, que resolveu o problema da cobertura estabelecendo uma restrição metodológica que potencialmente despreza informação relevante.

Neste trabalho, sugerimos uma medida de controvérsia capaz de refletir a quantidade de avaliações recebidas por um item. Além disso, propomos uma metodologia de experimentos mais flexível e mais próxima do que seria feito em um sistema real e investigamos sua aplicação em uma base que simula a de um sistema em início de operação. Todos esses pontos serão tratados em detalhes no próximo capítulo.



Capítulo 5

Propostas

5.1. Introdução

No capítulo anterior, vimos como os conceitos de controvérsia e de popularidade de um item podem ser utilizados como critérios de seleção em aprendizagem ativa para sistemas de filtragem colaborativa. Vimos ainda como o método *ActiveCP* os combina para gerar um critério de seleção unificado e capaz de fornecer melhorias significativas na qualidade das predições.

Apesar dos resultados positivos obtidos, mostramos como a metodologia experimental empregada no *ActiveCP* foi limitada pelas restrições inerentes à utilização da variância como medida de controvérsia. Neste trabalho investigamos a utilização de uma medida de controvérsia que não apresenta as limitações da variância. Com isso buscamos elaborar uma metodologia de experimentos mais abrangente, capaz de tirar o máximo proveito de toda a informação disponível na base de dados do sistema. Além disso, buscamos validar a aplicabilidade da nova metodologia em uma base de dados restrita, com poucos usuários e com filmes com quantidades pequenas de avaliações, simulando assim uma base em início de formação.

5.2. Nova medida de controvérsia

No método *ActiveCP* original, Teixeira usa a variância como medida de controvérsia, dada pela equação 5.1:

$$c_i = \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_i)^2}{n} \quad (\text{Eq. 5.1})$$

Onde $r_{u,i}$ é a avaliação do item i feita pelo usuário u e \bar{r}_i é a média de todas as avaliações feitas para o item i e n o total de avaliações. Como discutido anteriormente, o problema em utilizar a variância como medida de controvérsia de um item está no fato de que a quantidade de avaliações recebidas pelo item não influenciam no resultado da medida. Na Seção 4.3.5 vimos que, para resolver o problema da “cobertura versus intensidade”, a quantidade de avaliações utilizadas no cálculo da variância era fixada em um valor que chamaremos de I (nos experimentos de Teixeira I foi fixado com o valor

100). Isso implicava no fato de que, para que um item fosse seja passível de ser utilizado nos experimentos, era necessário que ele tivesse recebido pelo menos I avaliações. Essa deficiência da variância era responsável pelas restrições metodológicas e de aplicação prática do *ActiveCP* que discutimos no final do capítulo anterior. Por esse motivo, a primeira etapa do nosso trabalho foi encontrar uma medida de controvérsia mais aplicável à aprendizagem ativa em sistemas de FC.

Analisando a fórmula da variância dada pela equação 5.2, percebemos que o somatório das dispersões é normalizado, ao final, pelo total de avaliações. No entanto, essa normalização elimina a influência que o número de avaliações do item teria na medida. Assim, para que a medida fosse capaz de refletir em seu resultado a quantidade de avaliações recebidas por um item, bastaria que alterássemos a fórmula original da variância eliminando a normalização. Como resultado, temos uma nova medida de controvérsia chamada de desvio (deviation) [Breiman 98], que é dada pela equação 5.2 abaixo.

$$c_i = \sum_{u=1}^n (r_{u,i} - \bar{r}_i)^2 \quad (\text{Eq. 5.2})$$

Onde $r_{u,i}$ é a nota do item i dada pelo usuário u e \bar{r}_i é a média de todas as n avaliações feitas para o item i .

Pela análise da equação 5.2 é fácil perceber que o valor do desvio cresce também com a quantidade de avaliações envolvidas em seu cálculo e não apenas com a dispersão dessas avaliações. Por esse motivo, ao se utilizar o desvio, não haveria mais a necessidade de fixar uma quantidade mínima I de avaliações para serem utilizadas no cálculo da controvérsia. Podemos então supor que o desvio é uma escolha mais adequada que a variância para medir a controvérsia de um item, uma vez que ele já resolve o problema da “cobertura versus intensidade” sem a necessidade de realizar experimentos para estimar um valor ótimo para o parâmetro I .

5.3. Nova Metodologia de Experimentos

Após as reflexões feitas nas seções anteriores sobre as limitações metodológicas do *ActiveCP* e depois de termos proposto a utilização do desvio como uma medida de controvérsia que não possui os problemas da variância, é necessário sugerir uma nova

metodologia de experimentos. Nosso objetivo é que essa nova metodologia tire proveito das vantagens oferecidas pela utilização do desvio, eliminando as restrições presentes nos experimentos realizados por Teixeira, e que tenha uma boa aplicabilidade prática.

Assim, a nova metodologia que propomos neste trabalho possui as seguintes características que a diferem da que foi discutida na Seção 4.3.4:

- i) Utiliza todos os usuários da base nos testes, sem restrições quanto ao número de avaliações que ele possua. É importante notar que na metodologia original foi estabelecida uma quantidade mínima de itens que um usuário precisava ter avaliado para poder ser incluído nos experimentos. Nos experimentos do *ActiveCP* esse valor foi de 100 avaliações, o que fez com que fossem utilizados apenas 1.000 usuários nos experimentos de um total de 10.000 disponíveis;
- ii) Permite que todos os itens que um usuário tenha avaliado possam ser selecionados nos testes, não importando quantas avaliações cada um deles tenha recebido. Na descrição dos experimentos realizados por Teixeira vimos que, dentre todos os itens avaliados por um usuário, apenas 100 deles eram incluídos nos experimentos;
- iii) Pelas características já discutidas da nova medida de controvérsia do desvio, na metodologia que estamos propondo não há mais a necessidade de estimar um valor para I .

Para tornar mais claras as características da nova metodologia, A Tabela 5.1 ilustra as diferenças entre as duas abordagens, supondo que as duas fossem aplicadas no contexto da mesma base de 10.000 usuários utilizada no *ActiveCP*.

Tabela 5.1 - Diferenças entre as metodologia do *ActiveCP* e a nova metodologia proposta.

Metodologia	Nº de usuários utilizados	Nº de itens utilizados por usuário	Nº de avaliações utilizadas por item
<i>ActiveCP</i>	1.000	100	100
Nova	10.000	Todos o que ele tivesse avaliado.	Todas as avaliações disponíveis.

5.4. Análise Experimental

Os experimentos realizados neste trabalho foram conduzidos para atingir três objetivos específicos:

- Analisar através de experimentos o problema da cobertura versus intensidade. Apesar de Teixeira ter previsto esse problema em seu trabalho, nenhuma análise experimental foi de fato realizada para verificar o seu efeito sobre o método *ActiveCP*;
- Avaliar a aplicabilidade da nova medida de controvérsia e comparar com os critérios de seleção já estudados (critérios aleatórios, controvérsia e popularidade usados isoladamente e combinados através do método *ActiveCP*);
- Avaliar os experimentos em uma base de dados pequena, em processo de formação onde o problema cobertura X intensidade teria teoricamente maior efeito.

Nas próximas subseções, discutimos os pontos mais importantes da análise experimental realizada neste trabalho.

