

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
ENGENHARIA E GESTÃO DO CONHECIMENTO**

**LEONARDO GOMES DE OLIVEIRA**

**SISTEMA DE RECOMENDAÇÃO DE MEIOS DE HOSPEDAGEM  
BASEADO EM FILTRAGEM COLABORATIVA E INFORMAÇÕES  
CONTEXTUAIS**

**Florianópolis**

**2007**

**LEONARDO GOMES DE OLIVEIRA**

**SISTEMA DE RECOMENDAÇÃO DE MEIOS DE HOSPEDAGEM  
BASEADO EM FILTRAGEM COLABORATIVA E INFORMAÇÕES  
CONTEXTUAIS**

Dissertação submetida à Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Engenharia e Gestão do Conhecimento.

Orientador:  
Prof. Dr. Rogério Cid Bastos.

Co-orientador:  
Prof. Dr. Neri dos Santos.

**Florianópolis**

**2007**

**LEONARDO GOMES DE OLIVEIRA**

**SISTEMA DE RECOMENDAÇÃO DE MEIOS DE HOSPEDAGEM  
BASEADO EM FILTRAGEM COLABORATIVA E INFORMAÇÕES  
CONTEXTUAIS**

Esta dissertação foi julgada e aprovada para a obtenção do grau de **Mestre em Engenharia e Gestão do Conhecimento** no **Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento** da Universidade Federal de Santa Catarina.

Florianópolis, 29 de maio de 2007.

---

Prof. Roberto Pacheco, Dr.  
Coordenador do Programa - UFSC

**Banca Examinadora:**

---

Orientador Prof. Rogério Cid Bastos, Dr.  
Engenharia e Gestão do Conhecimento – UFSC

---

Co-Orientador Prof. Neri dos Santos, Dr.  
Engenharia e Gestão do Conhecimento – UFSC

---

Prof. Fernando Alvaro Ostuni Gauthier, Dr.  
Engenharia e Gestão do Conhecimento – UFSC

---

Prof. Luis Alberto Gómez, Dr.  
Departamento de Engenharia Civil – UFSC

## AGRADECIMENTOS

*A todas as pessoas que lutam para que o Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da UFSC seja um curso de excelência, pois com isso estão gerando oportunidade de aprendizado de alto nível para centenas de alunos em uma universidade pública federal.*

*Ao professor Eduardo Meditsch por ter acreditado no meu projeto inicial de mestrado, aceitando ser meu orientador, e por ter me apoiado quando decidi mudar de linha de pesquisa. A sua participação foi indispensável para minha formação.*

*Ao professor Rogério Cid Bastos por ter aceitado ser meu orientador neste projeto de pesquisa. Suas orientações, sempre muito objetivas e precisas, foram extremamente relevantes e úteis para a elaboração dessa dissertação.*

*Ao professor Mauro Roisenberg, por ter sido tão prestativo nas conversas que tivemos sobre meu projeto.*

*À minha prima Márcia Barros, que me incentivou a ingressar no mestrado e me apoiou em diversas circunstâncias.*

*À Valdenise Schmitt, que começou como revisora e se tornou minha tutora de orientação. Nossas inúmeras discussões contribuíram muito para a lapidação desse material.*

*A todos que direta ou indiretamente me ajudaram com a dissertação, especialmente Gean, Thaís, Mário, Cinthia, Paula, Anderson, Marília, George, Demetrius, Gustavo e Maurélio.*

*À contribuição inestimável da Juliana, minha namorada. Sua presença tornou essa caminhada menos árdua. Seu apoio foi indispensável para que eu atingisse meu objetivo.*

*À minha família, que mesmo distante sempre foi muito presente em minha vida.*

*À Deus por permitir a realização desse trabalho.*

*“Ora, o que é a vida? É uma espécie de comédia contínua em que os homens, disfarçados de mil maneiras diferentes, aparecem em cena, desempenham seus papéis, até que o diretor, depois de tê-los feito mudar de disfarce e aparecer ora sob a púrpura soberba dos reis, ora sob os andrajos repulsivos da escravidão e da miséria, força-os finalmente a sair do palco. Em verdade, este mundo não é senão uma sombra passageira, mas assim é a comédia que nele representamos todos os dias”*

(Erasmus de Rotterdam – Elogio da loucura)

## RESUMO

OLIVEIRA, Leonardo Gomes. *Sistema de recomendação de meios de hospedagem baseado em filtragem colaborativa e informações contextuais*, 2007, 112 p. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, UFSC, Florianópolis.

Este trabalho apresenta um sistema de recomendação de meios de hospedagem, que utiliza filtragem colaborativa e informações contextuais, com a finalidade de apresentar resultados personalizados aos usuários que buscam informações na Internet para planejarem suas viagens. A concepção e o desenvolvimento do sistema seguem um procedimento metodológico proposto a partir de uma revisão de literatura. Considerando a complexidade e as particularidades do domínio de aplicação (viagens e turismo), a filtragem colaborativa não é utilizada com sua formulação básica, mas com uma nova abordagem que permite oferecer ao usuário recomendações de hotéis que pessoas com perfil e contexto de viagem similares ao seu preferiram no passado. Para isso, variáveis contextuais são identificadas e incorporadas ao sistema. O sistema proposto foi implementado na agência Tropix Turismo Inteligente, onde uma avaliação preliminar foi realizada. Os experimentos consistiram na comparação entre duas técnicas colaborativas (co-seno e ‘similaridade heurística’) e a recomendação dos hotéis mais populares (‘POP’). Os resultados mostraram que as técnicas colaborativas são mais eficientes que a exibição dos itens mais populares, pois permitem a apresentação de recomendações personalizadas de acordo com o perfil e o contexto de viagem do usuário. A aplicação de sistemas de recomendação em *sites* de turismo proporciona benefícios aos usuários, por facilitar suas buscas e apoiar suas decisões. E também contribui para a competitividade das empresas do setor, por meio da fidelização do cliente e o conseqüente aumento das vendas.

**Palavras-chave:** Sistemas de Recomendação. Filtragem Colaborativa. Personalização. Informações Contextuais. Planejamento de Viagem. Filtragem de Informação.

## ABSTRACT

*OLIVEIRA, Leonardo Gomes. Sistema de recomendação de meios de hospedagem baseado em filtragem colaborativa e informações contextuais, 2007, 112 p. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, UFSC, Florianópolis.*

*This study presents an accommodation recommender system based on collaborative filtering and contextual information in order to offer customized results for users who look for tips on the Internet to plan their trips. The system conception and development follow the methodological structure designed by the literature review. When the complexity and the particularities of the application domain (travels and tourism) are considered, a new approach of the collaborative filtering is used instead of its basic formulation. This new approach offers the user recommended hotels which were, at some time in the past, preferred by people with similar profiles who went to the same traveling contexts. This is possible when contextual variables are identified and incorporated to the system itself. The proposed system is implemented at the Tropix Turismo Inteligente travel agency, where a preliminary evaluation is carried on. The experiments consist of the comparison between two collaborative techniques (the co-seno and the 'heuristic similarity') and the recommendations of the most popular hotels ('POP'). Results show that collaborative techniques are more efficient than the exhibition of the most popular items since they allow the presentation of customized recommendations according to the user's profile and travel context. The application of recommender systems in travel websites are beneficial to the users for they make their search easier and help them make decisions. It also contributes to the competitiveness of the companies in this sector by increasing client loyalty and consequently raising the sales.*

**Keywords:** *Recommender Systems. Collaborative Filtering. Personalization. Context-aware Recommender Systems. Travel Planning. Information Filtering.*

## LISTA DE FIGURAS

Figura 1 – Recomendações personalizadas feitas por uma loja virtual.....	22
Figura 2 – Processo de recomendação.....	23
Figura 3 – Exemplo de avaliação explícita.....	24
Figura 4 – Lista dos eletrônicos mais vendidos.....	26
Figura 5 – Recomendações personalizadas .....	27
Figura 6 – Recomendação por associação.....	27
Figura 7 – Representação de uma matriz de avaliações .....	42
Figura 8 – Itens co-avaliados para medida de similaridade entre $i_i$ e $i_j$ .....	43
Figura 9 – Representação de avaliações de usuários.....	50
Figura 10 – Matriz transformada.....	50
Figura 11 – Tela de recomendação do <i>MovieLens</i> .....	52
Figura 12 – Tela de recomendação do <i>What to Rent</i> .....	53
Figura 13 – Tela inicial do <i>Entree</i> .....	54
Figura 14 – Tela de recomendação de restaurante .....	54
Figura 15 – Interface do Pandora .....	56
Figura 16 – Representação de uma matriz usuário-item .....	70
Figura 17 – Representação do SRMH.....	73
Figura 18 – Tela da Tropix Turismo Inteligente – Resultado da consulta .....	76
Figura 19 – Tela de consulta do SRMH .....	78
Figura 20 – Recomendações do SRMH .....	79
Figura 21 – Página com informações detalhadas do hotel .....	80

## LISTA DE QUADROS

Quadro 1 – Tipos potenciais de avaliações implícitas.....	25
Quadro 2 – Vantagens dos tipos de filtragem.....	37
Quadro 3 – Desvantagens dos tipos de filtragem.....	37
Quadro 4 – Matriz de similaridade para a faixa etária .....	68
Quadro 5 – Variáveis de entrada .....	71
Quadro 6 – Representação do usuário como vetor multi-dimensional.....	72
Quadro 7 – Modelo de usuário – vetor multi-dimensional.....	72
Quadro 8 – Variáveis de entrada .....	77
Quadro 9 – Representação do usuário .....	78
Quadro 10 – Classificação hoteleira.....	106
Quadro 11 – Variáveis selecionadas para o SRMH .....	108
Quadro 12 – Matriz de similaridade para variável Faixa etária .....	109
Quadro 13 – Matriz de similaridade para variável Preço baixo .....	109
Quadro 14 – Matriz de similaridade para variável Motivação .....	109
Quadro 15 – Matriz de similaridade para variável Grupo .....	110
Quadro 16 – Matriz de similaridade para variável Transporte.....	110
Quadro 17 – Matriz de similaridade para variável Período.....	110
Quadro 18 – Perfis de turistas para avaliação do SRMH .....	111
Quadro 19 – Hotéis disponíveis para recomendação .....	112
Quadro 20 – Recomendações dos especialistas.....	112

## **LISTA DE TABELAS**

Tabela 1 – Resultados obtidos com o conjunto de teste 1 .....	85
Tabela 2 – Resultados obtidos com o conjunto de teste 2 .....	86
Tabela 3 – Resultados obtidos com o conjunto de teste 3 .....	87

## **LISTA DE GRÁFICOS**

Gráfico 1 – Resultados obtidos com o conjunto de teste 1 .....	86
Gráfico 2 – Resultados obtidos com o conjunto de teste 2 .....	86
Gráfico 3 – Resultados obtidos com o conjunto de teste 3 .....	87

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>12</b>
1.1	Objetivos.....	14
1.2	Justificativas .....	14
1.3	Metodologia.....	16
1.4	Escopo do trabalho .....	17
1.5	Estrutura do trabalho .....	18
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA: SISTEMAS DE RECOMENDAÇÃO .....</b>	<b>19</b>
2.1	Conceito, origem e aplicações.....	19
2.2	Taxonomia.....	22
2.2.1	Dados de entrada .....	23
2.2.2	Saídas.....	25
2.2.3	Tipos de filtragem.....	28
2.2.4	Técnicas de recomendação .....	38
2.3	Exemplos de sistemas de recomendação .....	51
2.3.1	<i>MovieLens</i> .....	51
2.3.2	<i>What to Rent</i> .....	52
2.3.3	<i>Entree</i> .....	53
2.3.4	Pandora.....	55
2.4	Avaliação das recomendações.....	56
2.5	Sistemas de recomendação em turismo.....	58
2.5.1	Teoria de decisão em viagem .....	59
2.6	Considerações finais sobre o capítulo .....	60
<b>3</b>	<b>PROCEDIMENTO METODOLÓGICO: SISTEMA DE RECOMENDAÇÃO DE MEIOS DE HOSPEDAGEM .....</b>	<b>62</b>
3.1	Caracterização da pesquisa.....	62
3.2	Passos da pesquisa.....	63
3.2.1	Escolha da saída .....	64
3.2.2	Escolha do tipo de filtragem e técnica de recomendação.....	64
3.2.3	Determinação do modelo de usuário .....	69
3.2.4	Definição do processo de recomendação.....	72
3.3	Considerações finais sobre o capítulo .....	74
<b>4</b>	<b>APLICAÇÃO DO SISTEMA.....</b>	<b>75</b>
4.1	Tropix Turismo Inteligente.....	75
4.2	Implementação .....	77
4.3	Avaliação das recomendações.....	80

4.3.1	Método de avaliação.....	82
4.3.2	Resultados.....	85
4.3.3	Considerações finais sobre o capítulo .....	88
<b>5</b>	<b>CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....</b>	<b>89</b>
5.1	Conclusões.....	89
5.2	Limitações e trabalhos futuros.....	90
	<b>REFERÊNCIAS .....</b>	<b>92</b>
	<b>APÊNDICE A .....</b>	<b>98</b>
	<b>APÊNDICE B.....</b>	<b>109</b>
	<b>APÊNDICE C .....</b>	<b>111</b>

# 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 *sites*<sup>1</sup>.

Em meio à profusão de informações e serviços on-line estão os usuários da Internet - atualmente mais de um bilhão de pessoas ao redor do mundo<sup>2</sup>. O crescente número de usuários tem motivado empresas e instituições a se lançarem no mundo virtual, ao passo que novos usuários entram na rede motivados pelos novos serviços oferecidos.

À medida que aumenta o volume de informações na *Web*, torna-se mais difícil para o usuário encontrar a “informação certa” no “tempo certo” (O’DONOVAN; SMYTH, 2005). A grande quantidade de conteúdo gera sobrecarga de informação, problema que vem sendo contornado pelo uso de sistemas de recomendação (O’DONOVAN; SMYTH, 2005; SHARDANAND; MAES, 1995).

Em um típico sistema de recomendação, pessoas provêem recomendações como entrada e o sistema as reúne e direciona para os destinatários apropriados (RESNICK; VARIAN, 1997).

Para Burke (2002), 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, pode ser considerado um sistema de recomendação.

Os sistemas de recomendação vêm sendo bastante utilizados em *sites* de comércio eletrônico, que no intuito de aumentar a lucratividade, lançam mão de diferentes técnicas para oferecer aos clientes os produtos mais adequados ao seu perfil (REATEGUI; CAZELLA, 2005).

---

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

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

Considerando que tais sistemas precisam ser aprimorados para serem mais eficazes e utilizados em uma maior gama de aplicações práticas – incluindo recomendações de viagens e certos tipos de serviços financeiros para investidores –, Adomavicius e Tuzhilin (2005) sugerem que sejam incorporadas informações contextuais ao processo de recomendação e métodos mais adequados para a representação do perfil do usuário e da informação sobre os itens a serem recomendados.

Segundo Delgado e Davidson (2002), os *sites* de viagem têm sido extremamente eficientes ao aceitar transações eletrônicas, permitindo que os usuários comprem diretamente, com o mínimo de intervenção humana, serviços de transporte e hospitalidade. No entanto, esses mesmos *sites* têm sido relativamente pobres em auxiliar os usuários durante a fase de planejamento. Na maioria das vezes supõem que os usuários sabem exatamente onde querem ir, o que querem fazer e onde querem ficar.

Os sistemas de recomendação para viagem e hospitalidade são vistos por Delgado e Davidson (2002) como uma tentativa de simular a interação do usuário com agentes de viagens, cujo objetivo é prover recomendações de viagens customizadas para melhor atender às necessidades dos viajantes, baseadas nos seus interesses e preferências individuais.

Para esses autores, é importante considerar as informações contextuais em sistemas de recomendação voltados para o turismo, pois fatores como o período, distância, atividades e interesses individuais exercem forte influência nas decisões do turista.

Embora o comércio eletrônico em turismo seja responsável por grande parte do valor transacionado na Internet<sup>3</sup>, observa-se que poucos *sites* desse segmento oferecem recomendações personalizadas que considere o perfil do usuário e as particularidades de sua viagem. Diante disso, pode-se formular a seguinte pergunta de pesquisa: como desenvolver um sistema de recomendação de meios de hospedagem que ofereça recomendações personalizadas conforme o perfil do usuário e o contexto da sua viagem?

---

<sup>3</sup> Disponível em: <<http://www.camara-e.net/interna.asp?tipo=1&valor=2700>>. Acesso em: 19 mar. 2007.

## ***1.1 Objetivos***

O objetivo geral deste trabalho é desenvolver um sistema de recomendação de meios de hospedagem utilizando técnicas colaborativas e informações contextuais.

Para alcançar esse objetivo geral, são estabelecidos os seguintes objetivos específicos:

- identificar as principais características dos sistemas de recomendação, com foco nas escolhas e definições necessárias à sua implementação;
- propor um procedimento metodológico para implementação de sistemas de recomendação;
- discutir o desenvolvimento de um sistema de recomendação de meios de hospedagem, utilizando como base o procedimento metodológico proposto;
- identificar variáveis influentes que devem ser incorporadas ao sistema;
- implementar e avaliar o sistema de recomendação de meios de hospedagem.

## ***1.2 Justificativas***

Em meados dos anos 1990, os sistemas de recomendação despontaram como importante área de pesquisa, com a publicação dos primeiros artigos sobre filtragem colaborativa (ADOMAVICIUS; TUZHILIN, 2005).

Segundo Adomavicius e Tuzhilin (2005), o interesse nessa área continua alto. Primeiro, porque é um rico campo de pesquisa; segundo, porque existem inúmeras aplicações práticas que podem ajudar os usuários a lidar com a sobrecarga de informação, provendo a eles recomendações personalizadas de conteúdos e serviços.

No turismo, a competitividade entre as empresas e entre os diversos destinos turísticos em todo o mundo foi diretamente afetada pelo desenvolvimento das tecnologias da informação e comunicação (TICs), que têm contribuído para a mudança nos canais de distribuição dos produtos turísticos. Isso se explica por ser a atividade turística fortemente dependente da boa disseminação de informações e da distribuição e promoção eficazes do produto (BOGADO; TELES, 2003).

É com base na informação que o turista opta por visitar um determinado local, selecionar um pacote de viagem ou escolher um meio de hospedagem específico entre as opções existentes. Mas, ao mesmo tempo em que ela se torna facilmente acessível e abundante, o excesso de informação pode tornar o processo de decisão mais complexo e demorado, uma vez que o turista precisa processar uma grande quantidade de informação, das mais variadas fontes *on-line*, como por exemplo, os *sites* de viagem (DELGADO; DAVIDSON, 2002).

Para facilitar as escolhas dos usuários, alguns *sites* de viagem passaram a incorporar sistemas de recomendação que oferecem sugestões de serviços a eles, de acordo com seus gostos e interesses, facilitando assim sua busca e tomada de decisão. Delgado e Davidson (2002) afirmam que esses sistemas contribuem para a fidelização dos clientes e para o aumento nas vendas, já que com eles a taxa de conversão de vendas tende a ser maior (relação entre vendas efetivadas e número de consultas aos *sites*).

Grande parte dos sistemas de recomendação é voltada para recomendar filmes, livros, músicas, entre outros artigos comuns em lojas virtuais. Isso porque a filtragem colaborativa, que tem sido amplamente utilizada na abordagem de problemas de personalização na *Web*, pode ser aplicada facilmente a produtos que são comprados com frequência. No entanto, como observam Ricci e Del Missier (2004), a filtragem colaborativa baseada unicamente no histórico de avaliações do usuário não pode ser aplicada a produtos que são comprados esporadicamente, como é o caso de carros e viagens. Essa restrição ocorre porque o sistema não acumula uma lista apropriada de avaliações pessoais que possa fornecer conhecimento suficiente sobre as preferências do usuário, para assim predizer suas futuras escolhas.

Além do fator “frequência de compra”, é importante considerar outro aspecto extremamente relevante ao analisar o domínio de viagens: as inúmeras variáveis que são consideradas pelos turistas no processo de tomada de decisão.

Fesenmaier et al. (2003) sugerem que os desenvolvedores de sistemas de recomendação de viagem não considerem apenas as teorias das ciências da computação, mas também as pesquisas em teoria de decisão de viagem, área amplamente investigada em estudos sobre turismo (ver relatório técnico apresentado por FESENMAIER et al. 2002).

Embora seja crescente o número de *sites* que oferecem suporte ao viajante para a escolha de um destino de viagem ou serviço turístico (RICCI et al., 2002), poucos parecem

usar sistemas de recomendação para personalizar e aumentar a relevância dos resultados apresentados aos usuários.

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.

A abordagem utilizada para o sistema de recomendação de meios de hospedagem, apresentado neste trabalho, se baseia em técnicas colaborativas. Isso permite que ele “aprenda” com os usuários, aprimorando suas recomendações ao longo do tempo. Para contornar limitações intrínsecas à filtragem colaborativa, informações contextuais devem ser incorporadas ao processo de recomendação.

A condução desta pesquisa permite discutir a implementação de sistemas voltados à personalização de *sites* e, principalmente, de explorar as particularidades de um sistema de recomendação no domínio de viagens, considerando o uso de informações contextuais.

A aplicação do sistema de recomendação em turismo confere à pesquisa um caráter multidisciplinar, já que agrupa disciplinas diversas, de áreas como ciências da computação, turismo e engenharia e gestão do conhecimento. Esse aspecto contribui para o avanço das pesquisas e desenvolvimento de aplicações que se insiram em um ponto de convergência entre essas áreas de conhecimento.

### ***1.3 Metodologia***

A metodologia adotada neste trabalho é composta por três momentos sequenciais e complementares: contextualização, desenvolvimento e aplicação.

O primeiro momento abrange uma revisão bibliográfica, voltada a identificar as principais características dos sistemas de recomendação. Além de contextualizar a pesquisa, constitui o alicerce necessário para a definição de um procedimento metodológico para o desenvolvimento de um sistema de recomendação de meios de hospedagem.

Com relação aos objetivos propostos para realização desta pesquisa, pode-se classificar essa etapa como sendo uma pesquisa exploratória (GIL, 1991). Para o autor, a pesquisa exploratória é, de certa forma, uma característica de todo trabalho científico. Por meio dela é possível ter maior familiaridade com o problema, na intenção de torná-lo mais

explícito. A pesquisa exploratória tem por finalidade aprimorar idéias e levantar hipóteses sobre assuntos novos ou pouco explorados. Tipos mais comuns: pesquisa bibliográfica, documental e estudo de caso.

O segundo momento compreende a definição de um procedimento metodológico genérico para o desenvolvimento de sistemas de recomendação e, a partir deste, a concepção de um sistema de recomendação de meios de hospedagem, baseado em técnicas colaborativas e informações contextuais.

Conforme o método de abordagem utilizado, nessa etapa da pesquisa prevalece o método indutivo, caracterizado pelo processo de construção da teoria.

O terceiro e último momento da pesquisa corresponde à aplicação do sistema de recomendação de meios de hospedagem, visando analisar a sua adequação para a solução do problema. Neste momento, são descritos os resultados obtidos e as principais contribuições e limitações do sistema.

Nessa etapa adota-se a abordagem dedutiva, por se tratar do teste da teoria. Os procedimentos metodológicos utilizados para a avaliação do sistema são detalhados no capítulo 4.

