
**Uma Abordagem Baseada em Filtragem
Colaborativa Integrada a Mapas de Saliência
para a Recomendação de Imagens**

Emília Alves Nogueira



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia
2015

Emília Alves Nogueira

**Uma Abordagem Baseada em Filtragem
Colaborativa Integrada a Mapas de Saliência
para a Recomendação de Imagens**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Denise Guliato

Coorientador: Elaine Ribeiro de Faria Paiva

Uberlândia

2015

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

N778a Nogueira, Emília Alves, 1990-
2016 Uma abordagem baseada em filtragem colaborativa integrada a
mapas de saliência para a recomendação de imagens / Emília Alves
Nogueira. - 2016.

103 f. : il.

Orientadora: Denise Guliato.

Dissertação (mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-Graduação em Ciência da Computação.

Inclui bibliografia.

1. Computação - Teses. 2. Processamento de imagens - Teses. 3.
Sistemas de recomendação (filtragem da informação) - Teses. I. Guliato,
Denise. II. Universidade Federal de Uberlândia. Programa de Pós-
Graduação em Ciência da Computação. III. Título.

CDU: 681.3

À minha família, José Carlos Nogueira, Suelene Alves Lopes, Müller Alves Nogueira, Morgana Lima, Isadora Alves Lima Nogueira e ao meu noivo Bruno Moraes Rocha.

Agradecimentos

Agradeço primeiramente a Deus, por todas as minhas conquistas, sem ele me guiando e protegendo não estaria a caminho de mais uma.

À meus pais por toda dedicação, carinho e amor em todos os dias de minha vida e por quem serei grata eternamente.

À meu irmão, cunhada e sobrinha que estão sempre presentes em minha vida, me motivando e compartilhando experiências de vida.

À meu noivo, pela amizade, companheirismo e carinho, sem os quais essa jornada seria ainda mais difícil.

À meus amigos dentro e fora da UFU, Reny, Franciny, Walter, Myllene, Juan, Maria Cristina e Marcos Aurélio, por todas as conversas, risadas, apoio mútuo e momentos compartilhados.

À minha orientadora Denise Guliato, pelo apoio, paciência e orientação nesta pesquisa.

A professora Elaine Ribeiro de Faria Paiva, pela coorientação dessa pesquisa.

Ao meu grande amigo e companheiro de pesquisa Ernani Viriato de Melo, por todas as ajudas, conselhos e compartilhamento de experiências, que tornaram essa pesquisa possível e a quem sou muito grata.

Agradeço à CAPES pelo apoio financeiro, que possibilitou está inteiramente dedicada aos estudos.

“A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original.”
(Albert Einstein)

Resumo

Hoje em dia, a quantidade de consumidores que utilizam *sites* para realizar compras está aumentando muito, principalmente devido à facilidade e rapidez dessa forma de consumo. Os *sites*, diferentemente de lojas físicas, podem mostrar qualquer conteúdo existente para os clientes. Nesse contexto, Sistemas de Recomendação (SR) tornaram-se indispensáveis para ajudar os consumidores a encontrar produtos que podem, eventualmente, ser agradáveis ou úteis para eles. Esses sistemas, geralmente, usam técnicas de Filtragem Colaborativa (FC), cuja principal ideia é que os produtos são recomendados para um determinado usuário com base em informações sobre compras e avaliações realizadas no passado, por um grupo de usuários similares ao usuário que está pedindo recomendação. Um dos principais desafios enfrentados por tal técnica é a necessidade de o usuário fornecer algumas informações sobre suas preferências sobre os produtos, a fim de obter novas recomendações do sistema. Quando há itens que não têm avaliações ou que possuem poucas avaliações disponíveis, o sistema de recomendação executa mal. Esse problema é conhecido como problema de partida a frio de itens. Neste trabalho, propomos investigar em que ponto a informação sobre a atenção visual pode ajudar a produzir modelos mais precisos de recomendação. Nós apresentamos uma nova estratégia de FC, denominada IKB-MS, que utiliza atenção visual para caracterizar as imagens e minimizar o problema de partida a frio de itens. A fim de validar essa estratégia, criamos um banco de dados de imagens de roupas e usamos três algoritmos bem conhecidos para a extração da atenção visual dessas imagens. Um extenso conjunto de experimentos mostra que a nossa abordagem é eficiente e supera os SR de FC do estado-da-arte.

Palavras-chave: Sistema de Recomendação, Atenção Visual, Filtragem Colaborativa e partida a frio.

Abstract

Nowadays, the amount of customers using sites for shopping is greatly increasing, mainly due to the easiness and rapidity of this way of consumption. The sites, differently from physical stores, can make anything available to customers. In this context, Recommender Systems (RS) have become indispensable to help consumers to find products that may possibly pleasant or be useful to them. These systems often use techniques of Collaborating Filtering (CF), whose main underlying idea is that products are recommended to a given user based on purchase information and evaluations of past, by a group of users similar to the user who is requesting recommendation. One of the main challenges faced by such a technique is the need of the user to provide some information about her preferences on products in order to get further recommendations from the system. When there are items that do not have ratings or that possess quite few ratings available, the recommender system performs poorly. This problem is known as new item cold-start. In this paper, we propose to investigate in what extent information on visual attention can help to produce more accurate recommendation models. We present a new CF strategy, called IKB-MS, that uses visual attention to characterize images and alleviate the new item cold-start problem. In order to validate this strategy, we created a clothing image database and we use three algorithms well known for the extraction of visual attention these images. An extensive set of experiments shows that our approach is efficient and outperforms state-of-the-art CF RS.

Keywords: *Recommender Systems, Visual Attention, Collaborating Filtering e cold-start.*

Lista de ilustrações

Figura 1 – Fluxo de dados do IKB-MS.	25
Figura 2 – Atenção visual baseada em características <i>top-down</i> e <i>bottom-up</i> , adaptada de (THEEUWES, 1992). (a) distratores variam em relação à forma, (b) existe um único distrator de cor vermelha, (c) vários distratores de outra cor e (d) entre os vários distratores, um pode ter forma diferente.	32
Figura 3 – Exemplos de tarefas de busca visual, adaptada de (WOLFE; HOROWITZ, 2004). (a) contraste de cores, (b) e (c) demonstram a importância da orientação e cor e (d) informação de intersecção.	33
Figura 4 – Três MCAV foram aplicados em uma imagem. (a) imagem original obtida em http://blog.mundi.com.br/ , (b) MS de (ITTI; KOCH; NIEBUR, 1998), (c) MS de (HAREL; KOCH; PERONA, 2006) e (d) MS da (JUDD et al., 2009).	34
Figura 5 – Arquitetura do método de Itti, adaptada de (LOPERA et al., 2012).	35
Figura 6 – Extração dos 4 canais de cores. (a) imagem de entrada, (b) extração do canal vermelho, (c) extração do canal verde, (d) extração do canal azul e (e) extração do canal amarelo.	36
Figura 7 – Pirâmide Gaussiana de 9 níveis.	37
Figura 8 – Exemplo de contraste na orientação. Barra vertical inserida em um ambiente com barras horizontais, representando a diferença nas orientações, e por isso, se torna atenta.	37
Figura 9 – As imagens das intensidades dos quatro <i>kernels</i> de Gabor são usadas para determinar a informação de orientação. (a) 0°, (b) 45°, (c) 90° e (d) 135°, adaptada de (SIKLOSSY, 2005)	38

Figura 10 – Extração das informações de orientação utilizando filtragem linear de Gabor. (a) imagem de entrada, (b) informações filtradas com 0° , (c) informações filtradas com 45° , (d) informações filtradas com 90° e (e) informações filtradas com 135° , adaptada de (SIKLOSSY, 2005) .	38
Figura 11 – Exemplo do comportamento do operador de normalização $\mathcal{N}(\cdot)$, adaptada de (ITTI; KOCH; NIEBUR, 1998).	40
Figura 12 – Mapa de saliência extraído utilizando o método de Itti.	41
Figura 13 – Arquitetura do método de Harel, adaptada de (HAREL; KOCH; PERONA, 2006)	42
Figura 14 – Criando G_A. (a) representação dos nós em um mapa (características ou conspicuidades) de entrada, (b) representação das arestas bidirecionais do grafo e (c) representação da ponderação das arestas no grafo direcionado tanto para a fase de Ativação quanto para a fase de Normalização.	43
Figura 15 – Mapa de Saliência extraído utilizando o método de Harel.	44
Figura 16 – Convolução Gaussiana realizada sobre as fixações dos usuários. (a) Fixações de um único usuário, (b) Fixações de todos os usuários e (c) Convolução Gaussiana das fixações.	46
Figura 17 – Treinamento do método de (JUDD et al., 2009). (a) extração das características, (b) convolução e escolha dos pontos positivos localizados nos 20% mais relevantes da imagem e negativos localizados nos 70% menos relevantes da imagem, (c) treinamento do SVM e (d) criação de um MCAV.	46
Figura 18 – Mapa de saliência extraído utilizando o método da Judd. (a) imagem original, (b) mapa resultante.	46
Figura 19 – Processo de recomendação, adaptada de (BURKE, 2002)	51
Figura 20 – Fatoração da matriz de utilidade usando o SVD. Neste exemplo u representa os usuários, i os itens e f os fatores latentes.	57
Figura 21 – Arquitetura geral do IKB-MS.	72
Figura 22 – Particionando cada imagem de acordo com a aplicação. (a) particionamento utilizado para boi de corte (b) particionamento utilizado para uma paisagem específica.	73
Figura 23 – Exemplo da construção de um vetor de atenção.	74
Figura 24 – Exemplos de imagens da base de dados UFU-CLOTHING. (a) roupas masculinas e (b) roupas femininas	78

Figura 25 – Imagens divididas em 12, 10 e 22 partições respectivamente. (a) superior, (b) inferior e (c) completo. No caso de (a) e (b) o vetor de atentividade também tem 22 posições mas são preenchidas as 12 e 10 primeiras posições, respectivamente sendo que as posições restantes ficarão com zero.	79
Figura 26 – Vetores de atentividade correspondendo aos MS extraídos de imagens exemplo. (a) 12 partições do MS, (b) 10 partições do MS e (c) 22 partições do MS.	80
Figura 27 – Comparando as diferentes medidas de similaridades utilizadas na SMS.	83
Figura 28 – Comparando diferentes valores de α.	84
Figura 29 – Comparando diferentes números de partições das imagens. . .	84
Figura 30 – Comparando IKB com IKB-MS e ponderando a predição com o cosseno. O RMSE é mostrado com vários tamanhos de vizinhança. .	85
Figura 31 – Comparando IKB com IKB-MS e ponderando a predição com a correlação de Pearson. O RMSE é mostrado com vários tamanhos de vizinhança.	86
Figura 32 – Top-N executado pelo IKB-MS.	89

Lista de tabelas

Tabela 1 – Características que podem guiar a atenção visual (WOLFE; HOROWITZ, 2004).	33
Tabela 2 – Representação de uma matriz de avaliações.	61
Tabela 3 – Matriz de avaliações, representa a correlação entre os itens 1 e 5.	62
Tabela 4 – Matriz de avaliações.	73
Tabela 5 – Matriz de avaliações atualizada com as predições.	75
Tabela 6 – Número de imagens separadas pelo gênero, classe do corpo e categoria.	78
Tabela 7 – Comparação da eficácia dos métodos em termos do RMSE e MAE.	86
Tabela 8 – Comparação da eficácia em termos do RMSE e MAE para o problema de partida a frio.	87
Tabela 9 – Média de precisão na recomendação Top-N.	88
Tabela 10 – p -valores da hipótese nula, que os sistemas são igualmente eficazes.	90

Lista de siglas

AP Precisão Média - *Average Precision*

FBC Filtragem Baseada em Conteúdo

FC Filtragem Colaborativa

FH Filtragem Híbrida

GBVS Saliência Visual Baseada em Grafo - *Graph-Based Visual Saliency*

IDF Inverso da Frequência nos Documentos - *Inverse Document Frequency*

IKB *ItemKNN+Baseline*

LSI Indexação Semântica Latente - *Latent Semantic Indexing*

MAE Erro Médio Absoluto - *Mean Absolute Error*

MCAV Métodos Computacionais de Atenção Visual

MS Mapa de Saliência

MSE Erro Médio Quadrático - *Mean Squared Error*

PTV Listagem de Televisão Pessoal - *Personal TeleVision listings*

RMSE Raiz Quadrada do Erro Médio Quadrático - *Root Mean Square Error*

SVD Decomposição de Valores Singulares - *Singular Value Decomposition*

SVM Máquina de Vetores de Suporte - *Support Vector Machines*

TF Frequência do Termo - *Term-Frequency*

WTA *Winner-Take-All*

WWW Rede Mundial de Computadores - *World Wide Web*

VC Visão Computacional

Sumário

1	INTRODUÇÃO	23
1.1	Objetivos e Hipótese	26
1.2	Contribuições	26
1.3	Organização da Dissertação	27
2	MÉTODOS COMPUTACIONAIS DE ATENÇÃO VISUAL .	29
2.1	Introdução	29
2.2	Atenção Visual	30
2.2.1	Modelos de atenção visual <i>top-down</i>	30
2.2.2	Modelos de atenção visual <i>bottom-up</i>	32
2.3	MCAV de Itti	34
2.3.1	Filtragem Linear	35
2.3.2	Diferenças centro-vizinhanças e normalização	38
2.3.3	Combinações variando escala e normalização	39
2.3.4	Combinações lineares	40
2.4	MCAV de Harel	41
2.4.1	Ativação	42
2.4.2	Normalização	44
2.5	MCAV de Judd	44
2.6	Considerações Finais	47
3	SISTEMAS DE RECOMENDAÇÃO	49
3.1	Introdução	49
3.2	Taxonomia	50
3.2.1	Dados de entrada	50
3.2.2	Saídas	52
3.3	Filtragem de Informações em SR	52
3.3.1	Filtragem Baseada em Conteúdo - FBC	53

3.3.2	Filtragem Colaborativa - FC	55
3.3.3	Filtragem Híbrida - FH	64
3.4	Medida de Avaliações	65
3.4.1	Medida de Exatidão da Previsão	66
3.4.2	Medida de Exatidão do <i>Ranking</i>	66
3.4.3	Validação Estatística	67
3.5	Considerações Finais	68
4	IKB-MS	71
4.1	Visão Geral da Proposta	71
4.1.1	Etapa offline	71
4.1.2	Etapa online	75
4.1.3	Recomendação de itens não avaliados	76
4.2	Considerações Finais	76
5	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	77
5.1	Introdução	77
5.2	Configurações dos Experimentos	77
5.2.1	Base de Dados	77
5.2.2	Critérios de Avaliação	80
5.2.3	Algoritmos Comparados	81
5.2.4	Configurações dos algoritmos utilizados	81
5.3	Experimentos	82
5.3.1	Escolha da medida de similaridade usada na SMS	82
5.3.2	Escolha do α usado na medida $sc(i, j)$	83
5.3.3	Escolha do número de partições	83
5.3.4	Escolha do tamanho de vizinhança para a predição	83
5.3.5	Minimizando a partida a frio de itens	86
5.3.6	Recomendação Top-N	88
5.3.7	Teste Estatístico	89
5.4	Considerações Finais	89
6	CONCLUSÃO	91
6.1	Conclusões	91
6.2	Limitações encontradas	91
6.3	Principais Contribuições	92
6.4	Trabalhos Futuros	92
6.5	Contribuições em Produção Bibliográfica	92
REFERÊNCIAS	95

Introdução

Com a disseminação da *Internet*, os serviços providos por ela estão sendo largamente utilizados, como é o caso do comércio eletrônico ou *e-commerce*, que inicialmente era usado para compra/venda de produtos modestos como CDs, livros, etc. Atualmente, esse meio de compra é utilizado para comercializar desde produtos com altos valores como carros, casas e barcos de luxo, a produtos que até pouco tempo eram inimagináveis suas vendas por esse meio, como são os casos dos alimentos, roupas e perfumes.

O comércio eletrônico tem proporcionado uma enorme comodidade aos consumidores e tem-se tornado um negócio multibilionário, mas, também fez com que seja mais difícil para os consumidores selecionarem produtos/itens que melhor atendam suas necessidades, em meio ao grande volume de informações disponíveis.

Nesse cenário, surgem os Sistemas de Recomendação (SR), que 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 da preferência dos usuários (O'DONOVAN; SMYTH, 2005; ADOMAVICIUS; TUZHILIN, 2005).

Com esses sistemas, ao tentar fazer uma escolha de produto, o usuário só recebe as informações que são de seu interesse. Alguns exemplos de SR são: recomendação de livros da Amazon¹, sugestão de vídeos do YouTube², recomendação de filmes da NETFLIX³ e sugestão de amigos do Facebook⁴.

De uma maneira geral, os SR são divididos de acordo com as técnicas de filtragem de informações utilizadas, como Filtragem Baseada em Conteúdo (FBC), Filtragem Colaborativa (FC) e Filtragem Híbrida (FH). A FBC (BALABANOVIĆ; SHOHAM, 1997) cria um perfil para cada item baseado nas suas características e um perfil de interesse para cada usuário. A recomendação consiste na combinação dos atributos do perfil do

¹ <http://www.amazon.com.br/>

² <http://www.youtube.com/>

³ <http://www.netflix.com/br/>

⁴ <https://www.facebook.com/>

usuário com os atributos dos perfis dos itens. Essa técnica avalia o histórico de compras dos usuários e recomenda itens semelhantes ao que ele já mostrou interesse.

A técnica FC (GOLDBERG et al., 1992) usa o comportamento passado do usuário (compras, navegação ou avaliação) e não requer informações sobre o conteúdo dos itens. Essa abordagem parte do princípio de que se duas pessoas compraram ou avaliaram itens de forma semelhante no passado, a probabilidade de elas continuarem concordando no futuro é grande. Já a FH procura combinar dois ou mais tipos de filtragem, podendo, assim, desfrutar das vantagens e minimizar os problemas de cada técnica.

Segundo (WIDISINGHE et al., 2010) a Rede Mundial de Computadores - *World Wide Web* (WWW) está se tornando uma grande fonte de imagens e, com isso, a procura por SR de imagens está aumentando rapidamente. Existem vários SR de imagens para as áreas comerciais, acadêmicas e de entretenimento, que lidam com a preferência dos usuários. Em (LEMOS et al., 2012) é apresentado o *Mobile Media to You* (MMedia2U), um SR de imagens, que sugere imagens previamente anotadas com informações contextuais. A fim de executar a recomendação, MMedia2U explora o contexto atual do usuário adquirido por seu dispositivo móvel. Este domínio é interessante, pois muitas das imagens disponíveis no Flickr⁵ e Picasa^{®6}, são captadas por dispositivos móveis que armazenam o local, data, hora e outras informações contextuais que podem ser exploradas para a recomendação.