5.4.1. Base de dados

Na Seção 4.3.5, levantamos a hipótese de que a exclusão de filmes e o desprezo de avaliações poderiam significar uma perda de informação significativa para o método de aprendizagem ativa do *ActiveCP*. Como consequência natural desse fato, em uma base de dados pequena, com poucos usuários e poucos filmes avaliados, o impacto da perda provavelmente seria mais severo. Para investigar essa suposição decidimos, para o escopo deste trabalho, adotar uma base de dados menor e mais restrita para executar os experimentos.

Nessa nova base foram executados os experimentos com a nova metodologia utilizando o desvio e também novos experimentos com a metodologia do *ActiveCP*. Isso porque a comparação das duas abordagens em um mesmo ambiente experimental fornece subsídios mais confiáveis para as análises de ganho de informação obtido com o uso da metodologia e aplicabilidade prática do método.

Nos experimentos realizados neste trabalho, decidimos partir da mesma base de usuários e avaliações de filmes [McJones 97] que foi usada por Teixeira, também por ser esta uma base utilizada em vários outros trabalhos de FC. Na Seção 4.3.4, vimos que Teixeira restringiu a base original de McJones a uma com 10.000 usuários, dos quais apenas 1.000 foram utilizados em seus experimentos.

Como um dos nossos objetivos era o de investigar o desempenho dos métodos de aprendizagem ativa (o do *ActiveCP* e o do novo método) em bases pequenas, decidimos restringir ainda mais esse universo. Selecionamos aleatoriamente 300 usuários do universo dos 10.000 iniciais para compor uma nova base que simulasse a de um sistema de FC em seu início. Assim, o processo de montagem gerou uma nova base com as seguintes características:

- i) 300 usuários cada um com pelo menos 30 filmes avaliados. A distribuição da quantidade de itens avaliados por usuário é mostrada pela Figura 5.1. É importante perceber que 40% dos usuários da base avaliaram menos de 50 filmes e mais de 80% possuem menos de 100 avaliações.

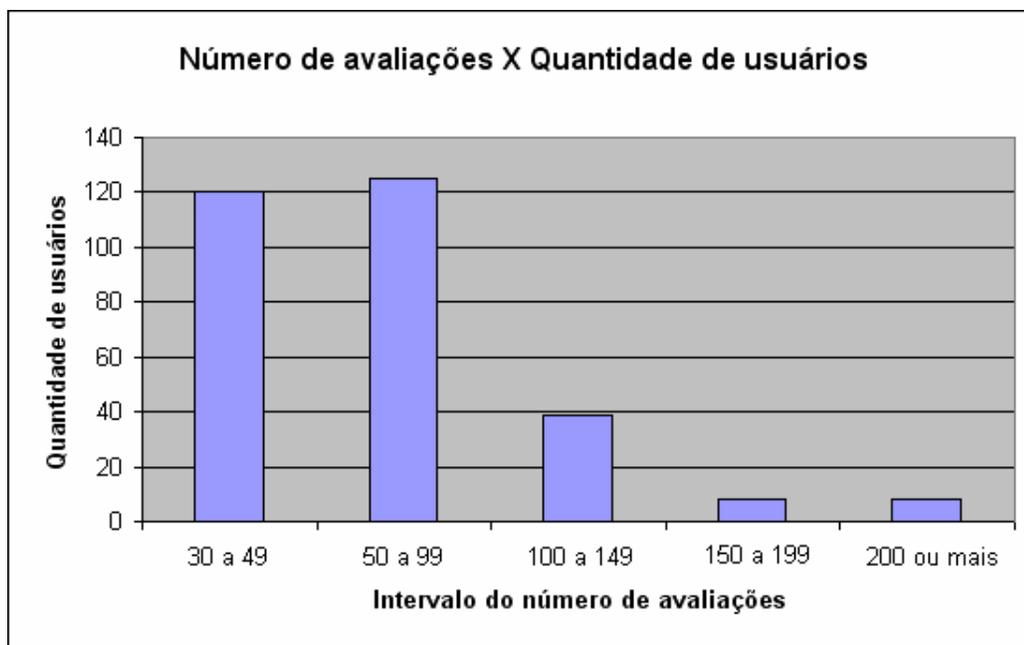


Figura 5.1- Distribuição da quantidade de itens avaliados por usuários.

- ii) 394 filmes com um total de 21518 avaliações, o que corresponde a uma média de 54,6 avaliações por filme. A distribuição da quantidade de avaliações recebidas por filmes é mostrada na Figura 5.2. Note que mais de 82% dos filmes da base receberam menos de 100 avaliações.

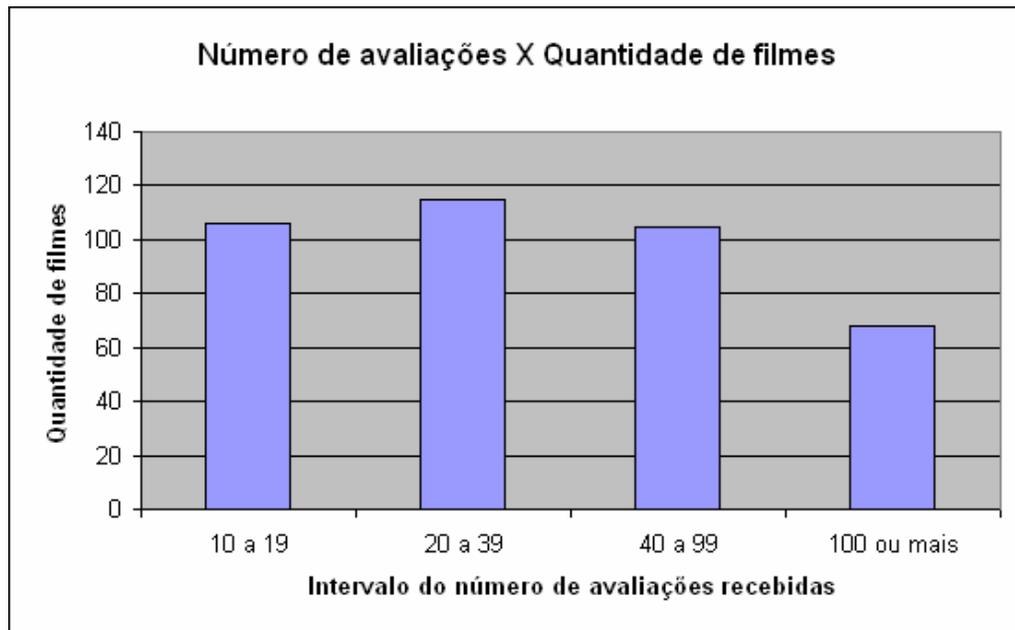


Figura 5.2 - Distribuição da quantidade de avaliações recebidas por filmes.

