

PROPOSAL OF THE METHODOLOGICAL STRUCTURE TO IMPLEMENT RECOMMENDER SYSTEMS

Leonardo Gomes de Oliveira (Santa Catarina Federal University – Post-Graduate Program in Knowledge Engineering and Management, Santa Catarina, Brazil) – leonardo@guiafloripa.com.br

Recommender Systems have been pointed out as an important tool to face the problem of information overload on the Web. Many studies have contributed to the number of customized services that are offered on the Internet. This research presents the results of a survey on the main features of recommender systems, that is, the used input data, the commonest outputs and the types and techniques of recommendation. It is on these grounds that this work suggests a general methodological structure to implement recommender systems.

Key-Words: Recommender systems. Information filtering. Methodological structure. Internet. Electronic commerce.

PROPOSTA DE UMA ESTRUTURA METODOLÓGICA PARA IMPLEMENTAÇÃO DE SISTEMAS DE RECOMENDAÇÃO

Os sistemas de recomendação vêm sendo apontados como importante ferramenta para enfrentar o problema de sobrecarga de informação na Web. Inúmeras pesquisas na área têm contribuído para a oferta de serviços personalizados na rede. Este trabalho apresenta um levantamento sobre as principais características dos sistemas de recomendação: os dados de entrada utilizados, as saídas mais comuns, os tipos e técnicas de recomendação. Com base nisso, propõe-se uma estrutura metodológica genérica para implementação de sistemas de recomendação.

Palavras-chave: Sistemas de recomendação. Filtragem de informação. Estrutura metodológica. Internet. Comércio eletrônico.

1 Introdução

A crise enfrentada pelas empresas ponto-com, em 2001, não intimidou o desenvolvimento da Internet. Em novembro de 2006, a *Word Wide Web*, ou simplesmente Web, alcançou a marca de 100 milhões de *websites*¹. Em meio à profusão de informações e serviços on-line estão os usuários da Internet - um contingente de mais de um bilhão de pessoas ao redor do mundo².

À medida que cresce o volume de informações na *Web*, aumenta a dificuldade dos usuários para encontrar a “informação certa” no “tempo certo” (O’Donovan; Smyth, 2005). A grande diversidade de conteúdo disponível gera sobrecarga de informação ao usuário. Os sistemas de recomendação vêm sendo apontados como uma importante ferramenta para enfrentar esse problema (O’Donovan; Smyth, 2005; Shardanand; Maes, 1995).

Em um típico sistema de recomendação, pessoas provêm recomendações como entrada e o sistema as reúne e direciona para os destinatários apropriados (Resnick; Varian, 1997), ampliando a capacidade e eficácia dos processos de indicação tão comuns nas relações sociais entre os seres humanos (Reategui; Cazella, 2005; Resnick; Varian, 1997).

Em meados dos anos 90, os sistemas de recomendação emergiram como uma área de pesquisa independente (Adomavicius; Tuzhilin, 2005). Os primeiros sistemas de recomendação divulgados foram: Tapestry (Goldberg et al., 1992), GroupLens (Resnick et al., 1994), Video Recommender (Hill et al., 1995) e Ringo (Shardanand; Maes, 1995).

Burke (2002) define sistemas de recomendação como qualquer sistema que produza recomendações individualizadas como saída, ou que tenha o efeito de guiar o usuário de forma personalizada a objetos interessantes e úteis, diante de uma grande variedade de opções.

Para Adomavicius e Tuzhilin (2005), o interesse em tal área é grande porque constitui um rico problema de pesquisa, além de ter inúmeras aplicações práticas que podem ajudar os usuários a lidar com a sobrecarga de informação, uma vez que tais sistemas oferecem recomendações personalizadas de conteúdos e serviços.

Reategui e Cazella (2005: 308) pontuam que “os websites de comércio eletrônico são atualmente o maior foco de utilização dos sistemas de recomendação, empregando diferentes técnicas para encontrar os produtos mais adequados para seus clientes e aumentar deste modo sua lucratividade”.

O desenvolvimento de um sistema de recomendação não é uma tarefa trivial, visto que requer tanto a análise do domínio de aplicação quanto a ponderação sobre os diversos aspectos que envolvem sua implementação.

Com base na literatura, procurou-se levantar as principais características dos sistemas de recomendação com o propósito de apresentar uma estrutura metodológica que relacione escolhas e definições necessárias à implementação. A estrutura está dividida nas seguintes etapas: escolha da estratégia de recomendação, escolha do tipo e técnica de recomendação, determinação do modelo do usuário, definição do processo de recomendação, avaliação das recomendações.

Na seção a seguir identificam-se os diversos aspectos que envolvem a implementação. Na seção 3 apresenta-se a estrutura metodológica proposta para

¹Considera o número de domínios registrados que possuem conteúdo publicado. Disponível em: <<http://www.netcraft.com>>. Acesso em: 16 jan. 2007.

²Disponível em: <<http://www.internetworldstats.com>>. Acesso em: 16 jan. 2007.

implementação de sistemas de recomendação. Por fim, são tecidas conclusões e recomendações.

2 Sistemas de Recomendação

De maneira geral, os sistemas de recomendação são formados por (Burke, 2002):

- dados prévios ou armazenados (*background data*): corresponde a toda informação que o sistema armazena para utilizar no processo de recomendação;
- dados de entrada ou do usuário: trata-se de informação fornecida pelo usuário para que o processo de recomendação possa ser iniciado;
- algoritmo de recomendação: é responsável por combinar dados prévios e dados de entrada para que sejam fornecidas as devidas recomendações.

Os dados prévios compreendem a relação de itens a serem recomendados (catálogo de produtos, documentos, páginas *web*, conteúdo multimídia etc.), podendo englobar também o conjunto de usuários já cadastrados e suas avaliações.

O processo de recomendação, de forma simplificada, está representado na Figura 1:

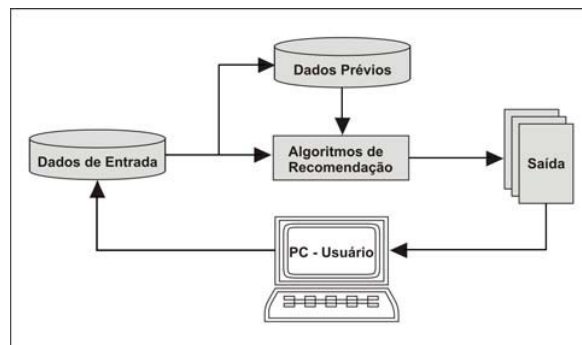


Figura 1- Processo de recomendação

Fonte: elaborada pelo autor.

A partir da visão geral do processo de recomendação serão discutidas as características de um sistema de recomendação. A seqüência de apresentação adotada é meramente para efeitos de compreensão. Destaca-se que os algoritmos não serão tratados em um tópico a parte, já que sua escolha depende da **técnica** a ser aplicada, e a técnica, por sua vez, está relacionada ao **tipo de recomendação**.

2.1 Dados de entrada

Os dados de entrada que o usuário fornece para que a predição seja possível, base para modelar seu perfil, podem ser coletados de forma explícita ou implícita (Reategui; Cazella, 2005). Segundo Breese, Heckerman e Kadie (1998), as avaliações explícitas se referem àquelas em que o usuário expressa conscientemente suas preferências, seja por meio de um questionário ou pelo voto declarado em itens (geralmente utilizando uma escala numérica discreta). Já as avaliações implícitas são automaticamente inferidas,

baseadas na interpretação das preferências do usuário segundo seu comportamento – navegação, histórico de compra etc.