Na proposta de (SANCHEZ et al., 2012) é criado um SR de imagens híbrido que combina a FBC com a FC para redes sociais móveis, em que os dispositivos têm capacidades limitadas em termos de tamanho de tela e poder de computação. Esse sistema ajuda a apresentar o conteúdo mais relevante para os usuários, economizando espaço de tela e processamento. Contudo, esses sistemas utilizam a FBC que necessita de informações dos itens para criar os perfis. Essas informações, muitas vezes, não estão disponíveis e os usuários podem não estar dispostos a fornecer informações sobre o seu perfil de interesse (SMYTH; COTTER, 2001). A FC é amplamente utilizada em diversos trabalhos (HU; ZHU; LI, 2014; XIAO-SONG et al., 2014; GOLDBERG et al., 1992; SU; KHOSHGOFTAAR, 2009; SCHAFER et al., 2007), por ser uma estratégia simples e eficaz, em que o interesse do usuário para um item é calculado por meio das avaliações de outros usuários ou itens similares. Com isso, a descrição dos itens para criar os perfis como ocorre na FBC não é necessária. Um desafio interessante é o uso de SR baseado em FC para recomendar imagens.

Segundo (WIDISINGHE et al., 2010), quando a preferência dos usuários nas imagens muda dependendo do contexto, os SR convencionais podem deixar de atingir o seu objetivo. Com isso, os autores propuseram um SR de imagens (picSEEK) que explora a FC para recomendar imagens apropriadas com relação ao contexto. Já em (BALOIAN et al., 2004), um SR baseado em FC é proposto, para sugerir material multimídia de

⁵ <http://www.flickr.com/>

⁶ Marca registrada da Google Inc ., Mountain View, EUA

aprendizado com base nas preferências do aluno, pois, de acordo com os autores, com a quantidade de material multimídia heterogêneo disponível, pode ser difícil para os alunos encontrarem um material que se encaixe às suas necessidades.

Vale ressaltar que um dos problemas presentes nos SR baseados em FC é que eles podem gerar recomendações ruins, quando um item é novo e não possui um número considerável de avaliações. Essa limitação é conhecida como problema de partida a frio de itens.

Com o objetivo de minimizar essa limitação, este trabalho propõe integrar ao SR baseado em FC a percepção visual humana modelada pelo Mapa de Saliência (MS). Um MS indica a conspicuidade de cada *pixel* da imagem, ou seja, a probabilidade com que partes da imagem irão atrair a atenção dos seres humanos. O mapa de saliência pode ser visto como uma imagem em níveis de cinza, em que o brilho de um *pixel* é proporcional à sua saliência (KOCH, 1985).

Para modelar a atenção visual dos itens, este trabalho propõe utilizar Métodos Computacionais de Atenção Visual (MCAV) como descrito em (ITTI; KOCH; NIEBUR, 1998; HAREL; KOCH; PERONA, 2006; JUDD et al., 2009). No SR proposto, denominado IKB-MS, as preferências dos usuários, representadas pelas avaliações, foram combinadas com informações de atenção visual dos usuários, obtidas a partir dos MS das imagens. Uma ilustração do fluxo dos dados do sistema proposto denominado IKB-MS: *ItemKNN+Baseline* (IKB) (KOREN, 2010) + Mapa de Saliência, é apresentado na Figura 1. A saída do algoritmo de recomendação é uma lista contendo os itens que possivelmente o usuário mais se interessaria. Nesta proposta, todo item é implicitamente representado por seu MS, independente se foi ou não avaliado pelos usuários.

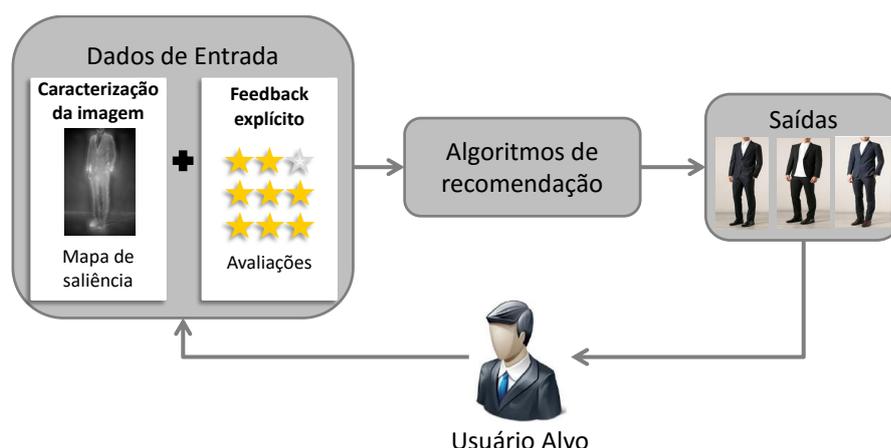


Figura 1 – Fluxo de dados do IKB-MS.

Um ponto forte dessa integração está no tratamento de novos itens, que ainda não foram avaliados pelos usuários ou que possuem poucas avaliações. Nos sistemas que utilizam a FC tradicional, enquanto um novo item não for avaliado por um número considerável de

usuários, o sistema não o recomendará. Com o IKB-MS, não será necessária a descrição do novo item e nem a avaliação do mesmo, para que o sistema possa recomendá-lo. Isso se deve ao fato de que, quando um item é novo e ainda não foi avaliado, ele ainda assim terá o seu MS, que determina os possíveis locais que chamam a atenção dos usuários.

Usando a similaridade entre os MS é possível determinar o quanto um novo item é similar aos itens já avaliados da base de dados, sendo assim possível determinar quais usuários poderão se interessar por esse novo item. Com isso, espera-se contribuir com a qualidade da recomendação de itens, especialmente àqueles em que as características visuais e atentivas são importantes para a definição da preferência dos usuários.

1.1 Objetivos e Hipótese

O objetivo geral desse trabalho é utilizar os métodos computacionais de atenção visual disponíveis na literatura, para melhorar a predição dos sistemas de recomendações baseados em filtragem colaborativa cujos itens são imagens. Para alcançar o objetivo geral deste trabalho, alguns objetivos específicos devem ser alcançados:

- Avaliar os métodos computacionais de atenção visual aplicados aos sistemas de recomendação de imagens;
- Minimizar o problema de partida a frio de itens em sistema de recomendação usando métodos computacionais de atenção visual;

A hipótese desse trabalho é que a atenção visual associada às preferências dos usuários melhora a precisão dos sistemas de recomendação de imagens baseados em filtragem colaborativa.

1.2 Contribuições

As principais contribuições do trabalho realizado são:

1. Proposta de uma medida de similaridade que pondera a atenção visual e as avaliações dos usuários.
2. Integração dos métodos computacionais de atenção visual aos sistemas de recomendação.
3. Minimização do problema de partida a frio dos itens na filtragem colaborativa.

A seguir, apresentamos como o presente trabalho está organizado.

1.3 Organização da Dissertação

Esta dissertação está organizada em seis capítulos, como mostrado a seguir:

- ❑ Capítulo 1 - apresenta a introdução e a motivação para realização do trabalho descrito nesta dissertação.
- ❑ Capítulo 2 - Descreve as principais diferenças entre os MCAV *top-down* e *bottom-up*. Os MCAV utilizados nesta pesquisa como (ITTI; KOCH; NIEBUR, 1998) (HAREL; KOCH; PERONA, 2006) e (JUDD et al., 2009) são descritos nesse capítulo.
- ❑ Capítulo 3 - Introduz os conceitos específicos sobre SR, apresentando também algumas técnicas de filtragem de informação e trabalhos que utilizam essas técnicas; o capítulo apresenta ainda algumas métricas de avaliação.
- ❑ Capítulo 4 - Apresenta todos os detalhes do método IKB-MS sendo proposto neste trabalho.
- ❑ Capítulo 5 - Discute os resultados dos experimentos realizados para validar a hipótese desta dissertação.
- ❑ Capítulo 6 - Conclui esta dissertação com algumas considerações e propostas para possíveis trabalhos futuros.

Métodos Computacionais de Atenção Visual

Este capítulo tem por objetivo apresentar os conceitos teóricos necessários sobre MCAV para o melhor entendimento deste trabalho. São apresentados detalhes de como a atenção visual pode contribuir para filtrar informações visuais em uma cena, e algumas técnicas propostas na literatura, como os MCAV de (ITTI; KOCH; NIEBUR, 1998) (HAREL; KOCH; PERONA, 2006) e (JUDD et al., 2009).

2.1 Introdução

O sistema visual é um dos principais sentidos para interação dos seres vivos com o ambiente. De acordo com (TORTORA; DERRICKSON, 2012), a visão é o sentido que responde a estímulos de luz. De forma geral, a visão pode ser descrita por fenômenos biológicos ou por métodos computacionais.

Uma característica muito importante na **visão biológica** é a seleção de informações relevantes de uma cena, pois o volume de informações que estimulam o campo visual dos sistemas biológicos é grande e, com seleção, a quantidade de informações a serem processadas é reduzida (TSOTSOS, 1990). Essa seleção de informações permite rápida detecção de predadores, perpetuação e evolução das espécies (ITTI; KOCH; NIEBUR, 1998). Com a estratégia de seleção apenas um subconjunto de estímulos é selecionado pelo cérebro a cada instante e esse subconjunto receberá um processamento rigoroso. Já o restante das informações visuais recebem apenas uma análise limitada (FISCHER; WEBER, 1993).

O funcionamento da visão biológica serve de inspiração para a **Visão Computacional (VC)**. A VC estuda e descreve sistemas de visão artificiais implementados por *hardware* ou *software* com objetivo de obter características específicas em uma determinada cena ou imagem (ULLMAN; POGGIO, 2010). Por esse motivo, inúmeras aplicações em diferentes áreas existem tais como, em medicina (AZEVEDO-MARQUES, 2001); na

indústria (MONGELO et al., 2011); na área militar (PAIVA; YEPES, 2014); na exploração espacial, entre outras, tornando a VC cada vez mais disseminada.

Outra aplicação da VC está no processamento de imagens em que a quantidade de informações a serem processadas é bem maior do que a capacidade de processá-las (TSOTSOS, 1990). Neste contexto, a atenção visual pode ser aplicada para lidar com a grande quantidade de informações visuais (FISCHER; WEBER, 1993). Mais detalhes sobre a atenção visual são apresentados a seguir.

2.2 Atenção Visual

Segundo (MILANESE, 1993), a atenção visual é a capacidade do ser humano de detectar rapidamente regiões de interesse na imagem retinal, a fim de reduzir a quantidade de dados para tarefas de processamento complexas, tais como reconhecimento de objetos.

Em (TSOTSOS, 1990), a complexidade computacional da atenção visual foi analisada e essa análise confirmou que a atenção visual é uma das mais importantes contribuições para otimizar o processamento de imagens. Essa técnica é uma alternativa para lidar com o grande volume de dados, pois é responsável por selecionar e processar rapidamente somente as regiões mais relevantes (chamadas ROI - Regiões de interesse) numa imagem, enquanto ignora o restante das informações (FISCHER; WEBER, 1993).

As regiões de atenção visual de um ser humano em uma imagem podem ser determinadas geralmente de duas formas: por meio de Mapas de Saliência (MS) obtidos com os MCAV, ou por dispositivos de rastreamento ocular¹.

Os rastreadores oculares são caros e não estão facilmente acessíveis a todos os usuários finais dos SR. Por esse motivo, utilizar os MCAV é uma boa alternativa, pois não requer nenhum *hardware* específico. Vários MCAV como os desenvolvidos por (KOCH, 1985) (NIEBUR; KOCH, 1996) (ITTI; KOCH; NIEBUR, 1998) (RAJASHEKAR et al., 2008) (HAREL; KOCH; PERONA, 2006) e (TORRALBA, 2003) tentam representar a atenção visual por meio do MS. Esses MCAV podem ser divididos de acordo com o seguinte aspecto: se o estímulo que guia a atenção é proveniente do cérebro ou do ambiente, nomeadamente atenção modelos *top-down* e atenção modelos *bottom-up*, respectivamente.

2.2.1 Modelos de atenção visual *top-down*

Com os modelos *top-down* a atenção visual pode ser direcionada para alvos específicos, de acordo com a importância atribuída pelo córtex cerebral que altera o foco da atenção.

¹ Os dispositivos de rastreamento ocular são um equipamento não invasivo de pequena escala, que registram os movimentos dos olhos. Os movimentos oculares são gerados de acordo com o interesse dos usuários ou estímulo sobre o conteúdo da imagem visualizada, que pode ser diferente de pessoa para pessoa.

Para isso, utilizam-se as características de alto nível das imagens como conhecimento obtido *a priori* para detectar regiões de maior interesse em uma imagem.

Por exemplo, se uma pessoa está procurando por um objeto com uma forma geométrica específica, aspectos de alto nível podem guiar o processo seletivo de atenção durante a busca e ignorar outras características visuais irrelevantes ou que não compõem o objeto em questão.

Porém, além das características de alto nível, outros conhecimentos também podem ser obtidos por um ser humano, que pode selecionar manualmente regiões de maior interesse numa imagem. São exemplos de modelos *top-down* para representar a atenção visual os modelos desenvolvidos por (MCCALLUM, 1996) (RAO et al., 2002) (LAAR; HESKES; GIELEN, 1997) (BORJI; AHMADABADI; ARAABI, 2011) (BACON; EGETH, 1994) e (THEEUWES, 1992).

Em (BACON; EGETH, 1994), alguns experimentos foram realizados com o intuito de demonstrar que, quando existe uma grande quantidade de elementos em uma cena, a atenção visual *top-down* pode não sofrer com a presença de um distrator² constante como, por exemplo, um distrator de cor diferente.

Já, no trabalho de (THEEUWES, 1992), são apresentados alguns experimentos baseados na busca por um alvo específico, como características conhecidas previamente pelos observadores, ou seja, uma busca a partir de um conhecimento prévio do usuário (*top-down*) ou por estímulos visuais (*bottom-up* descrito com mais detalhes na seção 2.2.2). A Figura 2 representa um desses exemplos, em que o alvo é previamente conhecido e possui a forma de um círculo na cor verde, contendo em seu interior uma linha com orientação vertical ou horizontal.

Na Figura 2 (a), os distratores podem se diferenciar do alvo em relação à forma, em que o alvo se encontra em meio a quadrados verdes. Na Figura 2 (b), além da forma diferente, os distratores podem variar a cor. Nesse caso, existe um único distrator de cor vermelha. Na Figura 2 (c), os distratores variam em relação à cor. Nesse caso, o alvo se encontra em meio a distratores de mesma forma, porém de cor diferente. Por fim, na Figura 2 (d), entre os distratores, pode existir um com uma forma diferente.

Com esse exemplo, os autores observaram que a procura por um alvo que contraste com os demais elementos não interfere no tempo necessário para a busca do alvo. Entretanto, em um ambiente no qual o alvo se difere dos distratores em relação à forma, a procura torna-se mais lenta na presença de outro distrator de cor diferente. Outro exemplo é em um ambiente no qual o alvo se difere dos distratores em relação a cor. A procura será mais lenta na presença de outro distrator de forma.

Esse experimento mostra que a atenção visual baseada exclusivamente em uma característica *top-down* (que neste caso é a forma), não é suficiente para guiar a atenção. Contudo, a atenção visual também pode ser direcionada de forma involuntária (*bottom-up*)

² Distratores são alvos destinados a distrair a atenção.

por elementos da imagem que se destacam dos demais em relação à cor. Essa abordagem está descrita com maiores detalhes a seguir.

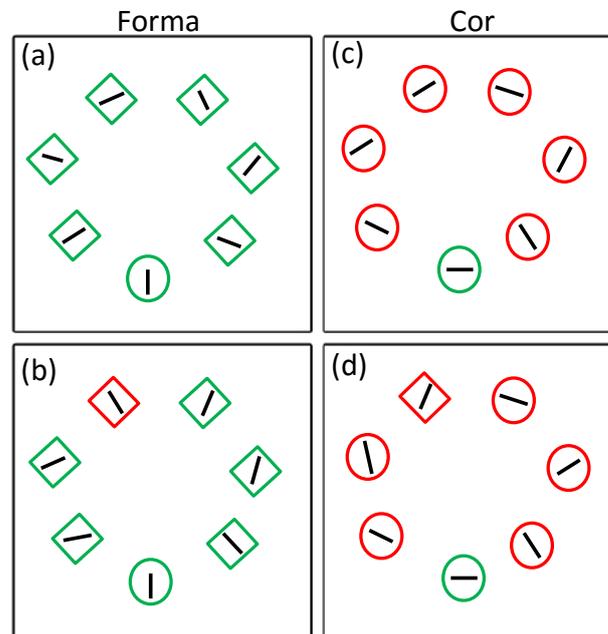


Figura 2 – Atenção visual baseada em características *top-down* e *bottom-up*, adaptada de (THEEUWES, 1992). (a) distratores variam em relação à forma, (b) existe um único distrator de cor vermelha, (c) vários distratores de outra cor e (d) entre os vários distratores, um pode ter forma diferente.

2.2.2 Modelos de atenção visual *bottom-up*

Os modelos *bottom-up* utilizam características de baixo nível das próprias imagens como cor, intensidade e orientação. Esse método atua de modo inconsciente, ou seja, o observador é levado a fixar sua atenção em determinadas regiões da imagem devido aos estímulos causados pelos contrastes entre características visuais presentes na imagem. Por exemplo, em uma imagem que contenha apenas objetos verdes e um único objeto vermelho, graças às características de cor, o objeto vermelho será um melhor candidato durante o processo de competição por atenção. Neste método nenhuma informação contextual é usada para definir a atenção visual (WOLFE; HOROWITZ, 2004). A Tabela 1 apresentada por (WOLFE; HOROWITZ, 2004) mostra outras características responsáveis por guiar o mecanismo biológico de atenção visual.

A busca por um ponto de maior saliência pode ser simples em alguns casos, porém não tão simples para outros. A Figura 3 (a) é um exemplo disso: encontrar o alvo vermelho, o maior alvo ou ainda o alvo inclinado tende a ser uma tarefa fácil. No entanto, em um primeiro momento, dificilmente é possível identificar o número “2” existente nesta imagem, sem que alguém tenha dito que há um número “2” em meio a vários “5”s. Depois do “2” ter

Tabela 1 – Características que podem guiar a atenção visual (WOLFE; HOROWITZ, 2004).