5.4.2. Métodos avaliados

Nesta seção são apresentados os métodos que foram comparados nos experimentos realizados.

a) Seleção Aleatória

Consiste em selecionar os itens sem nenhuma ordem particular. Será utilizada como referência na comparação dos métodos de seleção propostos, por ser a forma aproximada como os sistemas de recomendação tradicionais apresentam os itens para serem avaliados pelo usuário. A seleção aleatória aproxima o modelo de avaliação em um sistema de filtragem passivo mostrado na Figura 4.1.

b) Seleção baseada apenas na controvérsia

A seleção dos itens é feita utilizando-se apenas uma medida de controvérsia. Os experimentos com esse método de seleção foram realizados para verificar a eficiência da medida do desvio quando comparada à seleção aleatória e também à seleção utilizando a variância. É importante salientar que, ao utilizar a variância, o valor do parâmetro I (descrito na Seção 5.2) foi fixado em 20, 40 e 60 avaliações e não em 100 como feito no ActiveCP, devido ao tamanho mais restrito da base. Caso um filme não possuísse pelo menos I avaliações, ele não poderia ser selecionado pelo método. Adicionalmente também executamos experimentos com a variância sem estabelecer nenhuma restrição, permitindo a utilização de todas as avaliações de um item no cálculo da medida, ao que chamamos de “variância total”. O objetivo desse teste foi verificar a severidade do dilema “cobertura versus intensidade”, em relação ao parâmetro I .

c) Seleção baseada apenas na popularidade

São sugeridos os filmes que tenham recebido a maior quantidade de avaliações pelos usuários da base.

d) Seleção combinada (ActiveCP)

A seleção é feita combinando o critério da popularidade com as medidas de controvérsia listadas anteriormente (desvio, variância com I fixo em 20, 40, 60 e variância total). A combinação dos critérios foi feita da mesma maneira que no *ActiveCP*, como descrito na Seção 4.3.3 e com pesos iguais a 0.5 para ambos. Com a diferença que, quando a variância com I fixo é utilizada, caso um filme i não possua o mínimo de avaliações necessárias (20, 40 ou 60), apenas a popularidade é considerada. Em outras palavras, o termo vc_i da equação 4.3 para tal filme será zero. Na análise dos resultados, os métodos de seleção combinada serão identificados pelo prefixo CP (controvérsia + popularidade).

5.4.3. Métricas de Avaliação dos Resultados

As métricas utilizadas para medir a qualidade das recomendações produzidas foram ROC e Breese apresentadas no Capítulo 3. Para a Breese foram utilizadas as mesmas configurações do trabalho de Teixeira, ou seja, um *half-life* de 5 e o parâmetro d com valor igual a 3, escolhido por ser o valor médio no intervalo de avaliação sendo assim aquele que mais aproxima a neutralidade de preferência do usuário [Teixeira 02a]. Na métrica ROC foram consideradas as avaliações com valores 1, 2 e 3 como insatisfatórias e avaliações com valores de 4 e 5 como satisfatórias, também de acordo com o que considerou Teixeira em seu trabalho.

5.4.4. Descrição dos experimentos

Para testar a eficiência do método de seleção de itens, partimos de uma simplificação do que seria a aplicação real de aprendizagem ativa em um sistema de recomendação. Nos nossos experimentos, apenas itens cujas avaliações são conhecidas pelo usuário alvo são selecionados pelo algoritmo. Dessa forma, não analisamos o problema de o usuário não ser capaz emitir opinião sobre um item apresentado a ele na etapa de aprendizagem de suas preferências. Focamos neste trabalho, a exemplo do que foi feito por Teixeira [Teixeira 02a], apenas no quanto o sistema aprende ao escolher um item para o usuário avaliar, supondo que ele sempre consegue avaliá-lo.

Assim, os experimentos foram organizados de maneira semelhante aos descritos na Seção 4.3.4. A diferença é que, ao invés de utilizar uma quantidade fixa de itens (100 itens sorteados aleatoriamente naquele caso) e validação cruzada de 5 iterações, utilizamos todos os itens avaliados por cada usuário e fizemos a validação cruzada em 3 iterações. A diminuição do número de iterações da validação cruzada em relação aos experimentos do *ActiveCP* foi devido ao menor número de filmes que cada usuário da base possui.

5.4.5. Resultados Obtidos

Nesta seção apresentamos os resultados dos experimentos realizados utilizando vários métodos de seleção de itens. A Tabela 5.2 e Tabela 5.3 mostram o desempenho médio do sistema de recomendação segundo as métricas ROC e Breese para os métodos de

seleção isolados. Cada linha das tabelas contém os resultados para uma determinada quantidade de avaliações selecionada pelo método. Em outras palavras, cada linha é o resultado para um determinado tamanho de perfil montado pelo método de seleção. Nos experimentos realizados, variamos os tamanhos de perfil entre 2 e 10 por representar a etapa mais crítica de aquisição das preferências dos usuários.

Tabela 5.2 - Valores médios da medida ROC para as metodologias de seleção isoladas (apenas controvérsia e apenas popularidade).

ROC							
Número de avaliações selecionadas	Aleatório	Popularidade	Desvio	Variância Total	Variância I=20	Variância I=40	Variância I=60
2	0,62949171	0,66623032	0,66657121	0,59411625	0,61364081	0,62594161	0,63367222
4	0,69872589	0,71510414	0,71936312	0,68250156	0,69722815	0,70370425	0,71154438
6	0,72151952	0,72848397	0,73046094	0,71103909	0,71848190	0,72558968	0,73027651
8	0,72549591	0,73364871	0,73280623	0,72312890	0,72648767	0,73106415	0,73322317
10	0,72935453	0,73579770	0,73781886	0,73036044	0,73050183	0,73345042	0,73505378

Tabela 5.3 - Valores médios da medida Breese para as metodologias de seleção isoladas (apenas controvérsia e apenas popularidade).

Breese							
Número de avaliações selecionadas	Aleatório	Popularidade	Desvio	Variância Total	Variância I=20	Variância I=40	Variância I=60
2	62,16063303	65,81056526	65,43706824	58,64809500	60,70037177	61,98343890	62,67873348
4	69,09687785	70,47933727	70,70706827	66,79517785	68,28060990	69,25268462	69,81629551
6	70,79540144	71,69690310	71,95527074	69,94165672	70,32530309	71,27220408	71,60413908
8	71,31507035	72,07764278	72,20950649	71,04280688	71,07029162	71,90625429	71,76587023
10	71,74296304	72,33668479	72,55777116	71,58577367	71,52032372	72,30954076	71,53451812

Nas tabelas, as células destacadas em cinza escuro representam os melhores resultados para cada quantidade de avaliação selecionada, enquanto que as em cinza claro representam o pior resultado desconsiderando o da seleção aleatória.

Variância Total e com $I = 20, 40, 60$

Com base nos resultados da variância, é possível afirmar com mais segurança que efetivamente há um problema devido à medida não refletir a “cobertura” em seu resultado. Isso porque o pior resultado foi obtido com a variância total e os resultados tendem a melhorar com o aumento do tamanho de I .

Apesar de ter apontado a existência do problema em seu trabalho, Teixeira não investigou o real impacto que ele traria ao método. Tanto pela métrica ROC quanto pela Breese é fácil perceber que a utilização da variância puramente não fornece um bom critério de seleção de itens, uma vez que a “variância total” exibiu o pior de todos os resultados, tendo sido em muitos pontos pior até mesmo que a seleção aleatória.