Para Nichols (1997: 32), uma série de ações do usuário pode ser interpretada como avaliações implícitas, conforme mostra o Quadro 1.

| Ação | Notas |
|-------------------|---|
| Comprar | adquirir um item |
| Avaliar | considerar a atitude de avaliar ou não um item |
| Uso repetido | ex.: a busca pelo mesmo assunto |
| Salvar / Imprimir | artigos que o usuário deseja preservar |
| Apagar (deletar) | excluir um item |
| Referenciar | citar ou fazer menção a um item (artigo) |
| Responder | ex.: responder uma mensagem |
| Marcar | ex.: adicionar à lista de favoritos |
| Examinar / Ler | ex.: ler um artigo todo |
| Considerar | ex.: ler o resumo de um artigo |
| Dar uma olhada | ex.: ler o título de um artigo |
| Associar | relacionar itens aos que estão sendo examinados |
| Consultar | fazer associações entre os termos da consulta |

Quadro 1 - Tipos potenciais de avaliações implícitas

Fonte: Traduzido de Nichols (1997: 32).

As avaliações explícitas por voto declarado representam uma estratégia comum utilizada em *sites* de comércio eletrônico. Permitem que usuários opinem sobre itens oferecidos em lojas virtuais por meio de notas (geralmente, uma escala de um a cinco). A Figura 2 é um exemplo de como pode ser realizado esse tipo de coleta de dados.



Figura 2 - Exemplo de avaliação explícita por voto declarado

Fonte: <http://www.amazon.com>.

2.2 Saídas de um sistema de recomendação

Para Vozalis e Margaritis (2003), a saída de um sistema de recomendação pode ser uma Predição ou uma Recomendação.

- A *Predição* é expressa por um valor numérico que representa a opinião antecipada do usuário ativo u_a em relação ao item i . Ou seja, o sistema apresenta um item associado a uma escala que prediz sua relevância para o usuário, por exemplo, um filme apresentado com uma estrela significa “não recomendado ao usuário”, enquanto outro, com cinco estrelas representa a opinião “altamente recomendado”;
- A *Recomendação* é representada por uma lista de N itens mais relevantes para o usuário ativo. Normalmente, a lista apresenta itens não comprados, vistos ou avaliados pelo usuário. Este tipo de saída também é conhecido como *Recomendação Top-N*.

Portanto, a saída pode ser a apresentação do item mais recomendado ou de um conjunto de N itens. É possível, ainda, combinar essas duas saídas criando uma lista de itens recomendados, onde uma predição é associada a cada um deles.

Em *sites* de comércio eletrônico, conforme Reategui e Cazella (2005), é comum encontrar outros tipos de saída empregadas como estratégias de recomendação. Geralmente, são respostas com baixo nível de personalização e que não requerem conhecimento sobre o perfil do usuário, suas preferências e necessidades; baseiam-se unicamente no histórico de avaliações do grupo de consumidores. Ainda assim, esses tipos de estratégias, listadas na seqüência, podem ser considerados como sistemas de recomendação segundo a taxonomia proposta por Schafer, Konstan e Riedl (2001).

- *listas de recomendação*: listas contendo itens populares, seguindo uma organização temática, do tipo: “os eletrônicos mais vendidos”, “idéias para presente”, entre outras;
- “*usuários que compraram ‘X’ também compraram ‘Y’*”: lista de itens associados ao que está sendo analisado (ou comprado) pelo usuário ativo, baseado no histórico de compras de outros usuários;
- “*examinar itens semelhantes*”: relação de itens associados ao que está sendo analisado (ou comprado) pelo usuário ativo, baseado em suas características. Por exemplo: filmes do mesmo gênero, diretor, elenco etc.

2.3 Tipos de recomendação

As discussões a respeito da classificação dos sistemas de recomendação giram em torno das fontes de dados do sistema (dados prévios e de entrada) e da maneira como esses dados são utilizados para fornecer as recomendações (Burke, 2002). Os sistemas podem ser classificados nas seguintes categorias (Adomavicius; Tuzhilin, 2005; Burke, 2002):

- recomendação baseada em conteúdo ou filtragem baseada em conteúdo;
- recomendação colaborativa ou filtragem colaborativa;
- recomendação demográfica;
- recomendação baseada em utilidade;
- recomendação baseada em conhecimento.
- abordagem híbrida ou filtragem híbrida.

2.3.1 Recomendação baseada em conteúdo

Em um sistema de recomendação baseado em conteúdo, cada item de um conjunto I é definido por características associadas a ele (Burke, 2002). Por exemplo, um filme pode ter como características: título, diretor, elenco, gênero etc. Já em um texto, as próprias palavras que o compõe são consideradas como suas características. É com base nessas características que os itens podem ser comparados e a semelhança entre eles estabelecida.

Conhecendo o histórico de preferência do usuário por itens do conjunto I , e a semelhança entre esses itens, o sistema de recomendação baseado em conteúdo é capaz de recomendar para este usuário outros itens que possam ser de seu interesse (Reategui; Cazella, 2005).

Os sistemas de recomendação baseados em conteúdo têm origem nas técnicas de recuperação de informação (Balabanović; Shoham, 1997) e nas pesquisas de filtragem de informação (Adomavicius; Tuzhilin, 2005; Burke, 2002).

Conforme Adomavicius e Tuzhilin (2005), em virtude dos significativos avanços na área de recuperação de informação e filtragem, e dada à importância das inúmeras aplicações baseadas em texto, muitos dos sistemas baseados em conteúdo focaram na recomendação de itens que contêm informação textual, como documentos, *websites* e notícias. Para esses autores, a melhoria trazida pelos sistemas baseados em conteúdo, em relação à tradicional recuperação de informação, vem da utilização de perfis de usuários que contêm dados sobre seus gostos, preferências e necessidades.

De maneira geral, os sistemas baseados em conteúdo mantêm perfis de longo prazo: o modelo de usuário se baseia nas características dos itens avaliados por ele, sendo atualizado à medida que outros itens vão sendo avaliados (Burke, 2002).

De acordo com Adomavicius e Tuzhilin (2005), Balabanović e Shoham (1997) e Burke (2002) os sistemas baseados em conteúdo possuem as seguintes limitações:

Análise de conteúdo limitada

As técnicas baseadas em conteúdo são limitadas pelas características que explicitamente devem estar associadas aos itens que serão recomendados. Isso significa que as características do conteúdo devem estar estruturadas de forma que possam ser analisadas automaticamente pelo computador. A extração automática de características é de difícil aplicação a alguns domínios - como dados multimídia (por exemplo, vídeo e som) -, e atribuí-las manualmente pode ser inviável, em alguns casos, devido a limitações de recursos;

Superespecialização

Em virtude de o sistema se basear nos itens já avaliados pelo usuário para fazer suas recomendações, ele acaba limitando a recomendação a itens de grande semelhança àqueles já conhecidos pelo usuário. Exemplificando: uma pessoa que nunca experimentou comida árabe jamais receberia uma recomendação para ir ao melhor restaurante de comida árabe da cidade. A diversidade de recomendações é, frequentemente, uma característica desejável em sistemas de recomendação.