#### ***1.4 Escopo do trabalho***

Este trabalho envolve o estudo dos sistemas de recomendação a partir de uma visão geral do tema, desde aspectos históricos e conceituais até exemplos de aplicação e levantamento das características de tais sistemas. Além disso, contempla os aspectos de implementação de um sistema de recomendação de meios de hospedagem, bem como apresenta um protótipo funcional do mesmo.

Acredita-se que o desenvolvimento do protótipo pode criar uma condição favorável à realização de experimentos futuros nesta área. Sendo assim, essa proposta enfatiza, sobretudo, os aspectos construtivos, as escolhas possíveis, entre outros fatores que devem ser considerados em um sistema voltado para aumentar a relevância dos resultados apresentados aos usuários.

O protótipo a ser implementado tem como base um *site* já existente, voltado para reservas *on-line* de meios de hospedagem. Por isso, este trabalho não tem por objetivo

levantar e cadastrar informações sobre os meios de hospedagem que serão recomendados, e também não se atem aos aspectos da comercialização em si (disponibilidade, preços, reservas, forma de pagamento, atendimento ao cliente etc.).

A avaliação do usuário, após usufruir os serviços contratados, não será considerada. Questões relacionadas à escalabilidade e exigência computacional também não são vistas como relevantes à realização da pesquisa.

### ***1.5 Estrutura do trabalho***

O conteúdo do presente estudo segmenta-se em cinco partes distintas e complementares, além desta introdução:

- O Capítulo 2 apresenta informações sobre os sistemas de recomendação, definindo seus tipos e apresentando as principais estratégias e técnicas aplicadas ao seu desenvolvimento. As particularidades dos sistemas de recomendação em turismo são descritas. Alguns exemplos de sistemas de recomendação ilustram este capítulo.
- O Capítulo 3 identifica e descreve um procedimento metodológico genérico para o desenvolvimento de sistemas de recomendação. Com base nele, discute-se o desenvolvimento de um sistema de recomendação de meios de hospedagem, apresentando tal sistema e descrevendo detalhadamente suas etapas.
- O Capítulo 4 trata da aplicação do sistema, apontando as variáveis selecionadas para sua implementação. Apresenta a empresa onde foi aplicado o sistema de recomendação, uma descrição da aplicação, o método de avaliação e os resultados obtidos.
- Por fim, no Capítulo 5, apresentam-se as conclusões do trabalho e sugere-se a realização de novas pesquisas consideradas úteis à complementação desta e ao aprofundamento do tema.

## **2 FUNDAMENTAÇÃO TEÓRICA: SISTEMAS DE RECOMENDAÇÃO**

Este capítulo oferece uma visão geral sobre os sistemas de recomendação. As seções que o compõem abordam: os aspectos conceituais, históricos e áreas de aplicação; uma taxonomia que considera as principais características desses sistemas; exemplos representativos de sua utilização, tanto pelo mercado quanto pelo meio acadêmico; a avaliação das recomendações e as particularidades dos sistemas de recomendação no domínio de viagens e turismo.

### ***2.1 Conceito, origem e aplicações***

Os sistemas de recomendação podem ser entendidos 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 (BURKE, 2002).

Em um típico sistema de recomendação, pessoas provêem 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).

A origem dos sistemas de recomendação pode ser traçada a partir de trabalhos em ciências cognitivas, teorias de aproximação, recuperação de informação e teorias de previsão, bem como ter vínculos com a área de gestão e modelos de decisão do consumidor (ADOMAVICIUS; TUZHILIN, 2005).

Combinando idéias como perfis de usuários, filtragem de informação e aprendizagem de máquina (O'DONOVAN; SMYTH, 2005), os sistemas de recomendação emergiram como uma área de pesquisa independente em meados dos anos 90, quando os pesquisadores começaram a focar em problemas de recomendação que dependem fundamentalmente das avaliações dos usuários (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).

Os desenvolvedores do Tapestry – proposto para resolver o problema da sobrecarga de e-mails –, cunharam o termo “filtragem colaborativa”, que logo passou a ser adotado por muitos outros pesquisadores. Anos mais tarde, Resnick e Varian (1997) sugeriram a utilização do termo “sistemas de recomendação” por entenderem que tais sistemas não necessariamente precisam utilizar colaboração e que, além de filtrarem, podem sugerir itens interessantes aos usuários.

Normalmente as pessoas confiam em indicações de terceiros para efetuarem suas escolhas. Baseiam-se, por exemplo, em recomendações de amigos, listas de filmes, livros e músicas apresentadas em jornais e revistas, e em guias de restaurante, turismo, entre outros. Para Shardanand e Maes (1995), os sistemas de recomendação essencialmente automatizam o tradicional “boca-a-boca”.

Automatizar o processo de indicação é uma busca constante dos *sites* de comércio eletrônico – atualmente considerados o principal foco de aplicação de sistemas de recomendação (REATEGUI; CAZELLA, 2005). Entre os fatores que justificam esse interesse destacam-se:

a) grande disponibilidade de produtos: nas lojas virtuais o leque de produtos ofertados aos usuários é muito superior ao que normalmente se encontra nas lojas físicas. Isso porque o custo para manter produtos na *Web* é mínimo, diferente das lojas físicas que têm limitações de espaço, tendo que privilegiar os itens de maior saída. Para exemplificar, enquanto uma loja da Livraria Cultura em um shopping tem capacidade para abrigar 120 mil livros, na Internet o catálogo da livraria é formado por cerca de 1,3 milhão de livros<sup>4</sup>.

b) acesso aos pequenos nichos de mercado: produtos considerados pouco populares representam boa parte das vendas das lojas virtuais por estarem acessíveis a pequenos, mas diversificados, segmentos de mercado. Na Amazon.com, por exemplo, cerca de 25% das vendas se situa fora dos 100 mil principais títulos (ANDERSON, 2006).

Anderson (2006) observa que um número muitíssimo grande (dos produtos que se

---

<sup>4</sup> Disponível em: <<http://revistaepoca.globo.com/Revista/Epoca/0,,EDR74764-6012,00.html>>. Acesso em: 10 de

situam na Cauda Longa, ou seja, aqueles que estão fora das listas dos mais vendidos) multiplicado por um número relativamente pequeno (os volumes de vendas de cada um) ainda é igual a um número muito grande.

A ampliação da oferta de bens e serviços e a facilidade de acesso a uma grande diversidade de nichos de mercado, têm levado as lojas virtuais a customizar a apresentação de produtos para os usuários, a fim de reduzir os “custos de busca”<sup>5</sup>, atender os interesses dos diversos perfis de consumidores, e com isso aumentar as vendas (SCHAFER; KONSTAN; RIEDL, 2001).

Segundo Schafer, Konstan e Riedl (2001), os sistemas de recomendação contribuem com o comércio eletrônico de três maneiras:

- **convertendo visitantes em compradores:** muitas vezes uma loja virtual é acessada por pessoas que estão pesquisando, conhecendo e analisando produtos, mas que ainda não se decidiram pela compra. Os sistemas de recomendação podem facilitar sua escolha, diminuindo a sobrecarga de informação e de interação, apresentando produtos que venham ao encontro dos seus interesses, contribuindo assim para a venda;
- **incrementando as vendas cruzadas:** os sistemas de recomendação melhoram as vendas cruzadas, sugerindo aos consumidores produtos adicionais para compra. Quanto melhor a sugestão, maior a probabilidade de um incremento na venda. A recomendação de produtos adicionais pode estar baseada, por exemplo, nos itens já escolhidos pelo cliente;
- **fidelizeando o cliente:** os sistemas de recomendação contribuem para a fidelização do cliente, criando valor por meio da personalização do *site* de comércio eletrônico. As lojas virtuais investem para aprender sobre os consumidores, suas necessidades e hábitos de compra. Os sistemas de recomendação operacionalizam esse aprendizado, oferecendo ao usuário uma interface personalizada, com sugestões que atendam seus interesses. O

---

mar. 2007.

<sup>5</sup> Em economia, custo de busca é qualquer coisa que interfira na descoberta do que se tem em mira, como perda de tempo, estresse ou pagar preço excessivo por não encontrar alternativas mais baratas (ANDERSON, 2006).

relacionamento com o usuário contribui para a fidelização, já que o consumidor gosta de voltar a um lugar onde se sentiu bem atendido.

A Figura 1 é um exemplo de recomendação sugerida por uma loja virtual com base no perfil do usuário.



Figura 1 – Recomendações personalizadas feitas por uma loja virtual

Fonte: www.submarino.com.br

## 2.2 Taxonomia

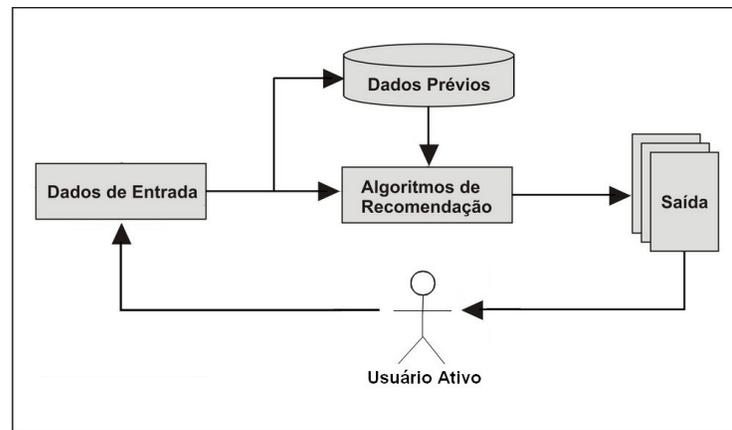
As discussões a respeito da classificação dos sistemas de recomendação geralmente giram em torno das fontes de dados do sistema (dados prévios e de entrada) e da maneira como esses dados são utilizados para fornecer as recomendações (BURKE, 2002).

Outras propostas de classificação consideram a forma como as recomendações são apresentadas aos usuários e como diferentes técnicas podem ser combinadas em um único sistema (TORRES, 2004, p. 41). Schafer, Konstan e Riedl (2001), por sua vez, propõem uma taxonomia baseada nas aplicações encontradas no comércio eletrônico.

A classificação aqui adotada aborda as principais características dos sistemas de recomendação, tendo como ponto de partida a análise dos elementos que compõem esses sistemas, que segundo Burke (2002) são:

- **dados prévios ou armazenados (*background data*):** corresponde a toda informação que o sistema armazena para utilizar no processo de recomendação. Compreendem a relação de itens a serem recomendados (catálogo de produtos, documentos, páginas *Web*, conteúdo multimídia etc.) e englobam ainda o modelo (ou perfil) de usuário;
- **dados de entrada ou do usuário:** trata-se da informação que deve ser fornecida pelo usuário para que o processo de recomendação possa ser iniciado;
- **algoritmo de recomendação:** responsável por combinar dados prévios e dados de entrada para que sejam fornecidas as devidas recomendações.

O processo de recomendação, considerando a apresentação dos resultados ao usuário e sua interação com o sistema, pode ser representado como mostra a Figura 2.



**Figura 2 – Processo de recomendação**  
Fonte: elaborada pelo autor.

Com base nisso, propõe-se que os sistemas de recomendação sejam classificados segundo: (a) os dados de entrada; (b) as saídas; (c) os tipos de filtragem; (d) as técnicas adotadas na implementação (referentes aos algoritmos).

### 2.2.1 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 onde 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.

As avaliações explícitas por voto declarado representam uma estratégia muito comum em *sites* de comércio eletrônico. A Figura 3 ilustra como pode ser feita a coleta de avaliações explícitas.



**Figura 3 – Exemplo de avaliação explícita**

Fonte: [www.amazon.com](http://www.amazon.com)

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

<b>Ação</b>	<b>Notas</b>
Comprar	adquirir um item
Avaliar	considerar a atitude do usuário avaliar ou não um item
Uso repetido	ex.: a busca pelo mesmo assunto
Salvar / Imprimir	preservar artigos de interesse
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, p. 32).

### 2.2.2 Saídas

Segundo 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 em relação a um item do catálogo, ou seja, o sistema apresenta um item associado a uma escala que prediz sua relevância para o usuário. Por exemplo, um filme pode ser apresentado a um usuário com uma estrela, caso o sistema considere que ele não gostaria do filme, e a outro usuário com cinco estrelas, considerando que este gostaria.

A recomendação, também conhecida como *Recomendação Top-N*, é representada por uma lista de  $N$  itens considerados interessantes para o usuário ativo. Normalmente, a lista apresenta itens não comprados, vistos ou avaliados pelo usuário.

É possível também combinar essas duas saídas criando uma lista de itens recomendados e associando uma predição a cada um deles.

Em *sites* de comércio eletrônico é comum encontrar outros tipos de saída. Os “modelos de aplicação” identificados por Schafer, Konstan e Riedl (2001), e as “estratégias de recomendação” relatadas por Reategui e Cazella (2005), são tomados como referência para descrever algumas das formas adotadas pelas lojas virtuais para apresentar as recomendações aos usuários. Entre elas, destacam-se:

## Listas de recomendação

Essa saída está fundamentada na publicação de listas de itens em uma organização temática, do tipo: “os eletrônicos mais vendidos”, “idéias para presente”, entre outras. A Figura 4 mostra uma aplicação de listas de recomendação.



**Figura 4 – Lista dos eletrônicos mais vendidos**

Fonte: www.amazon.com

Apesar de oferecer recomendações de forma automática e contribuir com a escolha dos usuários, essa estratégia não oferece personalização. No caso das listas de “mais vendidos”, por exemplo, o histórico de compra de todo o grupo de usuários é considerado para a sugestão.

## Suas recomendações

As avaliações dos usuários, sejam elas implícitas ou explícitas, são utilizadas pelo sistema para identificar suas preferências, servindo de base para a realização de recomendações personalizadas.

As sugestões geralmente são apresentadas aos usuários em uma lista de itens, acompanhados da avaliação média de outros usuários, conforme Figura 5.

Click me for cool deals! **amazon.com** Leonardo's Amazon.com See All 36 Product Categories Your Account | Cart | Your Lists | Help

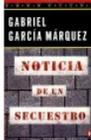
Your Browsing History | Recommended For You | Rate These Items | Improve Your Recommendations | Your Profile | Learn More

Search Amazon.com GO Gifts & Gift Certificates Web Search

Leonardo's Amazon.com™ > Recommended for you  
(If you're not Leonardo, [click here.](#))

**Recommendations Based on Activity**  
These recommendations are based on [items you own](#) and more.  
view: **All** | [New Releases](#) | [Coming Soon](#) [More results](#)

**Recommendations by Category**  
[Apparel & Accessories](#)  
[Baby](#)  
[Beauty](#)  
[Books](#)  
[Camera & Photo](#)  
[Computer & Video](#)  
[Games](#)  
[Computers & PC](#)

1.  **Noticia de un secuestro**  
by Gabriel Garcia Marquez  
Average Customer Review: ★★★★★  
In Stock  
Publication Date: July 1, 1996  
**Our Price: \$11.70**  
**Used & new from \$3.93**  
[Add to cart](#)  
[Add to Wish List](#)

I Own It  Not interested x|★★★★★ Rate it  
Recommended because you rated [Vivir Para Contarla / To Live to Tell It \(Spanish\)](#) ([edit](#))

Figura 5 – Recomendações personalizadas

Fonte: www.amazon.com

“Usuários que se interessaram por ‘X’ também se interessaram por ‘Y’”

Esse tipo de saída faz associações baseadas nos interesses dos usuários. É considerada com baixo grau de personalização porque não se baseia no histórico de avaliações do usuário que recebe a recomendação, mas no histórico de avaliações de outros usuários. Ainda assim, é importante ressaltar que a recomendação está associada a um item que o usuário demonstrou interesse, conforme mostra o exemplo apresentado na Figura 6.

**American History X (1998)**  
Starring: [Edward Norton](#), [Edward Furlong](#) Director: [Tony Kaye](#) Rating [R](#)  
★★★★★ (624 customer reviews)

List Price: \$14.98  
Price: **\$9.99** & eligible for **FREE Super Saver Shipping** on orders over \$25. [Details](#)  
You Save: **\$4.99 (33%)**

**Customers who bought this item also bought**  
[Requiem for a Dream \(Director's Cut\) DVD](#) ~ Burstyn  
[Blow \(Infinifilm Edition\) DVD](#) ~ Tony Amendola  
[Primal Fear DVD](#) ~ Richard Gere  
[Fight Club DVD](#) ~ Edward Norton  
▶ [Explore similar items](#) : DVD (49)

Figura 6 – Recomendação por associação

Fonte: www.amazon.com

### Examinar itens semelhantes

Também é possível fazer recomendações com base nas características dos itens, buscando associações entre itens semelhantes. Da mesma forma como apresentado acima, é possível fazer uma recomendação com baixo nível de personalização, sugerindo itens semelhantes àqueles que estão sendo analisados pelo usuário.

#### **2.2.3 Tipos de filtragem**

Normalmente, os sistemas de recomendação encontram-se classificados em categorias, conforme a abordagem utilizada para realizar as recomendações, ou seja, o tipo de filtragem adotado pelo sistema. Os tipos de filtragem mais referenciados na literatura são (BALABANOVIĆ; SHOHAM, 1997; BURKE, 2002; ADOMAVICIUS; TUZHILIN, 2005):

- filtragem baseada em conteúdo: o usuário receberá recomendação de itens similares àqueles que preferiu no passado;
- filtragem colaborativa: o usuário receberá recomendação de itens que pessoas com gostos e preferências similares ao seu preferiram no passado;
- filtragem demográfica: o usuário receberá recomendações conforme o grupo demográfico a que pertence;
- filtragem baseada em utilidade: o usuário receberá recomendações de acordo com uma função de utilidade determinada por ele;
- filtragem baseada em conhecimento: o usuário receberá recomendações baseadas no conhecimento que o sistema possui a respeito dos usuários e dos itens a serem recomendados;
- abordagem híbrida: este método procura combinar dois ou mais tipos de recomendação.

##### **2.2.3.1 Filtragem 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). Um livro, por exemplo, pode ter como características: título, autor, gênero, resenha etc. Já em um texto, geralmente, 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.

Formalizando (ADOMAVICIUS; TUZHILIN, 2005): a utilidade  $f(u, i_i)$  do item  $i_i$  para o usuário  $u$  é estimada com base nas utilidades  $f(u, i_j)$  atribuídas pelo usuário  $u$  para os itens  $i_j$  que são similares ao item  $i_i$ , considerando  $i_i \neq i_j$ .

Uma maneira de trabalhar com a filtragem baseada em conteúdo é solicitar ao usuário que avalie itens disponíveis no sistema, de forma que este manifeste seus gostos e preferências, indicando, por exemplo, se um item é interessante ou não. Uma vez realizada a avaliação, o sistema procura por itens semelhantes aos avaliados, recomendando aqueles com características próximas aos itens classificados como interessantes e desconsiderando aqueles que se assemelham aos não interessantes (REATEGUI; CAZELLA, 2005).

Os sistemas de recomendação baseados em conteúdo tem raízes nas técnicas de recuperação de informação e nas pesquisas de filtragem de informação (BALABANOVIĆ; SHOHAM, 1997; BURKE, 2002; ADOMAVICIUS; TUZHILIN, 2005).

Conforme Adomavicius e Tuzhilin (2005), em virtude dos significativos avanços na área de recuperação de informação e filtragem, e dada a importância das inúmeras aplicações baseadas em texto, muitos dos sistemas baseados em conteúdo focaram na recomendação de itens que contém informação textual, como documentos, páginas *Web* 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 contém dados sobre seus gostos, preferências e necessidades.

De maneira geral, os sistemas baseados em conteúdo mantêm perfis de longo prazo: o perfil do usuário é modelado conforme as características dos itens avaliados por ele, sendo atualizado à medida que novos itens são avaliados, mantendo-se como base para futuras recomendações (BURKE, 2002).

De acordo com Adomavicius e Tuzhilin (2005), Balabanović e Shoham (1997) e

Burke (2002), os sistemas baseados em conteúdo possuem algumas limitações, descritas na seqüência.

### 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, como é o caso dos textos, ou serem atribuídas manualmente a cada um dos itens.

O problema é que 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.

Outro problema é que se dois itens diferentes forem representados pelo mesmo conjunto de características, como é possível acontecer com textos, eles não serão diferenciados pelo sistema. Além disso, um texto que contenha palavras-chave de interesse do usuário pode não ser um bom texto, já que o sistema não distingue um texto bem escrito de um texto mal escrito.

### 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 expressou preferência por comida árabe, jamais receberia uma recomendação para ir ao melhor restaurante de comida árabe da cidade. No entanto, a diversidade de recomendações é, freqüentemente, uma característica desejável para os sistemas.

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 deve 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, por exemplo, 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).

#### **2.2.3.2 Filtragem colaborativa**

Segundo Burke (2002), a filtragem 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). Conseqüentemente, não exige a compreensão ou reconhecimento das características dos itens para realizar as recomendações.

Nos sistemas colaborativos, a essência está na troca de experiências entre as pessoas que possuem interesses comuns (REATEGUI; CAZELLA, 2005). Formalizando (ADOMAVICIUS; TUZHILIN, 2005): a utilidade  $f(u_a, i)$  do item  $i$  para o usuário  $u_a$  é estimada com base nas utilidades  $f(u_j, i)$  atribuídas para o item  $i$  por outros usuários  $u_j$  que são similares ao usuário  $u_a$ , considerando  $u_j \neq u_a$ .

Um perfil típico de usuário em um sistema colaborativo consiste em um vetor de itens avaliados, que aumenta sua dimensão à medida que o usuário interage com o sistema (BURKE, 2002). Em alguns casos, a avaliação feita pelo usuário pode ser binária, ou seja, somente é possível saber se o item agradou ou não; em outros, a avaliação pode ser representada por um número real que indica o grau de preferência do usuário.

Um exemplo de ambiente baseado em filtragem colaborativa é o sistema de recomendação de filmes *MovieLens* – detalhado na seção 2.5 (GOOD et al., 1999). Nele o usuário atribui nota a filmes que já assistiu e o sistema utiliza essas avaliações para encontrar pessoas com gostos similares e fazer suas previsões.

Segundo Adomavicius e Tuzhilin (2005), os primeiros sistemas a utilizarem algoritmos de filtragem colaborativa para automatizar a previsão de itens foram *GroupLens* (KONSTAN et al., 1998; RESNICK et al., 1994), *Bellcore Video Recommender* (HILL et al., 1995) e *Ringo* (SHARDANAND; MAES, 1995).

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, descritas a seguir (BALABANOVIĆ; SHOHAM, 1997; BURKE, 2002; ADOMAVICIUS; TUZHILIN, 2005).

### O problema do novo usuário

Assim como nos sistemas baseados em conteúdo, na filtragem colaborativa é necessário primeiramente aprender as preferências do usuário, a partir de suas avaliações, para então poder fazer recomendações mais precisas. Logo, esses sistemas também apresentam problemas de “elasticidade *versus* plasticidade”.

### O problema do novo item

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

Se o número de usuários é 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, por exemplo, aqueles que forem avaliados por poucos usuários são raramente recomendados, mesmo que esses atribuam pontuações elevadas.

### Usuário incomum

Um usuário cujo gosto é incomum, comparado à maioria da população, terá dificuldade 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 uma consequência das avaliações esparsas.

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

Na literatura, o problema de “novo usuário” e de “novo item”, ocasionados pela escassez de dados no sistema, também são conhecidos como problemas de *ramp-up* (KONSTANT et al., 1998 apud BURKE, 2002).