Com certeza	Provavelmente	Possivelmente	Talvez
Cor	Luminância	Direção da Iluminação	Novidade
Orientação	Profundidade	<i>Aspect ratio</i>	Categoria alfanumérica
Tamanho	Terminação de Linha	Número	Tipo de Letra

seu encontrado não há dificuldade alguma em encontrá-lo novamente. Contudo, quando maior for a quantidade de “5”s, mais difícil será encontrar o “2”. Os pares de triângulos na horizontal da Figura 3 (b) são difíceis de serem encontrados, mas na Figura 3 (c) essa tarefa é facilitada, pois existe um contraste de cores entre os retângulos pretos e os retângulos azuis. Já na Figura 3 (d), a busca por cruzeiros é complexa, devido ao fato de que as informações de intersecção não guiam a atenção visual (WOLFE; HOROWITZ, 2004).

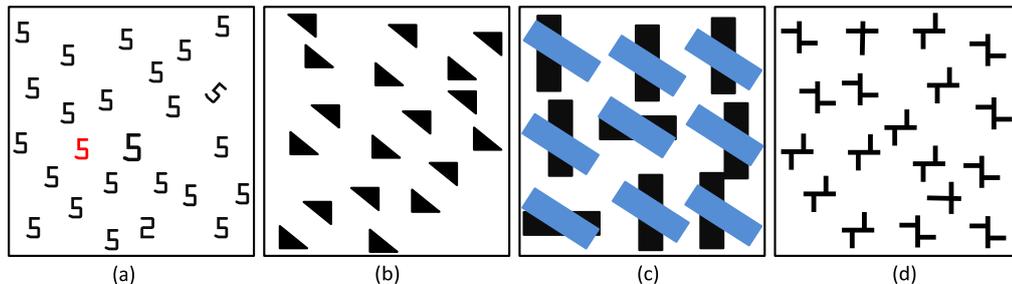


Figura 3 – Exemplos de tarefas de busca visual, adaptada de (WOLFE; HOROWITZ, 2004). (a) contraste de cores, (b) e (c) demonstram a importância da orientação e cor e (d) informação de intersecção.

O modelo *bottom-up* vem sendo amplamente utilizado por diversos trabalhos, como (TORRALBA, 2003) (MA et al., 2005) (PRIVITERA; STARK, 2000) (LUO; SINGHAL, 2000) (ITTI; KOCH; NIEBUR, 1998) (HAREL; KOCH; PERONA, 2006) e (WANG et al., 2011). (KOCH, 1985) propõe a utilização do mapa de saliência, como forma de reunir em um único mapa topográfico a conspicuidade de partes do campo visual. (ZHENG et al., 2007) propõe o uso de dois mapas de saliência que são baseados em diferenças de contraste globais. Ambas as técnicas produzem um MS em alta resolução e comparável a outros métodos de saliências.

Outra abordagem *bottom-up* é baseada no domínio da frequência. (GUO; ZHANG, 2010) propõem um método de detecção de saliências espaço-temporal denominado *Phase Spectrum of quaternion Fourier Transform* (PQFT). Uma vantagem desse método é o uso de aspectos temporais, o que permite que o método seja aplicado não somente em imagens estáticas, mas também em vídeos.

Em (MELO; AMO; GULIATO, 2014) vários MCAV para a recuperação em diferentes domínios visuais foram analisados, como (ITTI; KOCH; NIEBUR, 1998) (HAREL;

KOCH; PERONA, 2006) e (JUDD et al., 2009). Cada um desses métodos trabalha de forma distinta, mas independente do processo, a saída é um MS, como representado na Figura 4. É possível notar que o MS resultante não tem distinção de cor, textura e outras características, mas essas informações são consideradas durante a extração do MS. Com isso, se uma característica da imagem chama mais atenção do que outra, essa diferença será representada no MS.

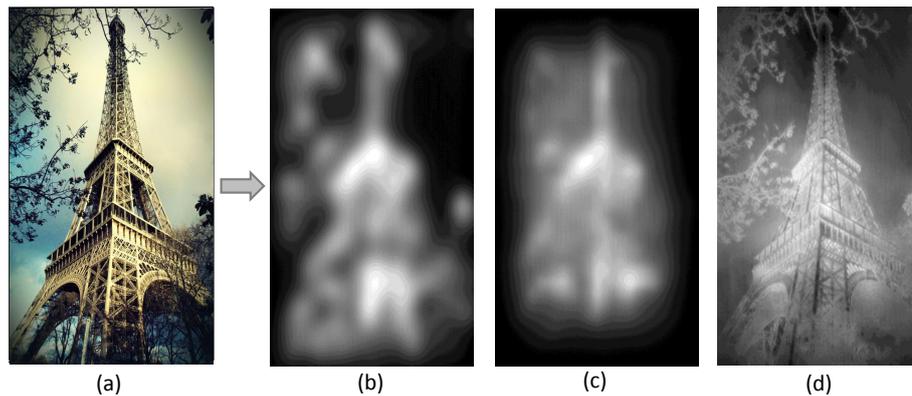


Figura 4 – **Três MCAV foram aplicados em uma imagem.** (a) imagem original obtida em <http://blog.mundi.com.br/>, (b) MS de (ITTI; KOCH; NIEBUR, 1998), (c) MS de (HAREL; KOCH; PERONA, 2006) e (d) MS da (JUDD et al., 2009).

Nas próximas seções são descritos dois MCAV *bottom-up* proposto por (ITTI; KOCH; NIEBUR, 1998) e (HAREL; KOCH; PERONA, 2006). Outro MCAV descrito neste trabalho é um híbrido *bottom-up* e *top-down* desenvolvido por (JUDD et al., 2009).

2.3 MCAV de Itti

O método proposto por (ITTI; KOCH; NIEBUR, 1998) baseia-se no comportamento do sistema visual dos primatas e é formado pela decomposição de mapas de características, baseando em aspectos do sistema visual humano como cor, intensidade e orientação para a extração do MS.

O propósito desse MS é a modelagem da atenção visual, baseando-se na ideia de que a atenção é direcionada para a diferença de contraste local dos atributos visuais. O método proposto por (ITTI; KOCH; NIEBUR, 1998) é uma das técnicas mais usadas na literatura para extrair a atenção visual das imagens. Esse método está dividido nas seguintes etapas: **filtragem linear; diferenças centro-vizinhanças e normalização; combinações variando escala e normalização; combinação linear.** Cada uma dessas etapas são descritas a seguir e estão ilustradas na Figura 5, que apresenta a arquitetura geral desse método.

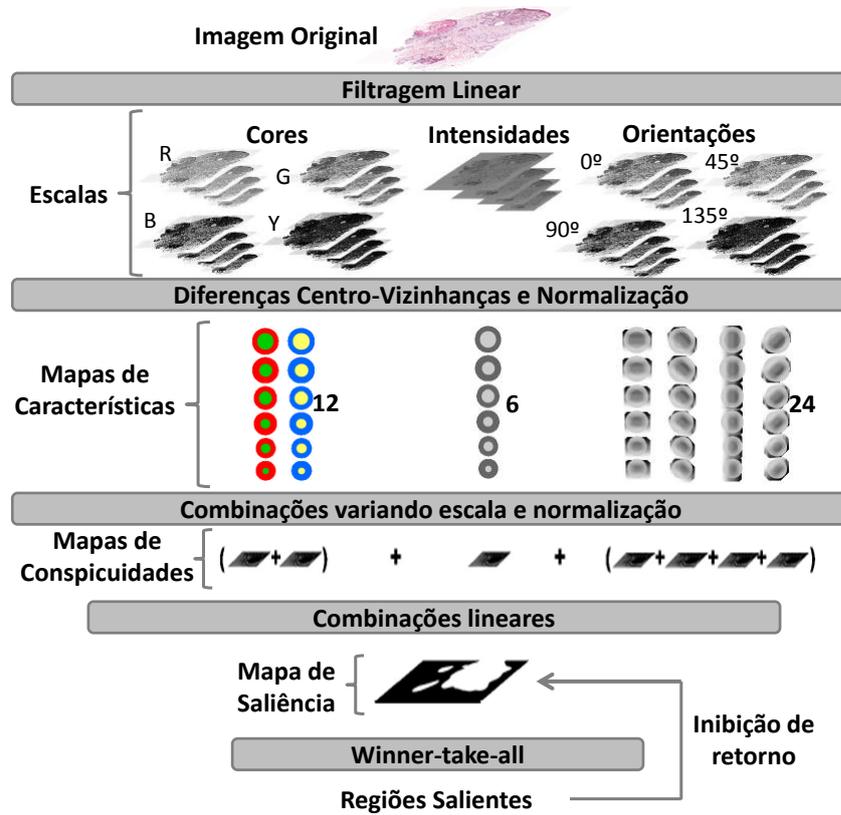


Figura 5 – Arquitetura do método de Itti, adaptada de (LOPERA et al., 2012).

2.3.1 Filtragem Linear

De um modo geral, o método é alimentado inicialmente por uma **imagem de entrada** e, em seguida, são extraídas as seguintes características visuais: cor, intensidade e orientação. Com a extração dos canais r , g e b (*red*, *green* e *blue*) da imagem de entrada, a intensidade da imagem é obtida por $I = (r + g + b)/3$. Os canais r , g e b são normalizados a partir de I . O objetivo dessa normalização é inibir regiões que apresentem baixos valores de luminosidade, ou seja, que não são salientes. Em seguida, os quatro canais de cores são criados da seguinte forma: vermelho (R - representado na Equação 1), verde (G - representado na Equação 2), azul (B - representado na Equação 3) e amarelo (Y - representado na Equação 4), os resultados obtidos com esses canais estão ilustrados na Figura 6 (ITTI; KOCH; NIEBUR, 1998).

$$R = r - \frac{g + b}{2} \quad (1)$$

$$G = g - \frac{r + b}{2} \quad (2)$$

$$B = b - \frac{r + g}{2} \quad (3)$$

$$Y = \frac{r + g}{2} - \frac{|r - g|}{2} - b \quad (4)$$

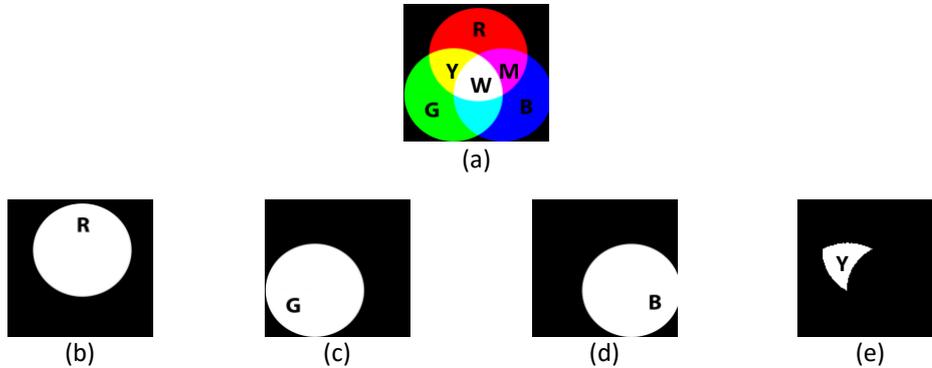


Figura 6 – **Extração dos 4 canais de cores.** (a) imagem de entrada, (b) extração do canal vermelho, (c) extração do canal verde, (d) extração do canal azul e (e) extração do canal amarelo.

A imagem de intensidade é representada por I , tanto ela, quanto os canais R , G , B e Y são utilizados para criar as pirâmides Gaussianas (BURT; ADELSON, 1983), em que $\sigma \in \{0..8\}$ é o fator de escala e corresponde aos níveis das pirâmides. Essas pirâmides são criadas com o objetivo de obter amostras da imagem em que detalhes indesejados e ruídos sejam reduzidos e as características importantes sejam realçadas. A Figura 7 mostra um exemplo de pirâmide Gaussiana. Essa pirâmide é composta por nove níveis.

No modelo de pirâmide tradicional, uma imagem é reamostrada em níveis diferentes de resolução, a partir da aplicação de algum filtro, que pode ser, por exemplo, o filtro gaussiano, formando uma estrutura similar a uma pirâmide (de imagens). A base da pirâmide é uma imagem com a mesma resolução e tamanho que a imagem original, já as imagens seguintes sofrem uma redução de tamanho, mas mantendo a mesma área de abrangência da imagem original (toda a imagem é reamostrada). Conseqüentemente, as imagens menores possuem resoluções menores, pois cobrem a mesma área que a base da pirâmide, mas com uma quantidade menor de *pixels* (UHR, 1972) (WITKIN, 1983) (BURT; ADELSON, 1983) (SCHWARTZ; GREVE; BONMASSAR, 1995).

As informações locais são características importantes para o desenvolvimento da atenção visual. Segundo (GREENSPAN et al., 1994) a extração destas características pode ser obtida por meio da aplicação de filtros direcionais sobre a imagem. Por esse motivo, além da pirâmide Gaussiana, o filtro de Gabor descrito em (GABOR, 1946), também é utilizado na filtragem linear. A Figura 8 apresenta um exemplo em que o contraste na orientação pode guiar a atenção visual.

Os mapas de orientações $O_\sigma(\theta)$ são criados por meio da convolução do mapa de intensidade I_σ , utilizando o filtro de Gabor, para quatro orientações $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

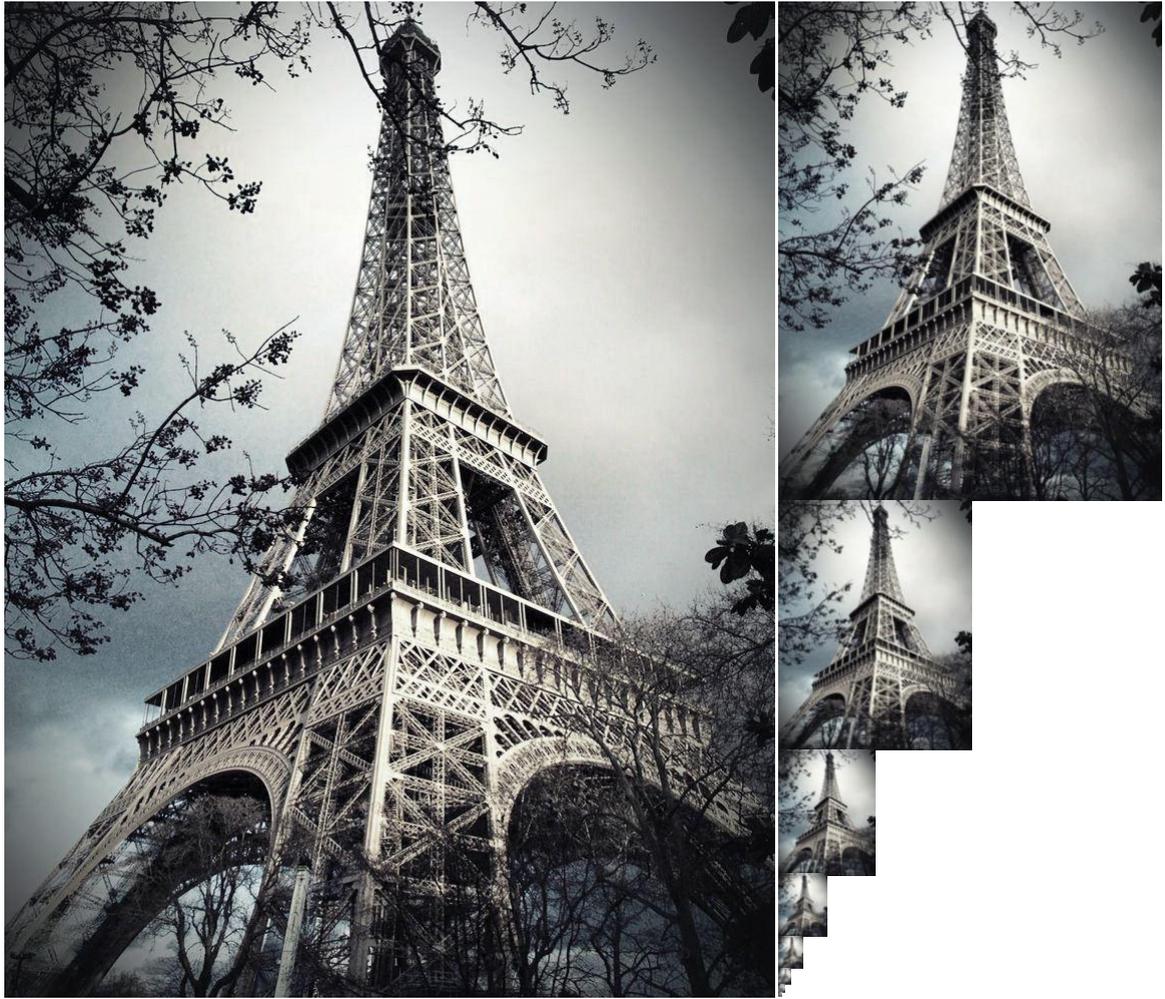


Figura 7 – Pirâmide Gaussiana de 9 níveis.

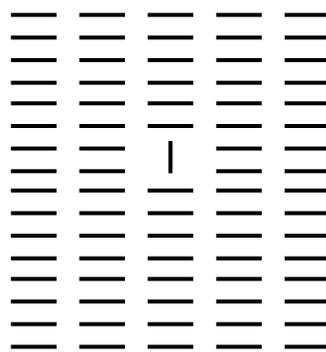


Figura 8 – **Exemplo de contraste na orientação.** Barra vertical inserida em um ambiente com barras horizontais, representando a diferença nas orientações, e por isso, se torna atenta.

O intuito desses filtros é identificar as barras ou bordas em uma determinada direção, como ilustrado na Figura 9.

Para demonstrar o processo de filtragem linear de Gabor, na Figura 10 é apresentado

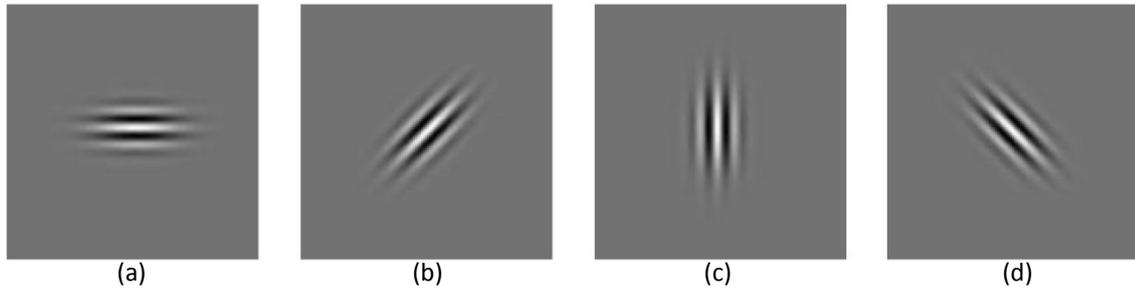


Figura 9 – As imagens das intensidades dos quatro *kernels* de Gabor são usadas para determinar a informação de orientação. (a) 0° , (b) 45° , (c) 90° e (d) 135° , adaptada de (SIKLOSSY, 2005)

o resultado deste processo em uma imagem sintética para os quatro *kernels* considerados. Como pode ser observado na Figura 10 (c), em que aplica-se o filtro linear com orientação 45° , a informação irrelevante é totalmente suprida, sendo destacadas somente as barras com orientação de 45° .