Desvio, Popularidade e Variância Isolados

A medida do desvio se mostrou bastante competitiva em relação aos outros critérios usados sem combinação. Em quase todos os tamanhos de perfil, o critério de desvio obteve os melhores resultados, com exceção do tamanho de perfil 2 pela métrica Breese e o tamanho de perfil 8 pela métrica ROC onde a popularidade foi superior. Em comparação com a variância, o desvio foi superior independente do valor de I , para todos os tamanhos de perfil e para as duas medidas de avaliação. É importante ressaltar ainda que para nenhum tamanho de perfil a seleção usando o desvio apresentou o pior resultado comparado com as demais. A Figura 5.3 mostra o gráfico do desempenho das metodologias de seleção isoladas correspondentes aos dados da Tabela 5.2 e Tabela 5.3.

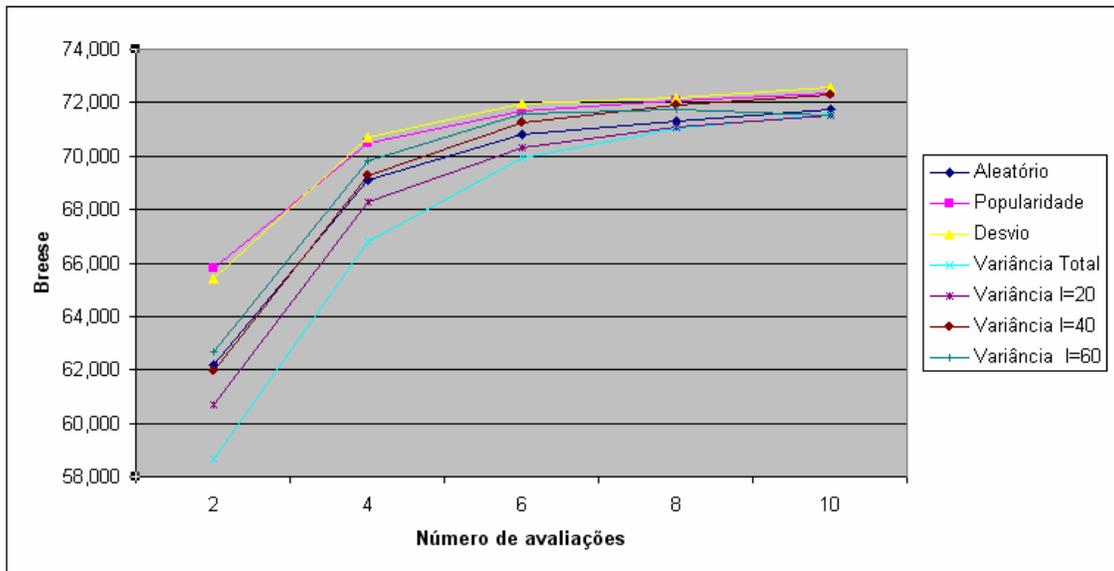
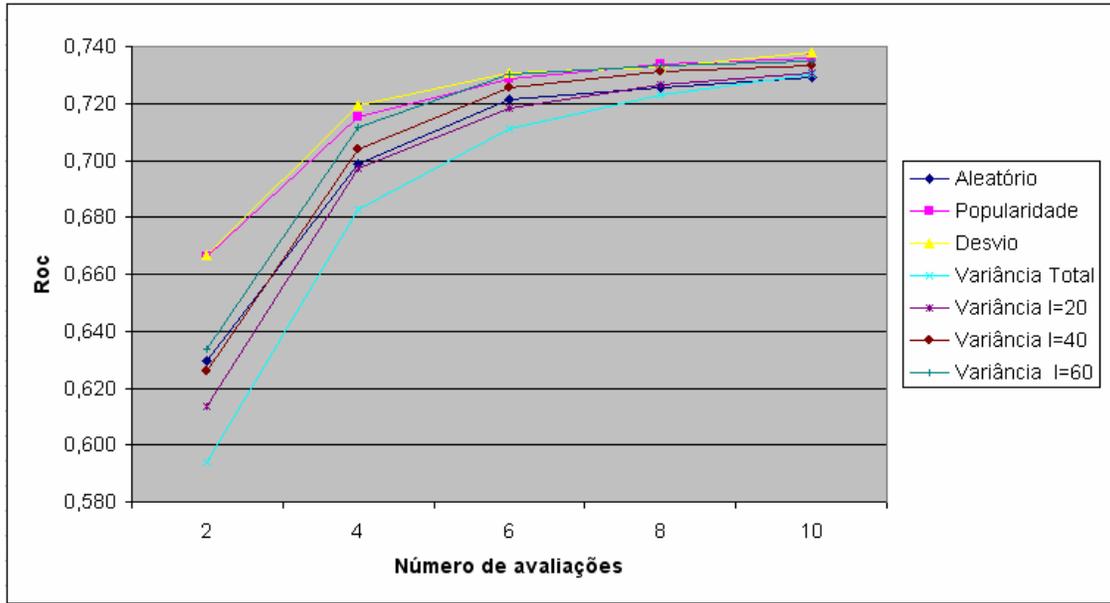


Figura 5.3 - Gráficos da performance das predições usando perfil do usuário formado pelos itens selecionados com os critérios isolados.

Tabela 5.4 - Valores médios da medida ROC para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas).

ROC						
	Aleatório	CP Desvio	CP Variância Total	CP Variância I=20	CP Variância I=40	CP Variância I=60
2	0,62949171	0,67001127	0,66459015	0,65790615	0,66612753	0,66255154
4	0,69872589	0,71895702	0,71831262	0,71514149	0,72038664	0,71964563
6	0,72151952	0,73080622	0,72831031	0,72708291	0,72995942	0,72861612
8	0,72549591	0,73446640	0,73470451	0,73344664	0,73256763	0,73488839
10	0,72935453	0,73616342	0,73844662	0,73702955	0,73620147	0,73737185

Tabela 5.5 - Valores médios da medida Breese para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas).

Breese						
	Aleatório	CP Desvio	CP Variância Total	CP Variância I=20	CP Variância I=40	CP Variância I=60
2	62,16063303	66,05922885	65,13285649	64,83634154	65,54427807	65,49914802
4	69,09687785	70,85743469	70,61704589	70,63538295	71,20912567	70,69997855
6	70,79540144	71,82031380	71,73810071	71,72597720	72,02107436	71,56031624
8	71,31507035	72,22177851	72,35374435	72,21725686	72,36189034	72,12941700
10	71,74296304	72,46935094	72,54404994	72,59585809	72,40239312	72,48713031

CP Desvio X CP Variância

Os resultados exibidos na Tabela 5.4 e Tabela 5.5 mostram que os métodos que combinam controvérsia e popularidade apresentam desempenho bastante parecidos entre si. Embora tenham se tornado mais próximos, os métodos que utilizam a variância apresentaram melhora no desempenho, aproximando-se dos resultados do método com o desvio, como é possível perceber na Figura 5.4.