Outra consequência da superespecialização é que, em alguns casos, certos itens não deveriam ser recomendados justamente por serem muito semelhantes aos já avaliados pelo usuário, como notícias que descrevem o mesmo fato. O sistema *DailyLearner*

(Billsus; Pazzani, 2000), por exemplo, emprega um limite de similaridade para excluir as notícias muito semelhantes às já lidas pelo usuário;

O problema do novo usuário

Para que o sistema de recomendação baseado em conteúdo possa realmente compreender as preferências e apresentar recomendações confiáveis, o usuário tem que avaliar um número suficiente de itens. Um novo usuário, que avaliou poucos itens, conseqüentemente, receberá recomendações pouco precisas.

Elasticidade *versus* plasticidade

Ao contrário do problema do novo usuário, o problema conhecido como “estabilidade *versus* plasticidade” (Burke, 2002) surge quando uma quantidade substancial de avaliações é feita e o perfil do usuário é consolidado. Com isso, torna-se difícil alterar suas preferências. Caso um apreciador de churrasco venha a se tornar vegetariano, ele continuará recebendo recomendações de churrascarias por um bom tempo, até que suas novas avaliações sejam suficientes para alterar seu perfil.

A fim de contornar esse problema, alguns sistemas procuram incorporar algum tipo de variável temporal que aumente a relevância das avaliações mais recentes. Em contrapartida, assumem o risco de perder informações sobre interesses de longo prazo que são manifestados esporadicamente (Burke, 2002).

2.3.2 Recomendação colaborativa

Segundo Burke (2002), a recomendação colaborativa é provavelmente a mais familiar, a mais utilizada e a que apresenta tecnologias mais consolidadas. Os sistemas de recomendação baseados em filtragem colaborativa procuram prever a utilidade de um item para um usuário particular com base em itens previamente avaliados por outros usuários (Adomavicius; Tuzhilin, 2005). Diferencia-se da filtragem baseada em conteúdo exatamente por não exigir a compreensão ou reconhecimento das características dos itens.

Nos sistemas colaborativos, a essência está na troca de experiências entre as pessoas que possuem interesses comuns (Reategui; Cazella, 2005). Um perfil típico de usuário em um sistema colaborativo consiste em um vetor de itens e suas avaliações, que aumenta continuamente à medida que o usuário interage com o sistema (Burke, 2002).

Um exemplo de ambiente baseado em filtragem colaborativa é o sistema de recomendação de filmes *MovieLens*³ (Good et al., 1999). Nele o usuário atribuiu nota aos filmes que já assistiu e o sistema utiliza essas avaliações para encontrar pessoas com gostos similares e fazer suas predições.

Os sistemas colaborativos puros resolvem as principais deficiências dos sistemas baseados em conteúdo. Usando recomendações de outros usuários é possível tratar qualquer tipo de conteúdo e recomendar itens, mesmo que esses não sejam semelhantes aos itens já avaliados pelo usuário. No entanto, os sistemas colaborativos têm suas próprias limitações (Adomavicius; Tuzhilin, 2005; Balabanović; Shoham, 1997; Burke, 2002):

O problema do novo item

³Disponível em: <<http://www.movielens.org>>. Acessado em: 16 jan. 2007.

Novos itens são regularmente adicionados aos sistemas de recomendação. Como os sistemas colaborativos dependem unicamente da avaliação dos usuários para que possam fazer recomendações, enquanto um novo item não for avaliado por um número considerável de usuários, o sistema não o recomendará.

Avaliações esparsas

Caso o número de usuários seja pequeno em relação ao volume de informações no sistema, existe um grande risco das avaliações tornarem-se muito esparsas, devido ao pequeno número de avaliações, comparado ao necessário, para se fazer recomendações. Em um sistema de recomendação de filmes, só para exemplificar, aqueles que forem avaliados por poucas pessoas são raramente recomendados, mesmo que essas atribuam pontuações elevadas.

Similaridade (usuário incomum)

Um usuário cujo gosto seja incomum, comparado à maioria da população, terá dificuldades para encontrar usuários com gostos semelhantes ao seu, e, por isso, suas recomendações podem se tornar pobres (Reategui; Cazella, 2005). Esse problema é também encontrado na literatura com o nome de “*gray sheep*”, ou “ovelha negra” (Claypool et al., 1999). Para Adomavicius e Tuzhilin (2005), esse problema é considerado um problema de esparsidade.

Na opinião de Burke (2002), um sistema colaborativo trabalha melhor quando o usuário pertence a um grupo com muitos “vizinhos” de gostos similares ao seu.

Assim como os sistemas baseados em conteúdo, os sistemas colaborativos apresentam problemas de “novo usuário” e de “elasticidade *versus* plasticidade”.

Na literatura, os problemas de “novo usuário” e “novo item”, provocados pela escassez de dados no sistema, são também conhecidos como “problemas de *ramp-up*” (Konstan et al., 1998 apud Burke, 2002: 335).

2.3.3 Recomendação demográfica

Os sistemas de recomendação demográfica têm como objetivo classificar o usuário em função de seus atributos pessoais (idade, sexo, renda etc.). Realizam a recomendação com base em classes demográficas (Burke, 2002).

Para Adomavicius e Tuzhilin (2005), a abordagem demográfica é uma extensão dos sistemas colaborativos, uma vez que ambos se baseiam na similaridade entre perfis de usuários. A diferença está na forma como o perfil do usuário é construído: enquanto a filtragem demográfica calcula a similaridade com base nos dados demográficos do usuário, a filtragem colaborativa utiliza o histórico de avaliações. Conseqüentemente, a recomendação demográfica apresenta as principais desvantagens encontradas no método colaborativo. Contudo, como não depende das avaliações para comparar usuários, sofre menos com o problema de esparsidade.

Burke (2002) menciona que todas as abordagens baseadas em aprendizagem (baseada em conteúdo, colaborativa e demográfica) possuem, de uma forma ou de outra, o problema de *ramp-up*, porque dependem de um número suficiente de avaliações para aprimorar o processo de recomendação.

2.3.4 Recomendação baseada em utilidade

Os sistemas baseados em utilidade fazem sugestões considerando um cálculo de utilidade de cada item para o usuário, sem que este precise ter um perfil de longo prazo identificado pelo sistema. A questão central consiste em criar uma função de utilidade que atenda aos interesses de cada usuário individualmente (Burke, 2002). O perfil do usuário é a própria função de utilidade que o sistema obtém como entrada de dados, e que será utilizada para encontrar a melhor opção de acordo com suas preferências.

As técnicas baseadas em utilidade requerem do sistema uma configuração que considere todas as características dos itens na criação da função de utilidade. Além das características próprias dos itens, o sistema pode incorporar outros fatores que contribuem para a análise de valor de um produto, tais como prazo de entrega e garantia. Em alguns casos, tais fatores podem ser decisivos para uma decisão de compra (Burke, 2002).

Como os sistemas baseados em utilidade não empregam processos que levem em consideração o histórico de avaliações do usuário, pode-se concluir que eles não enfrentam problemas típicos dos sistemas colaborativos e baseados em conteúdo, tais como: “novo usuário”, “novo item” ou “esparsidade”. Em contrapartida, neles, o usuário deve construir totalmente uma função de utilidade que determine suas preferências, o que implica considerar a importância de cada uma das características possíveis. Pode-se considerar isto, uma flexibilidade do sistema, mas também, em algum grau, um inconveniente, pois exige do usuário alto nível de interação (Burke, 2002).