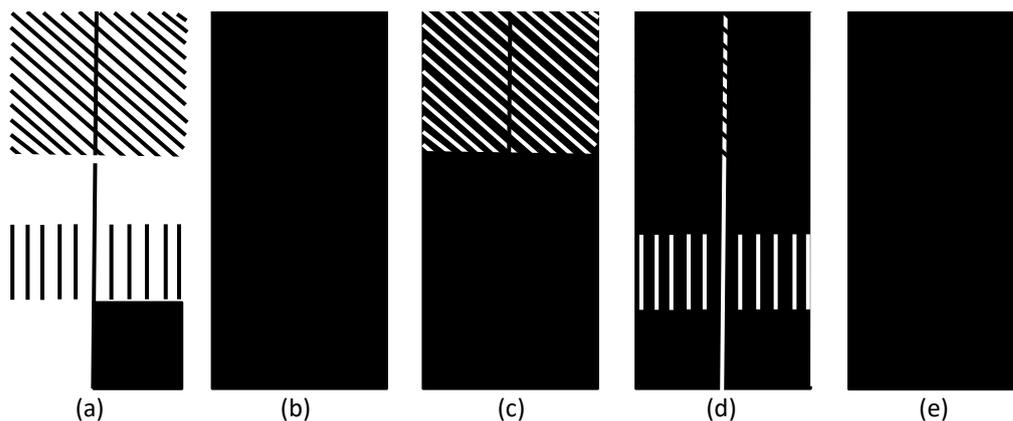


Figura 10 – Extração das informações de orientação utilizando filtragem linear de Gabor. (a) imagem de entrada, (b) informações filtradas com 0° , (c) informações filtradas com 45° , (d) informações filtradas com 90° e (e) informações filtradas com 135° , adaptada de (SIKLOSSY, 2005)

Com as pirâmides definidas, os mapas de características são obtidos efetuando-se a diferença entre os canais em diferentes escalas. Essa diferença é descrita a seguir.

2.3.2 Diferenças centro-vizinhanças e normalização

Este processo é conhecido como diferença centro-vizinhança³, definido por \ominus . O centro é um *pixel* na escala $c \in \{2, 3, 4\}$ (escalas finas) e sua vizinhança corresponde ao *pixel* na

³ O processo de diferença centro-vizinhança é inspirado biologicamente pela estrutura dos campos visuais. Segundo (ITTI, 2000), os neurônios dos campos visuais correspondem a uma pequena região do espaço visual central. Por sua vez, os estímulos apresentados nas regiões vizinhas inibem a resposta neural.

escala $s = c + \delta$ (escalas grossas), em que $\delta \in \{3, 4\}$.

O primeiro conjunto contendo os 6 mapas de características, é obtido com o contraste de intensidades $I(c, s)$, representado pela Equação 5.

$$\mathcal{I}(c, s) = |I(c) \ominus I(s)| \quad (5)$$

O segundo conjunto contém um total de 12 mapas e são construídos a partir dos canais de cores. Esses mapas são gerados com relação à rivalidade de cores, em que os neurônios no centro são excitados por uma cor e inibidos por outra, ocorrendo o inverso na vizinhança. Por meio dessa relação, são criados os mapas entre as rivalidades (*vermelho * verde*), (*verde * vermelho*), (*amarelo * azul*) e (*azul * amarelo*), conforme as Equações 6 e 7.

$$(\mathcal{RG})(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (6)$$

$$(\mathcal{BY})(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \quad (7)$$

Por fim, os 24 mapas restantes são obtidos a partir das informações de orientações, em que, $O_\sigma(\theta)$ e $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ é a orientação preferencial. Os mapas de orientações são criados com relação ao contraste da orientação entre as escalas do centro e da vizinhança, sendo gerados separadamente para cada orientação. A Equação 8 ilustra o cálculo para gerar os mapas de orientações.

$$\mathcal{O}(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (8)$$

Com a geração dos mapas de características, obtêm-se 6 mapas para intensidade, 12 mapas para cores e 24 mapas para orientação, totalizando 42 mapas. Esta combinação dos mapas faz com que os objetos sejam vistos mais nitidamente. Para proporcionar uma melhor saliência, as combinações variando escala e normalização devem ser realizadas sobre esses mapas de características.

2.3.3 Combinações variando escala e normalização

Para combinar os mapas de características em somente um MS, é necessária a utilização de um operador $\mathcal{N}(\cdot)$ de normalização. O objetivo desse operador de normalização é o de que uma região que apresente um nível de saliência **contrastante** com as demais seja amplificada e que, por outro lado, regiões salientes **não contrastantes** sejam mutualmente inibidas. A Figura 11 demonstra o que acontece com a normalização $\mathcal{N}(\cdot)$.

O operador de normalização é calculado da seguinte forma:

1. Normalização dos mapas de características;

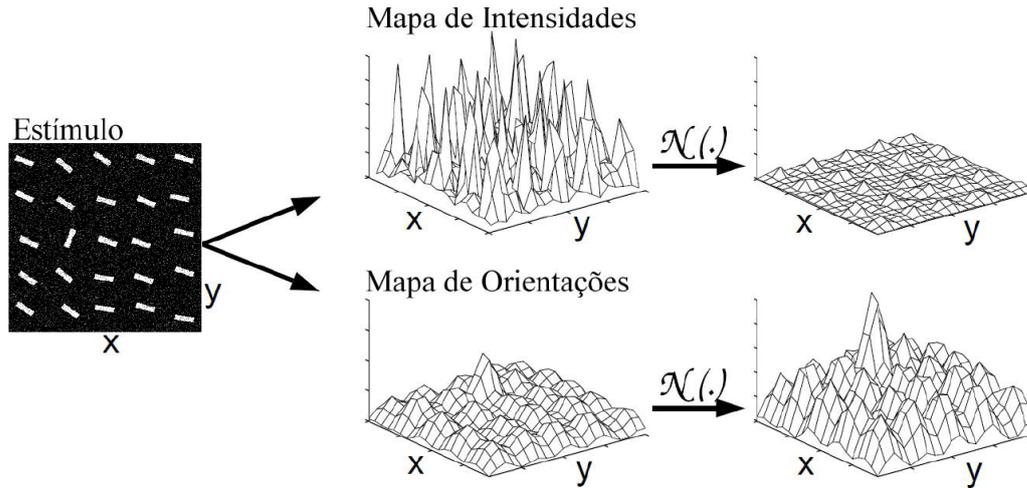


Figura 11 – Exemplo do comportamento do operador de normalização $\mathcal{N}(\cdot)$, adaptada de (ITTI; KOCH; NIEBUR, 1998).

2. Para cada mapa normalizado, encontra-se o máximo global M e a média \bar{m} de todos os outros máximos locais;
3. Multiplica-se globalmente o mapa por $(M - \bar{m})^2$.

Os mapas de características gerados são, por sua vez combinados em três mapas de conspicuidades, $\bar{\mathcal{I}}$ para intensidade (Equação 9), $\bar{\mathcal{C}}$ para cor (Equação 10) e $\bar{\mathcal{O}}$ para orientação (Equação 12). Nessas equações $\mathcal{N}(\cdot)$ representa a normalização e o operador \oplus representa a adição em escala, que consiste na redução de cada mapa a escala quadro com $\sigma = 4$. A motivação para a criação de três canais separados ($\bar{\mathcal{I}}$, $\bar{\mathcal{C}}$ e $\bar{\mathcal{O}}$) é a hipótese de que características similares competem pela saliência, enquanto que características diferentes contribuem independentemente para o MS.

$$\bar{\mathcal{I}} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{I}(c, s)) \quad (9)$$

$$\bar{\mathcal{C}} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(\mathcal{RG}(c, s)) + \mathcal{N}(\mathcal{BY}(c, s))] \quad (10)$$

$$\bar{\mathcal{O}} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} \mathcal{N} \left(\bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{O}(c, s, \theta)) \right) \quad (11)$$

2.3.4 Combinações lineares

Para a construção do MS final (\mathcal{S}), os três mapas de conspicuidade são normalizados, somados e uma média é efetuada.

$$\mathcal{S} = \frac{1}{3}(\mathcal{N}(\bar{\mathcal{I}}) + \mathcal{N}(\bar{\mathcal{C}}) + \mathcal{N}(\bar{\mathcal{O}})) \quad (12)$$

Para evitar que uma mesma região seja determinada como sendo mais saliente mais de uma vez, e para que seja possível determinar várias regiões salientes (mesmo que tais regiões possuam *pixels* de mesmo valor), utiliza-se o princípio da inibição de retorno. Com isso, o MS gerado alimenta uma rede neural *Winner-Take-All* (WTA) (KOCH, 1985). Uma WTA detecta e direciona a atenção para o ponto mais saliente que logo em seguida é suprimido por um mecanismo de inibição de retorno. Conseqüentemente, a atenção é direcionada para o próximo ponto mais saliente. A Figura 12 mostra um exemplo de uma imagem e seu respectivo MS obtido com o método descrito.

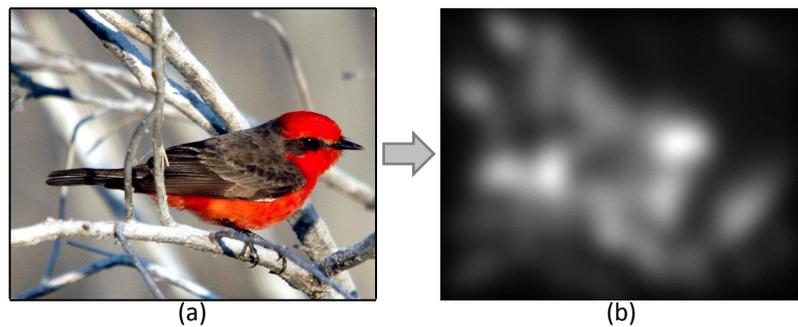


Figura 12 – Mapa de saliência extraído utilizando o método de Itti.

Outro MCAV *bottom-up* assim como o método de (ITTI; KOCH; NIEBUR, 1998) é descrito na seção seguinte.

2.4 MCAV de Harel

Em (HAREL; KOCH; PERONA, 2006), foi proposto um outro método *bottom-up* de extração da atenção visual baseado em grafos - Saliência Visual Baseada em Grafo - *Graph-Based Visual Saliency* (GBVS). Esse método propôs uma alternativa ao método de (ITTI; KOCH; NIEBUR, 1998) para as fases de **diferenças centro-vizinhanças e normalização** e **combinações variando escala e normalização**, conforme ilustrado na Figura 5. Essas novas fases foram chamadas de **ativação** e **normalização/combinação** e estão representadas na Figura 13 em azul.

Para os dois processos citados, ativação e normalização/combinação, são construídos grafos direcionais com pesos nas arestas a partir dos mapas de característica. Os grafos são tratados como uma cadeia de Markov⁴ para calcular a distribuição de equilíbrio (HAREL; KOCH; PERONA, 2006).

⁴ Na cadeia de Markov os estados anteriores são irrelevantes para a predição dos estados seguintes, desde que o estado atual seja conhecido. Por exemplo, em uma máquina de estados finitos, se você está no estado y no tempo n , então a probabilidade de que você se mova para o estado x no tempo $n + 1$ não depende de n e somente depende do estado atual y em que você está (SIMON; BLUME; DOERING, 2004).

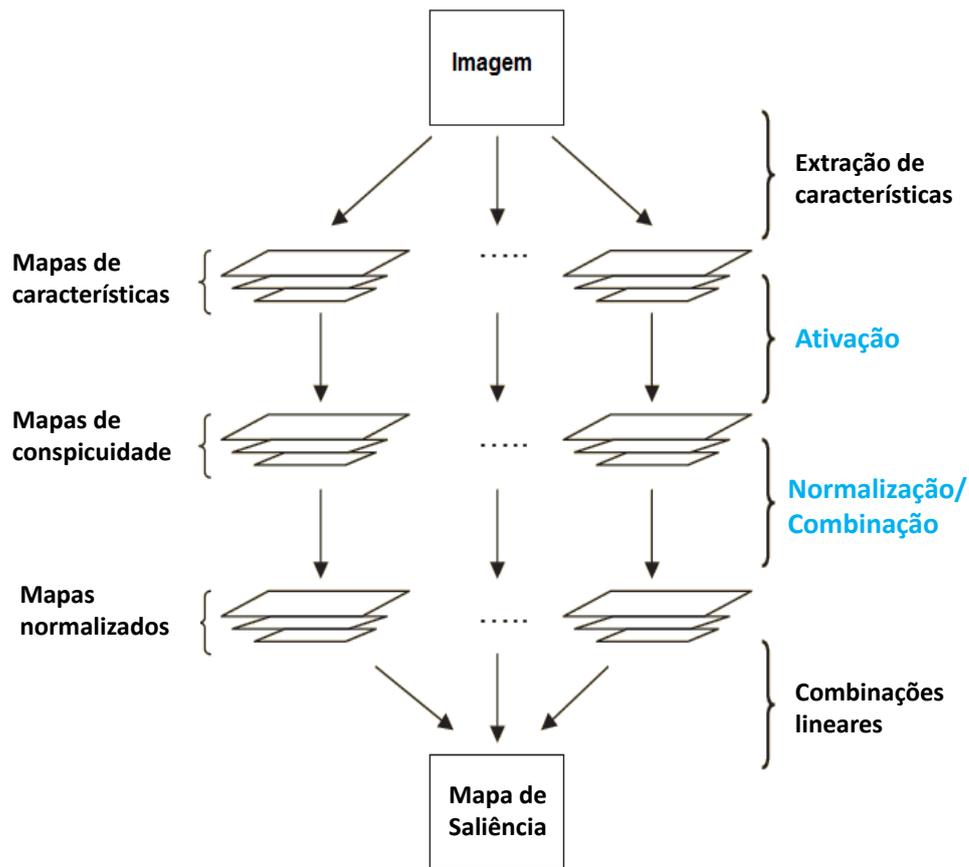


Figura 13 – Arquitetura do método de Harel, adaptada de (HAREL; KOCH; PERONA, 2006)

2.4.1 Ativação

Assim como no método de (ITTI; KOCH; NIEBUR, 1998), esta fase tem como entrada os mapas de características (gerados pela fase de extração de características), conforme mostra a Figura 13. Os mapas são representados como $M : [n]^2 \rightarrow \mathbb{R}$, sendo n o número de nós. O objetivo dessa fase de Ativação é gerar o mapa de conspicuidade $A : [n]^2 \rightarrow \mathbb{R}$, de tal forma que os valores de $M(i, j)$ que forem incomuns em sua vizinhança corresponderão a valores elevados em A . Dessa forma, para se obter A , é criado um grafo direcionado G_A totalmente conectado.

1. Um nó é instanciado em todas as regiões do mapa (característica ou conspicuidade) de entrada, conforme ilustrado na Figura 14 (a).
2. Criam-se arestas bidirecionais em todos os nós do grafo tornando-o totalmente conectado, conforme mostra a Figura 14 (b).
3. Ponderam-se as arestas, conforme mostra a Figura 14 (c). A aresta direcionada do nó (i, j) para o nó (p, q) , terá sua ponderação calculada conforme a Equação 13.

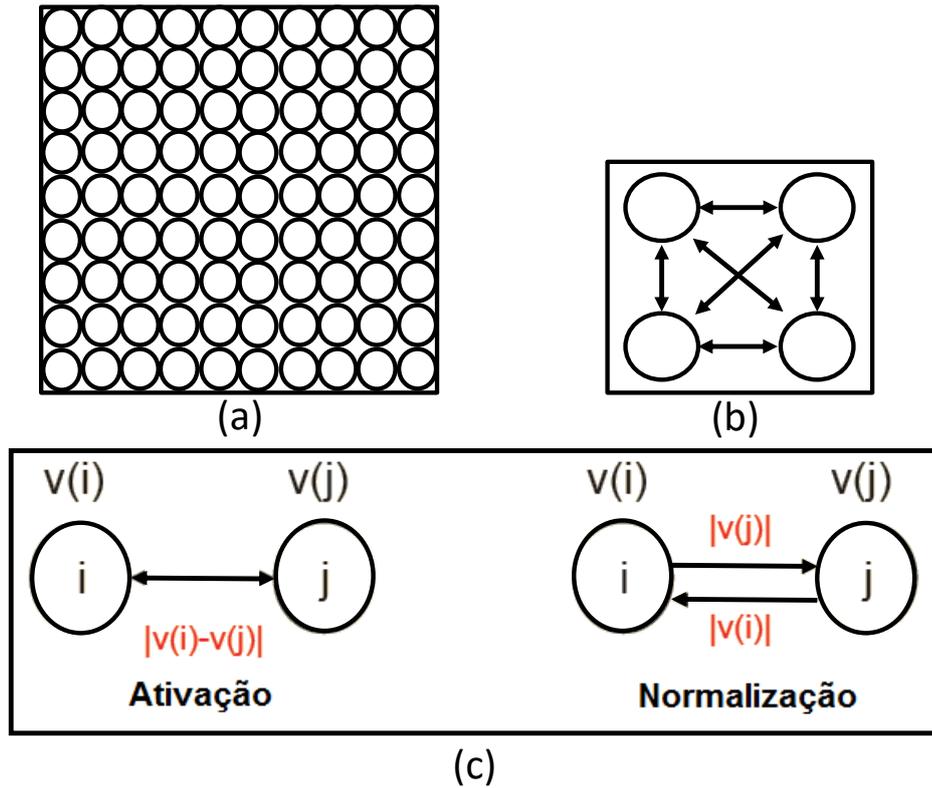


Figura 14 – **Criando** G_A . (a) representação dos nós em um mapa (características ou conspicuidades) de entrada, (b) representação das arestas bidirecionais do grafo e (c) representação da ponderação das arestas no grafo direcionado tanto para a fase de Ativação quanto para a fase de Normalização.

$$w_1((i, j), (p, q)) \triangleq d((i, j) || (p, q)) F(i - p, j - q) \quad (13)$$

em que $d((i, j) || (p, q))$ é dado pela Equação 14 e $F(i - p, j - q)$ é dado pela Equação 15, em que $a = i - p$ e $b = j - q$.

$$d((i, j) || (p, q)) \triangleq \left| \log \frac{M(i, j)}{M(p, q)} \right| \quad (14)$$

$$F(a, b) \triangleq \exp\left(-\frac{a^2 + b^2}{2\sigma^2}\right) \quad (15)$$

Com as ponderações para G_A calculadas, a distribuição de massa pode ser realizada. A cada passo, cada nó soma as massas de entrada e, em seguida, passa as medidas dessa massa para seus vizinhos de acordo com os pesos calculados nas arestas. Esse mesmo processo acontece nos nós simultaneamente e dá origem a uma distribuição equilibrada de massa, formando então o mapa de conspicuidade A .