#### **2.2.3.3 Filtragem demográfica**

A filtragem demográfica tem como objetivo classificar o usuário em classes demográficas, em função de seus atributos pessoais (idade, sexo, renda etc.), e oferecer recomendações conforme a classe a qual o usuário pertence (BURKE, 2002).

Um exemplo desse tipo de filtragem é o sistema *Grundy* (RICH, 1979), considerado por alguns autores como o primeiro sistema de recomendação. O sistema recomendava livros baseado em informações pessoais que eram armazenadas por meio de um diálogo interativo. As respostas dos usuários serviam como base para a seleção de uma biblioteca de estereótipos compilada previamente, de forma manual (BURKE, 2002). Usando os estereótipos o sistema *Grundy* podia construir modelos de usuário e fazer recomendações.

Adomavicius e Tuzhilin (2005) consideram a filtragem demográfica uma extensão da filtragem colaborativa, uma vez que ambas se baseiam na similaridade entre 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, a filtragem colaborativa utiliza o histórico de avaliações. Isso faz com que a filtragem demográfica apresente as principais desvantagens encontradas no método colaborativo. Contudo, como não depende das avaliações para comparar usuários, sofre menos com o problema de avaliações esparsas.

Alguns exemplos de aplicação da filtragem demográfica são encontrados nos trabalhos de Krulwich (1997) e Pazzani (1999).

#### **2.2.3.4 Filtragem 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 usuário é representado pela função de utilidade que o sistema obtém como entrada de dados, e que será utilizada para encontrar a melhor opção segundo suas preferências.

As recomendações 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, esses 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 não enfrentam problemas típicos dos sistemas colaborativos e baseados em conteúdo, tais como: “novo usuário”, “novo item” e “avaliações esparsas”. 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. Isso pode ser visto como uma flexibilidade do sistema, mas também, em algum grau, um inconveniente, pois exige do usuário alto nível de interação (BURKE, 2002).

### 2.2.3.5 Filtragem 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 para ele (BURKE, 2002).

Diferente dos sistemas baseados em utilidade, nos sistemas baseados em conhecimento não se exige que os usuários considerem todas as características dos itens para receber 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 elas está o Raciocínio Baseado em Casos (RICCI et al., 2002).

A principal vantagem dessa abordagem é aumentar a precisão e evitar as 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, ou seja, quando consultam o sistema raramente na busca de ter uma necessidade específica atendida.

O principal inconveniente dos sistemas baseados em conhecimento consiste, justamente, na necessidade de adquirir o conhecimento – gargalo de muitas aplicações que utilizam inteligência artificial (ADOMAVICIUS; TUZHILIN, 2005).

### 2.2.3.6 Abordagem híbrida

Vários sistemas combinam dois ou mais tipos de filtragem 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).

Adomavicius e Tuzhilin (2005) analisaram as principais formas adotadas para combinar filtragem baseada em conteúdo e colaborativa em um sistema híbrido. São elas:

- 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 é o próprio sistema selecionar, entre os dois métodos, aquele que oferecer “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 filtragem 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.

### **2.2.3.7 Vantagens e desvantagens dos tipos de filtragem**

Todos os tipos de filtragem possuem vantagens e desvantagens, discutidas anteriormente e sintetizadas no Quadro 2 e Quadro 3, baseados em Burke (2002). A notação utilizada é: Filtragem colaborativa (FC); Filtragem baseada em conteúdo (CN); Filtragem demográfica (DM); Filtragem baseada em utilidade (UT); Filtragem baseada em conhecimento (BC).

VANTAGENS	FC	CN	DM	UT	BC
Possui capacidade para recomendar itens diferentes daqueles que o usuário já avaliou ou que necessita no momento, mas que podem ser de seu interesse	X		X		
Não necessita conhecimento do domínio	X	X	X		
É adaptável, ou seja, sua qualidade melhora com o tempo de utilização	X	X	X		
Apenas as avaliações dos usuários são suficientes como dados de entrada	X	X			
Não apresenta problemas de <i>ramp-up</i>				X	X
É sensível às mudanças de preferência dos usuários				X	X
Pode considerar características que não são próprias dos itens, mas associadas a eles				X	X
Possui capacidade de relacionar diretamente as necessidades dos usuários com os itens a recomendar					X

**Quadro 2 – Vantagens dos tipos de filtragem**

Fonte: Adaptado de Burke (2002).

DESvantagens	FC	CN	DM	UT	BC
Apresenta problema de <i>ramp-up</i> para novos usuários	X	X	X		
Apresenta problema de <i>ramp-up</i> para novos itens	X				
Apresenta limitações para usuários com gostos incomuns	X		X		
A qualidade depende do histórico de utilização – acúmulo de dados	X	X	X		
Apresentam problema de “estabilidade <i>versus</i> plasticidade”	X	X	X		
Deve obter informação demográfica			X		
O usuário deve entrar com uma função de utilidade				X	
Não possui capacidade de aprender sobre as preferências do usuário				X	X
Requer conhecimento do domínio					X
Apresenta problema de avaliações esparsas	X				
Possui limitação quanto à análise de conteúdo		X			
Apresenta problemas de superespecialização		X			

**Quadro 3 – Desvantagens dos tipos de filtragem**

Fonte: Adaptado de Burke (2002).

Observa-se que as recomendações colaborativas e demográficas são as únicas que possuem a capacidade de recomendar itens diferentes daqueles que o usuário já avaliou, mas que possam ser de seu interesse. Essa característica, também conhecida como “*cross-genre*” ou “*outside the box*”, é o oposto da superespecialização, uma desvantagem dos sistemas baseados em conteúdo.

Todas as abordagens baseadas em aprendizagem (colaborativa, baseada em conteúdo e demográfica) possuem de uma forma ou de outra o problema de *ramp-up*, ou seja, dependem de um número suficiente de avaliações para aprimorar o processo de recomendação. Nota-se que os sistemas baseados em utilidade e baseados em conhecimento

não enfrentam esse problema, pois não utilizam perfis de longo prazo.

#### 2.2.4 Técnicas de recomendação

Tendo como base Shardanand e Maes (1995), é possível apresentar a idéia básica de um sistema de recomendação em três etapas.

1. O sistema mantém um perfil de usuário e seus interesses sobre itens específicos.
2. Ele compara este perfil com outros perfis, ou diretamente com outros itens (se baseado em conteúdo), e avalia o grau de similaridade entre eles.
3. Finalmente, ele considera um conjunto de perfis (ou itens) mais similares, e utiliza a informação relacionada a eles para recomendar itens ao usuário.

São várias as técnicas utilizadas para calcular a similaridade e realizar as recomendações. Dependendo da técnica selecionada é possível implementá-la utilizando algoritmos específicos.

Os algoritmos de filtragem colaborativa foram apresentados por Breese, Heckerman e Kadie (1998) divididos em duas categorias: baseados em memória (*memory-based*) e baseados em modelo (*model-based*). Essa mesma classificação foi adotada por Adomavicius e Tuzhilin (2005) para os algoritmos de filtragem baseada 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 aprendizado de máquina podem ser utilizados, como 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 inferior em relação aos sistemas que empregam técnicas baseadas 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, algumas delas são abordadas.

#### **2.2.4.1 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.

##### **2.2.4.1.1 TF-IDF**

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

Considerando que  $k_i$  é uma palavra que aparece em  $n_i$  documentos, e que  $f_{i,j}$  é o

número de vezes que a palavra  $k_i$  aparece em um documento específico  $d_j$ . O cálculo de  $TF_{ij}$  é obtido pela Equação 1.

$$TF_{ij} = \frac{f_{ij}}{\max_z f_{zj}} \quad 1$$

O termo  $f_{z,j}$  considera a frequência de cada palavra do documento  $d_j$ , tomando como base o maior valor encontrado.

Como as palavras muito comuns, que aparecem em muitos documentos, tornam-se pouco úteis para distinguir seu conteúdo, a medida do inverso da frequência em documentos (*IDF*) é bastante utilizada. Com ela, o peso das palavras muito comuns é reduzido.

Assumindo que  $N$  é o número total de documentos e que  $n_i$  é o número de documentos em que a palavra  $k_i$  aparece, o cálculo de  $IDF_i$  pode ser obtido pela Equação 2.

$$IDF_i = \log \frac{N}{n_i} \quad 2$$

O peso de uma palavra em um dado documento é calculado como sendo o produto das duas medidas (*TF-IDF*), conforme mostrado na Equação 3.

$$w_{ij} = TF_{ij} \times IDF_i \quad 3$$

Assim, o documento pode ser representado como um vetor de palavras e seus pesos, do tipo:  $d_j = (w_{1j}, \dots, w_{kj})$ .

Como nos sistemas baseados em conteúdo o modelo de usuário é obtido com base no conteúdo dos itens previamente vistos e avaliados por ele, seu perfil também pode ser representado como um vetor de pesos de palavras. Por isso, é possível medir a similaridade entre um item e uma instância do perfil do usuário calculando o co-seno entre os vetores, conforme apresentado na Equação 4.

$$\cos(\vec{w}_i, \vec{w}_j) = \frac{\sum_{k=1}^K w_{ki} w_{kj}}{\sqrt{\sum_{k=1}^K w_{ki}^2} \sqrt{\sum_{k=1}^K w_{kj}^2}} \quad 4$$

Observa-se que a técnica *TF-IDF* é voltada para aplicações onde os itens a serem recomendados possuem atributos textuais. Para itens cujas características são representadas por outro tipo de atributo, outras técnicas podem ser aplicadas, tais como *kNN* e Meta-protótipo (BEZERRA et al., 2002).

### 2.2.4.1.2 Correlação

A técnica da correlação fundamenta-se em calcular a similaridade entre dois usuários com base em suas avaliações. Normalmente, para isso, é utilizado o coeficiente de correlação de Pearson.

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):

$$w(a,u) = \frac{\sum_i (v_{a,i} - \bar{v}_a)(v_{u,i} - \bar{v}_u)}{\sqrt{\sum_i (v_{a,i} - \bar{v}_a)^2 \sum_i (v_{u,i} - \bar{v}_u)^2}}, \quad 5$$

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:

$$\bar{v}_a = \frac{1}{|I_a|} \sum_{i \in I_a} v_{a,i}. \quad 6$$

O peso  $w(a,u)$  pode refletir tanto a similaridade quanto a dissimilaridade entre o usuário  $u$  e o usuário ativo  $a$ . Uma vez definida a similaridade entre os usuários, a predição de um item  $i'$  para o usuário ativo  $a$  pode ser calculada como:

$$p_{a,i'} = \bar{v}_a + k \sum_{u \in \hat{U}} w(a,u)(v_{u,i'} - \bar{v}_u), \quad 7$$

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). A predição também pode ser determinada pelas Equações 8 e 9 (ADOMAVICIUS; TUZHILIN, 2005).

$$p_{a,i'} = \frac{1}{N} \sum_{u \in \hat{U}} v_{u,i'}. \quad 8$$

$$p_{a,i'} = k \sum_{u \in \hat{U}} w(a,u)v_{u,i'}. \quad 9$$

O termo  $k$  é um fator de normalização, geralmente selecionado como:

$$k = \frac{1}{\sum_{u \in \mathcal{U}} |w(a,u)|} \quad 10$$

Um problema com o uso da soma de pesos, como apresentado na Equação 9, é que ela não considera que os usuários podem utilizar escalas diferentes para avaliar os itens, ou seja, enquanto um usuário utiliza votos 2 e 3 em suas avaliações, outro pode utilizar votos de 1 a 5. O somatório de pesos *ajustado*, mostrado na Equação 7, tem sido largamente utilizado por considerar essa limitação. Isso significa que, ao invés de considerar os valores absolutos das avaliações, é considerado o desvio da avaliação média do usuário correspondente.

A Figura 7 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  $u_a$ .

	$i_1$	$i_2$	...	$i_i$	...	$i_n$
$u_1$	$v_{1,1}$	$v_{1,2}$		$v_{1,i}$		$v_{1,n}$
$u_2$	$v_{2,1}$	$v_{2,2}$		$v_{2,i}$		$v_{2,n}$
:						
$u_a$	$v_{a,1}$	$v_{a,2}$	...	$v_{a,i}$	...	$v_{a,n}$
:						
$u_m$	$v_{m,1}$	$v_{m,2}$		$v_{m,i}$		$v_{m,n}$

Figura 7 – Representação de uma matriz de avaliações  
Fonte: elaborada pelo autor.

#### 2.2.4.1.3 Correlação baseada em item

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 nas avaliações que os itens receberam de outros usuários. Em outras palavras, 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).

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 – conjunto  $U$  de usuários (ver Figura 8). O cálculo da correlação baseada em item é apresentado na Equação 11 (SARWAR et al., 2001).

$$w_{i,j} = \frac{\sum_{u \in U} (v_{u,i} - \bar{v}_i)(v_{u,j} - \bar{v}_j)}{\sqrt{\sum_{u \in U} (v_{u,i} - \bar{v}_i)^2} \sqrt{\sum_{u \in U} (v_{u,j} - \bar{v}_j)^2}} \quad (11)$$

Com as similaridades correspondentes computadas [ $w_{i1}, w_{i2}, \dots, w_{ij}$ ], seleciona-se os  $N$  itens mais similares [ $i_1, i_2, \dots, i_n$ ] para realizar a predição. O cálculo da predição considera as avaliações do usuário  $u_a$  e o peso de cada item  $I_a$  em relação a  $i$  ( $w_{i,Ia}$ ), conforme mostra a Equação 12 (SARWAR et al., 2001).

$$p_{a,i} = \frac{\sum_N w_{i,N} v_{a,N}}{\sum_N |w_{i,N}|} \quad (12)$$

	$i_1$	$i_2$	...	$i_i$	$i_j$	...	$i_n$
$u_1$				$v_{1,i}$	$v_{1,j}$		
$u_2$				--	$v_{2,j}$		
:							
$u_a$				$v_{a,i}$	$v_{a,j}$		
:							
$u_{m-1}$				$v_{m-1,i}$	$v_{m-1,j}$		
$u_m$				$v_{m,i}$	--		

Figura 8 – Itens co-avaliados para medida de similaridade entre  $i_i$  e  $i_j$   
Fonte: elaborada pelo autor.

#### 2.2.4.1.4 Vetor de similaridade (co-seno)

O vetor de similaridade, também denominado de co-seno, tem suas raízes nos sistemas de recuperação de informação (*Information Retrieval*). Como apresentado na técnica *TF-IDF* (tópico 2.2.4.1.1), a similaridade entre documentos é, frequentemente, medida tratando cada documento como um vetor de frequência de palavras, e computando o co-seno do ângulo formado por dois vetores (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 frequência de palavras. A similaridade de cada usuário  $u$  em relação ao usuário ativo  $a$  é então calculada

como (BREESE; HECKERMAN; KADIE, 1998):

$$w(a, u) = \sum_i \frac{v_{a,i}}{\sqrt{\sum_{k \in I_a} v_{a,k}^2}} \frac{v_{u,i}}{\sqrt{\sum_{k \in I_u} v_{u,k}^2}}, \quad 13$$

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 Equação 7.

#### 2.2.4.1.5 Extensões para os algoritmos baseados em memória

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*)

Essa extensão para o algoritmo da correlação surgiu ao se observar que quando existem relativamente poucos votos, tanto do usuário ativo quanto do que esta sendo comparado a ele, o algoritmo pode não oferecer bons resultados, porque considera apenas os votos da interseção dos itens avaliados por ambos ( $I_a \cap I_u$ ).

Se um valor padrão for assumido como um voto para os itens que não tenham recebido um voto explícito, então é possível fazer a comparação sobre a união dos itens avaliados ( $I_a \cup I_u$ ). 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).

Além disso, pode-se assumir o mesmo valor do voto padrão  $d$  para um número  $k$  de itens adicionais que nenhum dos usuários tenha avaliado. Considera-se assim que existe uma quantidade de itens não avaliados pelos usuários, que eles poderiam concordar em suas avaliações. Na maioria dos casos, o valor  $d$  refletirá uma preferência neutra ou até mesmo negativa para os itens não avaliados. A equação de correlação pode ser escrita como (BREESE; HECKERMAN; KADIE, 1998):

$$w(\alpha, u) = \frac{(n+k)\left(\sum_i v_{\alpha i} v_{u i} + k d^2\right) - \left(\sum_i v_{\alpha i} + k d\right)\left(\sum_i v_{u i} + k d\right)}{\sqrt{\left((n+k)\left(\sum_i v_{\alpha i}^2 + k d^2\right) - \left(\sum_i v_{\alpha i} + k d\right)^2\right)\left((n+k)\left(\sum_i v_{u i}^2 + k d^2\right) - \left(\sum_i v_{u i} + k d\right)^2\right)}} \quad , \quad 14$$

onde o somatório  $i$  considera a união dos itens avaliados tanto por  $a$  quanto por  $u$  ( $Ia \cup Iu$ ), e  $n = |Ia \cup Iu|$ .

Sarwar (2001) propõe utilizar o voto médio do usuário em todos os itens ainda não avaliados por ele – esquema de voto médio do usuário –, ou utilizar o voto médio recebido pelo item para preencher a coluna de votos vazios – esquema de voto médio do item.

O voto padrão também pode ser calculado considerando o desvio da avaliação média do usuário correspondente, conforme mostra a Equação 15.

$$v_{\alpha j} = \bar{v}_\alpha + \frac{\sum_{i=1}^M (v_{ij} - \bar{v}_i)}{N} \quad . \quad 15$$

### Filterbots

*Filterbots* são agentes de *software* que automaticamente avaliam novos itens. São como usuários ordinários, diferindo-se destes por gerarem mais avaliações e por não exigirem predições para si mesmos (VOZALIS; MARGARITIS, 2003).

Uma questão interessante a respeito dos *filterbots* está na quantidade de “inteligência” que pode ser incorporada a eles. O tipo mais simples de *filterbots*, chamado *Genrebots*, é utilizado para avaliar filmes conforme o gênero. Por exemplo, um sistema de recomendação de filmes pode utilizar um “*TerrorBot*”, que avaliará positivamente os filmes de terror. Nesse caso devem existir *Genrebots* para outros gêneros.

Algumas aplicações de *filterbots* são encontradas em Sarwar et al. (1998); Good et al. (1999) e Sarwar et al. (2001).

Assim como o voto padrão, o uso de *filterbots* contribui para a redução da esparsidade da matriz usuário-item. Outra técnica utilizada para lidar com esse problema é a Decomposição em Valores Singulares (*Singular Value Decomposition – SVD*), avaliada por Sarwar et al. (2001) e Vozalis e Margaritis (2003).

### Inverse user frequency

De forma análoga ao *IDF*, discutido no tópico 2.2.4.1.1, a idéia é que os itens muito populares não sejam tão úteis para avaliar a similaridade entre os usuários quanto os itens menos comuns (BREESE; HECKERMAN; KADIE, 1998).

A função do inverso da frequência de usuários (*inverse user frequency*) pode ser calculada como:

$$f_i = \log \frac{N}{n_i}, \quad 16$$

onde  $n_i$  é o número de usuários que votaram no item  $i$  e  $N$  é o número total de usuários da base de dados. Nota-se que se todos os usuários votaram no item  $i$  então  $f_i$  é zero.

Desta forma, o cálculo da similaridade dos usuários pode ser modificado com a função  $f_i$ . No vetor de similaridade, a transformada é obtida multiplicando-se o voto original pelo fator  $f_i$ . Na correlação, um item que tenha um alto valor de  $f_i$  é assinalado com mais peso no cálculo da correlação. A fórmula da correlação com o inverso da frequência é apresentada como (BREESE; HECKERMAN; KADIE, 1998):

$$w_{(a,u)} = \frac{\sum_i f_i \sum_i f_i v_{a,i} v_{u,i} - \left(\sum_i f_i v_{a,i}\right) \left(\sum_i f_i v_{u,i}\right)}{\sqrt{UV}} \quad 17$$

Onde:

$$U = \sum_i f_i \left( \sum_i f_i v_{a,i}^2 - \left(\sum_i f_i v_{a,i}\right)^2 \right) \quad 18$$

$$V = \sum_i f_i \left( \sum_i f_i v_{u,i}^2 - \left(\sum_i f_i v_{u,i}\right)^2 \right) \quad 19$$

### Amplificação de caso (case amplification)

A 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. Os pesos podem ser transformados como segue (BREESE; HECKERMAN; KADIE, 1998):

$$w'_{a,u} = \begin{cases} w_{a,u}^\rho & \text{if } w_{a,u} \geq 0 \\ -(-w_{a,u}^\rho) & \text{if } w_{a,u} < 0 \end{cases} \quad 20$$

Um valor típico para  $\rho$  é 2,5.

### 2.2.4.2 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 (BREESE; HECKERMAN; KADIE, 1998):

$$P_{a,i} = E(v_{a,i}) = \sum_{m_i=0}^m m_i \times \Pr(v_{a,i} = m_i | v_{a,k}, k \in I_a), \quad 21$$

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.

#### 2.2.4.2.1 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 de treinamento e classes é conhecida (MARTÍN GUERRERO, 2004).

Com os algoritmos de *clustering* é possível realizar agrupamentos a partir de um conjunto de padrões  $n$ -dimensionais, onde  $n$  corresponde ao número de características ou variáveis consideradas para agrupar os semelhantes.

Quando utilizados com a finalidade de realizar previsões, os algoritmos atuam na análise das características comuns aos padrões que formam um grupo e que os diferenciam dos padrões que pertencem a outro grupo. Com isso, quando um novo padrão é adicionado aos dados, é possível determinar a qual grupo pertencerá com maior probabilidade. No caso de um novo usuário de um sistema, por exemplo, a partir do momento que ele é identificado como pertencente a um grupo, pode-se recomendar os itens preferidos pelas pessoas que compõem seu grupo (MARTÍN GUERRERO, 2004).

Os algoritmos de *clustering* podem ser classificados em diversas categorias, conforme a estratégia utilizada para se fazer o agrupamento. Algoritmos pertencentes à categoria de “*clustering* probabilístico” são bastante utilizados em sistemas de recomendação. Para os algoritmos desta categoria, cada padrão é considerado um vetor  $v$  o qual é associado a um cluster  $C_i$  seguindo um esquema de classificação Bayesiana. A associação se dará quando a probabilidade  $P(C_i | x)$  for máxima. O algoritmo *Expectation-Maximization* (E-M) é um exemplo dessa categoria (DEMPSTER; LAIRD; RUBIN, 1977).

Breese, Heckerman e Kadie (1998) apresentam como modelo *cluster* para filtragem colaborativa, um classificador Bayesiano com enfoque ingênuo (*naive Bayes*), ou seja, que assume a hipótese de independência condicional dos votos dada uma associação em uma classe  $C$ .

A idéia é que existem determinados tipos de usuários que compartilham um conjunto de preferências e gostos que podem ser agrupados em classes. Dada a classe, as preferências em relação a vários itens (expressa em votos) são independentes. O modelo combinando a probabilidade de classe e votos para tratar o conjunto de probabilidades totais e condicionais segue a fórmula padrão de Bayes ingênuo, apresentado na Equação 22.

$$P(C = c, v_1, \dots, v_n) = P(C = c) \prod_{i=1}^n P(v_i | C = c) \quad 22$$

Os parâmetros do modelo, as probabilidades associadas à classe  $P(C=c)$ , e as probabilidades condicionais dos votos, dada a classe,  $P(v_i|C=c)$ , 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 a fórmula básica das técnicas baseadas em modelo, conforme a Equação 21.