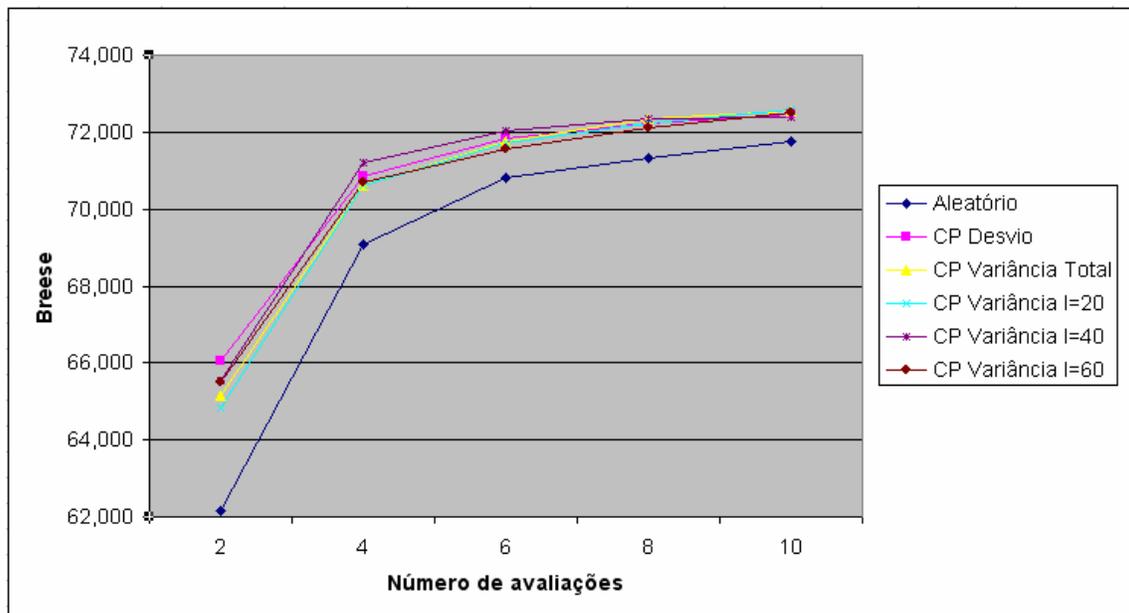
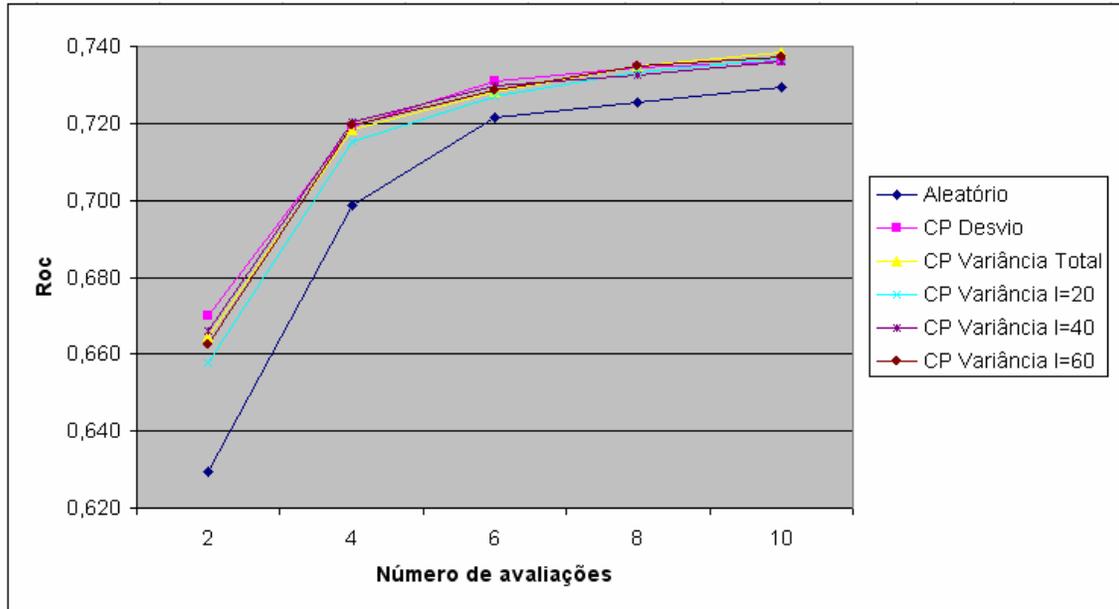


Figura 5.4 - Gráficos da performance das predições usando perfil do usuário formado pelos itens selecionados com os critérios combinados.

Nenhum dos métodos combinados se mostra significativamente superior aos demais. O método variância total apresentou o maior número de piores resultados. O método CP Desvio seria competitivo, com a vantagem de que não há necessidade de

estimar o parâmetro I com o objetivo de melhorar o desempenho do método. O método da variância combinado é mais instável em relação ao parâmetro I (nunca se sabe quem será o melhor). Nos experimentos isolados, se observa uma relação entre o desempenho da controvérsia e o tamanho do parâmetro I (variância total é significativamente o pior método). Já nos experimentos onde controvérsia é combinada com popularidade, o desempenho dos métodos CP Variância são menos dependentes de I. O método variância total não foi necessariamente o pior, o que indicaria que o dilema cobertura-intensidade no método combinado seria menos drástico.

Uma outra observação bastante interessante é que, quanto maior a quantidade de itens selecionados para compor o perfil do usuário, menos importa a estratégia de seleção utilizada. Nos gráficos da Figura 5.4 isso fica bastante claro na forma como o desempenho das estratégias de seleção combinadas se aproxima assintoticamente do desempenho da seleção aleatória. Isso reforça a idéia da importância que a utilização de uma estratégia de aprendizagem ativa tem para a aquisição de informação sobre as preferências de usuários novos.

5.4.6. O Desvio Como Estratégia de Seleção “Combinada”

O fato da medida do desvio refletir em seu resultado a quantidade de avaliações que um item recebeu implica que popularidade do item já é considerada em seu cálculo. De fato, se analisarmos a fórmula do desvio dada pela equação 5.1, veremos que ela se trata da variância (dispersão das notas) multiplicada pela quantidade de avaliações do item, isto é, sua popularidade:

$$c_i = n * \left(\frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_i)^2}{n} \right) \quad (\text{Eq. 5.2})$$

De fato, da comparação dos resultados fornecidos pelo método do desvio isolado com o do desvio combinado com a popularidade, observamos apenas uma pequena diferença de desempenho entre eles. Isso significa que a adição do critério da popularidade ao desvio isolado pouco acrescentou em termos qualidade das recomendações produzidas como mostram os resultados na Tabela 5.6 e Tabela 5.7 que incluem as metodologias combinadas incluindo o desvio.

Tabela 5.6 - Valores médios da medida ROC para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas) e o desvio.

ROC							
	Aleatório	CP Desvio	Desvio	CP Variância Total	CP Variância I=20	CP Variância I=40	CP Variância I=60
2	0,62949171	0,67001127	0,66657121	0,66459015	0,65790615	0,66612753	0,66255154
4	0,69872589	0,71895702	0,71936312	0,71831262	0,71514149	0,72038664	0,71964563
6	0,72151952	0,73080622	0,73046094	0,72831031	0,72708291	0,72995942	0,72861612
8	0,72549591	0,73446640	0,73280623	0,73470451	0,73344664	0,73256763	0,73488839
10	0,72935453	0,73616342	0,73781886	0,73844662	0,73702955	0,73620147	0,73737185

Tabela 5.7 - Valores médios da medida Breese para as metodologias de seleção combinadas (controvérsia + popularidade, com pesos iguais para ambas) e o desvio.