A próxima etapa é a normalização, em que o mapa de conspicuidade A é normalizado com o objetivo de concentrar suas massas.

2.4.2 Normalização

Considerando que $A : [n]^2 \rightarrow \mathbb{R}$, então o grafo direcionado G_N com n^2 nós é construído. Para cada nó (i, j) e todos os nós (p, q) (incluindo (i, j)) ao qual é conectado, é gerada a aresta de (i, j) para (p, q) com o peso calculado pela Equação 16:

$$w_2((i, j), (p, q)) \triangleq A(p, q)F(i - p, j - q) \quad (16)$$

Assim, normalizando os pesos das arestas de saída de cada nó e tratando o resultado do grafo como uma cadeia de Markov, é possível calcular a distribuição de equilíbrio sobre todos os nós. A massa então é acumulada justamente nos nós com alta ativação e, com isso, a saliência é obtida. Um exemplo de MS extraído com esse método pode ser encontrado na Figura 15.

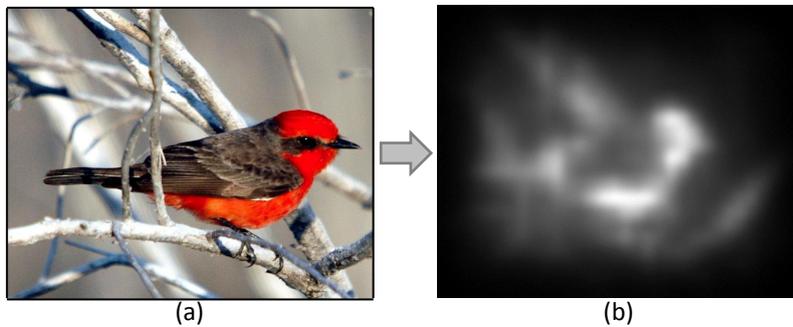


Figura 15 – Mapa de Saliência extraído utilizando o método de Harel.

Assim como Itti e Harel extraem a atenção visual das imagens e as representam pelo MS, na próxima seção detalhes são apresentados de como o método de (JUDD et al., 2009) também realiza esse processo usando uma combinação das estatísticas *top-down* e *bottom-up*.

2.5 MCAV de Judd

Segundo (JUDD et al., 2009), para diversas aplicações como, gráficos, interação homem-computador, publicidade, entre outras, é essencial entender onde os seres humanos olham em uma imagem. Tanto os dispositivos de rastreamentos oculares como os MCAV podem ser usados para prever possíveis locais de fixações de uma imagem.

Nesse sentido (JUDD et al., 2009) propuseram um novo MCAV, que utiliza um rastreador ocular durante a fase de treinamento, para determinar em que local de fato os seres humanos costumam olhar nas imagens. Para isso, foram coletados dados de rastreamento ocular de 15 usuários em 1003 imagens retiradas aleatoriamente do Flickr e LabelMe (RUSSELL et al., 2008). Essas imagens foram usadas no treinamento e teste, para a construção de um novo MCAV, com base nas características de baixo, médio e

alto nível das imagens. Este banco de dados de rastreamento dos olhos está disponível publicamente⁵.

Segundo os autores, o trabalho de (JUDD et al., 2009) tem duas contribuições: Criação de um banco de dados de rastreamento dos olhos com rótulos e análises. Desenvolvimento de um método de aprendizagem supervisionada de saliência que combina o *bottom-up* (informações de saliência) e *top-down* (informações semânticas).

Como os dados coletados pelos autores indicavam uma forte tendência das fixações humanas em estarem perto ou no centro das imagens, este comportamento foi atribuído ao fato de que os usuários foram colocados centralmente em frente ao monitor e de que os fotógrafos tendem a colocar objetos de interesse no centro das fotografias. Além do centro, outra característica que os autores observaram analisando os dados é de que os usuários fixam os olhos em faces, pessoas e texto. Outras fixações estão em partes do corpo, como olhos e as mãos, carros e animais. Observando esse comportamento, os autores utilizaram 33 características para cada *pixel* das imagens redimensionadas em 200x200:

- ❑ Nível baixo: cor, intensidade e orientação (ITTI; KOCH; NIEBUR, 1998).
- ❑ Nível médio: horizonte (GIST) (OLIVA; TORRALBA, 2001).
- ❑ Nível alto: detector de face, pessoas e carros (VIOLA; JONES, 2004) e (FELZENSZWALB; MCALLESTER; RAMANAN, 2008).
- ❑ *Center prior*: distância para o centro (JUDD et al., 2009).

O método proposto por (JUDD et al., 2009) é baseado em máquinas de aprendizagem, em que fixações de olhares e características visuais das imagens são utilizadas no treinamento. O treinamento desse método consiste inicialmente em obter todas as fixações dos usuários em cada imagem e a convolução gaussiana é realizada em cada imagem como exemplificado na Figura 16.

Nesta imagem convoluída são escolhidos aleatoriamente 10 pontos que fazem parte da região de interesse (pontos vermelhos) e 10 pontos que não fazem parte da região de interesse (pontos amarelos). Para treinar e testar o método, o conjunto de imagens foi dividido em 903 imagens de treinamento e 100 imagens de teste, produzindo um conjunto de treino de 18.060 amostras ($903 * 20$) e 2.000 amostras ($100 * 20$) de teste. Essa imagem convoluída com seus respectivos pontos em conjunto com as características de cada imagem compõem a entrada para o Máquina de Vetores de Suporte - *Support Vector Machines* (SVM), representada pela Figura 17. Com isso, o SVM aprenderá os pesos para cada característica que melhor predizem as imagens convoluídas.

Depois que o método está treinado basta utilizá-lo para extrair o MS. A Figura 18 mostra um exemplo de uma imagem e seu respectivo MS.

⁵ <http://people.csail.mit.edu/tjudd/WherePeopleLook/index.html>

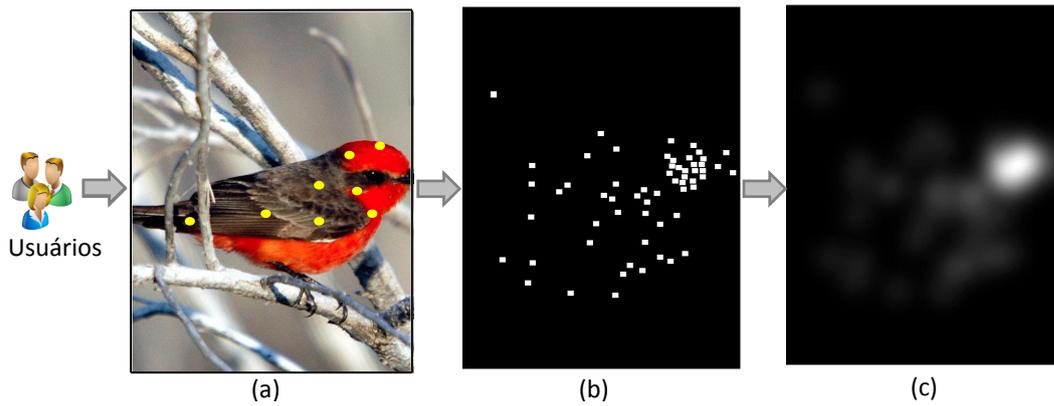


Figura 16 – **Convolução Gaussiana realizada sobre as fixações dos usuários.** (a) Fixações de um único usuário, (b) Fixações de todos os usuários e (c) Convolução Gaussiana das fixações.

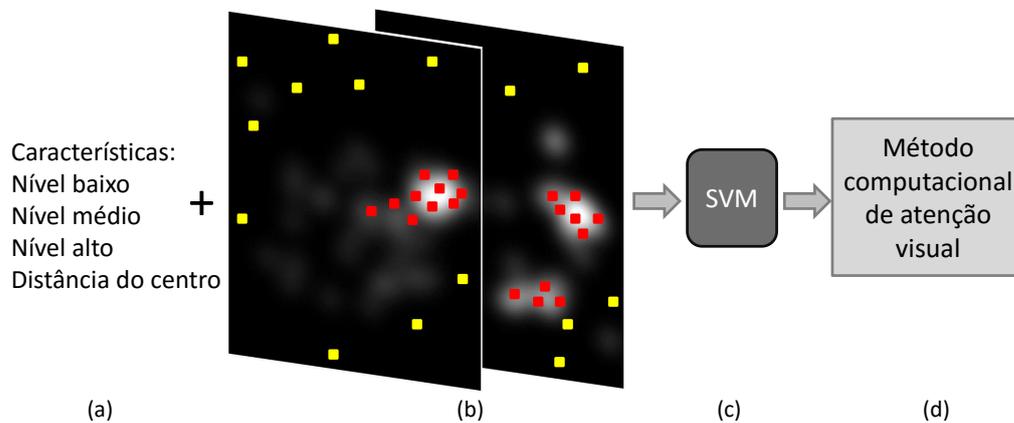


Figura 17 – **Treinamento do método de (JUDD et al., 2009).** (a) extração das características, (b) convolução e escolha dos pontos positivos localizados nos 20% mais relevantes da imagem e negativos localizados nos 70% menos relevantes da imagem, (c) treinamento do SVM e (d) criação de um MCAV.

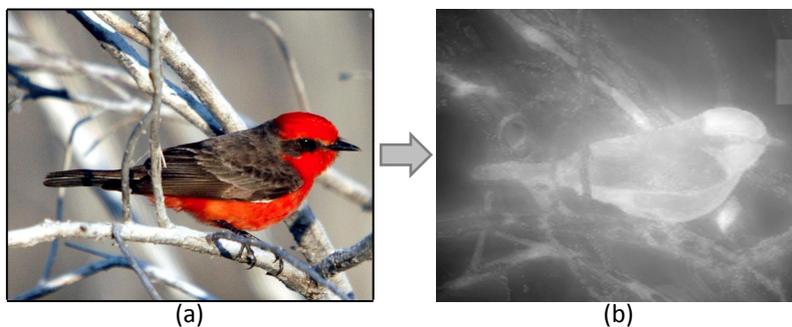


Figura 18 – **Mapa de saliência extraído utilizando o método da Judd.** (a) imagem original, (b) mapa resultante.

2.6 Considerações Finais

Neste capítulo foram apresentados os principais conceitos relativos aos MCAV, que são importantes para o desenvolvimento deste trabalho. A atenção visual é responsável por reduzir a quantidade de informações a serem processadas em uma imagem. Atenção visual é um fator fundamental para os diferentes MCAV apresentados. Uma aplicação bastante interessante para os MCAV é nos SR de imagens, pois a atenção visual pode caracterizar as imagens proporcionando mais informações ao SR, as informações sobre os itens da base de dados são fundamentais para que o SR gere boas recomendações. Esses MCAV serão empregados no SR sendo proposto e os resultados das recomendações serão comparados com outros SR comumente utilizados na área. Por esse motivo, outra área de pesquisa extremamente importante para o trabalho proposto é a dos SR, que é descrita no capítulo seguinte.

Sistemas de Recomendação

Neste capítulo são apresentados os conceitos básicos sobre os Sistemas de Recomendação (SR). Para isso, inicialmente apresenta-se a taxionomia desses sistemas, os tipos de filtragem como a FBC, FC e FH e alguns trabalhos que utilizam essas técnicas. Por fim, são apresentadas as principais medidas utilizadas para avaliar a eficácia dos SR.

3.1 Introdução

Cada vez mais informações vêm sendo disponibilizadas de forma facilitada, com auxílio da Internet, das redes sociais e aplicativos para celulares. Ao mesmo tempo em que as informações aumentam, a dificuldade para os usuários encontrarem conteúdos que são relevantes para eles aumentam proporcionalmente. Desta forma, surge a necessidade de identificar quais conteúdos são relevantes para cada usuário em meio a uma grande variedade de informações. Neste contexto, os SR se fazem extremamente importantes, pois possibilitam a filtragem das informações e a recomendação de itens que possam ser interessantes para o usuário (ADOMAVICIUS; TUZHILIN, 2005).

Pode-se dizer que, SR é uma técnica de filtragem de informação personalizada, usada para prever quais ou quantos itens um usuário poderia se interessar. Diversas empresas investem em SR, como é o caso do NETFLIX¹ para recomendação de filmes, Amazon² para recomendação de livros, roupas e outros produtos. Para recomendação de vídeos tem-se o YouTube³, Facebook⁴ sugere amigos, Submarino⁵ e Americanas⁶ recomendam produtos em geral, etc.

Os SR exploram as **relações entre os usuários e os itens**, a **relação entre os usuários** ou a **relação entre os itens**, tendo como o principal objetivo proporcionar as

¹ <http://www.netflix.com/br/>

² <http://www.amazon.com.br/>

³ <http://www.youtube.com/>

⁴ <http://www.facebook.com/>

⁵ <http://www.submarino.com.br/>

⁶ <http://www.americanas.com.br/>

melhores recomendações de itens aos usuários. Uma recomendação pode ser considerada relevante se o item oferecido ao usuário tiver alta probabilidade de ser utilizado.

Entretanto, para serem realizadas recomendações, é essencial conhecer quem é o usuário, para que as recomendações possam ser relevantes para ele. As informações sobre o usuário podem ser coletadas de forma **explícita** ou **implícita**, sendo descritas na seção 3.2.1. Posteriormente, essas informações serão utilizadas em um processo de recomendação ou predição, que visa identificar quais são os interesses do usuário e o seu padrão de comportamento.

Mais detalhes sobre o funcionamento de um SR são apresentados a seguir.

3.2 Taxonomia

Segundo (BURKE, 2002), um SR pode ser representado pela Figura 19, em que:

- **Dados prévios ou armazenados:** 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 perfil do 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. Essas informações podem variar de SR para SR e podem ser coletadas de forma explícita ou implícita.
- **Algoritmo de recomendação:** responsável por combinar dados prévios e dados de entrada para que sejam fornecidas as devidas recomendações.
- **Saída:** Uma lista contendo os itens que possivelmente o usuário mais se interessaria.

Os dados mais importantes nos SR são os dados de entrada e a saída obtida a partir desses dados (BURKE, 2002). Geralmente, quanto mais dados de entradas são fornecidos, melhor será a recomendação. As próximas seções descrevem em detalhes os dados de entrada e saída.

3.2.1 Dados de entrada

Os SR utilizam informações sobre o usuário tais como dados demográficos (ex: idade, ocupação e sexo), preferências, avaliações de itens, dentre outros, para selecionar itens que provavelmente serão de seu interesse. Desta forma, torna-se necessário coletar e armazenar as informações sobre o usuário para que posteriormente sejam utilizadas no processo de recomendação.

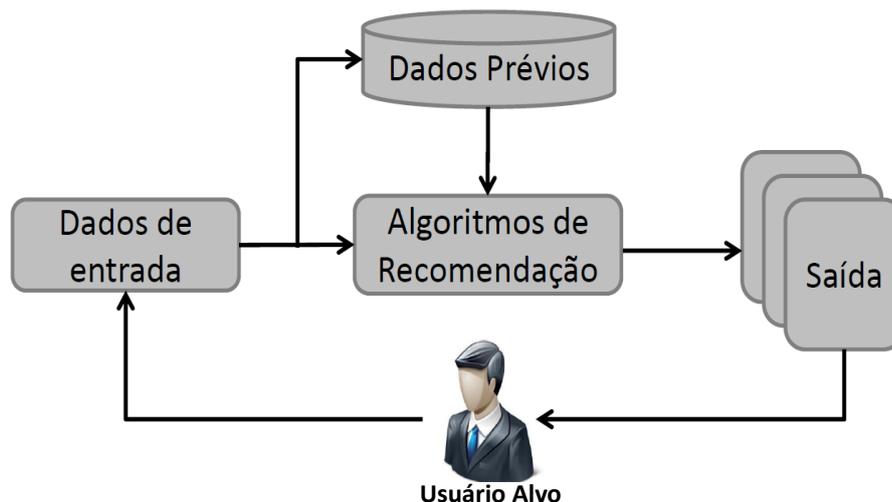


Figura 19 – Processo de recomendação, adaptada de (BURKE, 2002)

Alguns métodos podem ser empregados para representar o perfil do usuário como, vetores de características, modelos baseados em ontologia (indivíduos, classes, atributos e relacionamentos), matrizes de avaliações, etc. Cada método é utilizado de acordo com o domínio de aplicação e os requisitos do SR (FERNÁNDEZ, 2007).

Nas informações coletadas de forma explícita, o usuário é incentivado a manualmente especificar as informações que descrevem os seus interesses, por meio de avaliações sobre os produtos ou com a utilização de uma interface gráfica como um formulário. O usuário deve fornecer alguns dados como, por exemplo, idade, sexo, profissão e preferências. Esta técnica, segundo (ZIMMERMAN et al., 2004), assegura certa precisão na representação do perfil do usuário e pode gerar uma alta qualidade nas recomendações. As desvantagens desse tipo de coleta é que exige paciência e atenção dos usuários.

No entanto, de acordo com (SMYTH; COTTER, 2001) devido ao fato de alguns usuários não especificarem os seus dados completamente, deve ser empregado um mecanismo complementar, que permita atualizar o perfil do usuário e, principalmente, suas preferências de forma implícita. Nesta forma, o sistema infere os interesses do usuário por meio da análise das ações de interação ou a partir do seu histórico de uso. As informações são armazenadas em um histórico de uso, à medida que o usuário avalia um item, realiza uma consulta, uma compra ou efetua a leitura de um conteúdo.

A coleta de forma implícita tem a vantagem de não ser intrusiva, pois é possível descobrir o interesse do usuário sem que seja necessário que o mesmo forneça seus dados. Contudo esse tipo de coleta de informações também possui limitações e não é indicado para auxiliar na compra de presentes, ou ainda, se o acesso é realizado em vários locais ou por várias pessoas o perfil não será verídico.

Geralmente, boas recomendações (saídas) estão diretamente ligadas com os dados de entrada fornecidos ao SR.

3.2.2 Saídas

De acordo com (VOZALIS; MARGARITIS, 2003), a saída de um SR pode ser uma **predição** ou uma **recomendação**. A predição é expressa por um valor numérico que representa a opinião antecipada do usuário alvo (que pede recomendações) em relação a um item. 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 do filme, ou seja, o sistema apresenta um item associado a uma escala que prediz sua relevância para um determinado usuário.

Já a recomendação, é geralmente representada por uma lista de N itens considerados interessantes para o usuário alvo, obtidos com a recomendação Top N . 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.

Para que os SR apresentem bons resultados, as técnicas de filtragem de informações são fundamentais. Isso se deve ao fato de que um dos principais problemas dos SR é a sobrecarga de informações. A filtragem de informação é usada para que somente conteúdos interessantes e relevantes apareçam para o usuário.