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.

#### 2.2.4.2.2 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. Também é possível incluir estados correspondentes a “não voto” para os domínios onde não há nenhuma interpretação para dados faltantes (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 em Chickering, Heckerman e Meek (1997).

#### **2.2.4.2.3 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,  $C1 =$  gosta e  $C2 =$  não gosta.

Considerando que a esparsidade da matriz é o grande número de espaços vazios, dado que muitos usuários avaliam somente um pequeno número de itens dentre os possíveis, a tarefa de predição pode ser vista como a melhor forma de preencher os campos vazios da matriz usuário-item (matriz  $R$ ).

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

**Figura 9 – Representação de avaliações de usuários**

Fonte: Billsus e Pazzani (1998, p. 48).

A proposta inicial dessa abordagem é transformar o conjunto de dados das avaliações, correspondente à matriz  $R$  (Figura 9), em um formato que algoritmos de aprendizagem supervisionada possam operar, a fim de induzir uma função  $f : E \rightarrow C$ , que classifique os itens na classe apropriada.

	$E_1$	$E_2$	$E_3$
$U_1$ 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 10 – Matriz transformada**

Fonte: Billsus e Pazzani (1998, p. 48).

A Figura 10 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 para realizar uma predição para o usuário  $U_4$ .

As avaliações do usuário  $U_4$  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. Observa-se que as avaliações 1 e 2 são associadas à classe “não gosta”, enquanto as de valor 3 e 4 correspondem à classe “gosta”.

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 transformada (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 supervisionada (Redes Neurais Artificiais), oferece ganhos significativos em desempenho e

precisão, embora seja computacionalmente mais dispendiosa que outras técnicas, como a correlação.

### ***2.3 Exemplos de sistemas de recomendação***

Atualmente existem muitas aplicações de sistemas de recomendação, tanto no âmbito acadêmico quanto no comercial. Nesta seção são apresentados quatro projetos que adotam diferentes tipos de filtragem. O primeiro, *MovieLens*, foi escolhido por trata-se de uma iniciativa acadêmica que tem servido de base para diversos estudos na área; o segundo, *What to Rent*, por seu método *sui generis* de recomendação; o terceiro, *Entree*, por apresentar soluções interessantes para o desenvolvimento do sistema proposto nesta pesquisa; o quarto, Pandora™, por tratar-se de uma aplicação comercial bem sucedida, que tem recebido muita atenção da mídia.

#### ***2.3.1 MovieLens***

O MovieLens<sup>6</sup> é uma plataforma experimental do *GroupLens*, grupo de pesquisa do Departamento de Ciência da Computação e Engenharia da Universidade de Minnesota. A proposta do sistema é recomendar filmes aos usuários, utilizando filtragem colaborativa (GOOD et al., 1999).

Para que o usuário comece a receber recomendações do sistema, é necessário que ele “construa” seu perfil avaliando no mínimo 15 filmes. A partir disso, o *MovieLens* passa a sugerir filmes predizendo as avaliações do usuário (ver Figura 11) – as predições são apresentadas em vermelho, enquanto as avaliações do usuário aparecem na cor azul. À medida que o usuário avalia mais filmes, corrigindo ou aceitando as predições feitas pelo *site*, seu perfil é atualizado e, conseqüentemente, as predições são ajustadas.

---

<sup>6</sup> Disponível em: <<http://www.movielens.org>>. Acessado em: 18 mar. 2007.

The screenshot shows the MovieLens website interface. At the top, the logo 'movielens' is displayed in red, with the tagline 'helping you find the right movies'. A user is logged in as 'leonardo@egc.ufsc.br', having rated 78 movies and being the 30th visitor in the past hour. A legend on the right explains the star rating system: 5 stars for 'Must See', 4 for 'Will Enjoy', 3 for 'It's OK', 2 for 'Fairly Bad', and 1 for 'Awful'. The navigation bar includes links for Home, Find Movies, Discussion Forums, Preferences, and Help. The main content area shows search results for 'Prediction', with 8570 movies found. It lists related tags like 'classic', '70mm', 'Bibliothek', 'Oscar (Best Picture)', and 'erlend's DVDs'. The results are sorted by 'Prediction'. A table lists three movies: 'Departed, The (2006)', 'Shawshank Redemption, The (1994)', and 'Fubar (2002)'. Each entry includes a star rating, a 'Not seen' dropdown, a movie title, genre, and a 'Wish List' checkbox. The page is on 1 of 572 pages.

Figura 11 – Tela de recomendação do *MovieLens*

Fonte: <http://www.movielens.org>

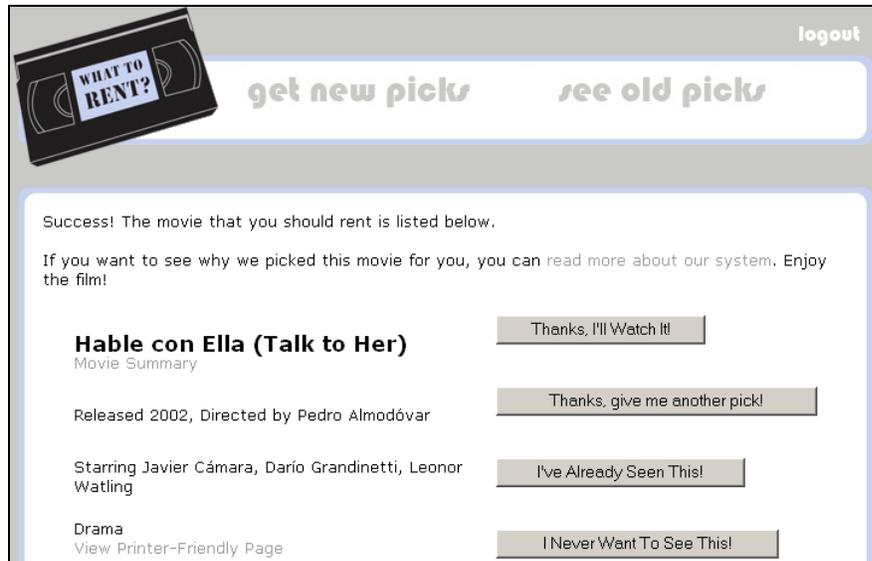
O *GroupLens* tem contribuído com o desenvolvimento de pesquisas na área de sistemas de recomendação, não só pelas intensas investigações e publicações realizadas, mas também por disponibilizar duas bases de dados do *MovieLens* para que outros pesquisadores realizem estudos. Uma delas contém 100.000 avaliações, feitas por 943 usuários, para 1.682 filmes. A outra possui aproximadamente um milhão de avaliações, realizadas por 6.040 usuários, para 3.900 filmes.

### 2.3.2 *What to Rent*

Outros *sites* de recomendação de filmes estão disponíveis na Web. O *What to Rent*<sup>7</sup> (ver Figura 12), por exemplo, solicita que os usuários respondam um questionário com questões sobre seu comportamento, a fim de modelar sua personalidade. Tendo respondido o questionário, o usuário solicita recomendações conforme seu “estado de espírito”, respondendo duas perguntas complementares.

Segundo os diretores do *site*, cada filme na base de dados é avaliado e analisado como se fosse uma pessoa, ou melhor, um modelo de personalidade.

<sup>7</sup> Disponível em: <<http://www.whattorent.com>>. Acessado em: 18 mar. 2007.



**Figura 12 – Tela de recomendação do *What to Rent***

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

### 2.3.3 *Entree*

O *Entree* é um sistema de recomendação de restaurantes que combina a filtragem colaborativa com a filtragem baseada em conhecimento – utilizando Raciocínio Baseado em Casos (RBC) (BURKE, 2002).

Para receber recomendações, o usuário deve selecionar características que correspondam ao seu interesse – tipo de cozinha, faixa de preço, estilo – ou solicitar que o sistema indique um restaurante semelhante a outro que ele já conheça (ver Figura 13).

Figura 13 – Tela inicial do Entree

Fonte: Burke (2002)

Ao receber a recomendação o usuário interage com o sistema, podendo “criticar” a sugestão selecionando uma opção que refine seu interesse. Por exemplo, pode solicitar um restaurante que seja mais barato que o recomendado ou um mais tranquilo (ver Figura 14). Suas interações são tomadas como avaliações implícitas influenciando as próximas recomendações.

Figura 14 – Tela de recomendação de restaurante

Fonte: Burke (2002).

### 2.3.4 Pandora

O Pandora<sup>8</sup> é um dos sistemas de recomendação de música mais populares da Internet. O sistema se baseia nas características das músicas para oferecer recomendações, sendo um bom exemplo de aplicação da filtragem baseada em conteúdo.

Observa-se que nesse tipo de sistema é necessário cadastrar as características dos itens para que eles possam ser comparados e a recomendação realizada. Como as características musicais são de difícil extração automática, é necessário que elas sejam levantadas e cadastradas manualmente.

No Pandora, uma música pode ser identificada por até 400 características. A tarefa de levantar as características das músicas é realizada por quarenta e cinco analistas, muitos deles graduados em música, que avaliam cerca de 15.000 canções por mês. Por isso o projeto é denominado de *Music Genome Project*<sup>TM</sup>.

O banco de dados do Pandora é composto de mais de 500.000 músicas, abrangendo mais de 20.000 artistas. Aproximadamente quatro milhões de usuários criam suas rádios personalizadas e recebem recomendações com base em suas escolhas.

Para começar a receber recomendações do Pandora basta que o usuário busque por uma música ou um artista. Com essa informação, o sistema cria uma estação que irá explorar canções e artistas que tenham características musicais similares ao que foi escolhido. A partir disso, o usuário pode avaliar cada música recomendada como sendo ou não do seu agrado (ver Figura 15). As avaliações são então utilizadas para aprimorar as recomendações.

---

<sup>8</sup> Disponível em: <<http://www.pandora.com>>. Acessado em: 18 mar. 2007.



**Figura 15 – Interface do Pandora**  
 Fonte: <http://www.pandora.com>

Ao recomendar “Aquele Abraço”, de Gilberto Gil (álbum “Unplugged”), o sistema identifica entre as características da música: “instrumentação acústica”, “raízes brasileiras”, “letra em português”, “teclado em tom maior”. Com base nesses atributos, o sistema recomenda também “Segue o Seco”, de Marisa Monte, por ter como características: “instrumentação acústica”, “letra em português”, “teclado em tom maior”, “influência brasileira”, entre outras. A qualquer momento o usuário pode questionar porque determinada música foi recomendada e o sistema informa as características que o levou a recomendá-la.

## 2.4 Avaliação das recomendações

Uma etapa fundamental no processo de 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, além de testar o funcionamento do *software*, é importante avaliar a qualidade de suas recomendações. 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 (*cold star problem*).

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 sistema *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, pois o sistema não é acessado por usuários reais.

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, para se fazer as avaliações 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):

$$MAE_i = \frac{\sum_{j=1}^{n_i} |ur_{ij} - r_{ij}|}{n_i} . \quad 23$$

O *MAE* total pode ser calculado pela média dos erros de cada usuário, como mostra a Equação 24 (VOZALIS; MARGARITIS, 2003).

$$MAE = \frac{\sum_{i=1}^m MAE_i}{m} . \quad 24$$

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, usada 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* (abrangência e precisão) e *Expected Utility* (utilidade esperada), assim como medidas para avaliar o desempenho computacional dos recomendadores: *tempo de resposta*, *memória requerida* e *complexidade computacional*.

## 2.5 *Sistemas de recomendação em turismo*

O produto turístico é um composto de bens e serviços diversificados e essencialmente relacionados entre si (ANDRADE, 2000). Por ser o turismo<sup>9</sup> uma atividade complexa e global, a informação é o seu elemento vital e a tecnologia tornou-se fundamental para que o setor opere de forma eficaz e competitiva (BOGADO; TELES, 2003).

Com o propósito de facilitar a busca de informações e apoiar a tomada de decisão dos turistas, alguns *sites* de viagem passaram a utilizar sistemas de recomendação (DELGADO; DAVIDSON, 2002).

Berka e Plößning (2004) citam como exemplos de sistemas de recomendação em turismo: o *Triplehops's TripMatcher* e o *VacationCoach's Me Print* – ambos classificados como sistemas baseados em conteúdo –, e o ITR (*Intelligent Travel Recommender*) – sistema baseado em conhecimento, apresentado por Ricci et al. (2002).

Para se desenvolver um sistema de recomendação em viagem e turismo é fundamental observar as particularidades desse domínio. Entre elas, destacam-se a pouca frequência de consumo dos produtos e serviços turísticos e a sua complexidade.

A pouca frequência de consumo afeta diretamente a abordagem a ser utilizada na recomendação. Enquanto a filtragem colaborativa, por exemplo, pode ser facilmente aplicável para a recomendação de itens como filmes, livros, músicas, entre outros artigos de varejo que são consumidos com frequência, os sistemas colaborativos apresentam limitações para a recomendação de produtos que são comprados esporadicamente, como é o caso de serviços turísticos (RICCI; DEL MISSIER, 2004). Isso porque a filtragem colaborativa pura se baseia no histórico de avaliações dos usuários para poder compor seu perfil e realizar as

---

<sup>9</sup> Algumas definições sobre turismo podem ser encontradas em Andrade (2000). Embora o termo viagem seja mais abrangente que turismo, tanto um quanto o outro são utilizados aqui com a mesma conotação.

recomendações.

Por outro lado, a complexidade dos produtos e serviços turísticos altera a forma como os usuários tomam suas decisões de compra. Normalmente os turistas consideram uma série de variáveis em seu processo de decisão, relacionadas principalmente ao contexto da sua viagem. Por isso, Delgado e Davidson (2002) afirmam que em viagem e lazer fatores como sazonalidade, distância, cenário da viagem, atividades e interesses individuais devem ser considerados nas recomendações.

Fesenmaier et al. (2003), por sua vez, sugerem que os desenvolvedores de sistemas de recomendação de viagem não foquem somente nas teorias das ciências da computação, mas que também considerem as pesquisas em teoria de decisão em viagem, área amplamente investigada em estudos sobre o turismo, conforme se pode verificar em Fesenmaier et al. (2002).

### **2.5.1 Teoria de decisão em viagem**

Fesenmaier e Jeng (2000, apud HWANG; GRETZEL; FESENMAIER, 2002) propõem um modelo de decisão em viagem que consiste em sub-decisões principais, secundárias e *en route*, aqui traduzida como “no roteiro”.

As sub-decisões principais são geralmente planejadas com mais antecedência, em detalhe, e inclui sub-decisões do destino preliminar, data de partida, duração da viagem, grupo de viagem, acomodações, roteiro e orçamento.

As sub-decisões secundárias incluem a escolha de outros destinos que podem ser visitados, a seleção de atividades, e atrações escolhidas para visitar. Apesar das decisões secundárias serem tomadas com certa antecedência, elas são moderadamente flexíveis, para acomodar mudanças possíveis de itinerário.

As sub-decisões “no roteiro” são as escolhas realizadas durante a viagem.

Hwang, Gretzel e Fesenmaier (2002), ao tratar de sistemas de apoio à decisão em viagem (*TDAS - Travel Decision-Aid Systems*), ressaltam que para se desenvolver um *TDAS* centrado no usuário é necessário entender a necessidade de informação, o processo de busca, o conteúdo da busca, a antecedência com que a procura é realizada e o processo de tomada de decisão de viagem. Assim, cinco questões podem ser formuladas:

1) Quem são os usuários do sistema?

- 2) Porque eles procuram informação?
- 3) Quando eles buscam a informação, no contexto do processo de tomada de decisão?
- 4) Que tipo de informação eles procuram?
- 5) Como a informação pode suprir sua necessidade específica para a tomada de decisão?

Com base nas questões apresentadas, o sistema centrado no usuário deve considerar: para quem é feito, o que apresentar e como apresentar.

Para os autores, os *TDAS* devem identificar duas questões essenciais para realizar recomendações eficientes: a natureza da viagem e da busca de informação. A natureza da viagem pode ser identificada pelas necessidades e benefícios procurados, bem como pelas restrições informadas. A natureza da busca de informação pode ser apreendida segundo a especificidade do que se procura para a tomada de decisão. Tanto a natureza da viagem quanto a da busca de informação, sofrem influência direta das características pessoais do usuário (aspectos sócio-demográficos, personalidade etc.).

Ricci et al. (2002) classificaram em dois grupos os fatores que influenciam a tomada de decisão dos turistas: fatores pessoais e características da viagem. O primeiro engloba fatores sócio-econômicos (idade, nível de escolaridade, renda etc.) e psicológico-cognitivos (experiência, personalidade etc.). O segundo abrange a proposta da viagem, os acompanhantes, o tempo de permanência, entre outros.

As teorias de decisão em viagem constituem o alicerce para a implementação de sistemas de recomendação em viagem e turismo, principalmente no que se refere à natureza das informações que influenciam o processo de tomada de decisão.

## ***2.6 Considerações finais sobre o capítulo***

Os sistemas de recomendação, surgidos na década de 1990, se destacam como importante área de pesquisa, com muitas aplicações no comércio eletrônico. Além de contribuir para minimizar os problemas da sobrecarga de informação, diminuindo o custo de busca do usuário, esses sistemas têm permitido que os usuários tenham uma experiência rica e nova de acesso a serviços personalizados, que “aprendem” sobre suas preferências, tornando-se melhores com o passar do tempo.

Muitos sistemas de recomendação estão disponíveis na Internet, podendo ser

classificados segundo seus principais aspectos: dados de entrada, saída, tipo de filtragem e técnica adotada (algoritmos utilizados).

Para implementar um sistema de recomendação cada um desses aspectos deve ser ponderado. Além disso, a análise do domínio de aplicação é imprescindível, pois suas particularidades devem ser consideradas na recomendação. No turismo, os sistemas de recomendação devem considerar as variáveis qualitativas que influenciam a tomada de decisão do turista.

Com base na análise do domínio de viagem e turismo, e nas características dos sistemas de recomendação, apresenta-se, no próximo capítulo, um sistema de recomendação de meios de hospedagem (SRMH).

### **3 PROCEDIMENTO METODOLÓGICO: SISTEMA DE RECOMENDAÇÃO DE MEIOS DE HOSPEDAGEM**

Este capítulo descreve o sistema de recomendação de meios de hospedagem. O desenvolvimento do sistema segue um procedimento metodológico que é apresentado na primeira seção do capítulo. Com base nele define-se os dados de entrada a serem utilizados, a saída do sistema, os algoritmos que serão implementados e a forma como os dados de entrada e os dados armazenados são combinados pelos algoritmos no processo de recomendação.

#### ***3.1 Caracterização da pesquisa***

Cada pesquisa tem um delineamento metodológico conforme o método de abordagem utilizado. O método de abordagem está relacionado ao tipo de raciocínio empregado: indutivo, dedutivo, hipotético-dedutivo e dialético (SANTAELLA, 2001).

Segundo Hyde (2000 *apud* SILVA, 2003), a razão indutiva é um processo de construção da teoria, enquanto a razão dedutiva é um processo de teste da teoria. Silva (2003) considera importante que um projeto de pesquisa tenha um estágio indutivo seguido de um dedutivo.

Nesta pesquisa o método indutivo é utilizado para a concepção do sistema de recomendação, detalhado nesse capítulo. A pesquisa apresenta também um estágio dedutivo, que se refere à aplicação e testes do sistema proposto. Este último estágio é detalhado no capítulo 4.

Com base em seus objetivos gerais, esta pesquisa pode ser classificada como pesquisa exploratória, porque busca proporcionar maior familiaridade com o problema, aprimorar idéias ou descobrir intuições (GIL, 2002).

Os passos adotados para concepção do sistema de recomendação proposto são apresentados a seguir.

### 3.2 *Passos da pesquisa*

Antes de implementar um sistema de recomendação é necessário definir os passos que devem ser percorridos para atingir tal objetivo. Com base nos aspectos discutidos no capítulo anterior é proposto um procedimento metodológico para a implementação de sistemas de recomendação, constituído dos seguintes passos<sup>10</sup>:

- **Escolha da saída:** refere-se à forma de apresentação das recomendações ao usuário.
- **Escolha do tipo de filtragem e técnica de recomendação:** consiste em definir, considerando as particularidades do problema, o tipo de filtragem mais adequado. Com isso, define-se uma das técnicas possíveis e os algoritmos a serem implementados.
- **Determinação do modelo de usuário:** corresponde à seleção dos dados de entrada utilizados pelo sistema para representar o usuário.
- **Definição do processo de recomendação:** consiste em definir a forma como os dados de entrada e os dados armazenados pelo sistema devem ser combinados pelos algoritmos para gerar recomendações. A partir disso, o sistema de recomendação pode ser representado por um modelo lógico.
- **Implementação e avaliação das recomendações:** uma vez implementado, o sistema deve ser avaliado para confirmar a eficiência de suas recomendações.

O sistema de recomendação de meios de hospedagem é desenvolvido com base neste procedimento metodológico. As primeiras quatro etapas são discutidas e apresentadas a seguir. A quinta, implementação e avaliação das recomendações, será tratada no capítulo 4.

---

<sup>10</sup> Um artigo com a proposta da estrutura metodológica foi aprovado para apresentação no 4º CONTECSI “*International Conference on Information Systems and Technology Management*” (OLIVEIRA, 2007).

### 3.2.1 Escolha da saída

Uma das primeiras definições necessárias à concepção de um sistema de recomendação é a forma como as recomendações serão apresentadas aos usuários. Caso se deseje adotar uma saída do tipo “usuários que se interessaram por  $X$  também se interessaram por  $Y$ ”, por exemplo, todas as demais definições do projeto serão influenciadas por essa escolha (dados de entrada, tipo de filtragem, técnica utilizada, e processo de recomendação).

Para o sistema de recomendação de meios de hospedagem (SRMH) deseja-se gerar como saída uma lista de  $N$  meios de hospedagem, ordenados segundo sua relevância ao usuário. Isto significa que se pretende predizer a relevância de cada um dos meios de hospedagem do catálogo para o usuário ativo (ver seção 2.2.2).

Seguindo essa lógica, todos os meios de hospedagem que compõem o catálogo podem ser apresentados ao usuário, dispostos em uma lista conforme o nível de relevância, ficando a seu critério selecionar ou não aqueles que são mais recomendados pelo sistema.

### 3.2.2 Escolha do tipo de filtragem e técnica de recomendação

Antes de definir os algoritmos que serão implementados, é preciso deixar claro qual a abordagem que será adotada na resolução do problema, ou seja, os tipos de filtragem que serão utilizados no sistema. Após definir os tipos de filtragem deve-se selecionar as técnicas de recomendação que serão aplicadas, e só então, os algoritmos a serem implementados.

#### Tipos de filtragem

Em princípio, qualquer tipo de filtragem (ver seção 2.2.3) poderia ser utilizado para desenvolver o sistema de recomendação de meios de hospedagem – lembrando que cada um deles apresenta vantagens e limitações específicas.

No entanto, a filtragem colaborativa e a baseada em conteúdo possuem técnicas mais consolidadas, com diversas aplicações práticas e amplo reconhecimento da comunidade científica.

O especial interesse na filtragem colaborativa, para o desenvolvimento do SRMH, deve-se ao fato desta ser uma tecnologia que permite aprender as preferências do usuário e realizar recomendações baseadas em dados do usuário e da comunidade de usuários (DAS et

al., 2007). Considera-se que essa “visão” colaborativa é a principal tendência das aplicações *Web* hoje em dia.