2.3.5 Recomendação baseada em conhecimento

Assim como os sistemas baseados em utilidade, os sistemas de recomendação baseados em conhecimento não utilizam perfis de longo prazo, já que suas recomendações fundamentam-se na análise de correspondência entre as necessidades do usuário e o conjunto de opções disponíveis a ele (Burke, 2002).

Nos sistemas baseados em conhecimento, ao contrário dos sistemas baseados em utilidade, não se exige que os usuários explicitem todas as suas necessidades para fazer recomendações. Na abordagem baseada em conhecimento, o sistema utiliza efetivamente o conhecimento a respeito dos usuários e produtos para fazer inferências sobre suas preferências (Burke, 2002).

Conforme Adomavicius e Tuzhilin (2005), os sistemas de recomendação podem ser melhorados com técnicas baseadas em conhecimento. Entre essas técnicas está o Raciocínio Baseado em Casos (RBC) (Ricci et al., 2002).

A principal vantagem desta abordagem é aumentar a precisão e evitar limitações intrínsecas aos sistemas colaborativos e baseados em conteúdo. Outra vantagem, a ser citada, é que estes sistemas são apropriados para aplicações onde os usuários são esporádicos; onde raramente consultam o sistema para ter uma necessidade específica atendida. O principal inconveniente dos sistemas baseados em conhecimento consiste, justamente, na necessidade de adquirir o conhecimento (Adomavicius; Tuzhilin, 2005).

2.3.6 Filtragem híbrida

É a combinação de dois ou mais tipos de recomendação. Vários sistemas de recomendação combinam diferentes abordagens em uma estrutura híbrida. O principal objetivo é evitar limitações apresentadas em sistemas que aplicam apenas uma abordagem (Adomavicius; Tuzhilin, 2005; Balabanović; Shoham, 1997; Burke, 2002).

De acordo com Adomavicius e Tuzhilin (2005), as principais formas adotadas para combinar filtragem baseada em conteúdo e colaborativa em um sistema híbrido são:

- implementar os métodos colaborativos e baseados em conteúdo separadamente e combinar suas predições: desta forma é possível combinar as avaliações obtidas individualmente em cada um dos métodos para oferecer uma recomendação final. Outra possibilidade consiste em que o próprio sistema escolha, entre dois métodos, aquele que oferece “melhor” resultado, baseado em alguma métrica de “qualidade”;
- incorporar algumas características baseadas em conteúdo em uma abordagem colaborativa: a exemplo do *Fab System* (Balabanović; Shoham, 1997), o sistema pode manter os perfis de usuários baseados em conteúdo, comparar diretamente os perfis para determinar os usuários semelhantes e então utilizar uma recomendação colaborativa. Assim, o usuário ativo recebe não só as recomendações de itens que foram bem avaliados por usuários com perfis semelhantes, mas também itens que sejam semelhantes àqueles já avaliados positivamente por ele;
- incorporar algumas características colaborativas em uma abordagem baseada em conteúdo: o mais comum nessa categoria é a utilização de uma técnica de *redução de dimensionalidade* (por exemplo, *latent semantic indexing*) para criar uma “visão” colaborativa de um grupo de perfis baseados em conteúdo;
- construir um modelo unificado que incorpore características das abordagens baseada em conteúdo e colaborativa: são inúmeras as pesquisas que tem sido desenvolvidas utilizando essa abordagem, onde várias técnicas são empregadas com o objetivo de se ter recomendações mais precisas.

Uma tabela comparativa com as vantagens e desvantagens de cada tipo de recomendação pode ser encontrada em Burke (2002: 335). Conhecendo as limitações de cada tipo de sistema é possível escolher com mais segurança o tipo de recomendação a ser utilizada em uma determinada aplicação prática. Definido o tipo de recomendação, faz-se necessário escolher técnicas e algoritmos capazes de tratar os dados e prover recomendações.

2.4 Técnicas de recomendação

Os algoritmos de filtragem colaborativa foram apresentados por Breese, Heckerman e Kadie (1998) em duas categorias: baseados em memória (*memory-based*) e baseados em modelo (*model-based*). Essa mesma classificação é utilizada por Adomavicius e Tuzhilin (2005) para as técnicas baseadas em conteúdo.

Os algoritmos baseados em memória mantêm uma base de dados com todos os usuários e suas preferências, e para cada recomendação operam sobre toda a base de usuários (ou itens). Já os algoritmos baseados em modelo, primeiro compilam as preferências do usuário em um modelo descritivo de itens, usuários e/ou suas avaliações para que as predições sejam geradas a partir desse modelo (Pennock et al., 2000).

Algoritmos baseados em memória empregam técnicas estatísticas na avaliação de

similaridade e predição. Estas são as mais populares e as mais utilizadas na prática (Sarwar et al., 2001). Entre as que merecem destaque, citam-se: *TF-IDF*, Correlação e Vetor de Similaridade (Adomavicius; Tuzhilin, 2005; Breese; Heckerman; Kadie, 1998).

Os algoritmos baseados em memória tendem a produzir recomendações de melhor qualidade (Deshpande; Karypis, 2004), podendo-se adicionar novos dados facilmente de forma incremental (Pennock et al., 2000). Contudo, podem sofrer problemas de escalabilidade, já que a exigência computacional tende a crescer à medida que a base de dados de usuários e/ou itens aumenta (Deshpande; Karypis, 2004).

Algoritmos baseados em modelo utilizam uma aproximação probabilística para computar o valor esperado de uma predição para o usuário, dadas as suas preferências por outros itens. Para construir o modelo, diferentes algoritmos de aprendizagem de máquina podem ser utilizados, entre os quais, redes *bayesianas*, *clustering*, redes neurais (Sarwar et al., 2001). Esses algoritmos procuram reconhecer padrões ou correlações entre os dados para fazer inferências.

De maneira geral, algoritmos baseados em modelo tendem a produzir recomendações mais rápidas, embora necessitem de um bom tempo de processamento para a modelagem (Deshpande; Karypis, 2004). Ressalta-se que o tempo e a carga computacional podem ser proibitivos, principalmente se, para acrescentar novos dados, for necessário uma recompilação total do modelo (Pennock et al., 2000). Em muitos casos, sistemas com tais algoritmos apresentam qualidade menor em relação à qualidade apresentada por sistemas que empregam métodos baseados em memória (Deshpande; Karypis, 2004).

Diferentes técnicas têm sido utilizadas para abordar os mais diversos problemas de pesquisa e aplicações práticas que têm surgido na área de sistemas de recomendação (Adomavicius; Tuzhilin, 2005). Na seqüência, abordam-se algumas delas.

2.4.1 Técnicas baseadas em memória

As técnicas baseadas em memória, freqüentemente referenciadas na área de sistemas de recomendação, são *TF-IDF*, correlação, correlação baseada em itens e vetor de similaridade. A primeira é um exemplo de algoritmo aplicado a sistemas baseados em conteúdo (*TF-IDF*). As demais são utilizadas em sistemas colaborativos.

TF-IDF

Uma das medidas mais conhecidas para especificar o peso das palavras em sistemas de recuperação de informação é a técnica *term frequency/inverse document frequency (TF-IDF)* (Adomavicius; Tuzhilin, 2005).

A técnica se baseia em calcular o número de vezes que uma palavra i aparece em um documento j ($TF_{i,j}$). Entretanto, as palavras muito comuns, que aparecem em muitos documentos, tornam-se pouco úteis para distinguir seu conteúdo. Por isso, a medida do inverso da freqüência em documentos (*IDF*) é bastante utilizada. Com ela, o peso das palavras comuns é reduzido.