3.3 Filtragem de Informações em SR

Com os usuários estando constantemente sobrecarregados de informações, a solução que parece mais interessante é filtrar as informações, para que somente conteúdos de seu interesse sejam apresentados. Muitos sistemas de filtragem atuais são baseados em construir um perfil de usuário a partir do conteúdo das informações a que ele teve acesso. Esses sistemas tentam extrair padrões do comportamento do usuário para prever quais informações seriam relevantes ou irrelevantes para ele.

Os sistemas de filtragem de informação são uma ótima solução para os usuários que sabem realmente o que desejam buscar, porém muita informação pode ser perdida pelo fato de o usuário não ter conhecimento de outros conteúdos que também seriam interessantes. Com isso, cada vez mais são exploradas técnicas de filtragem de informações para a recuperação de informações em SR (HERLOCKER; KONSTAN; RIEDL, 2000).

Normalmente, os SR encontram-se classificados em categorias, conforme a abordagem utilizada para realizar as recomendações, ou seja, o tipo de filtragem adotada pelo sistema. Dentre os mais referenciados na literatura (GOLDBERG et al., 1992) (BALABANOVIĆ; SHOHAM, 1997) (BURKE, 2002) e (ADOMAVICIUS; TUZHILIN, 2005), estão:

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 Híbrida: este método procura combinar dois ou mais tipos de filtragem. Nas próximas seções cada uma dessas técnicas será descrita com mais detalhes.

3.3.1 Filtragem Baseada em Conteúdo - FBC

Na FBC, a recomendação de um item a um usuário baseia-se na semelhança com itens adquiridos ou vistos por este usuário no passado (ADOMAVICIUS; TUZHILIN, 2005). Um livro, por exemplo, pode ser caracterizado pelo tema, autor, editora, título, etc. Já em um texto, geralmente, as próprias palavras que o compõem são consideradas como suas características. É com base nessas características que os itens podem ser comparados e a semelhança entre eles estabelecida.

Formalizando, a preferência $f(u, i_a)$ do usuário u , para o item i_a é estimada com base nas preferências $f(u, i_b)$ atribuídas pelo usuário u para o item i_b que é similar ao item i_a , considerando que $i_a \neq i_b$ (ADOMAVICIUS; TUZHILIN, 2005).

A FBC é normalmente empregada em SR, em que seu principal foco é recomendar itens contendo informações textuais, tais como *sites*, *blogs*, artigos científicos, fóruns, notícias, dentre outros. Dessa forma, o sistema aprende as preferências do usuário e filtra, entre os novos itens, aqueles que mais se adequam às preferências do usuário.

A maneira de representar o perfil do usuário depende das técnicas de recuperação de informação e aprendizagem de máquina, como: **Árvores de decisão**, que são treinadas de acordo com um conjunto de treino (exemplos previamente classificados) e posteriormente, outros exemplos são classificados de acordo com essa mesma árvore. **Redes neurais**, que são um conjunto de unidades (ou neurônios) com entradas e saídas conectadas para formar uma rede na qual cada conexão tem um peso associado. **Classificadores Bayesianos**, que são classificadores estatísticos que classificam um objeto numa determinada classe baseando-se na probabilidade deste objeto pertencer a essa classe. Têm-se ainda as **regras de associação**, que têm como premissa básica encontrar relacionamentos ou padrões frequentes entre conjuntos de dados (HAN; KAMBER, 2006).

Outra técnica que é bastante utilizada é o algoritmo **K-vizinhos Mais Próximos** (**K-Nearest Neighbors - KNN**). A fim de estimar o grupo de um novo usuário u , o KNN calcula os K-vizinhos mais próximos a u e classifica-o como sendo do grupo que aparece com maior frequência dentre os seus K-vizinhos.

O algoritmo KNN (COVER; HART, 1974), pode ser adaptado para a FBC e representar o interesse do usuário, assim como proposto em (ARYA, 1996; COTTER; SMYTH, 2000; KRULWICH; BURKEY, 1996; BALABANOVIĆ; SHOHAM, 1997). Para isso, um vetor de palavras e pesos é criado, sendo que o peso determina a relevância da palavra. Como os principais atributos da FBC são textuais, o Modelo de Espaço Vetorial e a medida Frequência do Termo - *Term-Frequency* (TF) são ideais, pois são voltados para aplicações que possuam atributos textuais. Nessa técnica, um conjunto de documentos/itens $I = i_1, i_2, \dots, i_N$, contém um conjunto de termos $T = t_1, t_2, \dots, t_n$. Cada item

pode ser representado por um vetor do espaço n -dimensional de pesos dos termos de T nos itens I , tal que $i_b = w_{1b}, w_{2b}, \dots, w_{nb}$, sendo w_{kb} representa o peso do termo t_k no item i_b .

A recomendação de um item é feita pela comparação das características dos itens. O *ranking* dos itens é obtido com a similaridade de suas características. Por exemplo, para uma situação de recomendação de um livro, os **termos** são as informações que se têm sobre o livro como título, gênero, autor e resumo.

Para calcular o Inverso da Frequência nos Documentos - *Inverse Document Frequency* (IDF) dos itens, basta usar o total de itens N pela quantidade de itens em que o termo t_k aparece, e é representado por n_k , desta forma: $IDF = \log \frac{N}{n_k}$. Já o TF representa o número de vezes que o termo t_k apareceu em i_b , $TF = freq(t_k, i_b)$, a Equação 17 representa essa ponderação.

$$w_{k,b} = TF \times IDF(t_k, i_b) = freq(t_k, i_b) \times \log \frac{N}{n_k} \quad (17)$$

Assim, o item pode ser representado como um vetor de palavras e seus pesos: $i_b = (w_{1b}, \dots, w_{kb})$. Com essa informação, a cada item que se deseja recomendar ao usuário, é feita a similaridade pelo cosseno (representada na Equação 18) entre os vetores dos itens que o usuário já possui e os vetores que se deseja recomendar (vetores de consulta).

Como resultado dos cálculos de similaridade, obtém-se um conjunto de itens ordenados pelo grau de similaridade. Cada item tem sua relevância em relação à consulta realizada. Com esse modelo, é possível restringir o conjunto ordenado obtido a um número máximo de itens desejados na busca. É possível também restringir a quantidade de itens recuperados definindo um limite mínimo para o valor obtido com a similaridade.

$$sim_{i_a, i_b} = \frac{\sum_k (w_{ka} w_{kb})}{\sqrt{\sum_k (w_{ka})^2} \sqrt{\sum_k (w_{kb})^2}} \quad (18)$$

Por fim, são recomendados ao usuário os itens mais similares aos itens que ele interagiu, pois estes contêm características parecidas.

Um SR com FBC é aquele em que são feitas recomendações para um usuário com base no perfil construído pela análise do conteúdo dos itens que o usuário tenha avaliado no passado. Exemplos de tais sistemas são Info Finder (KRULWICH; BURKEY, 1996), News Weeder (LANG, 1995), etc.

Mesmo sendo utilizada por vários trabalhos, esse tipo de filtragem possui alguns pontos fracos. Segundo (ADOMAVICIUS; TUZHILIN, 2005) e (BURKE, 2002) são eles:

- **Análise limitada do conteúdo:** essa filtragem é limitada pelas características que estão explicitamente associadas ao conteúdo. Tal conteúdo deve ser do tipo (por exemplo, texto) que possa ser “carregado” automaticamente por um computador, ou as características devem ser atribuídas aos itens manualmente, como, por exemplo, filmes ou músicas. Essa atividade pode ser dispendiosa e sujeita a erros em ambientes

dinâmicos e muito grandes. Por esse motivo, é menos eficaz empregar essa técnica em conteúdos como imagens, vídeos e áudios do que em um conteúdo textual.

- ❑ **Novo usuário:** para que o sistema 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 confiáveis.
- ❑ **Criação do perfil:** para que seja realizada a recomendação, essa técnica cria um perfil para cada item e usuário. No entanto, geralmente, as bases de dados reais não possuem informações disponíveis sobre todos os itens, podendo gerar perfis incompletos e recomendações imprecisas.

Por outro lado, um dos pontos fortes do SR com FBC é o de não sofrer com a partida a frio de itens como acontece na FC (descrita a seguir). Esse problema ocorre quando novos itens são inseridos na base de dados e, como esses itens possuem poucas ou nenhuma avaliação, o sistema não irá recomendá-los. A FBC não sofre dessa limitação, pois o critério de recomendação baseia-se apenas na comparação das características do item com os demais, independentemente se o item foi ou não avaliado por mais alguém (BURKE, 2002).

3.3.2 Filtragem Colaborativa - FC

O termo FC foi cunhado por (GOLDBERG et al., 1992), que foi um dos pioneiros no uso desta técnica. A FC foi proposta para atender pontos que estavam em aberto na FBC (CAZELLA, 2006) (HERLOCKER; KONSTAN; RIEDL, 2000) e (ANSARI; ESSEGAIER; KOHLI, 2000). 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.

A principal característica da FC é de que as recomendações feitas a um usuário são baseadas nos itens que lhe são desconhecidos mas que foram relevantes para usuários com gostos semelhantes. Se conhecermos alguém que tenha os mesmos gostos que nós, é muito provável que uma nova aquisição dessa pessoa seja também do nosso interesse.

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

Os algoritmos de FC foram apresentados por (BREESE; HECKERMAN; KADIE, 1998) e divididos em duas classes: **baseados em modelo** descritos na seção 3.3.2.1 e **baseados em memória** descritos na seção 3.3.2.2. Os algoritmos baseados em modelo

utilizam regras pré-definidas e, com isso, produzem as recomendações mais rapidamente, porém necessitam de tempo para criação das regras e as recomendações têm menos qualidade, se comparadas com as dos métodos baseados em memória (ou vizinhança). Essa abordagem será descrita com mais detalhes a seguir.

3.3.2.1 Técnicas baseadas em modelo

As técnicas de FC baseadas em modelo, diferentemente das baseadas em memória, usam o conjunto de avaliações para aprender um modelo, que é usado então para fazer as predições e recomendações. Em vez de usarem medidas de similaridades, estes algoritmos são caracterizados pela criação de modelos (BREESE; HECKERMAN; KADIE, 1998).

De forma resumida, os modelos são entidades que sintetizam o comportamento dos dados. Esses tipos de sistemas analisam a estrutura da matriz que relaciona usuários e itens, para encontrar relações entre os itens. A ideia por trás dessa abordagem é que os usuários se interessam por itens similares aos bem avaliados por ele no passado. Em contrapartida eles não se interessam pelos itens similares aos itens que ele não gostou no passado (LINDEN; SMITH; YORK, 2003) e (SARWAR et al., 2001).

Outra questão é que essa técnica não precisa identificar a vizinhança de usuários similares. Esse é o gargalo de desempenho dos algoritmos baseados em memória. Em consequência, essa técnica tende a produzir recomendações mais rapidamente, depois que os modelos já estão treinados. No entanto, a acurácia dessas recomendações são inferiores às dos sistemas baseados em memória (LINDEN; SMITH; YORK, 2003) e (SARWAR et al., 2001).

Diversos algoritmos podem ser usados para obter os modelos, tais como métodos de agrupamento, redes Bayesianas, regras de associação ou modelos de fatoração de matrizes (BREESE; HECKERMAN; KADIE, 1998).

Vários algoritmos de FC são baseados em modelos de fatoração de Matrizes. A maioria deles são baseados na fatoração da matriz de utilidade utilizando a técnica Decomposição de Valores Singulares - *Singular Value Decomposition* (SVD)⁷. A ideia da fatoração de matrizes é projetar usuários e itens para um espaço de baixa-dimensionalidade. Na maneira mais tradicional, esses modelos associam cada usuário u a um vetor de fatores latentes $p_u \in \mathbb{R}^M$, e cada item i a um vetor de fatores latentes $q_i \in \mathbb{R}^M$, em que M é o número de fatores latentes. A predição de avaliação $\hat{r}_{u,i}$ é obtida pelo produto escalar dos

⁷ A **decomposição em valores singulares** (ou *Singular Value Decomposition* - SVD) nada mais é do que o modelo tradicional de fatoração de uma matriz. Formalmente, a SVD de uma matriz M , de tamanho $m \times n$ é uma fatoração na forma: $M = U\Sigma V^*$. Em que U é uma matriz unitária de tamanho $m \times m$, Σ é uma matriz retangular diagonal $m \times n$ com números reais não-negativos na diagonal e V^* (matriz transposta de V) é uma matriz unitária $n \times n$. As entradas diagonais $\Sigma_{i,i}$ de Σ são os chamados **valores singulares** de M , representados por σ . As m colunas de U e as n colunas de V são os chamados vetores singulares à esquerda e vetores singulares à direita de M , respectivamente. A expressão $U\Sigma V^*$ pode ser interpretada intuitivamente como uma sequência de três transformações geométricas: uma rotação, uma escala e outra rotação (SZELISKI, 2010).

vetores p_u e q_i , conforme representado pela Equação 19.

$$\hat{r}_{u,i} = p_u^T \cdot q_i \quad (19)$$

As avaliações são extremamente influenciadas por um conjunto de fatores que são muito específicos de acordo com o domínio (por exemplo, tamanho da manga ou tipo do decote no caso de roupas; tipo de cadarço e sola para calçados). Em geral, estes fatores não são óbvios. Nós somos capazes de pensar em alguns deles, mas é difícil estimar o impacto de cada fator nas avaliações. O espaço latente tenta explicar as avaliações por meio da caracterização automática dos usuários e itens em fatores. Para cada item i , os elementos de q_i medem em que grau o item possui os fatores latentes. Para cada usuário u , os elementos de p_u medem o grau de interesse que o usuário u tem em itens com alta correspondência de fatores latentes. O processo de aprendizado para obter $p_u \in P$, e $q_i \in Q$ é um problema de otimização, em que dada uma matriz de utilidade esparsa R deve-se encontrar as matrizes P e Q tal que $P \cdot Q$ é uma matriz muito próxima de R , como ilustrado na Figura 20 e descrito em (KOREN; BELL, 2011).

$$\begin{array}{c}
 \begin{array}{ccccc}
 & i_1 & i_2 & i_3 & i_4 & i_5 \\
 u_1 & \boxed{1} & \boxed{3} & \boxed{5} & \boxed{2} & \boxed{1} \\
 u_2 & & \boxed{4} & & \boxed{3} & \\
 u_3 & \boxed{2} & & \boxed{2} & \boxed{3} & \boxed{5}
 \end{array} \\
 R = \text{Matriz de utilidade}
 \end{array}
 =
 \begin{array}{c}
 \begin{array}{ccccc}
 & f_1 & f_2 & f_3 & \dots & f_n \\
 u_1 & \boxed{} & \boxed{} & \boxed{} & \dots & \boxed{} \\
 u_2 & \boxed{} & \boxed{} & \boxed{} & \dots & \boxed{} \\
 u_3 & \boxed{} & \boxed{} & \boxed{} & \dots & \boxed{}
 \end{array} \\
 P
 \end{array}
 \times
 \begin{array}{c}
 \begin{array}{ccccc}
 f_1 & \boxed{} & \boxed{} & \boxed{} & \boxed{} \\
 \dots & \boxed{} & \boxed{} & \boxed{} & \boxed{} \\
 f_n & \boxed{} & \boxed{} & \boxed{} & \boxed{}
 \end{array} \\
 Q
 \end{array}$$

Figura 20 – **Fatoração da matriz de utilidade usando o SVD.** Neste exemplo u representa os usuários, i os itens e f os fatores latentes.

Trabalhos baseados em modelo

(BREESE; HECKERMAN; KADIE, 1998) listaram e testaram diversos algoritmos para a FC baseada em modelos. Eles propuseram uma alternativa para encontrar dependências entre artigos usando um classificador Bayesiano e um modelo de rede Bayesiana. De acordo com os autores, uma rede Bayesiana é aprendida para cada usuário e utilizada na classificação dos itens como relevante ou não-relevantes. Porém dois problemas foram encontrados: (1) a necessidade de calcular as probabilidades condicionais de todas as avaliações possíveis para um dado artigo e todas as avaliações possíveis de outros artigos, o que é computacionalmente trabalhoso e (2) não consegue estimar a qualidade da predição que está fazendo.

Já (BILLSUS; PAZZANI, 1998) apresentaram um modelo para FC em que empregavam redes neurais. Para cada usuário do sistema, era treinada uma rede que serviria de modelo na geração de predições. Os exemplos de treinamento eram construídos com os itens que o usuário já tinha avaliado. A entrada da rede eram as avaliações do item feitas por todos os outros usuários do sistema e a saída a avaliação do usuário alvo. O

problema dessa abordagem reside no fato de que o tamanho da entrada pode ser muito grande dependendo da quantidade de usuários do sistema.

Em (LATHAUWER; MOOR; VANDEWALLE, 2000; BILLSUS; PAZZANI, 1998) para tratar o problema da dimensionalidade dos dados (existente na FC baseada em memória), os autores utilizaram SVD. Contudo, ao fazer isso, algumas informações são perdidas, podendo comprometer a qualidade das recomendações. Já o método SVD++, além da fatoração de matrizes, considera o *feedback* implícito dos usuários para realizar a predição (KOREN, 2008).

No trabalho de (HERLOCKER; KONSTAN; RIEDL, 2000) foram utilizadas redes neurais junto a técnicas de redução de dimensionalidade, tais como SVD ou ganho da informação⁸. Com isso, a dimensão dos dados avaliados foi reduzida e, então, as redes neurais foram aplicadas nos dados simplificados para construir um modelo de recomendação.

Uma característica importante desse tipo de filtragem é a facilidade de lidar com os dados dispersos e ser escalável. No entanto, a construção do modelo pode ser custosa e muitas informações importantes podem ser perdidas com a redução da dimensionalidade.

Além da FC baseada em modelo, existe também a FC baseada em memória. Os algoritmos baseados em memória têm grande destaque pelo pioneirismo e também são conhecidos como baseados em vizinhança. Esse algoritmo se baseia na ideia de que uma pessoa pertence a uma comunidade de indivíduos com interesses similares. Desta forma, os itens avaliados pelos indivíduos dessa comunidade, podem ser utilizados como base para recomendar itens a um indivíduo pertencente à mesma comunidade. Mais detalhes dessa abordagem são descritos a seguir.

3.3.2.2 Técnicas baseadas em memória ou vizinhança

Os SR baseados em memória aproximam as pessoas com interesses similares para fazer a recomendação de um item que o usuário possivelmente se interessaria. Nessa abordagem, os dados são guardados em memória e são calculadas medidas de similaridade entre usuários e/ou entre itens, a cada vez que é solicitada uma recomendação.