Mas antes de partir para a escolha das técnicas de recomendação é importante destacar alguns pressupostos para o desenvolvimento do SRMH. O primeiro é que os usuários fazem apenas uso esporádico do sistema, ou seja, não têm perfil de longo prazo e não acumulam avaliações suficientes para que suas preferências sejam conhecidas com base nelas. O segundo é que o contexto da viagem é mais relevante para a recomendação que o perfil sócio-econômico do usuário ou suas avaliações anteriores – considera-se que o contexto da viagem é composto por dados demográficos do usuário, características da viagem, restrições e preferências quanto aos meios de hospedagem.

Ainda que o sistema não aprenda as preferências do usuário com base em suas escolhas anteriores – até porque uma escolha anterior não é tão relevante, já que o contexto da viagem pode mudar a cada consulta – ele pode aprender com as escolhas feitas em cada contexto e oferecer recomendações com base no histórico gerado pela comunidade de usuários. Por isso, a filtragem colaborativa não deve ser implementada em sua formulação típica (usando perfis de longo prazo, constituídos pelo histórico de avaliações), e sim com base na comparação de contextos de viagem, que expressem as necessidades e preferências dos usuários em uma situação particular.

Como o sistema se baseia no histórico das escolhas feitas em diversos contextos de viagem, ele tende a melhorar sua qualidade com o tempo de utilização, característica desejável para o SRMH. Segundo Burke (2002), apenas os sistemas baseados em aprendizagem (colaborativo, baseados em conteúdo e demográfico) são adaptáveis, ou seja, podem melhorar com o tempo.

Nota-se ainda que cada consulta deve ser armazenada como instâncias do modelo de usuário, devendo conter informações contextuais e as escolhas de hospedagem relacionadas ao contexto. Um novo usuário, ao entrar com as informações contextuais sobre sua viagem, deve ser comparado a instâncias do modelo de usuário. As instâncias mais parecidas servirão de base para as recomendações.

Tendo clara a abordagem utilizada, é preciso selecionar uma técnica que suporte essas restrições.

## Técnicas de recomendação

A partir da escolha do tipo de filtragem, é possível definir qual técnica será utilizada, e assim selecionar os algoritmos a serem implementados.

As técnicas utilizadas para desenvolver sistemas de recomendação colaborativos podem ser divididas em dois grupos: baseadas em memória (*memory-based ou heuristic-based*) e baseadas em modelo (*model-based*) (ver seção 2.5). Adomavicius e Tuzhilin (2005) destacam as técnicas comumente adotadas em cada grupo:

- baseadas em memória: *nearest neighbor* (co-seno e correlação), *clustering* e teoria dos grafos;
- baseadas em modelo: redes bayesianas, *clustering*, redes neurais artificiais, regressão linear e modelos probabilísticos.

Para selecionar a técnica adequada é importante considerar alguns aspectos do projeto. O principal aspecto a ser considerado no desenvolvimento do SRMH é que será desenvolvido o primeiro protótipo funcional do sistema. Isso significa que não existem dados armazenados referentes a consultas realizadas pelos usuários. Além disso, o foco principal do projeto é gerar recomendações de qualidade em um domínio onde os usuários são esporádicos e as informações contextuais são altamente relevantes. Não é dado ênfase para a velocidade das recomendações, onde a escalabilidade e exigência computacional devem ser consideradas.

Sendo assim, a escolha recai sobre as técnicas baseadas em memória que, segundo Deshpande e Karypis (2004), tendem a produzir recomendações de melhor qualidade, além de serem as mais populares e mais utilizados na prática, empregando técnicas estatísticas na avaliação de similaridade e predição (SARWAR et al., 2001).

Para gerar as recomendações é preciso primeiro calcular a similaridade entre o usuário ativo e as instâncias do modelo de usuário, e com isso determinar a vizinhança, formada pelo grupo de usuários mais similares. Os vizinhos mais próximos são utilizados para o cálculo da predição (nesse caso, a relevância do meio de hospedagem), e as recomendações são então apresentadas ordenadas do mais para o menos relevante.

As duas formas mais comuns de se medir a similaridade são a *correlação* e o *co-seno* (vetor de similaridade) (SARWAR et al., 2000; TORRES, 2004). Também é possível medir a similaridade com base em uma análise heurística, como propõe Burke (2002).

Para que dois usuários tenham a similaridade calculada usando-se o co-seno, é preciso que eles sejam representados por vetores multi-dimensionais, formados pelas informações contextuais (*ic*) associadas a eles (ver Quadro 7).

$$\cos(A, B) = \frac{\sum_{ic} V_A \times V_B}{\sqrt{\sum_{ic} V_A \times \sum_{ic} V_B}} \quad 25$$

A correlação, por sua vez, baseia-se nas avaliações dos usuários (avaliações implícitas, relacionadas às escolhas dos meios de hospedagem) e pode ser calculada utilizando o coeficiente de correlação de Pearson.

$$cor(m, j) = \frac{\sum_i (A_{m,i} - \bar{A}_m)(A_{j,i} - \bar{A}_j)}{\sqrt{\sum_i (A_{m,i} - \bar{A}_m)^2 \sum_i (A_{j,i} - \bar{A}_j)^2}} \quad 26$$

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

$$\bar{A}_m = \frac{1}{|I_m|} \sum_{i \in I_m} A_{m,i} \quad 27$$

A medida de similaridade do co-seno é que será utilizada de fato para a formação da vizinhança, já que o usuário ativo não possui histórico de avaliações, necessárias para o cálculo da correlação. Uma vez calculada a vizinhança, a instância mais similar ao usuário ativo deve ser tomada como base para o cálculo da correlação, para que as escolhas feitas pelos vizinhos mais semelhantes sejam consideradas no cálculo de similaridade e predição.

Baseado em Vozalis e Margaritis (2004), o cálculo da similaridade ( $sim_{a,j}$ ) entre o usuário  $a$  e os demais usuários  $j$ , pode considerar tanto as informações contextuais quanto a correlação baseada em avaliações, conforme apresentado na Equação 28.

$$sim_{a,j} = \cos_{a,j} + \cos_{a,j} \times cor_{a,j} \quad 28$$

Para a formação da vizinhança, o cálculo do co-seno pode ser substituído por uma medida de similaridade heurística. Como o co-seno utiliza uma representação binária, ele desconsidera qualquer grau de semelhança, ou proximidade, entre dois atributos de uma variável. Por exemplo, se uma pessoa pertence a uma faixa etária entre 25 a 31 anos, e outra entre 32 a 40 anos, utilizando o co-seno eles teriam semelhança zero.

Utilizando uma medida de similaridade heurística, é possível considerar níveis de

semelhança entre os atributos de uma variável. Para isso, as variáveis devem ter seus atributos representados em uma tabela, onde o índice de similaridade entre cada um deles será assinalado, seguindo regras que se baseiam na análise do senso-comum ou na percepção de especialistas.

Conforme Burke (2002), a similaridade heurística é pouco sensível à magnitude dos valores escolhidos, mas altamente sensível ao seu significado, ou seja, é importante representar atributos opostos utilizando índices extremos de dissimilaridade. Sugere-se a utilização dos seguintes índices:

- 1: máximo de similaridade – considerando que sempre um atributo tem índice máximo em relação a si mesmo;
- 0,5: similares;
- 0: neutro;
- -0,5: não similares;
- -1: máximo de dissimilaridade – atributos que se opõem em similaridade.

O Quadro 4 apresenta um exemplo de como os atributos de uma variável, no caso a faixa etária, podem ser comparados utilizando uma heurística.

18 a 24	25 a 31	32 a 40	41 a 50	51 a 59	60 ou mais	
1	0,5	0,5	0	-0,5	-1	18 a 24
	1	0,5	0	-0,5	-1	25 a 31
		1	0,5	0	-0,5	32 a 40
			1	0,5	0	41 a 50
				1	0,5	51 a 59
					1	60 ou mais

**Quadro 4 – Matriz de similaridade para a faixa etária**

Fonte: elaborada pelo autor.

A similaridade heurística pode substituir o co-seno na formação da vizinhança, e assim ser utilizada no cálculo da similaridade apresentado na Equação 28.

A medida de similaridade heurística é determinada pela média das distâncias  $d(a,b)$  entre os atributos das variáveis de entrada (informações contextuais) – ver Equação 29. Assim, se uma pessoa pertence a uma faixa etária entre 25 a 31 anos, e outra entre 32 a 40 anos, elas teriam similaridade heurística igual a 0,5, conforme mostra o Quadro 4.

$$\text{similaridade heurística}(A,B) = h(A,B) = \frac{\sum_{t \in T} d(a,b)}{|T|}, \quad 29$$

onde  $T$  é o total de variáveis de entrada que representa os usuários.

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

$$r_{ai} = \frac{\sum_{j \in \hat{U}} sim_{aj} \times (A_{ji} - \bar{A}_j)}{\sum_{j \in \hat{U}} |sim_{aj}|}, \quad 30$$

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).

A próxima seção apresenta como deve ser o modelo de usuário para que as técnicas possam ser devidamente aplicadas.

### 3.2.3 Determinação do modelo de usuário

Os usuários devem ser representados pelos dados de entrada que fornecem ao sistema. Nos sistemas de recomendação, tais dados geralmente são tomados como avaliações ou votos, podendo ser obtidos de forma implícita ou explícita (ver seção 2.2.1).

Enquanto nas avaliações explícitas o usuário expressa sua opinião conscientemente, nas avaliações implícitas suas preferências são inferidas segundo sua interação com o sistema. No SRMH tanto as avaliações implícitas quanto as avaliações explícitas devem ser utilizadas como dados de entrada.

#### Avaliações implícitas

Uma interpretação de como as ações dos usuários podem ser adotadas como avaliações implícitas foi proposta por Nichols (1997). Para o SRMH, duas ações são particularmente interessantes para inferir o interesse do usuário por determinado item: a primeira é quando o usuário seleciona um meio de hospedagem para examinar informações mais detalhadas e a segunda é quando ele decide efetuar a reserva em uma das opções sugeridas. Neste contexto, sugere-se a seguinte escala para as avaliações implícitas:

- Solicitação de reserva. Avaliação = 1;
- Seleção do hotel para consulta de mais informações. Avaliação = 0.5;
- Hotéis que não foram selecionados ou reservados. Avaliação = 0.

A Figura 16 representa uma matriz de avaliações, ou matriz usuário-item, onde nas linhas estão distribuídos os usuários e nas colunas os meios de hospedagem. Os elementos da

matriz são as avaliações dos usuários. A relevância ( $r_{a,i}$ ) de cada meio de hospedagem do catálogo para o usuário ativo  $u_a$  é determinada com base nas avaliações feitas por usuários semelhantes.

	$h_1$	$h_2$	...	$h_i$	...	$h_n$
$u_1$	1	0		0.5		0
$u_2$	0	0.5		1		0.5
:						
$u_a$						
:						
$u_m$	0	1		0.5		1

**Figura 16 – Representação de uma matriz usuário-item**  
Fonte: elaborada pelo autor.

O SRMH não utiliza um esquema de voto explícito do usuário nos meios de hospedagem, principalmente por considerar baixa a frequência de compra dos usuários. O voto explícito implicaria que o usuário já conhecesse o meio de hospedagem, oferecendo um *feedback* sobre sua qualidade. Como levaria muito tempo para acumular esse tipo de voto explícito, esse projeto não o considera nos cálculos.

Por outro lado, acredita-se que utilizando avaliações implícitas é possível melhorar o processo de recomendação, visto que cada usuário, com seus interesses particulares, analisará uma lista de meios de hospedagem contendo informações preliminares sobre eles, tais como: localização, preço, descrição sucinta e foto. Portanto, suas escolhas representam um filtro que corresponde a suas preferências.

Apesar de as avaliações dos usuários serem fundamentais para o processo de recomendação colaborativa, no SRMH cada consulta será tratada como um novo usuário, com interesses específicos e, portanto, sem histórico de avaliações. Com isso, a similaridade não pode ser calculada com base nas avaliações, mas em variáveis de entrada (avaliações explícitas, representadas por informações contextuais) que representem os interesses dos usuários.

### Variáveis de entrada

As variáveis de entrada, aqui denominadas de informações contextuais, devem ser selecionadas considerando-se as particularidades dos sistemas de recomendação no domínio de viagens e turismo (seção 2.5).

A escolha das variáveis fundamenta-se na literatura sobre turismo (Andrade, 2000), nos modelos de decisão em viagem (HWANG; GRETZEL; FESENMAIER, 2002; FESENMAIER et al., 2002, 2003) e também nos dados estatísticos sobre o turismo doméstico no Brasil<sup>11</sup>.

Seguindo a proposta de Ricci et al. (2002), dois grupos de fatores devem ser considerados no SRMH. O primeiro se refere às características pessoais (idade, renda etc.), o segundo às características da viagem (acompanhantes, tempo de permanência etc.). Além desses, pode-se incluir outro grupo de fatores que exercem influência na tomada de decisão do turista: as características do meio de hospedagem (localização, preço etc.).

O Quadro 5 apresenta os três grupos de variáveis de entrada selecionados para a implementação do SRMH. Para fins de simplificação, o conjunto de variáveis de entrada (dados demográficos, contexto da viagem e restrições quanto à hospedagem) é denominado de informação contextual (*IC*).

<b>VARIÁVEIS DE ENTRADA</b>	
Características pessoais:	— <i>p1</i>
Características da viagem:	— <i>v1</i> — <i>v2</i> — <i>v3</i> — <i>v4</i>
Características da hospedagem:	— <i>h1</i>

**Quadro 5 – Variáveis de entrada**

Fonte: Elaborado pelo autor.

O modelo de usuário deve ser representado por um vetor  $V$ , que considere todas as informações contextuais.

O Quadro 6 mostra um exemplo, sugerido por Vozalis e Margaritis (2004), de como o usuário pode ser representado por um vetor de características.

---

<sup>11</sup> Caracterização e Dimensionamento do Turismo Doméstico no Brasil 2006. Disponível em:

<[http://200.189.169.141/site/arquivos/dados\\_fatos/demanda%20turistica/demanda\\_turismo\\_domestico\\_setembro\\_2006.pdf](http://200.189.169.141/site/arquivos/dados_fatos/demanda%20turistica/demanda_turismo_domestico_setembro_2006.pdf)>. Acesso em: 13 fev. 2007.

VARIÁVEIS	ATRIBUTOS DAS VARIÁVEIS	COMENTÁRIOS
1	idade $\leq$ 18	cada usuário pertence somente a um grupo, que deve ser tomado com o valor 1 (verdadeiro), enquanto os outros atributos da variável “idade” são marcados com 0 (falso).
2	18 < idade $\leq$ 29	
3	29 < idade $\leq$ 49	
4	idade > 49	
5	homem	da mesma forma, o atributo que descreve o sexo do usuário é assinalado com valor 1, o outro com 0.
6	mulher	
7 - 27	ocupação	um único atributo que descreve a ocupação do usuário é marcado com 1, o restante é 0.

**Quadro 6 – Representação do usuário como vetor multi-dimensional**

Fonte: Vozalis e Margaritis (2004: p. 4)

A representação do perfil de dois turistas, André e Bruna, como um vetor multi-dimensional, é ilustrada no Quadro 7.

	<i>p1</i>		<i>v1</i>		<i>v2</i>	<i>v3</i>		<i>v4</i>	<i>h1</i>
	<i>p1,1</i>	<i>p1,2</i>	<i>v1,2</i>	<i>v1,4</i>	<i>v2,3</i>	<i>v3,1</i>	<i>v3,3</i>	<i>v4,2</i>	<i>h1,4</i>
(V <sub>A</sub> ) André	1	0	1	0	1	1	0	1	1
(V <sub>B</sub> ) Bruna	0	1	0	1	1	0	1	1	1

**Quadro 7 – Modelo de usuário – vetor multi-dimensional**

Fonte: Elaborado pelo autor.

Tendo os usuários representados por um vetor multi-dimensional, pode-se calcular a similaridade entre eles aplicando a fórmula do co-seno (Equação 25).

A escolha das variáveis será discutida no capítulo 4, que trata da implementação do SRMH.

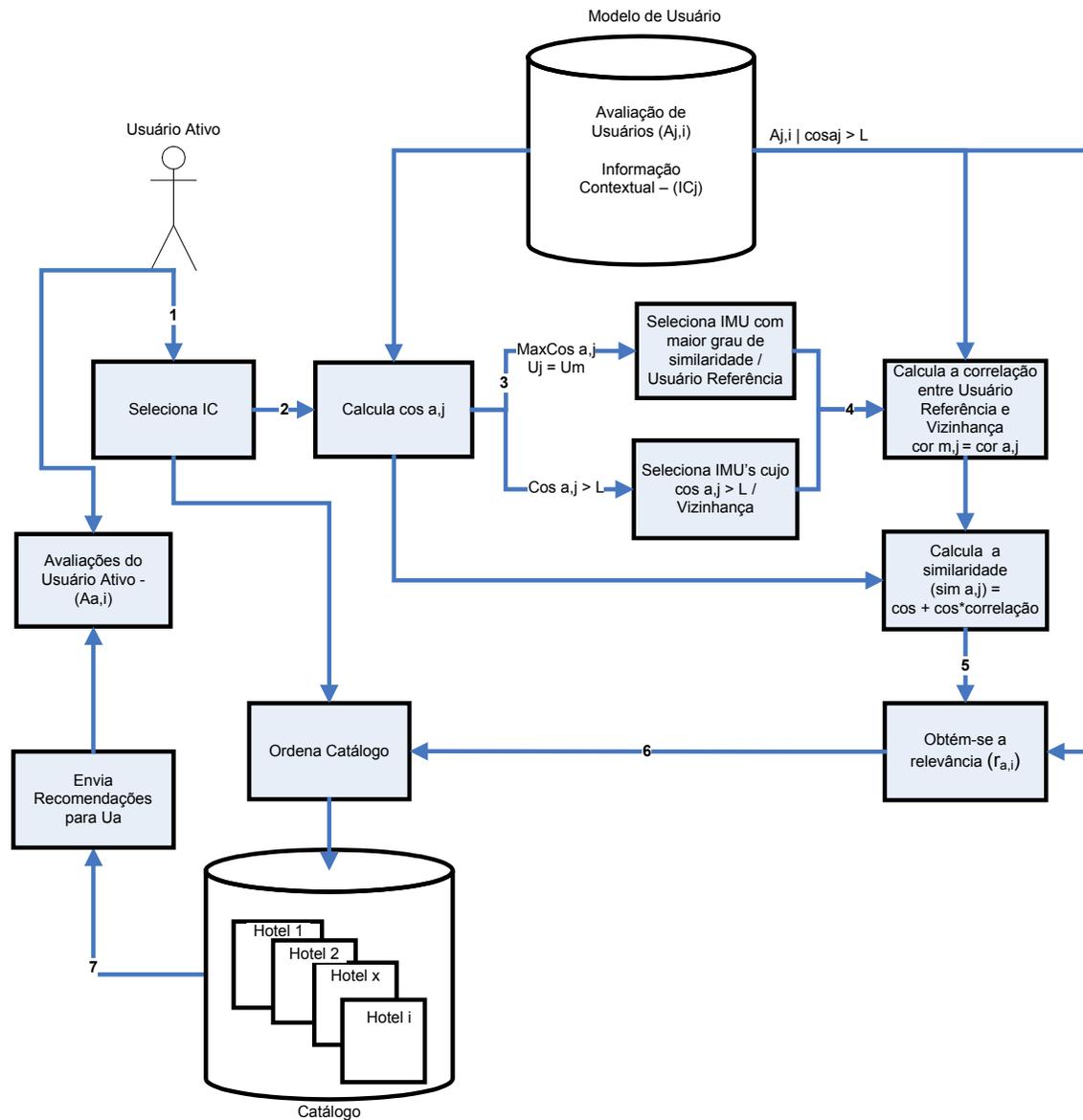
### 3.2.4 Definição do processo de recomendação

O processo de recomendação é definido pela maneira como os dados de entrada são combinados com os dados armazenados para medir a relevância de cada item para o usuário ativo.

No caso do SRMH, têm-se como dados de entrada: as informações contextuais (*IC*), em um primeiro momento, e as avaliações dos usuários (*A<sub>j,i</sub>*), após a interação desses com as recomendações recebidas. Os dados armazenados, por sua vez, englobam o catálogo de meios

de hospedagem e o modelo de usuário, composto por instâncias que contêm informações contextuais e avaliações.

A Figura 17 representa o processo de recomendação do sistema de recomendação de meios de hospedagem.



**Figura 17 – Representação do SRMH**

Fonte: Elaborada pelo autor.

As etapas do processo são descritas a seguir:

1. O processo se inicia quando o usuário seleciona as ICs de acordo com seu perfil, preferências e interesse.

2. Usando o cálculo do co-seno (Equação 25), se estabelece a similaridade entre o usuário ativo e as instâncias do modelo de usuário (IMU). Para as próximas etapas, considera-se apenas os vizinhos mais próximos, ou seja, as instâncias que tiverem um grau de similaridade acima de um limite  $L$ . A formação da vizinhança também pode ser feita substituindo o co-seno pela similaridade heurística.
3. A instância mais similar ao usuário ativo é tomada como um “usuário referência” ( $U_m$ ), assim a correlação entre este usuário e outras instâncias ( $cor_{m,j}$ ) pode ser calculada com base nas avaliações acumuladas por eles (Equação 26).
4. A  $cor_{m,j}$  é igualada a  $cor_{a,j}$  para que a correlação baseada nas avaliações dos usuários seja considerada no cálculo da similaridade ( $sim_{a,j}$ ), embora tenha menos importância que o co-seno, como pode ser visto na Equação 28.
5. Aplica-se a Equação 30 para calcular o índice de relevância de cada meio de hospedagem ( $r_{a,i}$ ) com base na similaridade ( $sim_{a,j}$ ) e nas avaliações ( $A_{j,i}$ ) dos vizinhos mais próximos.
6. Tanto o índice de relevância ( $r_{a,i}$ ) quanto as restrições impostas pelo usuário em relação às características da hospedagem, são utilizadas para ordenar os meios de hospedagem do catálogo.
7. Os  $N$  meios de hospedagem mais recomendados ao usuário são apresentados em uma lista. O processo de recomendação termina quando o usuário avalia as recomendações recebidas fazendo suas escolhas. Neste momento, seu perfil passa a integrar o modelo de usuário do sistema.

### ***3.3 Considerações finais sobre o capítulo***

O procedimento metodológico proposto orientou as escolhas e definições para a concepção do SRMH. Após concluir as quatro primeiras etapas do procedimento metodológico foi possível partir para a implementação do sistema e avaliação das suas recomendações, que correspondem à quinta etapa, discutida no próximo capítulo.

## 4 APLICAÇÃO DO SISTEMA

Este capítulo trata da aplicação do sistema de recomendação de meios de hospedagem. Um protótipo funcional do SRMH foi implementado na agência Tropix Turismo Inteligente. Uma avaliação preliminar do sistema foi realizada e os resultados são discutidos no final do capítulo.

### 4.1 Tropix Turismo Inteligente

A Tropix Turismo Inteligente é uma agência especializada na comercialização on-line de meios de hospedagem. A empresa foi fundada em 2004, com o propósito inicial de ser o canal de comércio eletrônico do portal Guia Floripa<sup>12</sup> – *site* de informações turísticas da cidade de Florianópolis, lançado em 1996.