Breese							
	Aleatório	CP Desvio	Desvio	CP Variância Total	CP Variância I=20	CP Variância I=40	CP Variância I=60
2	62,16063303	66,05922885	65,43706824	65,13285649	64,83634154	65,54427807	65,49914802
4	69,09687785	70,85743469	70,70706827	70,61704589	70,63538295	71,20912567	70,69997855
6	70,79540144	71,82031380	71,95527074	71,73810071	71,72597720	72,02107436	71,56031624
8	71,31507035	72,22177851	72,20950649	72,35374435	72,21725686	72,36189034	72,12941700
10	71,74296304	72,46935094	72,55777116	72,54404994	72,59585809	72,40239312	72,48713031

Os dados nas tabelas mostram que o desvio possui resultados bastante competitivos em relação às metodologias combinadas. Tanto os resultados pela métrica ROC quanto pela Breese mostram que o desvio sozinho não foi o melhor dos métodos em nenhum dos tamanhos de perfil, mas também não foi o pior em nenhum ponto. A Figura 5.5 mostra a comparação entre o desvio, o CP com desvio e a seleção aleatória.

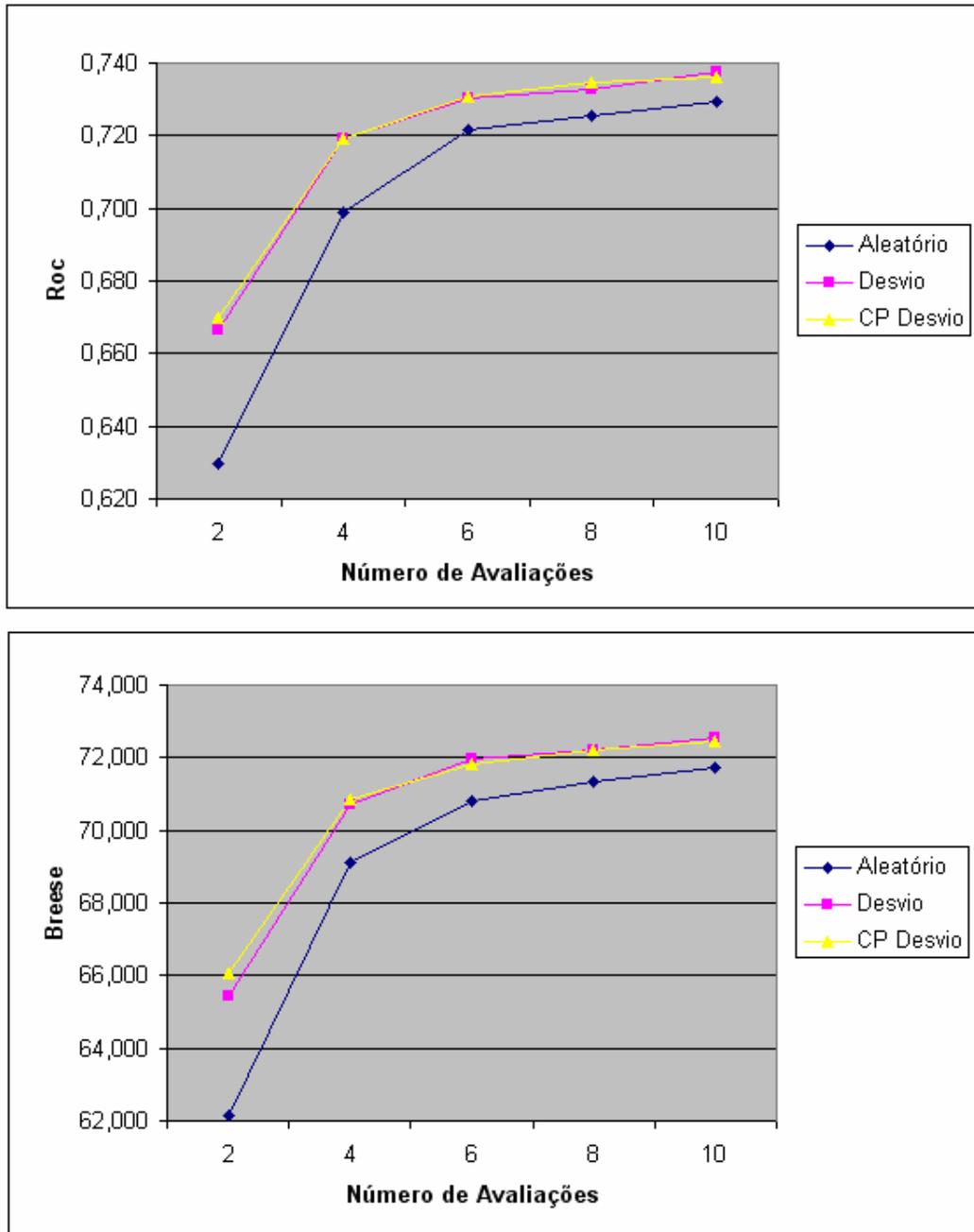


Figura 5.5 - Comparação entre Desvio isoladamente e o Desvio combinado com a Popularidade.

Assim, é possível afirmar que a utilização do desvio já constitui um método de seleção para aprendizagem ativa em sistemas de filtragem colaborativa que fornece um ganho na qualidade das recomendações equiparável aos obtidos com o método *ActiveCP*. A vantagem do desvio é o fato de ele ser mais simples de implementar e mais leve por não

envolver a combinação de duas listas como no *ActiveCP* (Seção 4.3.3). Além disso, sua utilização é direta e abrangente, não envolve o estabelecimento de uma quantidade mínima de avaliações que um item precisa ter para poder ser utilizado (parâmetro I).

5.5. Considerações Finais

Neste capítulo sugerimos uma nova medida de controvérsia para resolver as restrições metodológicas impostas pela variância que vimos no capítulo anterior. Assim, propusemos uma medida que chamamos de desvio cujo resultado já reflete a quantidade de avaliações que um item tenha recebido. Apresentamos também uma nova metodologia experimental, eliminando as restrições presentes na que foi originalmente utilizada nos experimentos do *ActiveCP*. Essa metodologia tenta tirar proveito das vantagens da utilização do desvio.

Em seguida apresentamos os resultados obtidos pelos experimentos realizados em uma base de 300 usuários que simula a base de um sistema de FC em início de operação. Foram realizados experimentos utilizando critérios de seleção de itens baseados na popularidade e controvérsia utilizados isoladamente e também os combinando. Como medidas de controvérsia, utilizamos o desvio e a variância. No caso da variância, executamos os experimentos em quatro configurações:

- Utilizando todas as avaliações de um item
- Utilizando 20 avaliações de um item ($I = 20$)
- Utilizando 40 avaliações de um item ($I = 40$)
- Utilizando 60 avaliações de um item ($I = 60$)

Os resultados dos experimentos realizados forneceram subsídios que nos permitiram afirmar que o fato da variância não refletir a quantidade de avaliações de um item no seu cálculo, efetivamente se constitui em um problema quando se deseja utilizá-la como método de seleção. A possibilidade da ocorrência desse problema havia sido apontada por Teixeira em seu trabalho [Teixeira 02a] sem, no entanto ter sido investigado de maneira mais aprofundada.

Finalmente, vimos que o desvio sozinho pode ser utilizado como um critério de seleção de itens em um sistema de filtragem colaborativa, por já combinar em seu cálculo a

dispersão das notas de um item e a sua popularidade. Os resultados experimentais mostraram que o desvio fornece um ganho na qualidade das recomendações bastante semelhante aos obtidos pelo *ActiveCP*, sendo um método mais leve e de implementação bastante simples.