O algoritmo de FC baseado em memória utiliza as seguintes etapas para gerar recomendações:

1. Calcula a similaridade (descrita a seguir no item (a) Medidas de Similaridade), que reflete a distância ou correlação, entre dois usuários ou itens.
2. Produz uma previsão para o usuário alvo pela média ponderada de todas as avaliações do usuário (descrita a seguir no item (b) Predição Baseada em Usuário) ou

⁸ O ganho da informação de um atributo x é definido como a diferença entre a incerteza anterior e a posterior utilizando x . O atributo x é preferido, ao invés do atributo y , se o ganho de informação de x for maior que o ganho de informação y (LIU; YU, 2005).

item (descrita a seguir no item (c) Predição Baseada em Item) em um determinado item ou usuário.

3. Gera recomendação Top-N (descrita na seção 3.4.2). Para isso, é necessário encontrar usuários ou itens (vizinhos mais próximos) após calcular as similaridades e, em seguida, os vizinhos são reunidos para se obter o Top-N como recomendação.

Algumas medidas que podem ser utilizadas para calcular a similaridade entre os usuários ou itens são: a correlação de Pearson (RESNICK et al., 1994), cosseno (GUNAWARDANA; SHANI, 2009), correlação de Spearman, correlação de Kendal, diferença média quadrática e entropia (MELVILLE; SINDHWANI, 2010).

(a) Medidas de Similaridade

As medidas de similaridade definem os níveis de semelhança entre os usuários ou itens. Isto permite que se encontre a vizinhança do usuário ativo/alvo. Ao mesmo tempo, elas também fornecem uma ponderação para o cálculo da predição da recomendação dos itens aos usuários. Uma das formas para se obter a similaridade entre dois itens é feita pelas avaliações que os usuários deram aos itens (informações explícitas). Estas avaliações podem ser notas de 1 a 5 estrelas como já é feito em vários *sites* como: YouTube⁹, Amazon¹⁰, NETFLIX¹¹, entre outros.

Cosseno: com essa medida, os usuários ou itens são mapeados em um espaço n -dimensional. No caso da similaridade entre usuários, cada usuário tem a lista de itens com as avaliações feitas. Para se obter a similaridade, basta calcular o cosseno, entre os dois vetores de usuários r_{ui} e r_{vi} resultando em $sim(u, v)$ representado pela Equação 20.

$$sim(u, v) = \cos(r_{ui}, r_{vi}) = \frac{r_{ui} \cdot r_{vi}}{\|r_{ui}\| \cdot \|r_{vi}\|} = \frac{\sum_{i=1}^n r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i=1}^n (r_{ui})^2} \cdot \sqrt{\sum_{i=1}^n (r_{vi})^2}} \quad (20)$$

Já no caso da similaridade entre itens, cada item tem uma lista de usuários que o avaliaram e, para se obter a similaridade, calcula-se o cosseno entre os dois vetores de itens r_{ui} e r_{uj} , como na Equação 21.

$$sim(i, j) = \cos(r_{ui}, r_{uj}) = \frac{r_{ui} \cdot r_{uj}}{\|r_{ui}\| \cdot \|r_{uj}\|} = \frac{\sum_{u=1}^m r_{ui} \cdot r_{uj}}{\sqrt{\sum_{u=1}^m (r_{ui})^2} \cdot \sqrt{\sum_{u=1}^m (r_{uj})^2}} \quad (21)$$

em que r_{ui} representa a avaliação do item i pelo usuário u e m o número de usuário que avaliaram i e j .

⁹ <http://www.youtube.com/>

¹⁰ <http://www.amazon.com.br/>

¹¹ <http://www.netflix.com/br/>

Correlação de Pearson: a correlação de Pearson (RESNICK et al., 1994) mede a linearidade entre duas variáveis. Segundo (HERLOCKER et al., 2004) as técnicas de correlação apresentam algumas desvantagens. A principal é que, se dois usuários não avaliarem nenhum item em comum, o método de correlação não consegue calcular similaridade entre eles, mesmo se compartilharem de interesses comuns. Para se obter as correlações $sim(u, v)$ entre os usuários u e v , são feitas as correlações das avaliações que ambos os usuários fizeram aos mesmos itens, representadas na Equação 22.

$$sim(u, v) = \frac{\sum_{i=1}^n (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i=1}^n (r_{ui} - \bar{r}_u)^2} \cdot \sqrt{\sum_{i=1}^n (r_{vi} - \bar{r}_v)^2}} \quad (22)$$

em que n é o número de itens avaliados tanto pelo usuário u quanto por v . A média $\frac{1}{n} \cdot \sum_{i=1}^n r_{ui}$ envolvendo a avaliação r_{ui} é denotada por \bar{r}_u .

Analogamente, é possível obter a correlação $sim(i, j)$ para dois itens i e j a partir das avaliações que os mesmos usuários fizeram a eles, como mostra a Equação 23.

$$sim(i, j) = \frac{\sum_{u=1}^m (r_{ui} - \bar{r}_i)(r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u=1}^m (r_{ui} - \bar{r}_i)^2} \cdot \sqrt{\sum_{u=1}^m (r_{uj} - \bar{r}_j)^2}} \quad (23)$$

em que m é o número de usuários que avaliaram tanto o item i quanto o item j . A média $\frac{1}{m} \cdot \sum_{u=1}^m r_{ui}$ envolvendo a avaliação r_{ui} é denotada por \bar{r}_i .

Tanto a cosseno como a correlação de Pearson podem ser usadas para fornecer uma ponderação no cálculo de predição das recomendações. A seguir, é descrito como essa predição pode ser realizada, tanto para usuários quanto para itens.

(b) Predição baseada em usuário

O primeiro passo para realizar a predição baseada em usuários é obter as avaliações dos usuários para os itens. Essas avaliações podem ser representadas por uma matriz de avaliações $A : U \times I$, em que U representa os usuários e I os itens, conforme mostra a Tabela 2. Nessa tabela, os valores numéricos representam as avaliações dadas pelo usuário para o item. O valor 5 indica que os usuários gostaram muito dos itens indicados na intersecção das respectivas células, já o valor 1 indica que eles não gostaram. Neste caso, o **usuário alvo** (que pede recomendações) é o usuário 4, pois o símbolo “?” representa a avaliação a ser determinada. Já a diferença de cor nas linhas indica os usuários a serem comparados nesse exemplo.

Com a matriz de avaliações, é possível calcular a similaridade entre os usuários por meio da correlação de Pearson ou cosseno, como nas Equações 20 e 22. A correlação de Pearson produz valores no intervalo de $[-1, 1]$ e somente os itens avaliados pelos dois usuários são usados no cálculo da correlação entre eles.

Tabela 2 – Representação de uma matriz de avaliações.

	item 1	item 2	item 3	item 4	item 5
usuário 1	2		3	2	3
usuário 2	4		4		5
usuário 3	1	4	3		
usuário 4	?	2		4	5

A matriz de similaridades armazena a correlação entre os usuários. Depois que as correlações foram obtidas é preciso selecionar os vizinhos que farão parte do cálculo de recomendações. Os vizinhos podem ser escolhidos por meio de filtros na matriz de similaridades. Alguns exemplos de filtros são: os 5 usuários com as maiores correlações, ou ainda, todos os usuários com correlação maior do que 0.3. Esses filtros podem variar de acordo com cada aplicação.

A predição da avaliação $\hat{r}_{u,i}$ do usuário alvo para o item, é calculada com base nas avaliações feitas para esse item por todos os usuários vizinhos ao usuário alvo. Desta forma, o valor da predição $\hat{r}_{u,i}$ do item pelo usuário é feita pela média ponderada das avaliações dos k itens similares, essa predição está representada na Equação 24.

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{u \in \mathbb{U}(v,i,k)} sim(u,v)(r_{v,i} - \bar{r}_v)}{\sum_{u \in \mathbb{U}(v,i,k)} |sim(u,v)|} \quad (24)$$

em que \bar{r}_u e \bar{r}_v são as médias das avaliações dos usuários u e v , respectivamente, $r_{v,i}$ é a avaliação do usuário v para o item i . Já $\mathbb{U}(v,i,k)$ é o conjunto de k vizinhos obtidos com a correlação $sim(u,v)$, é possível prever a avaliação que o usuário alvo faria para o item que ele ainda não avaliou. A correlação do usuário é utilizada para dar peso em suas recomendações.

Como exemplo, supondo que a predição da avaliação $\hat{r}_{u,i}$ do usuário alvo (usuário 4) para o item (item 1), é calculada com base nas avaliações feitas para esse item por todos os usuários vizinhos ao usuário 4, que neste caso são, usuário 1 e o usuário 3 representados pela cor azul na Tabela 2. Desta forma, o valor da predição $\hat{r}_{u,i} = 3$ do item 1 pelo usuário 4 é feita pela média ponderada das avaliações dos k itens similares, representada na Equação 24.

O outro tipo de predição além da predição de usuários é a predição de itens, descrita a seguir.

(c) Predição baseada em item

Com o grande aumento do número de usuários e de itens, o processo de recomendação *online* torna-se cada vez mais demorado para a FC baseada em usuário. Isso acontece, pois cada vez que é solicitada uma recomendação, é necessário comparar um usuário com todos os outros e depois comparar todos os itens que cada usuário avaliou. Além disso, é normal que possa haver poucos itens em comum o que torna

difícil a tarefa de encontrar usuários semelhantes (MELVILLE; SINDHWANI, 2010). Para minimizar este problema (MELVILLE; SINDHWANI, 2010) propuseram a FC baseada em item. Em vez da similaridade ser calculada entre os usuários, é calculada entre os itens, de acordo com as avaliações que lhes são dadas pelos usuários.

Esta predição, analisa o conjunto de itens avaliados pelo usuário alvo e calcula o quanto similar eles são ao item que se pretende prever, com base nas avaliações que os itens receberam de outros usuários, como mostra a Tabela 3. Em outras palavras, para prever ao usuário alvo o quanto ele pode gostar de um item, o sistema examina todos os itens já avaliados pelo usuário alvo e verifica a correlação de cada um deles com o item que se pretende prever. Apenas os usuários que avaliaram ambos os itens são considerados (SARWAR et al., 2001). Esta correlação é descrita na Equação 25.

Tabela 3 – Matriz de avaliações, representa a correlação entre os itens 1 e 5.

	item 1	item 2	item 3	item 4	item 5
usuário 1	2		3	2	3
usuário 2	4		4		5
usuário 3	1	4	3		
usuário 4	?	2		4	5

$$\hat{r}_{u,i} = \frac{\sum_{j \in \mathbb{I}(u,i,k)} \text{sim}(i,j)(r_{u,j})}{\sum_{j \in \mathbb{I}(u,i,k)} |\text{sim}(i,j)|} \quad (25)$$

em que $\mathbb{I}(u,i,k)$ é o conjunto de k itens avaliados pelo usuário u e mais similares ao item i . Desta forma, o valor da avaliação $\hat{r}_{u,i}$ do item i que recebe o valor predito pelo usuário alvo u é feita pela média ponderada das avaliações dos k itens similares.

Por exemplo, para prever ao usuário 4 o quanto ele pode gostar do item 1, o sistema examina todos os itens já avaliados pelo usuário 4 (conjunto de itens $I_a = \{\text{item 2, item 4 e item 5}\}$) e verifica a correlação de cada um deles com o item 1. Para realizar a correlação de Pearson do item 1 com o item 5, como mostra a Tabela 3, apenas os usuários 1 e 2 são considerados e são representados pelo conjunto $U = \{\text{usuário 1 e usuário 2}\}$. Desta forma, o valor da avaliação $\hat{r}_{u,i} = 3$ do item i (item 1) pelo usuário u (usuário 4) é feita pela média ponderada das avaliações dos k itens similares.

Trabalhos baseados em vizinhança

A FC baseada em memória é utilizada pelo GroupLens que impulsionou essa área de pesquisa e foi citado por diversos outros trabalhos. O GroupLens desenvolveu o Usenet para recomendação de notícias avaliadas quantitativamente por notas de 1 a 5 pelos usuários, relacionando o quanto o usuário gostou daquela notícia. O MovieLens também

foi desenvolvido pelo GroupLens, mas para recomendação de filmes (RESNICK et al., 1994).

Outras aplicações que estão utilizando a FC baseada em vizinhança são: Bellcore, que realiza recomendação de vídeo (HILL et al., 1995); Ringo, utilizado para recomendação de músicas (SHARDANAND; MAES, 1995); para recomendação de livro um exemplo é a Amazon (LINDEN; SMITH; YORK, 2003); já o PHOAKS ajuda as pessoas a encontrarem informações relevantes no *World Wide Web* (TERVEEN et al., 1997); o sistema Jester recomenda piadas (GOLDBERG et al., 2001); no ambiente da TV Digital Interativa, a FC foi adotada por alguns SR de programas de TV tais como, PTV (COTTER; SMYTH, 2000) e AVATAR (FERNÁNDEZ et al., 2006).

Algoritmos baseados em memória (SI; JIN, 2003; SARWAR et al., 2000; MARLIN, 2003) são essencialmente heurísticos e fazem previsões de avaliações com base em toda a coleção de itens previamente avaliado pelos usuários, ou seja, o valor desconhecido da avaliação é normalmente calculado com base nas avaliações dos outros usuários (geralmente, os mais semelhantes).

No trabalho de (KOREN, 2010) é descrito o IKB, este método é o *baseline* da FC baseada em memória. A forma como esse método calcula a correlação entre dois itens i e j , descrita na Equação 26, é calculada baseando-se nas avaliações passadas.

$$sim(i, j) = \frac{\sum_{u \in \mathbb{U}(i, j)} (r_{ui} - b_{ui})(r_{uj} - b_{uj})}{\sqrt{\sum_{u \in \mathbb{U}(i, j)} (r_{ui} - b_{ui})^2} \cdot \sqrt{\sum_{u \in \mathbb{U}(i, j)} (r_{uj} - b_{uj})^2}} \quad (26)$$

em que $\mathbb{U}(i, j)$ é o conjunto de usuários que avaliaram tanto o item i quanto o item j . Já a predição de uma avaliação desconhecida está representada na Equação 27.

$$\hat{r}_{u, i} = b_{u, i} + \frac{\sum_{j \in \mathbb{I}(u, i, k)} sim(i, j)(r_{u, j} - b_{u, j})}{\sum_{j \in \mathbb{I}(u, i, k)} |sim(i, j)|} \quad (27)$$

em que a única diferença em relação a Equação 25 é o *bias* $b_{u, i}$ que é usado para ajustar os efeitos de avaliação, por exemplo um usuário dar notas mais altas que os demais e para alguns itens que recebem melhor avaliação do que os outros.

Uma característica importante da FC baseada em memória é a facilidade na implementação e em adicionar novos dados, contudo a escalabilidade dessa técnica é limitada. Em (MELVILLE; SINDHWANI, 2010) foi proposto a predição baseada nos itens, com o intuito de tornar a FC mais escalável.

Algumas das vantagens da FC, segundo (BALABANOVIĆ; SHOHAM, 1997) (BURKE, 2002) e (ADOMAVICIUS; TUZHILIN, 2005) são:

- A possibilidade de apresentar aos usuários recomendações automáticas.
- Não exige um processamento das características (descrições) dos conteúdos.
- A possibilidade de formação de comunidade de usuários pela identificação de seus gostos e interesses similares.

- ❑ É independente do domínio, indicada para recomendações de conteúdo que não pode ser descrito adequadamente por metadados.

Algumas das desvantagens deste tipo de filtragem, segundo (BALABANOVIĆ; SHOHAM, 1997) (BURKE, 2002) e (ADOMAVICIUS; TUZHILIN, 2005) são:

- ❑ **Novo item:** novos itens são regularmente adicionados aos SR. Os sistemas que utilizam a FC dependem unicamente da avaliação dos usuários para que possam fazer recomendações, ou seja, enquanto um novo item não for avaliado por um número considerável de usuários, o sistema não o recomendará.
- ❑ **Novo usuário:** este problema é similar ao que ocorre nos sistemas que utilizam a FBC. Para fazer recomendações precisas, o SR deve primeiramente aprender as preferências do usuário, a partir de suas avaliações.
- ❑ **Esparsidade:** o sucesso de um SR baseado em FC depende da disponibilidade de uma grande quantidade de usuários. Se o número de usuários é pequeno em relação ao volume de informações no sistema, existe um grande risco de as avaliações tornarem-se muito esparsas, devido ao pequeno número de avaliações, comparado ao necessário para se fazer recomendações.
- ❑ **Ovelha negra:** usuário com gosto incomum, terá dificuldade de encontrar usuários com gostos semelhantes, quando comparado com a maioria da população. Esse problema é específico dos usuários. Com isso as recomendações geradas para esse usuário podem ser pobres.

Pensando nas limitações apresentadas da FC e da FBC a FH combina essas duas técnicas, com o objetivo de maximizar as vantagens apresentadas em cada abordagem e minimizar as desvantagens de cada uma, como descrito a seguir.

3.3.3 Filtragem Híbrida - FH

O caso mais comum da FH é a junção da FBC com a FC (MELVILLE; SINDHWANI, 2010) (ADOMAVICIUS; TUZHILIN, 2005) e (BALABANOVIĆ; SHOHAM, 1997). Existem diferentes formas de agregar esses dois tipos de filtragem:

- ❑ **Incorporar algumas características da FBC na FC:** o sistema pode manter os perfis dos usuários, comparar diretamente os perfis para determinar os usuários semelhantes com a FBC e então utilizar a FC. Esta estratégia permite reduzir os problemas da escassez de itens comuns e dos novos itens, pois podem ser sugeridos com base no perfil do usuário. Assim, o usuário alvo 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 da FC na FBC:** o mais comum nessa categoria é a utilização de uma técnica de redução da dimensionalidade aplicada ao grupo de perfis com FBC. Para isso, o Indexação Semântica Latente - *Latent Semantic Indexing* (LSI)¹² é usado para criar perfis de usuários tipo FC. Neste caso, LSI é aplicada sobre o conjunto de perfis (tanto de FC como de FBC) em vez de ser aplicado no conjunto de documentos, com o objetivo de obter a semelhanças entre os perfis (ADOMAVICIUS; TUZHILIN, 2005).
- **Desenvolvimento de um modelo unificado de recomendação:** um modelo unificado que incorpora características das abordagens baseada em conteúdo e colaborativa. Um exemplo é a junção de um sistema de FBC com um classificador baseado em regras, objetivando a obtenção de recomendações confiáveis
- **Implementar os métodos FBC e FC separadamente e combinar suas recomendações:** nesta forma, existem dois cenários: combinar as avaliações obtidas individualmente em cada um dos métodos para oferecer uma recomendação final; avaliar em cada recomendação qual dos sistemas gera recomendações mais exatas, determinando o grau de confiança de cada uma e escolher o maior ou o mais consistente com as avaliações realizadas pelo usuário no passado.