O peso de uma palavra em um dado documento é calculado como sendo o produto das duas medidas (*TF-IDF*). Assim, o documento pode ser representado como um vetor de palavras e seus pesos. Como nos sistemas baseados em conteúdo o modelo de usuário é obtido com base nas características dos itens previamente vistos e avaliados por ele, seu perfil também pode ser representado como um vetor de palavras (Adomavicius; Tuzhilin,

2005).

Com o usuário e o documento representados como vetores é possível medir a similaridade, calculando o cosseno entre eles (Adomavicius; Tuzhilin, 2005).

Além da técnica *TF-IDF*, a *kNN* e a Meta-protótipo são citadas na literatura como técnicas aplicadas a sistemas baseados em conteúdo (Bezerra et al., 2002).

Correlação

A técnica da correlação usa o coeficiente de correlação de Pearson e suas variações para calcular a similaridade entre dois perfis, com base em suas avaliações em determinados itens (ver Figura 3).

Segundo Breese, Heckerman e Kadie (1998), sua formulação estatística para filtragem colaborativa foi apresentada primeiramente por Resnick et al. (1994). A correlação entre o usuário ativo a e outro usuário da base u , é definida como (Breese; Heckerman; Kadie, 1998):

$$w(a,u) = \frac{\sum_i (v_{a,i} - \bar{v}_a)(v_{u,i} - \bar{v}_u)}{\sqrt{\sum_i (v_{a,i} - \bar{v}_a)^2 \sum_i (v_{u,i} - \bar{v}_u)^2}} \quad (1)$$

onde o somatório corresponde ao conjunto de itens i que ambos os usuários avaliaram; $v_{a,i}$ a avaliação (ou voto) do usuário a sobre o item i ; \bar{v}_a o voto médio do usuário a , definido como:

$$\bar{v}_a = \frac{1}{|I_a|} \sum_{i \in I_a} v_{a,i} \quad (2)$$

Uma vez definida a similaridade entre os usuários, a predição de um item i' para o usuário ativo a pode ser calculada como:

$$p_{a,i'} = \bar{v}_a + k \sum_{u \in \hat{U}} w(a,u)(v_{u,i'} - \bar{v}_u) \quad (3)$$

onde \hat{U} denota o conjunto de N usuários que são mais similares ao usuário a e que tenham avaliado o item i' (N pode variar de 1 ao número total de usuários). O peso $w(a,u)$ pode refletir tanto a similaridade quanto a dissimilaridade (distância) entre cada usuário u e o usuário ativo a . O termo k é um fator de normalização, geralmente selecionado como (Adomavicius; Tuzhilin, 2005):

$$k = \frac{1}{\sum_{u \in \hat{U}} |w(a,u)|} \quad (4)$$

A Figura 3 representa uma matriz de avaliações, ou matriz usuário-item. Nas linhas estão distribuídos os usuários e nas colunas os itens. Os elementos da matriz são os votos dos usuários nos itens avaliados. Nota-se que $v_{a,j}$ é a predição que se pretende fazer do item j ao usuário ativo u_a .

| | i_1 | i_2 | \dots | i_i | \dots | i_n |
|----------|-----------|-----------|---------|-----------|---------|-----------|
| u_1 | | | | | | |
| u_2 | | | | | | |
| \vdots | | | | | | |
| u_a | $v_{a,1}$ | $v_{a,2}$ | \dots | $v_{a,i}$ | \dots | $v_{a,n}$ |
| \vdots | | | | | | |
| u_m | | | | | | |

Figura 3 - Representação de uma matriz usuário-item
 Fonte: elaborada pelo autor.

Correlação baseada em itens

Diferentemente da abordagem baseada em usuário discutida anteriormente, a correlação baseada em item (*item-based*) analisa o conjunto de itens avaliados pelo usuário ativo u_a e calcula o quão similar eles são ao item i que se pretende predizer, com base em todas as avaliações que cada um deles recebeu.

Isso significa que para predizer ao usuário u_a o quanto ele pode gostar do item i , o sistema examina todos os itens já avaliados por u_a (conjunto I_a de itens) e verifica a correlação de cada um deles com o item i (Sarwar et al., 2001). A predição é, portanto, uma média das avaliações do usuário u_a , considerando o peso de cada item I_a em relação a i (w_{i,I_a}) (Shardanand; Maes, 1995).

O cálculo da similaridade pode ser implementado com base no algoritmo de correlação de Pearson, considerando apenas os casos de co-avaliação, isto é, para comparar o item i ao item j , consideram-se apenas os usuários que avaliaram ambos (ver a Figura 4).

| | i_1 | i_2 | \dots | i_i | i_j | \dots | i_n |
|-----------|-------|-------|---------|-------------|-------------|---------|-------|
| u_1 | | | | $v_{1,i}$ | $v_{1,j}$ | | |
| u_2 | | | | -- | $v_{2,j}$ | | |
| \vdots | | | | | | | |
| u_a | | | | $v_{a,i}$ | $v_{a,j}$ | | |
| \vdots | | | | | | | |
| u_{m-1} | | | | $v_{m-1,i}$ | $v_{m-1,j}$ | | |
| u_m | | | | $v_{m,i}$ | -- | | |

Figura 4 - Itens co-avaliados para medida de similaridade entre i_i e i_j
 Fonte: elaborada pelo autor.

Vetor de similaridade

O vetor de similaridade tem suas raízes nos sistemas de recuperação de informação (*Information Retrieval*). Como apresentado na técnica *TF-IDF*, a similaridade entre documentos é freqüentemente medida tratando cada documento como um vetor de freqüência de palavras, e computando o cosseno do ângulo formado por dois vetores de freqüência (Breese; Heckerman; Kadie, 1998). Analogamente, na filtragem colaborativa, os usuários podem ser vistos como documentos, os itens como palavras e as avaliações (ou votos) como a freqüência de palavras. Os itens não avaliados recebem voto zero. A similaridade de cada usuário em relação ao usuário ativo é então calculada como (Breese; Heckerman; Kadie, 1998):

$$w(a, u) = \sum_i \frac{v_{a,i}}{\sqrt{\sum_{k \in I_a} v_{a,k}^2}} \frac{v_{u,i}}{\sqrt{\sum_{k \in I_u} v_{u,k}^2}}, \quad (5)$$

sendo, I_a o conjunto de itens avaliados pelo usuário a , e I_u os itens avaliados pelo usuário u . A raiz do denominador serve para normalizar os votos, para que os usuários que votaram em mais itens não sejam *a priori* mais similares que outros. A normalização também pode ser feita de outras formas (Breese; Heckerman; Kadie, 1998).

Uma vez calculada a similaridade de cada usuário, a predição pode ser calculada com base na fórmula 3.

Diversas extensões propostas para melhorar o desempenho dos algoritmos baseados em memória são encontradas na literatura, entre as quais se destacam:

- voto padrão (*default voting*): um valor de voto padrão é atribuído aos itens não avaliados. Com isso, os usuários são comparados com base em um número maior de itens e a esparsidade da matriz usuário-item é reduzida (considerando esparsidade como os campos vazios da matriz) (Breese; Heckerman; Kadie, 1998);
- uso de *Filterbots*: agentes de *software* que avaliam automaticamente novos itens. São como usuários ordinários, diferem desses por gerarem mais avaliações e por não exigirem predições para si mesmos. Também contribuem para reduzir a esparsidade da matriz usuário-item (Vozalis; Margaritis, 2003);
- *inverse user frequency*: de forma análoga ao *IDF*, a idéia é que os itens extremamente populares não sejam tão úteis para avaliar a similaridade entre os usuários quanto os itens menos comuns (Breese; Heckerman; Kadie, 1998);
- amplificação de caso: tem por objetivo valorizar as similaridades de valor mais alto e punir as de valores mais baixos, usando para isso uma transformada (Breese; Heckerman; Kadie, 1998).