Florianópolis, capital de Santa Catarina, é um dos principais destinos turísticos do Brasil. Na temporada de verão de 2006/2007 a cidade recebeu cerca de 780 mil turistas<sup>13</sup>. Também é crescente o número de turistas que vêm a Florianópolis participar de eventos, feiras, congressos ou a negócios. Recentemente a capital entrou na lista das cidades brasileiras que mais cediam eventos internacionais<sup>14</sup>.

A rede hoteleira de Florianópolis é composta de aproximadamente 62 hotéis e 20 pousadas, totalizando 10.400 unidades habitacionais e 28.500 leitos<sup>15</sup>.

O *site* da Tropix oferece um catálogo de cerca de 40 meios de hospedagem, incluindo hotéis, pousadas, chalés, entre outros. Na relação estão incluídas as principais opções de hospedagem de Florianópolis, abrangendo todas as regiões da cidade, vários portes de estabelecimentos e grande diversidade de preços e serviços.

---

<sup>12</sup> Disponível em: <<http://www.guiafloripa.com.br>>. Acesso em: 23 abr. 2007.

<sup>13</sup> Disponível em: <<http://www.sol.sc.gov.br/santur/FrameDemanda2007.asp?Link=Floripa.htm>>. Acesso em: 23 abr. 2007.

<sup>14</sup> Disponível em: <[http://www.revistafator.com.br/ver\\_noticia.php?not=8197](http://www.revistafator.com.br/ver_noticia.php?not=8197)>. Acesso em: 23 abr. 2007.

<sup>15</sup> Disponível em: <<http://www.sindicatohrbs-fpolis.org.br/index.php?codwebsite=&codpagina=00000415>>. Acesso em: 23 abr. 2007.

No *site* o usuário pode consultar e reservar o hotel<sup>16</sup> de seu interesse. Para isso, é necessário informar o número de pessoas, a data de chegada e a data de saída (dados obrigatórios); outras preferências também podem ser adicionadas à consulta, como localização, tipo de hospedagem, nome e serviços especiais. O resultado da consulta traz uma relação de meios de hospedagem, cada um contendo nome, foto ilustrativa, breve descrição, endereço e as unidades habitacionais com suas respectivas tarifas para o período solicitado.

Os hotéis são listados em uma única página, ordenados pelo nome – ordem alfabética (ver Figura 18). Diante das opções, o usuário pode solicitar a reserva em uma das unidades habitacionais de um determinado hotel ou acionar o botão “mais detalhes”, caso deseje ver informações detalhadas do hotel antes de efetuar a reserva. Vale lembrar que os hotéis podem ser reordenados pelo menor preço, pela localização ou pelo tipo de hospedagem (hotel, pousada, chalé etc.).

The screenshot shows the Tropix Turismo Inteligente website interface. At the top, there is a navigation menu with links for HOME, EMPRESA, FLORIANÓPOLIS, CONSULTA (highlighted), and ATENDIMENTO. Below the navigation is the Tropix logo and a search form. The search form includes fields for Check-in (27 abr 2007), Check-out (01 mai 2007), adultos (02), and crianças. There is also a 'CONSULTAR' button. To the right of the search form is a 'PREFERÊNCIAS' section with checkboxes for 'serviços especiais' (aceita animais, acessibilidade, ar refrigerado, estacionamento, internet, piscina). Below the search form is the 'RESULTADO DA CONSULTA' section, which includes a 'Listar por:' dropdown menu and an 'ENVIAR ESTE RESULTADO P/ UM E-MAIL' button. The search results show 'Foram encontrados 40 estabelecimentos para os seguintes parâmetros de consulta: Data de check-in: 27/04/2007, Data de check-out: 01/05/2007, Número de adultos: 2, Número de crianças: 0'. The first result is 'BAÍA NORTE OTHON CLASSIC HOTEL' with a photo and address: 'Rua Almirante Lamego, 220 - Centro'. Below the hotel name is a table of 'UNIDADES HABITACIONAIS' with two options: 'APARTAMENTO STANDARD FRENTE PISCINA' and 'APARTAMENTO STANDARD FRENTE MAR'. Each option has a table of daily rates for sexta, sábado, domingo, and segunda, and a 'TOTAL' column. The total price for both is 473,95. There is a 'RESERVAR' button next to each total.

Figura 18 – Tela da Tropix Turismo Inteligente – Resultado da consulta  
Fonte: www.tropix.com.br

<sup>16</sup> O termo “hotel” será usado com um sentido mais amplo, significando meios de hospedagem.

A implementação do SRMH na Tropix deve permitir uma consulta diferenciada, onde o usuário entrará com informações contextuais sobre sua viagem e receberá uma lista dos hotéis mais recomendados para ele, ordenados pelo nível de relevância.

A aplicação do sistema de recomendação na Tropix será detalhada na próxima seção.

## 4.2 Implementação

Tendo claras as definições a respeito das características do SRMH, discutidas no capítulo 3, é possível partir para a implementação do sistema. Porém, antes disso, é preciso identificar as variáveis de entrada que influenciam o processo de tomada de decisão do turista, para que estas sejam incorporadas ao sistema.

Os fundamentos para a escolha das variáveis são apresentados no Apêndice A. O Quadro 8 apresenta as variáveis selecionadas.

VARIÁVEIS DE ENTRADA	
Características pessoais:	— <i>Idade</i>
Características da viagem:	— <i>Motivação da Viagem</i> — <i>Grupo</i> — <i>Transporte no destino</i> — <i>Período</i>
Características da hospedagem:	— <i>Serviços e instalações</i> — <i>Localização</i> — <i>Preço</i>

**Quadro 8 – Variáveis de entrada**

Fonte: Elaborado pelo autor.

Com as variáveis de entrada definidas é possível criar a interface para a consulta de recomendações e gerar as tabelas necessárias para armazenar os dados dos usuários. Também é possível estabelecer a similaridade heurística entre os atributos de cada uma das variáveis, como mostra o Apêndice B.

O Quadro 9 ilustra a representação de dois usuários no sistema.

	Idade		Motivo		Grupo	Transporte		Período	Serviços		Localização		Preço
	25 a 31	41 a 50	Sol e praia	Visita a parentes	Com parceiro (a)	Veículo alugado	Veículo próprio	Dez / Fev	Internet	Ar-condicionado	Centro	Praia Brava	Pouco importante
André	1	0	1	0	1	1	0	1	1	0	1	0	1
Bruna	0	1	0	1	1	0	1	1	0	1	0	1	1

**Quadro 9 – Representação do usuário**

Fonte: Elaborado pelo autor.

Para receber recomendações de hospedagem, o usuário deve entrar com as informações que correspondem ao seu perfil, sua viagem e seus interesses. A Figura 19 apresenta a tela de consulta do SRMH.

Cosseno Preferências:

Idade: Nenhum anos. Período: Nenhum

Motivação da viagem: Nenhum

Grupo de viagem: Nenhum

Transporte utilizado no destino: Nenhum Preço baixo: Nenhum

TOP-N 5 Limpar Buscar

**Figura 19 – Tela de consulta do SRMH**

Fonte: Elaborado pelo autor.

Ao acionar o botão “buscar” o sistema gera a lista dos hotéis mais recomendados ao usuário, apresentando primeiramente os *TOP-N* (o valor *default* para *N* é cinco), ordenados conforme seu nível de relevância – porcentagem que sinaliza o quanto o hotel é indicado ao usuário. A apresentação das recomendações é mostrada na Figura 20<sup>17</sup>.

<sup>17</sup> Protótipo de tela da lista de recomendações.

**tropix**  
Turismo Inteligente

**RESULTADO DA CONSULTA**  
• Suas Recomendações

ENVIAR ESTE RESULTADO P/ UM E-MAIL  
seu e-mail  OK

**MAJESTIC PALACE HOTEL**  
Com localização privilegiada, na esquina da Av. Beiramar com a Av. Mauro Ramos, ao lado do Beira Mar Shopping, este hotel, reúne sofisticação e conforto a uma vista deslumbrante e a facilidade de deslocamento.  
Rua Bocaiúva, 2304 - Centro  
Hotel : Região Central : Centro  
A partir de R\$ 220,00  
MAIS DETALHES

**BLUE TREE TOWERS FLORIANÓPOLIS**  
Localizado ao lado do Beiramar Shopping, uma das regiões mais nobres da cidade, o Blue Tree está próximo ao centro, a bares, boates, cinemas, teatros, ao comércio em geral e pontos turísticos.  
Rua Bocaiúva, 2304 - Centro  
Hotel : Região Central : Centro  
A partir de R\$ 188,00  
MAIS DETALHES

**Figura 20 – Recomendações do SRMH**

Fonte: Elaborado pelo autor.

Diante das recomendações, o usuário pode ver informações sobre os hotéis, como localização, descrição, foto, preço mínimo, além da porcentagem que indica o quanto o hotel é relevante para ele. Se desejar ter mais informações sobre um hotel, ou efetuar a reserva, ele deve acionar o botão “mais detalhes”. Este conduzirá o usuário a uma página específica do hotel, onde o usuário terá acesso a informações detalhadas e poderá solicitar a reserva em uma das unidades habitacionais existentes.

A página específica do hotel, onde o usuário pode efetuar a reserva, é apresentada na Figura 21.

**tropix**  
Turismo Inteligente

HOME EMPRESA FLORIANÓPOLIS CONSULTA ATENDIMENTO

HOTEL MAJESTIC PALACE HOTEL

Com localização privilegiada, na esquina da Av. Beira Mar com a Av. Mauro Ramos, ao lado do Beiramar Shopping, este hotel, reúne sofisticação e conforto a uma vista deslumbrante e a facilidade de deslocamento.

São 261 apartamentos planejados e decorados para o bem-estar e conforto dos hóspedes. Possui um restaurante com cozinha internacional e um café com espaço especial para se contemplar o fim de tarde ou para uma reunião com amigos.

Possui um centro de eventos com capacidade para até 400 pessoas, quatro salas de reunião, sendo um auditório com capacidade para até 90 pessoas e um local especial para coquetéis e jantares, localizado na cobertura, com capacidade para até 300 pessoas. Além disso, possui um heliponto no topo do prédio.

**UNIDADES HABITACIONAIS**

**SUÍTE JUNIOR**  
**Descrição:** Amplas suítes com 30m<sup>2</sup>, localizadas no 14° e 15° andares, com vista para o mar. Cama de casal king size e sofá-cama, além de Jacuzzi são os destaques.  
**Capacidade padrão:** 1 pessoa(s) (máximo: 2 pessoa(s))  
**Características:** ar refrigerado, banheiro, cama extra, canais a cabo, frigobar, hidromassagem, internet banda larga, roupas c/m/b, secador de cabelo, telefone, televisão 25"

CH	segunda	terça	TOTAL	RESERVAR
	574,00	574,00	1.148,00	

**SUÍTE EXECUTIVA**  
**Descrição:** Amplas suítes com 45m<sup>2</sup>, localizadas do 11° ao 14° andares, de frente para o mar. São divididas em 2 ambientes, sendo quarto com cama de casal e sala com sofá-cama.  
**Capacidade padrão:** 1 pessoa(s) (máximo: 2 pessoa(s))  
**Características:** ar refrigerado, banheiro, cama extra, canais a cabo, frigobar, hidromassagem, internet banda larga, roupas c/m/b, secador de cabelo, telefone, televisão 25"

CH	segunda	terça	TOTAL	RESERVAR
	640,00	640,00	1.280,00	

**APARTAMENTO EXECUTIVO**  
**Descrição:** Amplas apartamentos com 30m<sup>2</sup>, localizados do 1° ao 10° andares, com vista para a

**CONSULTAR OUTRO PERÍODO**

Check-in: 30 abr 2007  
 Check-out: 02 mai 2007  
 adultos: 02 crianças: 0

**FOTOS DO ESTABELECIMENTO**

Figura 21 – Página com informações detalhadas do hotel

Fonte: www.tropix.com.br.

Os procedimentos referentes à reserva de hospedagem em si não são de interesse dessa pesquisa e por isso não serão detalhados.

Após concluir a consulta, efetuando ou não a reserva, os dados de entrada são armazenados, juntamente com as escolhas feitas pelo usuário – hotéis selecionados para ver mais detalhes e/ou efetuar reserva. Assim, a consulta passa a integrar o modelo de usuário do sistema, servindo de base para as próximas recomendações.

### 4.3 Avaliação das recomendações

Uma prática comum para avaliar recomendações *Top-N* baseia-se no uso de duas métricas largamente utilizadas em sistemas de Recuperação de Informação (*Information Retrieval - IR*), chamadas de *recall* (abrangência ou revocação) e *precision* (precisão). (SARWAR et al., 2000; VOZALIS; MARGARITIS, 2003).

Para o caso de sistemas que geram uma lista de recomendações *Top-N*, é necessário realizar alguns ajustes na definição padrão encontrada na área de *IR*. Nos sistemas de recomendação o objetivo é apresentar um número fixo de *N* itens relevantes ao usuário.

Para calcular as medidas de *abrangência* e *precisão*, primeiro é preciso dividir os dados armazenados no modelo de usuário em dois conjuntos – um conjunto de treinamento e um conjunto de teste.

O sistema opera somente com o conjunto de treinamento. Para cada nova entrada do conjunto de teste o sistema gera uma lista de recomendações, chamada de conjunto *top-N*. O principal objetivo da avaliação é comparar os itens presentes no conjunto de teste com aqueles exibidos na lista de recomendações. O conjunto dos itens que aparecem em ambos é denominado de conjunto *hit*.

Para os sistemas de recomendação *Top-N* pode-se definir *abrangência* e *precisão* como:

- *Abrangência (recall)*: a razão do número de itens do conjunto *hit* sobre o número de itens do conjunto de teste.

$$recall = \frac{|\text{conjunto hit}|}{|\text{conjunto teste}|} = \frac{|\text{teste} \cap \text{top} - N|}{|\text{teste}|} \quad 31$$

- *Precisão (precision)*: a razão do número de itens do conjunto *hit* sobre o número de itens do conjunto *top-N*.

$$precision = \frac{|\text{conjunto hit}|}{|\text{conjunto top} - N|} = \frac{|\text{teste} \cap \text{top} - N|}{N} \quad 32$$

Estas duas medidas são conflitantes em sua natureza. Isto é, o aumento de  $N$  normalmente resulta no aumento da *abrangência*, e ao mesmo tempo na diminuição da *precisão*. Como as duas medidas são importantes para avaliar a qualidade das recomendações *Top-N*, elas podem ser combinadas em uma medida denominada de *F1*. A métrica padrão da *F1* considera o mesmo peso para ambas as medidas: *abrangência* e *precisão*, como mostra a Equação 33.

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \quad 33$$

A medida *F1* pode ser calculada para cada usuário individualmente e a média sobre todos os usuários representa um valor geral para o sistema de recomendação.

### 4.3.1 Método de avaliação

O método adotado para fazer a avaliação preliminar do sistema, antes de disponibilizá-lo para o público, consiste nas seguintes etapas:

- Gerar um conjunto de 25 perfis de turistas.
- Solicitar a especialistas que indiquem as melhores opções de hospedagem para cada um desses perfis.
- Lançar os perfis com as avaliações dos especialistas no sistema, simulando consultas reais de usuários.
- Separar as instâncias do modelo de usuário em dois grupos: conjunto de treinamento e conjunto de teste.
- Manter o conjunto de treinamento armazenado no sistema e utilizar o conjunto de teste para simular novas consultas de usuários.
- Comparar as recomendações oferecidas aos usuários do conjunto de teste, com os itens previamente selecionados por estes usuários, utilizando as medidas *abrangência* e *precisão*.

Os 25 perfis de turistas foram gerados com base no histórico de atendimento da Tropix Turismo, e considera diversas possibilidades de perfis e contextos de viagem. Cada perfil é representado pela combinação de seis variáveis de entrada: faixa etária, preço baixo, motivo da viagem, grupo, transporte e período. Optou-se por não utilizar as variáveis “localização” e “serviços e instalações” na avaliação preliminar, a fim de privilegiar a análise subjetiva dos especialistas. O uso dessas variáveis para o método heurístico também se tornaria muito dispendioso, devido ao grande número de atributos destas. A relação dos perfis selecionados pode ser vista no Apêndice C.

A aplicação de uma análise subjetiva ao processo de previsão deve ser feita de uma maneira estruturada, pela utilização de métodos qualitativos. As previsões de métodos qualitativos são resultados da opinião de entrevistados ou da opinião de especialistas (LEMOS, 2006).

As estimativas dos especialistas são baseadas no julgamento, intuição, pesquisas, técnicas comparativas, conhecimento técnico, conhecimento sobre análise de dados e procedimentos de previsão, e/ou no conhecimento de relações de causa e efeito entre variáveis adquirido com a experiência em processos preditivos nas organizações (WEBBY; O’CONNOR, 1996; BALLOU, 2001 *apud* LEMOS, 2006).

As pesquisas qualitativas utilizam uma amostra pequena, muitas vezes não representativa dos respondentes, com o intuito de refinar um conceito, conhecer reações, aprender a linguagem do cliente ou explorar novas idéias e oportunidades (SILVA, 2003).

Há duas aproximações gerais para a razão que podem gerar aquisição de novos conhecimentos, nominalmente, a razão indutiva e a razão dedutiva. A razão indutiva é um processo de construção da teoria, começando com observações de eventos específicos, buscando estabelecer generalizações sobre o fenômeno sendo investigado. A razão dedutiva é um processo de teste da teoria, o qual inicia com uma teoria estabelecida ou generalização e procura ver se a teoria se aplica a eventos específicos (HYDE, 2000 *apud* SILVA, 2003). Segundo Silva (2003), o que é importante em qualquer programa de pesquisa não é um estágio qualitativo, seguido de um quantitativo, mas sim um estágio indutivo seguido de um dedutivo.

Para Rowe e Wright (2001 *apud* SILVA, 2003), a opinião de especialistas é freqüentemente necessária na tarefa de previsão, porque faltam informações apropriadas para usar procedimentos estatísticos. Segundo Armstrong (2001 *apud* SILVA, 2003), em muitas situações, o primeiro passo é perguntar aos especialistas. Algumas vezes isto é suficiente, pois os especialistas podem produzir excelentes previsões.

Morris (1977 *apud* SILVA, 2003) define especialista como “qualquer pessoa que tenha conhecimentos especiais sobre certo evento”. Para Silva (2003), a técnica de opinião de especialistas consiste em obter e agregar de alguma forma a opinião desses especialistas.

Segundo Armstrong (1986 *apud* SILVA, 2003), a opinião de muitos especialistas é melhor que a opinião de um, por isso, sugere que o número ótimo de especialistas seja entre cinco e vinte.

No método de avaliação proposto, os especialistas não avaliam o SRMH; eles analisam os perfis de turistas e procuram predizer suas escolhas, ou seja, indicam as opções de hospedagem mais adequadas para cada perfil. Desta forma, é possível simular consultas reais de usuários e prover o sistema com dados iniciais relevantes e coerentes, e com isso realizar uma avaliação preliminar das recomendações apresentadas pelo sistema.

O conjunto formado pelos 25 perfis de turistas foi apresentado a três especialistas da agência Tropix Turismo, que individualmente realizaram uma análise subjetiva para indicar as opções mais apropriadas para cada perfil. Os critérios que orientaram a avaliação dos especialistas são:

- Somente os meios de hospedagem presentes no catálogo da Tropix Turismo podem ser recomendados – 40 meios de hospedagem no total.
- Devem ser recomendados no mínimo 03 e no máximo 04 hotéis. O primeiro hotel a ser recomendado deve ser o mais indicado. Isto é, considera-se que o usuário selecionou todas as opções indicadas e reservou a primeira.
- Para as viagens motivadas por negócio, evento ou estudo, considerar que estes ocorrem na região central da cidade.
- Entender a variável “Preço Baixo” como uma pergunta feita ao cliente, do tipo: “é importante o hotel ter preço baixo?”. As respostas possíveis são: é indispensável, é importante, é pouco importante, é irrelevante.
- A variável “Transporte” considera a maneira como o turista se deslocará na cidade.
- O mesmo hotel pode ser indicado para diferentes perfis, desde que o especialista considere a recomendação coerente.

Ao final da avaliação dos especialistas, cada perfil de turista terá no máximo 12 indicações distintas (caso todos os especialistas indiquem o máximo de opções possíveis e estas não sejam coincidentes). As avaliações dos especialistas estão apresentadas no Apêndice C.

As respostas dos especialistas são lançadas no sistema, simulando a entrada de dados de usuários reais, de tal forma que as recomendações dos especialistas representem as escolhas feitas pelos usuários. Assim, são incluídas 75 instâncias no modelo de usuário do sistema (25 de cada especialista).

A avaliação do SRMH pode então ser realizada utilizando-se as métricas de *abrangência* e *precisão*. Para isso 80% das instâncias, 60 no total, são mantidas no sistema formando o conjunto de treinamento, enquanto 20% das instâncias, 15 no total, são separadas aleatoriamente em um conjunto de teste.

Cada instância do conjunto de teste é lançada no sistema como sendo um novo usuário (usuário ativo); o sistema gera as recomendações utilizando como *background* o conjunto de treinamento e os hotéis disponíveis; as recomendações oferecidas ao usuário ativo são comparadas com os itens previamente selecionados por ele (na entrada simulada), e com isso é possível ter as medidas de *abrangência* e *precisão*.

As instâncias de teste lançadas no sistema são incorporadas ao conjunto de

treinamento, de tal forma que a 15ª instância de teste seja comparada a 74 instâncias no conjunto de treinamento.

A avaliação realizada compara duas técnicas colaborativas para a formação da vizinhança: o co-seno e a similaridade heurística. Para atestar a qualidade das técnicas colaborativas é utilizada também a recomendação dos hotéis mais populares (POP), ou seja, os hotéis são ordenados com base na média de votos recebidos por eles, servindo assim como parâmetro para avaliar os outros algoritmos. Como afirmam Shardanand e Maes (1995), um algoritmo de recomendação colaborativa não é personalizado e nem preciso se não oferecer uma melhoria significativa sobre a recomendação dos itens mais populares. O mesmo parâmetro de comparação é usado por outros autores, entre eles Breese, Heckerman e Kadie (1998).

A mesma avaliação foi realizada com três conjuntos de teste distintos, selecionados aleatoriamente<sup>18</sup>, conforme apresentado no Apêndice C. Para cada conjunto de teste foram efetuadas medidas para as recomendações *Top-3*, *Top-5* e *Top-10*. Um ambiente de teste<sup>19</sup>, não acessível ao público, foi utilizado para a avaliação do sistema.

### 4.3.2 Resultados

Os resultados obtidos com o conjunto de teste 1 são apresentados na Tabela 1 e no Gráfico 1.