Capítulo 6

Conclusões e Trabalhos Futuros

6.1. Conclusões

Nos últimos anos, os sistemas de recomendação personalizada têm se tornado uma ferramenta bastante útil para usuários que necessitem de auxílio inteligente na busca por itens relevantes em meio ao enorme volume de informação disponível. Porém, antes de poder gerar recomendações de boa qualidade, o sistema precisa identificar as necessidades específicas de cada usuário.

Ao longo deste trabalho, vimos que a aprendizagem ativa tem sido empregada com sucesso para diminuir a quantidade de avaliações que um usuário novato precisa fazer até que o sistema consiga gerar boas recomendações para ele. Vimos como os conceitos de popularidade e controvérsia (entropia) foram empregados para selecionar os itens que serão avaliados pelos usuários na fase de aquisição de preferências, pelo método *ActiveCP* [Teixeira 02a].

Neste trabalho, nós estudamos os problemas de utilização prática da metodologia do *ActiveCP*, tal como havia sido proposta, devidos em grande parte à utilização da variância como medida de controvérsia. Propusemos então a utilização de uma nova medida de controvérsia, chamada desvio, juntamente com uma nova metodologia de experimentos mais flexível e mais facilmente aplicável na prática.

Os resultados dos experimentos realizados nos forneceram ainda bons indicativos de que a medida do desvio pode ser utilizada sozinha como um método de seleção que já combina em seu cálculo os conceitos de controvérsia e popularidade de um item. Com a vantagem de ser mais simples de implementar, computacionalmente mais leve e com ganhos na qualidade de recomendação compatíveis com os obtidos pelo *ActiveCP*.

6.2. Contribuições

Diante das propostas e resultados experimentais apresentados, nós acreditamos que este trabalho contribui com o campo da aprendizagem ativa em sistemas de recomendação com os seguintes pontos:

- Revisão do estado da arte nos assuntos investigados;

- Confirmação da utilidade prática dos critérios da controvérsia e popularidade na seleção de itens em um método de aprendizagem ativa para sistemas de FC com kNN. Em especial, a utilização dos critérios foi testada em uma situação de um sistema em início de operação com bons resultados;
- Os testes experimentais executados confirmam as limitações impostas pela utilização da variância como medida de controvérsia. Tal fato havia sido apontado por Teixeira em seu trabalho, porém sem uma investigação mais aprofundada;
- As propostas da utilização do desvio, uma nova medida de controvérsia que não possui as limitações da variância, e de uma nova metodologia de experimentos mais facilmente aplicável na prática e mais simples. Adicionalmente, encontramos indicativos de que o desvio pode ser utilizado sozinho como método de seleção que já inclui ambos os conceitos de entropia de um item (controvérsia) e popularidade (cobertura). As vantagens em utilizá-lo são a simplicidade de implementação e o desempenho computacional.

6.3. Limitações e Trabalhos Futuros

Ao aplicar uma estratégia de aprendizagem ativa para aprender as preferências de um usuário novo, é importante levar em consideração a possibilidade de o usuário não conhecer o item que o sistema pede que ele avalie. Caso isso ocorra, o sistema está fazendo com que o usuário perca tempo analisando itens que ele não consegue avaliar. Esse problema foi chamado por Rashid et al. de “esforço do usuário” (*user effort*) [Rashid 02], mas não foi investigado no contexto desse trabalho. Como trabalho futuro, um estudo relevante seria investigar o quanto os métodos de seleção sugeridos requerem de esforço por parte do usuário na etapa inicial de aquisição de preferências.

Uma vez que um dos objetivos do trabalho foi investigar o efeito do dilema cobertura versus intensidade, decidimos realizar todos os experimentos em uma base de dados pequena, onde se prevê que o dilema teria maior impacto. Em bases grandes, se espera que seu efeito seja menor uma vez que cada filme teria um grande número de avaliações para cálculo seguro da variância. No entanto, é necessário realizar novos

experimentos para investigar o dilema e verificar a maior utilidade do desvio sobre a variância em bases grandes (já consolidadas).

As conclusões obtidas no trabalho são válidas dentro do contexto de filtragem colaborativa, mais especificamente usando kNN. No entanto, é necessário realizar experimentos para verificar a utilidade dos métodos de aprendizagem ativa estudados em sistemas de recomendação que utilizem filtragem baseado em conteúdo ou mesmo filtragem híbrida. Ainda dentro do contexto de filtragem colaborativa, seria interessante investigar a utilização dos métodos de seleção em outros algoritmos de recomendação, não baseados no kNN.

Na Seção 4.3.1 mencionamos a possibilidade de utilizar outras medidas de dispersão para medir a controvérsia de um item. Esse estudo poderia ser conduzido investigando medidas de dispersão na literatura para tentar compor novas métricas combinadas dentro do método *ActiveCP*.

Referências

- [**Angluin 88**] Angluin, D. 1988. *Queries and concept learning*. Machine Learning, 6(1): p. 37-66.
- [**Baeza-Yates 99**] Baeza-Yates, R., Ribeiro-Neto, B. 1999. *Modern Information Retrieval*. 1a Edição, Addison Wesley Longman.
- [**Balabanovic 97**] Balabanovic, M., Shoham, Y. 1997 *Fab: Content-Based, Collaborative Recommendation*. Communications of the ACM, 40(3): p. 66-72.
- [**Belkin 92**] Belkin, N., Croft, B. W. 1992. *Information Filtering and Information Retrieval: Two Sides of the Same Coin?* Communications of the ACM 35(12): p.29-38.
- [**Bezerra 04**] Bezerra, B.L.D. 2004. *Uma Solução Em Filtragem de Informação Para Sistemas de Recomendação Baseada em Análise de Dados Simbólicos*. Dissertação de Mestrado, Universidade Federal de Pernambuco.
- [**Billsus 98**] Billsus, D., Pazzani, M. J. 1998. *Learning Collaborative Information Filters*. Proceedings of the 15th International Conference on Machine Learning, p. 46-54.
- [**Boutilier 03**] C. Boutilier, R. S. Zemel, and B. Marlin. 2003. Active collaborative filtering. In Proceedings of the Nineteenth Annual Conference on Uncertainty in Artificial Intelligence, p. 98-106.
- [**Breese 98**] Breese, J. S., Heckerman, D., Kadie, C. 1998. *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*. Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, p. 43-52.
- [**Breiman 98**] Breiman, L., Friedman, J.H., Olshen, R. A., Stone, C.J. 1984. *Classification and Regression Trees*. Chapman Hall.
- [**Burke 99**] Burke, R. 1999. *The Wasabi Personal Shopper: A Case-Based Recommender System*. Proceedings of the 11th National Conference on Innovative Applications of Artificial Intelligence, p. 844-849.
- [**Claypool 99**] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., Sartin, M. 1999. *Combining Content-Based and Collaborative Filters in an Online Newspaper*. Proceedings of ACM SIGIR Workshop on Recommender Systems. Disponível em <http://www.cs.umbc.edu/~ian/sigir99-rec>.