2.4.2 Técnicas baseadas em modelo

As técnicas baseadas em modelo, na filtragem colaborativa, podem ser vistas como o cálculo do valor esperado de um voto a partir de informações conhecidas sobre o usuário, utilizando uma perspectiva probabilística. Assumindo que os votos têm valor inteiro m_i , variando de 0 a m , tem-se:

$$P_{a,i} = E(v_{a,i}) = \sum_{m_i=0}^m m_i \times \Pr(v_{a,i} = m_i | v_{a,k}, k \in I_a), \quad (6)$$

onde a expressão apresenta a probabilidade do usuário ativo a avaliar o item i com o valor de voto m_i , dado que são conhecidos os votos de a em outros itens (I_a) já avaliados por ele.

A seguir descrevem-se algumas alternativas de modelos probabilísticos para filtragem colaborativa.

Modelo cluster

Os algoritmos de *clustering* encontram-se entre os sistemas baseados em aprendizagem não supervisionada, ou seja, aqueles utilizados quando não se conhece a classe a que pertencem os padrões de treinamento – de forma contrária, na aprendizagem supervisionada, a relação entre os padrões e classes é conhecida (Martín Guerrero, 2004).

Breese, Heckerman e Kadie (1998) apresentam como modelo *cluster* para filtragem colaborativa um classificador Bayesiano com enfoque ingênuo (*naive Bayes*). A idéia é que existem determinados tipos de usuários que compartilham um conjunto de preferências e gostos que podem ser agrupados em classes.

Os parâmetros do modelo, as probabilidades associadas à classe e as probabilidades condicionais dos votos, dada a classe, são estimadas a partir de um conjunto de treinamento de votos de usuários de uma base de dados. A predição do item i para o usuário ativo a , utiliza uma fórmula básica dos métodos baseados em modelo, como mostra a fórmula 6.

Na literatura sobre modelos *cluster* aplicados a sistemas de recomendações, destacam-se: Ungar e Foster (1998) e Martín Guerrero (2004), que comparam diversos algoritmos de clusterização; O'Connor e Herlocker (1999), que ressaltam o ganho de escalabilidade proporcionado pela técnica; Kohrs e Merialdo (1999), que focam no problema da esparsidade e confirmam a eficiência da técnica.

Modelo rede bayesiana

Uma formulação alternativa do modelo probabilístico para filtragem colaborativa é a rede Bayesiana, com os nós correspondendo aos itens e o estado de cada nó correspondendo ao valor do voto (Breese; Heckerman; Kadie, 1998). Essa abordagem se baseia na utilização de uma rede Bayesiana distinta para cada usuário que se deseja fazer a predição (Vozalis; Margaritis, 2003).

Para treinar a rede aplica-se um algoritmo de aprendizagem de redes Bayesianas aos dados de treinamento. Esse procura vários modelos de estrutura em termos de dependência de cada item. Na rede resultante, cada item terá um conjunto de itens “pais” (predecessores) de maior relação com ele, formando assim, a base para a predição. Cada tabela de probabilidade condicional é representada por uma árvore de decisão que codifica as probabilidades condicionais para um dado nó.

Dado um usuário ativo a , suas avaliações para os itens do domínio são utilizadas como entrada para a rede. Com base nesses dados, e após a rede ter sido treinada, é possível fazer a predição de um determinado item para esse usuário por meio de uma rede Bayesiana correspondente ao seu perfil.

Os algoritmos de redes Bayesianas do tipo árvore de decisão e sua generalização - conhecida como grafo de decisão -, são discutidos por Chickering et al. (1997).

Algoritmos de aprendizagem supervisionada

A recomendação de itens ao usuário, com base na avaliação feita por outros usuários, pode ser vista como um problema de classificação. Este pode ser solucionado utilizando-se técnicas de extração de características combinadas com algoritmos de aprendizagem de máquina (Billsus; Pazzani, 1998).

Baseado em um conjunto de avaliações de usuários em itens, Billsus e Pazzani (1998) propõem induzir um modelo para cada usuário com a finalidade de classificar itens não avaliados em duas ou mais classes C , por exemplo, $C_1 =$ gosta e $C_2 =$ não gosta.

A proposta inicial dessa abordagem é transformar o conjunto de dados das avaliações, correspondente à matriz usuário-item (Figura 5), em um formato que algoritmos de aprendizagem supervisionada possam operar, a fim de induzir uma função $f: E \rightarrow C$, e classificar os itens na classe apropriada (Billsus; Pazzani, 1998).

| | I_1 | I_2 | I_3 | I_4 | I_5 |
|-------|-------|-------|-------|-------|-------|
| U_1 | 4 | | 3 | | |
| U_2 | | 1 | | 2 | |
| U_3 | 3 | 4 | 2 | | 4 |
| U_4 | 4 | 2 | 1 | | ? |

Figura 5 - Matriz usuário-item

Fonte: Traduzido de Billsus e Pazzani (1998: 48).

A Figura 6 mostra o resultado de uma transformação das avaliações dos usuários em uma estrutura de representação *booleana*, de tal forma que um algoritmo de aprendizagem supervisionada possa tratar os dados a fim de realizar uma predição para o usuário U_4 . Nota-se que as avaliações 1 e 2 são associadas à classe “não gosta”, enquanto as de valor 3 e 4 correspondem à classe “gosta”.

| | E_1 | E_2 | E_3 |
|-----------------|-------|-----------|-----------|
| U_1 gosta | 1 | 0 | 1 |
| U_1 não gosta | 0 | 0 | 0 |
| U_2 gosta | 0 | 0 | 0 |
| U_2 não gosta | 0 | 1 | 0 |
| U_3 gosta | 1 | 1 | 0 |
| U_3 não gosta | 0 | 0 | 1 |
| Classe | gosta | não gosta | não gosta |

Figura 6 - Matriz transformada

Fonte: Traduzido de Billsus e Pazzani (1998: 48).

As avaliações do usuário 4 nos itens I_1 , I_2 e I_3 são utilizadas como exemplos de treinamento, sendo representados como vetores de características E , onde os usuários correspondem às características (Billsus; Pazzani, 1998).

Antes de aplicar um algoritmo de aprendizagem supervisionada é preciso uma etapa de pré-processamento que reduza a esparsidade da matriz, considerando esparsa a matriz com muitos elementos zero. Para isso, geralmente é computada a Decomposição em Valores Singulares (SVD) da matriz resultante (Vozalis; Margaritis, 2003).

Os experimentos realizados por Billsus e Pazzani (1998) mostram que a abordagem combinando a técnica de extração de característica (SVD) e o algoritmo de aprendizagem de máquina (Redes Neurais Artificiais), oferece ganhos significativos em desempenho e precisão, embora seja computacionalmente mais dispendiosa que outras técnicas, como a correlação.

Na próxima seção discute-se a avaliação das recomendações.

2.5 Avaliação das Recomendações

Uma etapa fundamental ao desenvolvimento de *software* é a validação do sistema. Essa consiste em realizar testes com o objetivo de identificar erros e garantir a qualidade do produto.