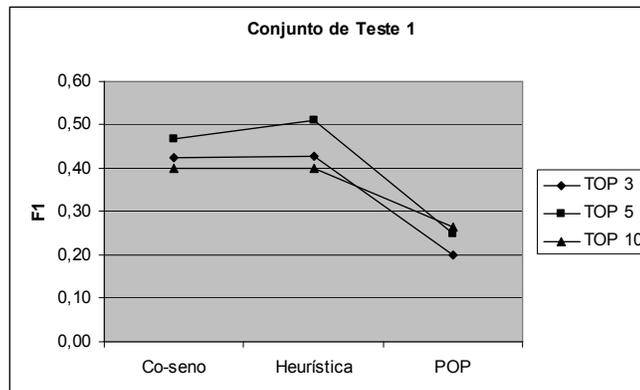
	Co-seno			Heurística			POP		
	<i>R</i>	<i>P</i>	<i>FI</i>	<i>R</i>	<i>P</i>	<i>FI</i>	<i>R</i>	<i>P</i>	<i>FI</i>
<b>TOP 3</b>	0,39	0,47	0,43	0,40	0,47	0,43	0,18	0,22	0,20
<b>TOP 5</b>	0,57	0,40	0,47	0,62	0,44	0,51	0,30	0,21	0,25
<b>TOP 10</b>	0,77	0,27	0,40	0,77	0,27	0,40	0,51	0,18	0,26

**Tabela 1 – Resultados obtidos com o conjunto de teste 1**

Fonte: Elaborado pelo autor.

<sup>18</sup> Foi utilizado um sistema de sorteio (Sorteio Virtual - Mais!) para a seleção dos conjuntos de teste. Disponível em: <<http://www.ihs.rg.com.br>>. Acesso em: 01 maio 2007.

<sup>19</sup> Disponível em: <<http://www.tropix.com.br/teste>>. Acesso em: 01 maio 2007.



**Gráfico 1 – Resultados obtidos com o conjunto de teste 1**

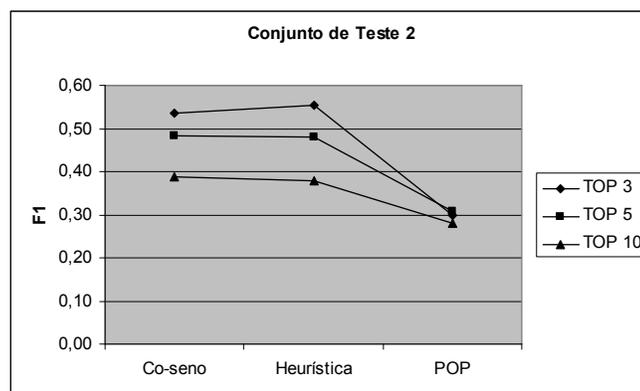
Fonte: Elaborado pelo autor.

Os resultados obtidos com o conjunto de teste 2 são apresentados na Tabela 2 e no Gráfico 2.

	Co-seno			Heurística			POP		
	<i>R</i>	<i>P</i>	<i>F1</i>	<i>R</i>	<i>P</i>	<i>F1</i>	<i>R</i>	<i>P</i>	<i>F1</i>
TOP 3	0,52	0,56	0,53	0,53	0,58	0,55	0,29	0,31	0,30
TOP 5	0,62	0,40	0,48	0,61	0,40	0,48	0,40	0,25	0,31
TOP 10	0,78	0,26	0,39	0,76	0,25	0,38	0,58	0,19	0,28

**Tabela 2 – Resultados obtidos com o conjunto de teste 2**

Fonte: Elaborado pelo autor.



**Gráfico 2 – Resultados obtidos com o conjunto de teste 2**

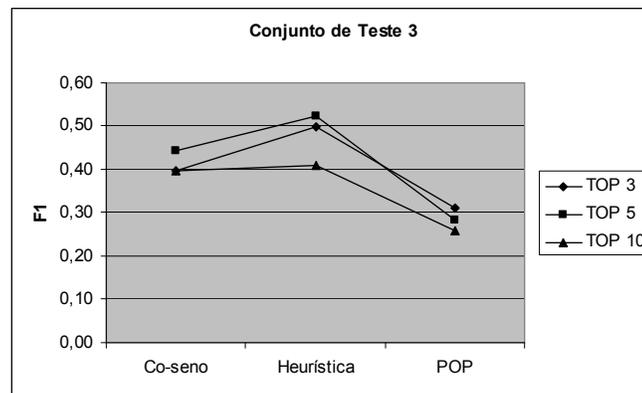
Fonte: Elaborado pelo autor.

Os resultados obtidos com o conjunto de teste 3 são apresentados na Tabela 3 e no Gráfico 3.

	Co-seno			Heurística			POP		
	<i>R</i>	<i>P</i>	<i>F1</i>	<i>R</i>	<i>P</i>	<i>F1</i>	<i>R</i>	<i>P</i>	<i>F1</i>
TOP 3	0,38	0,42	0,40	0,47	0,53	0,50	0,29	0,33	0,31
TOP 5	0,55	0,37	0,44	0,65	0,44	0,52	0,35	0,24	0,28
TOP 10	0,79	0,27	0,40	0,82	0,27	0,41	0,51	0,17	0,26

**Tabela 3 – Resultados obtidos com o conjunto de teste 3**

Fonte: Elaborado pelo autor.



**Gráfico 3 – Resultados obtidos com o conjunto de teste 3**

Fonte: Elaborado pelo autor.

Os três conjuntos de teste, embora apresentem pequenas variações nos resultados, possuem comportamento semelhante. Observa-se que as técnicas colaborativas, utilizadas com informações contextuais, permitem oferecer recomendações personalizadas aos usuários, apresentando significativa melhoria de qualidade em relação à recomendação dos itens mais populares.

A técnica baseada na similaridade heurística apresenta resultados ligeiramente melhores que o co-seno. Como a avaliação preliminar limita-se a poucos usuários no conjunto de treinamento e teste, não é possível afirmar que a similaridade heurística é superior ao co-seno.

Nota-se que o tempo de resposta para a recomendação baseada na similaridade heurística é maior que para o co-seno. O desempenho computacional dos algoritmos é um fator que se deve considerar em testes futuros.

Os resultados obtidos com a recomendação *Top-5* foram melhores que os demais, salvo no conjunto de teste 2, onde a recomendação *Top-3* mostrou-se superior.

### 4.3.3 Considerações finais sobre o capítulo

A aplicação do SRMH no *site* da Tropix Turismo permitiu avaliar a abordagem proposta e verificar que um sistema de recomendação de meios de hospedagem baseado em técnicas colaborativas e informações contextuais permite oferecer recomendações personalizadas conforme o perfil do usuário e o contexto da sua viagem.

No capítulo 5 encontram-se as conclusões sobre a pesquisa realizada e recomendações para trabalhos futuros.

## 5 CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Este trabalho abordou o desenvolvimento de um sistema de recomendação de meios de hospedagem, explorando o uso de informações contextuais e a aplicação de técnicas colaborativas no domínio do turismo.

Este capítulo apresenta as conclusões da pesquisa e sugestões para a continuidade deste trabalho.

### 5.1 Conclusões

Os experimentos realizados mostraram que é viável desenvolver sistemas voltados à personalização de *sites* de turismo utilizando filtragem colaborativa e informações contextuais. O sistema de recomendação de meios de hospedagem, quando aplicado a um *site* de reservas on-line de hotéis, pode trazer benefícios aos usuários, no sentido de oferecer a esses indicações que correspondam a seus interesses e necessidades. Conseqüentemente, tende a aumentar a taxa de efetivação de reservas, além de contribuir para a fidelização do cliente, gerando ganhos para a empresa.

Por considerarem que os usuários de um *site* de comércio eletrônico em turismo fazem apenas uso esporádico do sistema, os pesquisadores costumam contra-indicar a utilização de filtragem colaborativa em sua formulação típica, em que o perfil do usuário é representado por suas avaliações anteriores. Mas foi mostrado que é factível a aplicação de técnicas colaborativas nesse domínio, incorporando informações contextuais ao sistema, o que permite que o usuário receba recomendações de hotéis que pessoas com perfil e contexto de viagem similares ao seu preferiram no passado.

O desenvolvimento de um sistema de recomendação não é uma tarefa trivial, pois requer tanto o conhecimento das técnicas quanto do domínio de aplicação. A realização de uma ampla revisão bibliográfica sobre esse campo de pesquisa permitiu identificar as principais características dos sistemas de recomendação e os aspectos relacionados à sua implementação. Com isso, propôs-se um procedimento metodológico para o desenvolvimento de sistemas de recomendação. O procedimento, que é composto de cinco etapas, foi adotado

para a implementação do sistema de recomendação de meios de hospedagem, mostrando-se adequado para orientar as escolhas e definições do projeto, o que possibilitou validar o procedimento metodológico proposto.

Os desenvolvedores de sistemas de recomendação voltados para o turismo devem considerar as particularidades do domínio, principalmente o que diz respeito às teorias de decisão em viagem. Variáveis relacionadas às características pessoais, características da viagem e características da hospedagem, exercem grande influência nas escolhas dos usuários, afetando assim as recomendações. Tais variáveis foram incorporadas ao sistema proposto, sendo denominadas de informações contextuais. Com base nelas os usuários foram representados, tornando possível oferecer recomendações personalizadas utilizando filtragem colaborativa.

A implementação de um protótipo do sistema de recomendação de hospedagem na agência Tropix Turismo Inteligente permitiu avaliar a abordagem proposta. Diante da dificuldade de se avaliar um sistema de recomendação antes de colocá-lo em funcionamento – ou seja, sem dados iniciais –, foi proposto um método que combina a avaliação de especialistas, para simular acessos de usuários, com a avaliação quantitativa do sistema, utilizando métricas já consolidadas na área de Recuperação de Informação e sistemas de recomendação *Top-N* – conhecidas como *recall-precision*. Todos os testes realizados mostraram que a abordagem adotando técnicas colaborativas e informações contextuais melhora sensivelmente a eficiência do sistema.

Dentre os pontos fortes do SRMH destacam-se: não exige que o usuário se cadastre para receber recomendações; oferece recomendações com base em informações contextuais; tende a melhorar com o tempo de utilização; não apresenta problema do “novo usuário”; não depende de avaliações anteriores para realizar recomendações; permite considerar características dos itens a serem recomendados.

Os principais pontos fracos são: apresenta limitações para usuário incomum; a qualidade depende do acúmulo de dados (histórico de utilização); apresenta problema do “novo item” (hotéis ainda não avaliados).

## ***5.2 Limitações e trabalhos futuros***

O SRMH ainda encontra-se em estágio inicial de avaliação. O protótipo foi

construído, mas não foi disponibilizado ao público. Por isso, não possui um histórico de utilização que permita avaliações mais aprofundadas. A simulação de acessos, por meio da avaliação de especialistas, possibilitou a realização de uma avaliação preliminar, mas é necessário avaliar o sistema após a consulta de um número expressivo de usuários reais.

Para avaliar o protótipo foram descartadas duas características dos meios de hospedagem: *localização* e *serviços e instalações*. Em trabalhos futuros essas duas variáveis podem ser incorporadas ao sistema utilizando um componente baseado em conteúdo, que associado ao componente colaborativo já implementado pode contribuir para contornar os problemas de “novo item” e “usuário incomum”.

Para o cálculo da similaridade entre os usuários, além da técnica colaborativa utilizada para formar a vizinhança (co-seno e heurística), foi utilizada também a medida de correlação (baseada nas avaliações dos usuários), com o objetivo de melhorar a qualidade das recomendações. No entanto, não foi avaliado o impacto do uso da correlação no processo de recomendação. É importante avaliar esse aspecto.

O desempenho computacional dos algoritmos não foi avaliado, mas percebeu-se que a técnica baseada em similaridade heurística leva mais tempo para apresentar os resultados. Acredita-se que após longo período de utilização (com o acúmulo de dados), as técnicas baseadas em memória não sejam as mais adequadas para o SRMH, pois tendem a demandar muito tempo de processamento. Recomenda-se, para trabalhos futuros, uma análise considerando diversos algoritmos de recomendação aplicados ao SRMH, avaliando a qualidade das recomendações e o tempo de resposta.

## REFERÊNCIAS

- 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.
- ANDERSON, Chris. **A Cauda Longa: do mercado de massa para o mercado de nicho**. Trad. Afonso Celso da Cunha Serra. Rio de Janeiro: Elsevier, 2006.
- ANDRADE, José Vicente. **Turismo, Fundamentos e Dimensões**. 8ª Ed. São Paulo: Ed. Ática, 2000.
- BALABANOVIĆ, Marko; SHOHAM, Yoav. Fab: Content-based, collaborative recommendation. **Communications of the ACM**, New York, v.40, n. 3, p. 66-72, março 1997.
- BERKA, Tobias; PLÖBNING, Manuela. Designing Recommender Systems for Tourism. In: **ENTER 2004**, Kairo. Disponível em: <[http://www.salzburgresearch.at/research/gfx/enter\\_ploessnig.pdf](http://www.salzburgresearch.at/research/gfx/enter_ploessnig.pdf)>. Acesso em: 20 mar. 2007.
- 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.
- BOGADO, Claudia; TELES, Adonai. **A Tecnologia da Informação na indústria do turismo: fatos, perspectivas e uma visão brasileira**. Observatório de Inovação do Turismo – Área Temática: Tecnologia da Informação, 07 abr. 2003. Disponível em: <[http://200.189.169.141/site/arquivos/dados\\_fatos/observatorio/ATecnologiadaInformacaonai ndustriadoturismo.pdf](http://200.189.169.141/site/arquivos/dados_fatos/observatorio/ATecnologiadaInformacaonai ndustriadoturismo.pdf)>. Acesso em: 20 abr. 2007.
- 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.

DAS, Abhinandan et al. Google News Personalization: Scalable Online Collaborative Filtering. In: International World Wide Web Conference Committee (IW3C2) – WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.

DELGADO, Joaquin; DAVIDSON, Richard. Knowledge bases and user profiling in travel and hospitality recommender systems. In: **Proceedings of the ENTER 2002 Conference**, Innsbruck, Austria, Springer Verlag, p. 1-16, Jan. 2002.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. **Journal of the Royal Statistical Society**, Ser B, Vol. 39, p. 1-38, 1977.

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.

FESENMAIER, Daniel R. et al. DieToRecs: Travel Advisory for Multiple Decision Styles. In A. J. Frew, M. Hitz, & P. O'Connors (Eds.). **Information and Communication Technologies in Tourism 2003**, Springer-Verlag, New York, p. 232-241, 2003.

FESENMAIER, Daniel R. et al. **Tourist Decision Model**. Technical report, D2.2 DieToRecs IST-2000-29474, EU IST project, 2002. Disponível em: <<http://dietorecs.itc.it/PubDeliverables/D2.2-V1.0.pdf>>. Acesso em: 13 fev. 2007.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 3. ed. São Paulo: Atlas, 1991.

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.

HWANG, Yeong-Hyeon; GRETZEL, Ulrike; FESENMAIER, Daniel R. Behavioral foundations for human-centric travel decision-aid systems. In: **Proceedings of the ENTER 2002 Conference**, Innsbruck, Austria, Springer Verlag. Jan. 2002.

KOHR, 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.

KRULWICH, B. Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data. **Artificial Intelligence Magazine** 18 (2), p. 37-45, 1997.

LEMOS, Fernando de Oliveira. **Metodologia para seleção de métodos de previsão de demanda**. Universidade Federal do Rio Grande do Sul – Programa de Pós-Graduação em Engenharia de Produção. Dissertação Mestrado. Porto Alegre, 2006.

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. Tese (Doutorado em Engenharia Eletrônica) - Departament d'Enginyeria Electrònica, University of València, Espanha, 2004. Disponível em: <[http://www.uv.es/jdmg/tesis\\_jdmartin.pdf](http://www.uv.es/jdmg/tesis_jdmartin.pdf)>. Acesso em: 15 jan. 2007.

MORAES, Cláudia Corrêa de Almeida. **Turismo – Segmentação de Mercado: um Estudo Introductório**, in ANSARAH, Marília Gomes dos Reis (org.). *Turismo: Segmentação de Mercado*, São Paulo, Futura, 1999.

NICHOLS, David. M. Implicit Rating and Filtering. In: **Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering**, Budapeste, Hungria: ERCIM, 1997, p. 31-36. Disponível em: <<http://www.ercim.org/publication/ws-proceedings/DELOS5/nichols.pdf>>. Acesso em: 10 jan. 2007.

NICOLAU, Juan Luis; MÁ S, Francisco J. **Proceso de decisión jerárquico del Turista**. XVI Encuentro de Profesores Universitarios de Marketing. Alicante, Espanha, set. 2004. Disponível em: <<http://www.epum2004.ua.es/aceptados/223.pdf>>. Acesso em: 20 abr. 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.

OLIVEIRA, Leonardo G. Proposta de uma estrutura metodológica para implementação de sistemas de recomendação. **Anais do IV Congresso Internacional de Gestão de Tecnologia e Sistemas de Informação**. TECSI/FEA/USP, São Paulo. São Paulo: CONTECSI, 2007.

PAZZANI, Michael J. A Framework for Collaborative, Content-Based and Demographic Filtering. **Artificial Intelligence Review**, 13 (5/6), p. 393-408, 1999.

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 Kayfmann, 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://diatorecs.itc.it/Papers/eccbr2002final.pdf>>. Acesso em: 25 nov. 2006.

RICH, E. User Modeling via Stereotypes. **Cognitive Science** vol. 3, no. 4, p. 329-354, 1979.

SANTAELLA, Lucia. **Comunicação e Pesquisa**: projetos para mestrado e doutorado. São Paulo: Hacker Editores, 2001.

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.

SARWAR, Bradul et al. Analysis of recommendation algorithms for ecommerce. In **Electronic Commerce, 2000**. EC'00, October 17-20, 2000, Minneapolis, Minnesota.

SARWAR, Bardul M. et al. Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system. In: **Conference on Computer Supported Cooperative Work**, 1998.

SCHAFER, 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.

SILVA, Sívio Ceroni. **Alternativas para a previsão de demanda de gás natural: um estudo orientado ao estado do Rio Grande do Sul**. Universidade Federal do Rio Grande do Sul – Programa de Pós-Graduação em Engenharia de Produção. Tese Doutorado. Porto Alegre, 2003.

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

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, Konstantinus 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.

VOZALIS, Manolis; MARGARITIS, Konstantinos G. Collaborative Filtering enhanced by demographic correlation. In: **Proceedings of the AIAI Symposium on Professional Practice in AI**, part of the 18th World Computer Congress, Toulouse, France, 2004.

# APÊNDICES

## APÊNDICE A

### VARIÁVEIS DE ENTRADA SELECIONADAS

A escolha das variáveis que exercem influência na tomada de decisão do turista por um meio de hospedagem, baseia-se principalmente na literatura sobre modelos de decisão em viagem e em estudos sobre o turismo, embora considere também dados de pesquisas de mercado da Embratur e Sebrae.

Para cada grupo de variáveis, apresentado no capítulo 3 – características pessoais, características da viagem e características da hospedagem –, serão selecionadas as variáveis mais significativas para o SRMH.

#### 1. Características pessoais

As características pessoais que mais exercem influência nas decisões de viagem podem ser divididas em duas categorias (FESENMAIER et al., 2002): sócio-econômica e psicológica/cognitiva. Na primeira categoria estão variáveis como: idade, nível de instrução, renda e estado civil. Na outra: valores, experiências/conhecimento, personalidade etc.

As variáveis consideradas mais relevantes para estratificar os usuários são: idade e renda.

#### Idade

Uma das dimensões demográficas mais importantes que influenciam a demanda turística é a idade do turista (MIECZKOWSKI, 1990, apud NICOLAU; MÁZ, 2004). Por isso, ao segmentar a demanda turística, a idade normalmente é utilizada como critério. De acordo com a faixa etária, o turismo pode ser segmentado em: turismo infanto-juvenil, turismo de meia idade e turismo da melhor idade<sup>20</sup>.

---

<sup>20</sup> Disponível em: <[http://pt.wikipedia.org/wiki/Lista\\_de\\_segmentos\\_do\\_mercado\\_tur%C3%ADstico](http://pt.wikipedia.org/wiki/Lista_de_segmentos_do_mercado_tur%C3%ADstico)>. Acesso em: 20 abr. 2007.

Para facilitar o enquadramento do turista em um determinado segmento, será utilizada uma divisão por faixa etária, seguindo a classificação do perfil sócio-econômico do visitante, estabelecida pela Embratur<sup>21</sup>:

- 18 a 24 anos;
- 25 a 31 anos;
- 32 a 40 anos;
- 41 a 50 anos;
- 51 a 59 anos;
- 60 anos ou mais;

### Renda

Embora o nível de renda tenha grande peso nas decisões de consumo, e seja amplamente utilizado para a segmentação de nichos de mercado no turismo, quando se trata da escolha por um meio de hospedagem é importante considerar outros aspectos. Por exemplo, enquanto uma pessoa com maior poder aquisitivo pode estar interessado em uma opção econômica para sua viagem de negócio, outra, com menor nível de renda, pode estar disposta a gastar um pouco mais em sua viagem de lua-de-mel.

Andrade (2000) adverte que “existem ricos que fazem turismo econômico e consomem o mínimo possível, assim como há turistas de poder aquisitivo regular ou modesto que fazem gastos excessivos na aquisição de bens e serviços que os mais abastados evitam, porque de preços exorbitantes”. Por isso, ao invés de inserir uma variável que determine o estrato de renda do usuário, será utilizada uma outra que questione a relevância do preço para sua decisão.

Essa variável também pode ser entendida, ou tratada, como uma característica do meio de hospedagem, devendo ser apresentada ao usuário como uma questão fechada, do tipo: na sua escolha por uma hospedagem o preço baixo é:

- indispensável;
- importante;

---

<sup>21</sup> Disponível em:

<[http://200.189.169.141/site/arquivos/dados\\_fatos/demanda%20turistica/turismo\\_internacional\\_brasil\\_2004\\_2005\\_marco2007.pdf](http://200.189.169.141/site/arquivos/dados_fatos/demanda%20turistica/turismo_internacional_brasil_2004_2005_marco2007.pdf)>. Acesso em: 20 abr. 2007.

- pouco importante;
- irrelevante;

## 2. Características da viagem

Além das características pessoais, e alguns fatores diretamente relacionados ao destino (tais como, condição geográfica e grau de urbanização), pode-se dizer que a maioria dos critérios utilizados para definir os segmentos do turismo está relacionada às características da viagem, como destaca Moraes (1999):

- critério meio de transporte: turismo aéreo, rodoviário, ferroviário, marítimo, fluvial e lacustre;
- critério duração: turismo de curta duração, de média duração e de longa duração;
- critério tipo de grupo: turismo individual, de casais, de famílias, de grupos e de grupos especiais (single, GLS, terceira idade, naturalistas, portadores de deficiência etc.);
- critério do sentido do fluxo turístico: turismo emissivo e receptivo;
- critério do aspecto cultural: turismo étnico, religioso, histórico, antropológico, arqueológico, artístico e de acontecimentos programados;
- critério da motivação da viagem: turismo de negócios, de eventos, de entretenimento, de saúde, educacional, de aventuras, esportivo, de pesca, de descanso, de natureza e de interesse específico (ex.: observação de baleias);

Olhando sob a perspectiva do processo de tomada de decisão, as características de viagem que têm sido utilizadas para explicar e prever as escolhas dos turistas inclui (FESENMAIER et al., 2002):

- a finalidade da viagem;
- a natureza do grupo de viagem (acompanhantes);
- a duração da viagem (tempo disponível para a viagem);
- a distância do destino;
- a familiaridade com o destino;
- as características do destino e da origem;
- o transporte utilizado no destino;

Os critérios que geram mais impacto na escolha do destino que do meio de hospedagem não serão considerados como variáveis de entrada, já que este trabalho parte do princípio que o turista já escolheu o destino de sua viagem, e que utilizará o SRMH

especificamente para escolher uma hospedagem em Florianópolis.