- [**Claypool 01**] Claypool, M., Le, P., Waseda, M. e Brown, D. 2001. *Implicit Interest Indicators*. ACM IUI 01 Intelligent User Interfaces, p.33-40.
- [**Cohen 98**] Cohen, W. W., Basu, C., Hirsh, H. 1998. *Recommendation as Classification: Using Social and Content-Based Information in Recommendation*. Proceedings of the AAAI-98, p. 714 – 720.
- [**Cohn 94**] Cohn, D. A., Atlas, L., Lander, R. 1994. *Improving generalization with active learning*. Machine Learning, 15(2): p. 201-221.
- [**Cotter 00**] Cotter, P., Smyth, B. 2000. *PTV: Intelligent Personalised TV Guides*. Proceedings of IAAI 2000 conference, p. 957-964.
- [**Goldberg 92**] Goldberg, D., Nichols, D., Oki, B. M. e Terry, D. 1992. *Using Collaborative Filtering to Weave an Information Tapestry*. Communications of the ACM, 35(12): p. 61-70.
- [**Good 99**] Good, N., Schafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B., Herlocker, J., Riedl J. 1999. *Combining Collaborative Filtering with Personal Agents for Better Recommendations*. Proceedings of IAAI 99 conference, p. 439-446.
- [**Hanley 82**] Hanley, J. A.,McNeil, B. J., 1982. *The Meaning and use of the Area Under a Receiver Operating Characteristic (ROC) Curve*. Radiology, 143: p. 29-36.
- [**Hasenjager 98**] Hasenjager, M., Ritter, H. 1998. *Active Learning with Local Models*. Neural Processing Letters, 7(2): p.107-117.
- [**Herlocker 00**] Herlocker, J. L. 2000. *Understanding and Improving Automated Collaborative Filtering Systems*. Phd thesis, University of Minnesota.
- [**Herlocker 04**] Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J. 2004. *Evaluating Collaborative Filtering Recommender Systems*. Proceedings of ACM Transactions on Information Systems, 22 (1), p. 5–53.
- [**Hill 95**] Hill, W., Stead, L., Rosenstein, M., Furnas, G. W. 1995. *Recommending and Evaluating Choices in a Virtual Community of Use*. Proceedings of ACM CHI 95, p. 194-201.
- [**Konstan 97**] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., Riedl, J. 1997. *GroupLens: Applying Collaborative Filtering to Usenet News*. Communications of the ACM, 40(3). p. 77-87.

- [**Krukwich 96**] Krukwich, B., Burkey, C. 1996. *Learning user information interests through extraction of semantically significant phrases*. Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access.
- [**Lin 00**] Lin, W., Alvarez, S. A., Ruiz, C. 2000. *Collaborative Recommendation via Adaptive Association Rule mining*. ACM-SIGKDD 2000 Workshop on Web Mining for ECommerce. Disponível em <http://robotics.stanford.edu/~ronnyk/WEBKDD2000/papers/>
- [**Lindebaum 99**] Lindenbaum, M., Markovitch, S., Rusakov, D. 1999. *Selective Sampling for Nearest Neighbor Classifiers*. Proceedings of The Sixteenth National Conference on Artificial Intelligence AAAI-99, p. 366-371.
- [**Malone 87**] Malone, T. W., Grant, K. R., Turbak, F. A., Brobst, S. A., Cohen, M. D. 1987. *Intelligent Information-Sharing Systems*. Communications of the ACM, 30(5): p. 390-402.
- [**McJones 97**] McJones, P. 1997. Eachmovie collaborative filtering data set. <http://www.research.digital.com/SRC/eachmovie> (Disponível até Outubro de 2004). DEC System Research Center.
- [**Morita 94**] Morita, M., Shinoda, Y. 1994. *Information Filtering Based on User Behaviour Analysis and Best Match Text Retrieval*. Proceedings of ACM-SIGIR 94, p. 272-281.
- [**Nichols 97**] Nichols, D. M. 1997. *Implicit Ratings and Filtering*. 5th DELOS Workshop on Filtering and Collaborative Filtering. p. 10-12.
- [**Oard 98**] Oard, D., Kim, J. 1998. *Implicit FeedBack for Recommender Systems*. Proceedings of the AAAI WorkShop on Recommender Systems. Disponível em <http://www.glue.umd.edu/~oard/research.html>
- [**Pazzani 97**] Pazzani, M. and Billsus, D. 1997. *Learning and revising user profiles: The identification of interesting web sites*. Machine Learning 27(3): p. 313-331.
- [**Pennock 00**] Pennock, D. M. and Horvitz, E. 2000. *Collaborative Filtering by Personality Diagnosis: A Hybrid Memory and Model-Based Approach*. In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, p. 473-480.
- [**Queiroz 03**] Queiroz, S. R. M. 2003. *Group Recommendation Strategies Based On Collaborative Filtering*. Dissertação de Mestrado, Universidade Federal de Pernambuco.
- [**Rashid 02**] Rashid, A.M., Albert, I., Dosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., Riedl, J. 2002. *Getting to Know You: Learning New User Preferences in Recommender*

Systems. In: Proceedings of International Conference on Intelligent User Interface (IUI2002).

[**Resnick 94**] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. 1994. *GroupLens: An Open Architecture fo Collaborative Filtering of Netnews*. Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work, p. 175-186.

[**Sarwar 00**] Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. 2000. *Analysis of Recommendation Algorithms for E-Commerce*. Proceedings of the 2nd ACM Conference on Electronic Commerce (EC-00), p. 158-167.

[**Schafer 01**] Schafer, J.B., Konstan, J.A. e Riedl, J. 2001. *E-Commerce Recommendation Applications*. Data Mining and Knowledge Discovery, Volume 5, Issue 1-2, p. 115-153.

[**Shardanand 95**] Shardanand, U., Maes, P. 1995. *Social Information Filtering: Algorithms for Automating “Word of Mouth”*. Proceedings of CHI 95, p. 210-217.

[**Swets 63**] Swets, J. A., 1963. *Information Retrieval Systems*. Science, 141: p. 245-250.

[**Terveen 01**] Terveen, L., Hill, W. 2001. *Beyond Recommender Systems: Helping People Help Each Other*. Capítulo 22 de Human-Computer Interaction in the Millenium, Addison-Wesley, 2001.

[**Teixeira 02a**] Teixeira, I.R. 2002. *Um Método de Aprendizagem Ativa em Sistemas de Filtragem Colaborativa*. Dissertação de Mestrado, Universidade Federal de Pernambuco.

[**Teixeira 02b**] Teixeira, I. R., De Carvalho, F. A. T., Ramalho, G. L., Corruble, V. 2002. *Active CP: A method for Speeding up User Preferences Acquisition in Collaborative Filtering Systems*. In: 16th Brazilian Symposium on Artificial Intelligence (SBIA2002), p. 237-247.

[**Yao 95**] Yao, Y. Y., 1995. *Measuring Retrieval Effectiveness Based on User Preference of Documents*. Journal of the American Society for Information Science, 46(2): p. 133-145.

[**Yu 04**] Yu, K., Schwaighofer, A., Tresp, V., Xu, X., Kriegel, H.P. 2004. *Probabilistic Memory-based Collaborative Filtering*. In: IEEE Transactions On Knowledge and Data Engineering, Volume 16, Issue 1, p. 56 - 69.