Para os sistemas de recomendação, é importante avaliar a qualidade das recomendações para atestar o bom funcionamento do recomendador *Web*. A grande dificuldade dos pesquisadores é avaliar recomendações de um sistema em seu estágio inicial de funcionamento; quando ainda não existem dados suficientes para a realização de testes.

Uma idéia para validar sistemas de recomendação no estágio inicial é fazer uma avaliação piloto. Ricci e Del Missier (2004) utilizaram essa estratégia para avaliar o protótipo *NutKing* de sua autoria. A avaliação focou na análise da interação usuário-sistema e se limitou a 35 casos gerados por dois grupos distintos: um de especialistas e outro de estudantes.

Outra idéia para validar o sistema antes deste ser colocado em funcionamento é sugerida por Martín Guerrero (2004). O autor utiliza um simulador de acessos para testar o agrupamento de usuários e a efetividade das recomendações de seu recomendador *Web*. Diferente de Ricci e Del Missier (2004), sua metodologia não considera a influência da interface.

Quando o sistema de recomendação é colocado em funcionamento e acumula uma significativa massa de dados (por exemplo, usuários e suas avaliações), um número diverso de métricas são propostas para avaliar seu desempenho (Vozalis; Margaritis, 2003). De maneira geral, os dados acumulados são divididos em: (1) conjunto de treinamento, que representa os dados prévios do sistema e (2) conjunto de teste, que simula novas entradas de dados, permitindo comparar a predição feita pelo sistema (r_{ij}) com a avaliação efetivamente realizada pelo usuário (ur_{ij}) (Burke, 2002).

Uma métrica normalmente usada para avaliar a precisão da predição é o Erro Absoluto Médio (*Mean Absolute Error – MAE*) (Shardanand; Maes, 1995; Vozalis; Margaritis, 2003). Sendo r_{ij} a predição do item j para o usuário u_i calculada pelo sistema, e ur_{ij} a avaliação de fato feita pelo usuário u_i sobre o item j . Sendo n_i o número total de itens avaliados pelo usuário u_i . O *MAE* para o usuário u_i pode ser computado como (Vozalis; Margaritis, 2003):

$$MAE_i = \frac{\sum_{j=1}^{n_i} |ur_{ij} - r_{ij}|}{n_i} \quad (7)$$

O *MAE* total também pode ser calculado pela média dos erros de cada usuário (Vozalis; Margaritis, 2003).

Ressalta-se que quanto menor o erro absoluto médio, maior é a precisão da recomendação (Shardanand; Maes, 1995). A propósito, é possível ainda calcular o desvio padrão do *MAE*, sendo que o baixo valor do desvio padrão sinaliza que o esquema de precisão é mais consistente (Shardanand; Maes, 1995).

Outra medida de qualidade, para avaliar um sistema, é a *cobertura*. Essa representa a porcentagem de itens que o sistema é capaz de computar a predição, uma vez que os sistemas de recomendação podem não conseguir gerar predição para itens específicos (Shardanand; Maes, 1995; Vozalis; Margaritis, 2003).

Vozalis e Margaritis (2003) apresentam ainda métricas de avaliações específicas para *Recomendações Top-N*, tais como *Recall-Precision* e *Expected Utility* (*utilidade esperada*), assim como medidas para avaliar o desempenho computacional dos recomendadores: *tempo de resposta*, *memória requerida* e *complexidade computacional*.

A seção 3 sintetiza os aspectos abordados em uma estrutura metodológica para implementação de sistemas de recomendação.

3 Estrutura metodológica

Com base nos aspectos discutidos, uma estrutura metodológica para implementação de um sistema de recomendação é proposta. Esta estrutura, apresentada em

cinco etapas, representa escolhas e definições referentes às características desejáveis para o sistema.

- 1) *Escolha da estratégia de recomendação*: consiste em definir o tipo de saída que se espera do sistema de recomendação (predição, recomendação *Top-N* ou outras estratégias adotadas por lojas virtuais). A resposta desejada para o sistema está diretamente relacionada ao problema que se pretende resolver, por isso deve ser definida antes de se partir para as próximas etapas.
- 2) *Escolha do tipo e técnica de recomendação*: conhecendo os pontos fortes e fracos de cada tipo e técnica de recomendação deve-se selecionar aqueles mais adequados à resolução do problema. Os tipos de recomendação podem ser baseados em conteúdo, colaborativo, demográfico, baseado em utilidade, baseado em conhecimento ou híbrido e, as técnicas, baseadas em memória e em modelo. Esta etapa é concluída com a definição do algoritmo a ser implementado.
- 3) *Determinação do modelo do usuário*: após a escolha do tipo e técnica de recomendação é possível definir os dados de entrada que serão utilizados pelo sistema: avaliação implícita, avaliação explícita por voto declarado ou formulário. Com a definição dos dados de entrada o modelo do usuário deve ser estabelecido.
- 4) *Definição do processo de recomendação*: deve-se definir a forma como os dados de entrada e os dados armazenados serão combinados pelos algoritmos para gerar recomendações. A partir disso, o processo de recomendação pode ser representado por meio de uma arquitetura lógica.
- 5) *Avaliação das recomendações*: uma vez implementado, o sistema deve ser avaliado para confirmar a eficiência de suas recomendações.

4 Conclusões e Recomendações

A área de sistemas de recomendação constitui um campo prolífico de pesquisas e aplicações voltadas para a personalização de *sites*. Sua utilização oferece ganhos de competitividade para as empresas que possuem negócios baseados na Internet, assim como facilita a busca de conteúdos e serviços, economizando tempo e esforço do usuário.

Este artigo oferece noções gerais sobre essa área. A análise da literatura permitiu elaborar uma estrutura metodológica para implementação de sistemas de recomendação. Acredita-se que essa estrutura e o levantamento realizado possam ser úteis a desenvolvedores e pesquisadores iniciantes neste campo de estudo.

Pretende-se validar a metodologia proposta em um sistema de recomendação de meios de hospedagem. Por fim, sugere-se que a metodologia seja utilizada para implementação de outros sistemas de recomendação e que as experiências sejam relatadas, contribuindo para a validação da estrutura proposta.

Referências Bibliográficas

ADOMAVICIUS, Gediminas; TUZHILIN, Alexander. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. **IEEE Transactions on Knowledge and Data Engineering**, Piscataway, v. 17, n. 6, p. 734-749, jun. 2005.

BALABANOVIĆ, Marko; SHOHAM, Yoav. Fab: Content-based, collaborative recommendation. **Communications of the ACM**, New York, v.40, n. 3, p. 66–72, março 1997.

BEZERRA, Byron et al. Speeding up Recommender Systems with Meta-prototypes. In: **Proceedings of the 16th Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence**, November 11-14, 2002, p. 227-236.

BILLSUS, Daniel; PAZZANI, Michael J. **User Modeling for Adaptive News Access**. User-Modeling and User-Adapted Interaction 10 (2-3), 2000, p. 147-180.

BILLSUS, Daniel; PAZZANI, Michael J. Learning Collaborative Information Filters Proc. In: **Proceedings of The Fifteenth International Conference On Machine Learning**. Madison, WI: Morgan Kaufmann Publishers Inc., jul. 1998, p. 46- 54. Disponível em: <<http://citeseer.ist.psu.edu/billsus98learning.html>>. Acesso em: 26 dez. 2006.