A duração da viagem e a distância são apontadas como características que influenciam principalmente a escolha do destino principal e a possibilidade de se visitar destinos alternativos. Podem influenciar também na escolha das atividades que se pretende praticar e/ou atrações que se deseja conhecer, mas exercem pouca influência na seleção de serviços turísticos, como acomodações.

A familiaridade com o destino é um fator importante a se considerar em praticamente todos os níveis de decisão de viagem, até mesmo na escolha de uma hospedagem. Contudo, se o objetivo das variáveis de entrada é estabelecer semelhanças entre perfis, para então utilizá-las como base para a recomendação, não faz muito sentido que os turistas com menos experiência no destino só recebam recomendações dos menos experientes, e vice-versa.

Tal variável seria interessante se fosse implementado algum sistema de reputação que desse mais peso às recomendações dos mais experientes, mas não é essa a proposta deste trabalho.

Outra discussão que se faz necessária é sobre as variáveis “meio de transporte” e “transporte utilizado no destino”. O tipo de veículo utilizado para viajar tem forte relação com a distância e duração da viagem, e causa impacto considerável na escolha do destino, como comentado anteriormente. Por outro lado, o transporte usado pelo turista para locomoção no destino está diretamente relacionado com sua mobilidade (FESENMAIER et al., 2002), o que pode exercer uma influência considerável na escolha da sua hospedagem.

Outro fator ainda não relatado, mas que é considerado uma das principais características da demanda turística é a sazonalidade. “As épocas das temporadas ou as estações altas ou mais aprazíveis do ano, cada qual com suas características próprias, também se constituem em fatores importantes de influência no volume e na qualidade da demanda turística” (ANDRADE, 2000).

A alta temporada geralmente está relacionada a condições climáticas favoráveis e a períodos de férias escolares. Assim, o perfil do turista de baixa e alta temporada tende a variar. Além disso, um meio de hospedagem altamente recomendável para o verão, pode não ser indicado para o inverno.

Diante do que foi exposto, as variáveis selecionadas para compor as características da viagem são: motivação ou finalidade da viagem, grupo, transporte utilizado no destino e período.

## Motivação da viagem

A finalidade da viagem pode ser classificada, de maneira simplificada, em duas categorias: lazer e não lazer (ou negócio). Entretanto, essas categorias podem ser subdivididas em outras mais específicas, conforme a natureza da viagem (FESENMAIER et al., 2002). Andrade (2000) ressalta que:

não é fácil estabelecer distinções viáveis e válidas entre os vários tipos de turismo, porque as chamadas motivações [de viagens] principais justapõem-se umas às outras de tal forma que, se para os turistas não descaracterizam finalidades de viagens, para os técnicos e estudiosos criam problemas classificatórios diversos, pois a maioria das classificações fundamentais se alicerça em objetivos econômicos, administrativos e empresariais e não em relação às motivações que levam indivíduos e grupos a viajar.

Segundo as motivações de viagem, Andrade (2000) classifica os tipos de turismo em:

- turismo de férias: turismo balneário, turismo montanhês e turismo de repouso;
- turismo cultural: turismo científico e turismo de congresso ou de convenções;
- turismo de negócios: viagens de negócios e viagens de compras;
- turismo desportivo;
- turismo de saúde;
- turismo religioso;

Uma outra classificação, atenta à segmentação de mercado, é apresentada em um estudo de Sebrae/DF sobre o perfil do setor de agências de viagens e turismo do Distrito Federal<sup>22</sup>. Os principais tipos de turismo, identificados como nichos de mercado pelas agências, são:

- turismo rural;
- turismo de negócios;
- ecoturismo;
- turismo místico;
- turismo religioso;
- turismo cultural;
- turismo de sol e praia;
- turismo gastronômico;
- turismo de eventos;

---

<sup>22</sup> Disponível em:

<[http://www.df.sebrae.com.br/Downloads/desenvolvimento\\_setorial/agencia/agenciaweb.pdf](http://www.df.sebrae.com.br/Downloads/desenvolvimento_setorial/agencia/agenciaweb.pdf)>. Acesso em: 20 abr. 2007.

- turismo de saúde;
- GLS;

Dados estatísticos sobre o turismo doméstico no Brasil<sup>23</sup> apontam como principais motivos para a realização das viagens:

- visitar amigos e parentes;
- sol e praia;
- turismo cultural;
- eventos culturais / esportivos / sociais;
- saúde;
- compras pessoais;
- negócios;
- ecoturismo;
- eventos profissionais;
- turismo rural;
- estâncias climáticas / hidrominerais;
- religião;
- parques temáticos;
- resorts / hotéis fazenda;
- cursos / estudo;
- praticar esportes;

Procurando compor uma síntese das principais classificações apresentadas, sugere-se que o usuário informe a finalidade que melhor representa sua viagem, selecionando uma das opções abaixo:

- sol e praia;
- negócios / trabalho;
- eventos (profissionais, científicos, culturais, esportivos e sociais);
- lazer / descanso / diversão;
- visita a amigos / parentes;
- cursos / estudo;
- compras pessoais;
- saúde;

---

<sup>23</sup> Considera-se como turismo doméstico as viagens realizadas por pessoas dentro do seu próprio país de origem.

Disponível em:

<[http://www.braziltour.com/site/arquivos/dados\\_fatos/demanda%20turistica/demanda\\_turismo\\_domestico\\_setembro\\_2006.pdf](http://www.braziltour.com/site/arquivos/dados_fatos/demanda%20turistica/demanda_turismo_domestico_setembro_2006.pdf)>. Acesso em: 20 abr. 2007.

- prática de esportes / aventura;
- ecoturismo;
- turismo GLS;
- turismo religioso / místico;

### Grupo de viagem

O tipo de grupo também é reconhecido como um importante aspecto que define a natureza da viagem (FESENMAIER et al., 2002; MORAES, 1999). Os acompanhantes representam heterogeneidade de interesses, influenciam na mobilidade do grupo e, sobretudo, nas decisões de viagem.

As opções adotadas são:

- sozinho;
- com parceiro (a);
- com amigo/a (s);
- em família sem crianças;
- em família com crianças;
- em grupo (excursão);

### Transporte utilizado no destino

A relevância da mobilidade na escolha de um meio de hospedagem já foi discutida anteriormente. A seguir estão relacionadas as opções apresentadas aos usuários:

- veículo próprio;
- veículo alugado;
- táxi / serviços de traslado;
- a pé / transporte coletivo;
- ônibus de excursão;

### Período

A sazonalidade, especialmente em Florianópolis, é geralmente demarcada como alta e baixa temporada. A primeira associada ao verão e às férias escolares, a segunda compreendendo os outros meses do ano. Enquanto na alta temporada predomina o turismo de sol e mar, na baixa temporada tem destaque o turismo de eventos e negócios.

Embora usualmente o turismo seja dividido em apenas dois períodos distintos, o

clima subtropical garante ao estado de Santa Catarina estações do ano bem definidas, que permitem atividades específicas em cada uma delas. Por isso, sugere-se a divisão de períodos de acordo com as estações do ano:

- entre dezembro e fevereiro;
- entre março e maio;
- entre junho e agosto;
- entre setembro e novembro;

### **3. Características dos meios de hospedagem**

Andrade (2000) lembra que “é impossível satisfazer a natureza do homem sem atender às suas necessidades de proteção e repouso, de higiene e alimentação, de privacidade e de tranquilidade, requisitos que formam os fundamentos teóricos e ideais da moradia ou residência”.

Essas mesmas necessidades justificam uma das preocupações básicas de um viajante, que é encontrar um lugar adequado para se hospedar. Nesse sentido, o hotel pode ser visto como uma extensão ou substitutivo da própria residência do turista. O que não significa que essas necessidades básicas excluam uma série de outros fatores que pesam na escolha por um meio de hospedagem.

A empresa hoteleira possui características muito peculiares, devendo ser analisada tanto pela sua estrutura física (padrões de instalações) quanto pela sua estrutura humana (serviços que presta). Os inúmeros critérios que podem ser avaliados em cada uma dessas estruturas é que orientam sua classificação em diferentes categorias de conforto e atendimento<sup>24</sup>.

Ao Sistema Oficial de Classificação dos Meios de Hospedagem, regulamentado pelo Ministério do Turismo em 2002<sup>25</sup>, cabe classificar, categorizar, qualificar os meios de hospedagem, em território nacional, simbolizados por estrelas, de acordo com as condições de conforto, comodidade, serviços e atendimento que possuam. O Quadro 10 apresenta a atual

---

<sup>24</sup> Regulamentação do Sistema Oficial de Classificação de Meios de Hospedagem. Disponível em: <[http://www.abih.com.br/principal/classificacao/regulamento\\_sistema.php](http://www.abih.com.br/principal/classificacao/regulamento_sistema.php)>. Acesso em: 20 abr. 2007.

<sup>25</sup> Disponível em: <<http://www.abih.com.br/principal/classificacao/embratur.php>>. Acesso em: 20 abr. 2007.

classificação hoteleira aprovada pelo Ministério do Turismo.

<b>Categoria</b>	<b>Estrelas</b>
Super Luxo	★★★★★SL
Luxo	★★★★★
Superior	★★★★
Turismo	★★★
Econômico	★★
Simples	★

**Quadro 10 – Classificação hoteleira**

Fonte: Elaborado pelo autor.

Os critérios contemplados na avaliação estão distribuídos entre itens gerais e específicos<sup>26</sup>:

- itens gerais: posturas legais, segurança, saúde / higiene, conservação / manutenção, atendimento ao hóspede;
- itens específicos: portaria / recepção, acessos e circulação, setor habitacional, áreas sociais, comunicações, alimentos e bebidas, lazer, reuniões / escritório virtual, serviços adicionais, ações ambientais;

Embora o sistema de classificação hoteleira seja uma das formas mais adequadas para distinguir as diversas opções de hospedagem disponíveis em um destino turístico, a nova regulamentação alterou integralmente o processo de classificação dos meios de hospedagem, invalidando as classificações obtidas antes desta regulamentação. Isso significa que os hotéis que não passaram pela nova classificação, não podem (ou pelo menos não deveriam) adotar uma categoria simbolizada por estrelas. Considerando que até então um número muito reduzido de hotéis se sujeitaram à nova classificação<sup>27</sup>, esse critério, de grande importância, perde seu valor e significado, não podendo ser um parâmetro de distinção das opções de hospedagem – pelo menos a curto e médio prazo. Mesmo que todos os meios de hospedagem brasileiros se sujeitem à nova classificação, vale ressaltar que esta não é uma metodologia de classificação internacionalmente aceita.

<sup>26</sup> Matriz de Classificação Hoteleira. Disponível em:

<[http://www.abih.com.br/principal/downloads/matriz\\_classifica.pdf](http://www.abih.com.br/principal/downloads/matriz_classifica.pdf)>. Acesso em: 20 abr. 2007.

<sup>27</sup> Disponível em: <<http://www.abih.com.br/principal/classificacao.php>>. Acesso em: 20 abr. 2007.

### Serviços e instalações

Sendo a qualidade dos serviços um critério de difícil avaliação e categorização, devido às suas características intrínsecas, serão priorizados os aspectos relacionados às instalações dos meios de hospedagem. Dentre eles, alguns considerados mais relevantes como fator de decisão para o turista:

- acessibilidade a deficientes físicos (cadeirantes);
- ar condicionado;
- estacionamento;
- internet;
- piscina;
- aceita animais;

### Localização

Além desses aspectos, outro fator importante na escolha de um meio de hospedagem é sua localização. A localização pode ter forte relação com a motivação da viagem. Para um turista que visita uma cidade para participar de um evento, por exemplo, é interessante ficar hospedado próximo ao local onde o evento será realizado.

A localização também pode estar relacionada à forma como o turista se locomoverá pela cidade. Uma pessoa que venha com o objetivo de aproveitar sol e praia, e dependa de transporte público para se locomover, preferirá um lugar próximo da praia para se hospedar, ou pelo menos com fácil acesso a elas.

Enfim, a localização representa um fator importante na tomada de decisão do turista por uma hospedagem. Sendo assim, o turista poderá optar por um bairro (por exemplo, Canasvieiras) ou uma região (por exemplo, Região Norte, que abrange vários bairros do Norte da Ilha de Santa Catarina).

### Preço

Por fim, uma característica da hospedagem que não deve ser negligenciada é o preço, fator determinante na escolha de grande parte dos turistas. Como esse aspecto já foi tratado ao se analisar a renda do turista, não será abordado novamente. No entanto, optou-se por tratar essa variável como uma característica do meio de hospedagem, ao invés de uma característica pessoal.

O Quadro 11 apresenta uma visão geral das variáveis de entrada, selecionadas para a implementação do SRMH.

<b>Variáveis de Entrada</b>	
Características pessoais:	— Idade
Características da viagem:	— Motivação da Viagem — Grupo — Transporte no destino — Período
Características da hospedagem:	— Serviços e instalações — Localização — Preço

**Quadro 11 – Variáveis selecionadas para o SRMH**

Fonte: Elaborado pelo autor.

## APÊNDICE B

### SIMILARIDADE HEURÍSTICA

As variáveis utilizadas para o cálculo da similaridade heurística são: faixa etária, preço baixo, motivação, grupo, transporte e período. Para cada uma delas os atributos são comparados e apresentados nos quadros a seguir.

#### Faixa etária

18 a 24	25 a 31	32 a 40	41 a 50	51 a 59	60 ou mais	
1	0,5	0,5	0	-0,5	-1	18 a 24
	1	0,5	0	-0,5	-1	25 a 31
		1	0,5	0	-0,5	32 a 40
			1	0,5	0	41 a 50
				1	0,5	51 a 59
					1	60 ou mais

**Quadro 12 – Matriz de similaridade para variável Faixa etária**

Fonte: Elaborado pelo autor.

#### Preço baixo

indispensável	importante	pouco importante	irrelevante	
1	0,5	-0,5	-1	indispensável
	1	0	-0,5	importante
		1	0,5	pouco importante
			1	irrelevante

**Quadro 13 – Matriz de similaridade para variável Preço baixo**

Fonte: Elaborado pelo autor.

#### Motivação

sol e praia	negócios	eventos	lazer	visita parente	cursos	compras	saúde	esportes	ecoturismo	GLS	religioso	
1	-1	-0,5	0,5	0,5	-0,5	0	0	0,5	0,5	0,5	0	sol e praia
	1	0,5	-1	-0,5	0,5	0	0	-0,5	-0,5	-0,5	0	negócios
		1	-0,5	0	0,5	0	0	-0,5	-0,5	-0,5	0	eventos
			1	0,5	0	0,5	0	0,5	0,5	0,5	0	lazer
				1	0	0,5	0	0	0,5	-0,5	0	visita parente
					1	0	0	0	0	-0,5	0	cursos
						1	0	0	0	0	0	compras
							1	0	0	-0,5	0	saúde
								1	0,5	0	0	esportes
									1	0	0	ecoturismo
										1	-1	GLS
											1	religioso

**Quadro 14 – Matriz de similaridade para variável Motivação**

Fonte: Elaborado pelo autor.

## Grupo

sozinho	com parceiro (a)	com amigos	em família sem crianças	em família com crianças	em grupo	
1	0	0	-0,5	-1	-1	sozinho
	1	0,5	0	-0,5	-0,5	com parceiro (a)
		1	0	-0,5	0	com amigos
			1	0	0	em família sem crianças
				1	0	em família com crianças
					1	em grupo

**Quadro 15 – Matriz de similaridade para variável Grupo**

Fonte: Elaborado pelo autor.

## Transporte

veículo próprio	veículo alugado	táxi / serviços de traslado	a pé / transp. coletivo	ônibus de excursão	
1	1	-0,5	-0,5	-0,5	veículo próprio
	1	-0,5	-0,5	-0,5	veículo alugado
		1	0	0	táxi / serviços de traslado
			1	0	a pé / transp. coletivo
				1	ônibus de excursão

**Quadro 16 – Matriz de similaridade para variável Transporte**

Fonte: Elaborado pelo autor.

## Período

dez - fev	mar - mai	jun - ago	set - nov	
1	0	-0,5	0	dez - fev
	1	0	0,5	mar - mai
		1	0	jun - ago
			1	set - nov

**Quadro 17 – Matriz de similaridade para variável Período**

Fonte: Elaborado pelo autor.

## APÊNDICE C

### RECOMENDAÇÕES DOS ESPECIALISTAS

Os perfis de turistas selecionados para as recomendações dos especialistas são apresentados no Quadro 18. São 25 perfis, apresentados para três especialistas, resultando em 75 avaliações distintas.

Usuários	Faixa etária	Preço baixo	Motivo	Grupo	Transporte	Período
1-26-51	18 a 24	indispensável	lazer	amigo (s)	a pé / coletivo	set-nov
2-27-52	25 a 31	importante	lazer	parceiro (a)	veículo próprio	set-nov
3-28-53	32 a 40	importante	sol e praia	família c/ criança	veículo próprio	dez-fev
4-29-54	41 a 50	irrelevante	sol e praia	parceiro (a)	veículo alugado	dez-fev
5-30-55	51 a 59	pouco importante	sol e praia	família s/ criança	veículo alugado	dez-fev
6-31-56	60 ou mais	pouco importante	lazer	parceiro (a)	táxi/traslado	mar-mai
7-32-57	32 a 40	pouco importante	evento	sozinho	táxi/traslado	mar-mai
8-33-58	18 a 24	importante	sol e praia	em grupo (excursão)	ônibus de excursão	dez-fev
9-34-59	32 a 40	importante	negócio	sozinho	táxi/traslado	jun-ago
10-35-60	60 ou mais	pouco importante	sol e praia	em grupo (excursão)	ônibus de excursão	dez-fev
11-36-61	41 a 50	pouco importante	negócio	amigo (s)	táxi/traslado	jun-ago
12-37-62	51 a 59	irrelevante	lazer	parceiro (a)	veículo próprio	set-nov
13-38-63	51 a 59	importante	evento	sozinho	táxi/traslado	mar-mai
14-39-64	25 a 31	indispensável	evento	amigo (s)	a pé / coletivo	jun-ago
15-40-65	25 a 31	importante	esporte / aventura	amigo (s)	veículo próprio	mar-mai
16-41-66	18 a 24	importante	curso / estudo	sozinho	a pé / coletivo	mar-mai
17-42-67	32 a 40	irrelevante	lazer	parceiro (a)	veículo alugado	dez-fev
18-43-68	41 a 50	pouco importante	visita a parente	família c/ criança	veículo próprio	jun-ago
19-44-69	25 a 31	pouco importante	GLS	parceiro (a)	veículo próprio	dez-fev
20-45-70	18 a 24	importante	ecoturismo	parceiro (a)	veículo alugado	jun-ago
21-46-71	32 a 40	importante	evento	amigo (s)	a pé / coletivo	set-nov
22-47-72	25 a 31	indispensável	lazer	amigo (s)	veículo alugado	set-nov
23-48-73	60 ou mais	importante	sol e praia	família s/ criança	veículo próprio	dez-fev
24-49-74	51 a 59	importante	negócio	amigo (s)	táxi/traslado	jun-ago
25-50-75	41 a 50	pouco importante	evento	sozinho	táxi/traslado	set-nov

**Quadro 18 – Perfis de turistas para avaliação do SRMH**

Fonte: Elaborado pelo autor.

Os hotéis disponíveis para as recomendações dos especialistas são mostrados no Quadro 19.

HOTEL	NOME	HOTEL	NOME
1	BAÍA NORTE OTHON CLASSIC HOTEL	21	INTERCITY PREMIUM FLORIANÓPOLIS
2	BLUE TREE TOWERS FLORIANÓPOLIS	22	JURERÉ BEACH VILLAGE
3	BRISA MAR HOTEL	23	MAJESTIC PALACE HOTEL
4	BRISTOL MULTY CASTELMAR HOTEL	24	MAR DEL PLATA HOTEL
5	CABANAS DUNA SOL	25	MAREA APART HOTEL
6	CAMBIRELA HOTEL	26	OSCAR HOTEL
7	CECOMTUR EXECUTIVE HOTEL	27	PARADOR DA CACHOEIRA HOTEL
8	CHALÉS DO CANTO	28	PORTO INGLESES HOTEL
9	DEVILLE EXPRESS FLORIANÓPOLIS	29	POUSADA BIZKAIA
10	FLORIANÓPOLIS PALACE HOTEL	30	POUSADA COLINAS DA LAGOA
11	HOTEL DAIFA	31	POUSADA DAS PALMEIRAS
12	HOTEL ENGENHO ECO PARK	32	POUSADA DOS SONHOS
13	HOTEL FAIAL	33	POUSADA NATUR CAMPECHE
14	HOTEL MARIA DO MAR	34	POUSADA VILA TAMARINDO
15	HOTEL PORTO DA ILHA	35	PRAIA MOLE ECO VILLAGE
16	HOTEL QUINTA DA BICA D'ÁGUA	36	PRAIATUR HOTEL
17	HOTEL SÃO SEBASTIÃO	37	VALERIM CENTER
18	HOTEL TORRES DA CACHOEIRA	38	VALERIM PLAZA
19	INGLESES HOLIDAY RESORT	39	VICTORIA PALACE HOTEL
20	INGLESES PRAIA HOTEL	40	VILLAS DEL SOL Y MAR

**Quadro 19 – Hotéis disponíveis para recomendação**

Fonte: Elaborado pelo autor.

O Quadro 20 apresenta as recomendações dos especialistas para os usuários *Un*.

Especialista 1					Especialista 2					Especialista 3			
<i>U</i>	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>U</i>	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>U</i>	<i>H1</i>	<i>H2</i>	<i>H3</i>
1	37	7	26		26	29	5	24	37	51	11	30	26
2	7	30	5		27	30	29	17	7	52	30	29	8
3	17	5	8	14	28	14	17	30	20	53	17	39	18
4	22	36	32	31	29	31	22	32	40	54	32	40	35
5	36	19	28	19	30	22	19	36	28	55	25	28	22
6	19	22	36	20	31	32	31	40	33	56	23	2	1
7	1	21	7		32	21	1	2	7	57	21	7	10
8	17	5	19		33	24	39	36	11	58	24	27	28
9	7	26	1		34	7	37	26	1	59	7	26	10
10	18	20	36	19	35	36	19	25	28	60	22	36	28
11	1	21	2	7	36	21	1	2	7	61	1	4	9
12	23	31	22	2	37	31	32	40	22	62	14	35	12
13	7	26	1		38	7	37	26	11	63	7	10	26
14	37	26	7		39	7	37	26	11	64	26	7	11
15	30	29	17		40	12	30	5	8	65	35	17	12
16	7	26	37		41	16	7	37	26	66	7	11	26
17	31	22	23		42	32	31	40	2	67	40	31	34
18	12	2	1		43	19	17	14	34	68	14	8	17
19	35	22	30		44	35	40	30	29	69	33	22	40
20	17	29	30		45	30	8	33	5	70	12	35	8
21	7	26	37		46	7	37	26	11	71	7	26	11
22	37	29	30		47	29	5	24	37	72	24	27	5
23	17	5	27		48	22	19	36	20	73	20	25	35
24	7	26	1		49	7	37	26	11	74	7	26	11
25	1	21	2		50	21	1	2	7	75	2	23	1

**Quadro 20 – Recomendações dos especialistas**

Fonte: Elaborado pelo autor.