BREESE, John S.; HECKERMAN, David; KADIE, Carl. Empirical analysis of predictive algorithms for collaborative filtering. In: UAI 98. **Proceedings of The 14th Conference On Uncertainty in Artificial Intelligence**. Madison, WI: Morgan Kaufmann, jul. 1998, p. 43-52. Disponível em: <<http://citeseer.ist.psu.edu/breese98empirical.html>>. Acesso em: 02 jan. 2007.

BURKE, Robin. Hybrid recommender systems: Survey and experiments. **User Modeling and User Adapted Interaction**, v.12, n. 6, p. 331–370, nov. 2002. Disponível em: <<http://josquin.cti.depaul.edu/~rburke/pubs/burke-umuai02.pdf>>. Acesso em: 28 dez. 2006.

CHICKERING, David Maxwell; HECKERMAN, David; MEEK, Christopher. A Bayesian approach to learning Bayesian networks with local structure. In: **Proceedings of UAI-97**, Providence, RI, Morgan Kaufmann, p. 80-89, Aug. 1997.

CLAYPOOL, Mark et al. Combining Content-Based and Collaborative Filters in an Online Newspaper. In: **Proceedings of ACM SIGIR Workshop on Recommender Systems**, Aug. 1999.

DESHPANDE, Mukund; KARYPIS, George. Item-Based Top-N Recommendation Algorithms. **ACM Transactions on Information Systems**. New York, v. 22, n. 1, p. 143-177, jan. 2004.

GOLDBERG, David et al. Using Collaborative Filtering to Weave an Information Tapestry Comm. **Communications of the ACM**. New York, v. 35, n. 12, p. 61-70, dec.1992.

GOOD, Nathaniel et al. Combining collaborative filtering with personal agents for better recommendations. In: **Proceedings of AAAI-99**, AAAI Press. p. 439–446, 1999.

HILL, Will, et al. Recommending and evaluating choices in a virtual community of use. In: **CHI '95: Conference Proceedings on Human Factors in Computing Systems**, Denver, CO, p. 194-201, 1995.

KOHRN, Arnd; MERIALDO, Bernard. Clustering for collaborative filtering applications. In: **Computational Intelligence for Modelling**, Control Automation. IOS Press, 1999.

KONSTAN, Joseph et al. Recommender Systems: A GroupLens Perspective. In: **Recommender Systems: Papers from the 1998 Workshop (AAAI Technical Report WS-98-08)**. Menlo Park, CA: AAAI Press, 1998, p. 60 – 64.

MARTÍN GUERRERO, José David. **Determinación de tendencias en un portal web utilizando técnicas no supervisadas**: Aplicación a sistemas de recomendaciones basados en filtrado colaborativo. 2004. 210 f. Tese (Doutorado em Engenharia Eletrônica) - Departament d'Enginyeria Electrònica, University of València, Espanha, 2004. Disponível em: <http://www.uv.es/jdmg/tesis_jdmartin.pdf>. Acesso em: 15 jan. 2007.

NICHOLS, David. M. Implicit Rating and Filtering. In: **Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering**, Budapeste, Hungria: ERCIM, 1998, p. 31-36. Disponível em: <<http://www.ercim.org/publication/ws-proceedings/DELOS5/nichols.pdf>>. Acesso em: 10 jan. 2007.

O'CONNOR, Mark; HERLOCKER, Jon. Clustering items for collaborative filtering. In: **Proceedings of the ACM SIGIR Workshop on Recommender Systems**, Berkeley, CA, 1999.

O'DONOVAN, John; SMYTH, Barry. Trust in recommender systems. In: INTERNATIONAL CONFERENCE ON INTELLIGENT USER INTERFACES, 2005, San Diego. **Proceedings of the 10th international conference on Intelligent user interfaces**. Nova York: ACM Press, 2005, p. 167-174.

PENNOCK, David. M. et al. Collaborative Filtering by Personality Diagnosis: A Hybrid Memory And Model-Based Approach. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 1999, Estocolmo. **Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence**. São Francisco: Morgan Kaufmann, 2000, p. 473-480. Disponível em: <<http://dpennock.com/papers/pd-uai-00.pdf>>. Acesso em: 15 jan. 2007.

REATEGUI, Eliseo Berni; CAZELLA, Sílvio César. Sistemas de Recomendação. In: XXV CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2005, São Leopoldo. **Anais do Encontro Nacional de Inteligência Artificial**. São Leopoldo, 2005, p. 306 - 348. Disponível em: <<http://www.sbc.org.br/bibliotecadigital/download.php?paper=415>>. Acesso em: 25 dez. 2006.

RESNICK, Paul; VARIAN, Hal. R. Recommender Systems. **Communications of the ACM New York** v. 40, n. 3, março 1997, p. 56-58.

RESNICK, Paul et al. Grouplens: An open architecture for collaborative filtering of netnews. In: COMPUTER SUPPORTED COOPERATIVE WORK, 1994, Chapel Hill. **Proceedings of the 1994 ACM conference on Computer supported cooperative work**. New York: ACM Press, 1994, p. 175–186.

RICCI, Francesco; DEL MISSIER, Fabio. Supporting Travel Decision Making Through Personalized Recommendation. **Designing personalized user experiences in eCommerce**, Kluwer Academic Publishers, Norwell, MA, 2004. Disponível em: <<http://sra.itc.it/tr/RD04.pdf>>. Acesso em: 12 dez. 2006.

RICCI, Francesco et al. ITR: A case-based travel advisory system. In: 6TH EUROPEAN CONFERENCE ON CASE BASED REASONING, 2002, Aberdeen. **Proceedings of the seventh European Conference on Case Based Reasoning**. Heidelberg: Springer Berlin, 2002, p. 613-627. Disponível em: <<http://dietorecs.itc.it/Papers/eccbr2002final.pdf>>. Acesso em: 25 nov. 2006.

SARWAR, Bradul et al. Item-Based Collaborative Filtering Recommendation Algorithms Proc. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 2001, Hong Kong. **Proceedings of the 10th international conference on World Wide Web**. New York: ACM Press, 2001, p. 285-295.

SCHAFFER, J. Ben; KONSTAN, Joseph. A.; RIEDL, John. E-Commerce Recommendation Applications. **Data Mining and Knowledge Discovery**, Hingham, v. 5, n. 1-2, p. 115-153, jan/abril 2001. Disponível em: <http://www.grouplens.org/papers/pdf/ECRA.pdf>>. Acesso em: 25 out. 2006.

SHARDANAND, Uprenda; MAES, Pattie. Social information filtering: Algorithms for automating “word of mouth.” In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 1995, Denver. **Proceedings of the SIGCHI conference on Human factors in computing systems**. New York: ACM Press/Addison-Wesley Publishing Co., 1995, p. 210–217. Disponível em: <<http://www.cs.ubc.ca/~conati/532b/papers/chi-95-paper.pdf>>. Acesso em: 30 nov. 2006.

UNGAR, Lyle H.; FOSTER, Dean P. Clustering Methods for Collaborative Filtering. In: **Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence**. 1998.

VOZALIS, Emmanouil; MARGARITIS, Konstantinos G. Analysis of recommender systems algorithms. In: **Proceedings of the Sixth Hellenic-European Conference on Computer Mathematics and its Applications - HERCMA 2003**, 2003. Disponível em: <<http://macedonia.uom.gr/~mans/papiria/hercma2003.pdf>>. Acesso em: 10 nov. 2006.