Pensando nessas características, alguns autores optaram por utilizar a FH, em (WASFI, 1998) como no Listagem de Televisão Pessoal - *Personal TeleVision listings* (PTV) criado por (SMYTH; COTTER, 2001) tanto a FC como a FBC são empregadas para gerar dois conjuntos de recomendações independentemente, que são posteriormente combinados. (BALABANOVIĆ; SHOHAM, 1997) desenvolveram o Fab um SR que objetiva a construção de uma arquitetura híbrida. Os perfis de usuários são gerados a partir da análise de conteúdo e comparados para identificar usuários similares para recomendação colaborativa. Entretanto, os SR híbridos requerem a caracterização adequada dos produtos, que pode não estar disponível nas bases de dados reais.

Ambas FBC, FC e FH podem ser utilizadas para gerar recomendações e para avaliar a eficácia desses sistemas. Diversas medidas podem ser utilizadas, como descrito a seguir.

3.4 Medida de Avaliações

Segundo (HERLOCKER et al., 2004), a exatidão dos SR tem sido alvo de pesquisas desde 1994, sendo que diferentes medidas têm sido utilizadas para avaliar os SR. No estudo proposto por (HERLOCKER et al., 2004), os autores analisaram as medidas mais

¹² LSI: Permite descobrir padrões relativos à forma como as palavras são utilizadas num conjunto de documentos. Ao agrupar as palavras podem surgir coocorrências que caracterizam grupos de documentos.

populares e as classificaram em: **medida de exatidão da previsão** e **medida de exatidão do *ranking***, descritas a seguir.

3.4.1 Medida de Exatidão da Previsão

Nos SR, o objetivo da medida de exatidão da previsão é verificar o quanto as avaliações previstas pelo sistema estão próximas das verdadeiras avaliações realizadas pelo usuário. Por exemplo, o sistema MovieLens (DAHLEN et al., 1998) prevê o número de estrelas que um usuário deverá atribuir para cada filme. Com essa medida, a previsão será avaliada comparando o número de estrelas previstas com o número de estrelas realmente atribuída pelo usuário a um determinado filme.

Para verificar a exatidão das avaliações, as mesmas são divididas em: (1) conjunto de treinamento, que representa as avaliações realizadas pelos usuários r_{ui} e (2) conjunto de teste, que simula as novas avaliações \hat{r}_{ui} (BURKE, 2002).

Uma medida normalmente usada para este fim é o Erro Médio Absoluto - *Mean Absolute Error* (MAE) (SHARDANAND; MAES, 1995). O sistema gera a predição de avaliação \hat{r}_{ui} para um conjunto de teste \mathbb{T} de pares usuário-item (u, i) para as avaliações conhecidas r_{ui} . O MAE pode ser calculado, como mostra a Equação 28. Quanto mais próximo de zero for o MAE, mais precisas são as recomendações.

$$MAE = \sqrt{\frac{1}{\mathbb{T}} \sum_{(u,i) \in \mathbb{T}} |\hat{r}_{ui} - r_{ui}|} \quad (28)$$

Já o Raiz Quadrada do Erro Médio Quadrático - *Root Mean Square Error* (RMSE) é uma variação do Erro Médio Quadrático - *Mean Squared Error* (MSE) (HERLOCKER et al., 2004), correspondendo à sua raiz quadrada e é definido conforme a Equação 29.

$$RMSE = \sqrt{\frac{1}{\mathbb{T}} \sum_{(u,i) \in \mathbb{T}} (\hat{r}_{ui} - r_{ui})^2} \quad (29)$$

3.4.2 Medida de Exatidão do *Ranking*

A outra medida apresentada por (HERLOCKER et al., 2004) é a exatidão do *ranking*, que mede a habilidade do SR em apresentar uma listagem, em que os itens estão ordenados de acordo com a preferência do usuário, como o Top-N.

O Top-N é útil para os SR, pois proporciona ao usuário uma maior visibilidade das recomendações que estão nas primeiras posições da lista. As primeiras posições da lista têm um maior impacto na decisão do usuário em aceitar ou rejeitar uma recomendação. Outro ponto é a limitação visual para a recomendação, pois geralmente os *sites* têm o costume de mostrar até as 5 primeiras recomendações e, caso o usuário deseje, pode olhar as demais posições da lista.

Para o tarefa de recomendação TOP-N em que o sistema recomenda para o usuário os N primeiros itens mais relevantes para ele, os resultados são apresentados em termos de Média de Precisão ($AP@N$), descrita na Equação 30.

$$AP@N = \frac{\sum_{u \in \mathbb{T}} \frac{|\text{itens relevantes recomendados}|}{|N \text{ itens recomendados}|}}{|\mathbb{T}|} \quad (30)$$

\mathbb{T} é um conjunto de usuários de teste. Nos experimentos, se as avaliações não forem binárias, é necessário convertê-las para uma escala binária, pois os itens devem estar separados em duas categorias, relevantes e irrelevantes. No sistema MovieLens, por exemplo, se um item foi avaliado com 4 ou 5 estrelas, é considerado relevante, enquanto que, se um item foi avaliado entre 1 e 3, será considerado não relevante.

As medidas de exatidão da previsão e exatidão do *ranking*, descritas anteriormente, podem não ser suficientes para definir diferenças de eficácia entre os SR. Para esse propósito, o teste estatístico é aplicado, com o intuito de avaliar os resultados obtidos (DEMŠAR, 2006).

3.4.3 Validação Estatística

O teste de hipótese pode ser empregado em uma validação estatística, usando dados de um estudo científico. Geralmente, um conjunto de passos é utilizado como guia durante a execução do teste de hipóteses (DEMŠAR, 2006):

1. As hipóteses são formadas, ou seja, nessa fase é necessário definir a hipótese nula (H_0) e a hipótese alternativa (H_1) que é contraditória a H_0 . Normalmente, quando se trata de comparação de algoritmos a hipótese nula H_0 afirma que a eficácia dos algoritmos testados é equivalente. Após formular a hipótese nula, é conveniente formular uma hipótese alternativa H_1 , a qual pode ser aceita desde que a hipótese nula seja rejeitada.
2. Um nível de significância é determinado para o teste, isto é, a probabilidade de se cometer o erro do tipo I é fixada. Esse erro ocorre quando H_0 é rejeitada mesmo quando ela é verdadeira. Ao fixar a probabilidade de ocorrer o erro do tipo I , é possível determinar o valor crítico considerado (dado em tabela), o qual separará a região de rejeição da região de aceitação. Após o cálculo da estatística é possível saber se o resultado está na região crítica e então rejeitar H_0 , ou caso contrário, aceitar H_0 .

Para decidir entre a aceitação ou rejeição de uma determinada hipótese, é necessário calcular a estatística dos resultados gerados por metodologias de avaliação que comparem diferentes algoritmos em diversos conjuntos de dados. Nesse cenário, um teste não-paramétrico pode ser usado, pois não é muito exigente quanto à distribuição dos dados e

é geralmente utilizado em teste de hipóteses. Para tal, o teste de sinal aparece como uma solução adequada para realizar o teste estatístico (SHANI; GUNAWARDANA, 2013). Esse teste é apresentado a seguir.

3.4.3.1 Teste Sinal

O uso do teste estatístico é importante para criar conclusões confiáveis. O teste mais simples que pode ser usado é o teste do sinal, proposto por (DEMŠAR, 2006). (SHANI; GUNAWARDANA, 2013) utiliza o teste de sinal para avaliar a tarefa de previsão de classificação em um ambiente emparelhado usando o mesmo conjunto de teste.

Com base no teste de sinal, é possível então comparar os resultados que cada SR obteve e concluir qual o melhor, por exemplo, o sistema que tiver o RMSE menor. O RMSE de cada usuário u é calculado e o número de vezes que o sistema A atingiu um RMSE menor que o sistema B é contado, denotado por m_A . Como o interesse está em um único sistema, por exemplo, só será contabilizado em m_A quando o RMSE de A for menor do que o RMSE de B , em caso de empate, m_A não será contabilizado. O número de usuários no conjunto de teste para os quais foram feitas previsões é denotado por m .

Agora, a probabilidade pode ser calculada, de que, pelo menos m_A vezes o sistema A obteve uma pontuação melhor do que o sistema B . Nesse caso, a hipótese H_0 define que os dois sistemas são iguais e H_1 que os dois sistemas são diferentes, ambas usando o p -valor ou o nível de significância, obtido de acordo com a Equação 31.

$$p = (0.5)^n \sum_{i=m_A}^n \frac{n!}{i!(n-i)!}, \quad (31)$$

em que $n = m_A + m_B$. Quando o p -valor for inferior a algum valor pré-definido (normalmente, 0.05), H_0 é rejeitado, pois o sistema A não é igual ao sistema B com uma confiança de $(1 - p) * 100\%$. Com H_0 rejeitada, H_1 pode ser aceita.

3.5 Considerações Finais

Este capítulo apresentou os principais fundamentos sobre SR. Muitos desses conceitos serviram de base para o desenvolvimento da solução proposta neste trabalho. Os SR surgiram na década de 1990 e se destacam como importante área de pesquisa, com muitas aplicações em *e-commerce*. Além de minimizar os problemas da sobrecarga de informação, diminuindo o tempo de busca do usuário, esses sistemas têm permitido que os usuários tenham uma experiência nova de acesso a serviços personalizados e conheçam melhor suas preferências.

Geralmente, os SR estão classificados segundo seus principais aspectos: dados de entrada, saída e tipo de filtragem. Durante a implementação de um SR cada um desses

aspectos deve ser ponderado. Além disso, a análise do domínio de aplicação é extremamente importante, pois suas particularidades devem ser consideradas na recomendação.

A FBC, por exemplo, é uma técnica bem conhecida entre os SR, contudo pode não ser a melhor opção para bases de dados reais com pouca ou nenhuma informação associada aos itens, dificultando a criação dos perfis desses itens. Já os SR com FC não possuem essa limitação, pois a descrição dos itens não é importante, mas sim a avaliação dos mesmos. O foco dessa pesquisa será na FC. Neste trabalho será utilizada a FC baseada em memória, por não necessitar que se crie um modelo, ou que se atualize o mesmo, sempre que um conjunto de itens é adicionado na base de dados.

No entanto, a FC baseada em memória ainda precisa de aprimoramentos para melhorar a qualidade das recomendações. Esses aprimoramentos estão relacionados à forma de representar o comportamento do usuário, que seja menos intrusivo. É necessário que os SR que utilizam a FC incorporem mais características sobre os itens, para que, mesmo que não haja avaliações por parte dos usuários, o item possa ser comparado com os demais.

Os detalhes de como este trabalho incorpora mais características sobre os itens estão descritos no capítulo seguinte.

O método IKB-MS

Os SR surgiram da necessidade dos usuários de encontrar as informações de seu interesse em meio a uma grande quantidade de informações. Em um primeiro momento, o propósito desses sistemas foi alcançado, contudo, os SR apresentam limitações relevantes para o processo de recomendação. A necessidade de o usuário fornecer suas preferências e opiniões sobre os produtos/itens de forma explícita é uma delas. Outra limitação importante presente nos SR baseados em FC é a partida a frio de itens. Se o item não foi avaliado e não existe informações associadas a ele, como o mesmo poderá ser recomendado.

Neste capítulo, é descrito o SR que utiliza MS para caracterizar as imagens da base de dados, desenvolvido por este trabalho, denominado IKB-MS. O IKB-MS estende o método ItemKNN + Baseline (IKB) apresentado na seção 3.3.2.2. Os termos itens e imagens são usados indistintamente neste trabalho.

4.1 Visão Geral da Proposta

Com o objetivo de melhorar a eficácia do método IKB, este trabalho propõe modificar a similaridade $sim(i, j)$ da Equação 27, ponderando informações do comportamento passado do usuário e informações sobre a representatividade atenta das imagens. Uma visão geral da proposta deste trabalho é ilustrada na Figura 21, dividida em etapa *offline* (antes de recomendar uma imagem para o usuário) e etapa *online* (durante a recomendação). A etapa *offline* será melhor descrita pelo Algoritmo 1.

4.1.1 Etapa offline

(1) Construção da matriz de avaliações:

Nesta fase todas as avaliações dos itens fornecidas pelos usuários, variando de 1 (um) à 5 (cinco) são organizadas em uma matriz de avaliações. Já os itens ainda não avaliados contem um * (asterisco) no lugar da avaliação, como representado na Tabela 4.

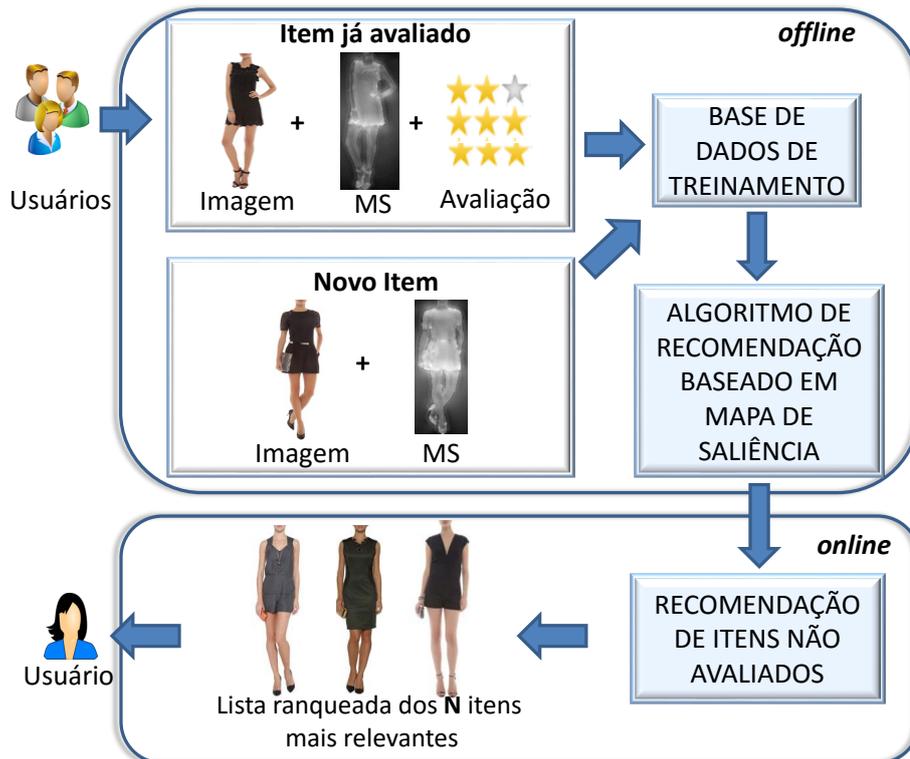


Figura 21 – Arquitetura geral do IKB-MS.

Algoritmo 1 IKB-MS (offline).

Entrada: Banco de imagens e o conjunto de avaliações para cada imagem

Saída: Matriz de avaliações atualizada

1. Constrói matriz de avaliações.
2. Para{cada imagem $i \in$ banco de imagens} {
 aplica-se o MCAV e a imagem i é particionada para gerar o vetor de atentividade.
 }
3. Combinando *sim* com *SMS* em uma medida *sc* que será utilizada na predição.
4. Atualiza a matriz de avaliações com as novas predições.

Fim algoritmo

- (2) Para{cada imagem $i \in$ banco de imagens} aplica-se o MCAV e a imagem i é particionada para gerar o vetor de atentividade:

Os MCAV descritos no capítulo 2 podem ser aplicados em qualquer imagem com o intuito de obter o MS daquela respectiva imagem. Já o particionamento das imagens é extremamente dependente da aplicação, pois pode ser diferente de uma imagem para outra, como ilustrado na Figura 22.

Esse particionamento é fundamental para determinar quais características de cada partição do MS chamam mais a atenção de um usuário e com isso recomendar itens

Tabela 4 – Matriz de avaliações.

	item 1	item 2	item 3	item 4	item 5
usuário 1	4	2	5	3	*
usuário 2	*	*	3	*	4
usuário 3	2	*	*	*	3
usuário 4	*	5	4	*	1

que possivelmente sejam de seu agrado.

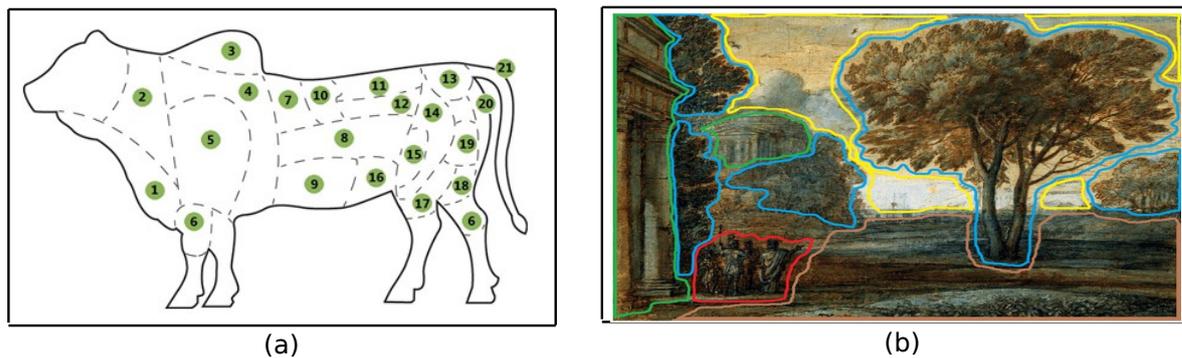


Figura 22 – **Particionando cada imagem de acordo com a aplicação.** (a) particionamento utilizado para boi de corte (b) particionamento utilizado para uma paisagem específica.

Por fim o vetor de atentividade é criado com o objetivo de determinar o quão atenta é cada região da imagem particionada. Para isso, a mediana¹ da atenção visual de cada região é calculada. Quanto maior for o valor da mediana, mais atenta é a respectiva região na imagem original. Após o cálculo da mediana uma normalização é realizada de forma que a soma de todas as coordenadas do vetor seja igual a 1. O tamanho desse vetor de atentividade é igual ao número de regiões da imagem particionada.

Um exemplo da construção de um vetor de atentividade é ilustrado na Figura 23. Primeiro aplica-se os MCAV para extrair o MS, depois esse MS é particionado em regiões distintas e a mediana de cada região é calculada e normalizada para se obter o vetor de atentividade.

(3) Combinando *sim* com *SMS* em uma medida *sc* que será utilizada na predição:

¹ Para determinar a mediana, os valores da amostra devem ser colocados em ordem crescente ou decrescente. Se a quantidade de valores da amostra for ímpar, a mediana é o valor central da amostra. Nesse caso, há a mesma quantidade de valores acima e abaixo desse valor. Se a quantidade de valores da amostra for par, é preciso tirar a média dos valores centrais para calcular a mediana. Nesse caso, 50% dos valores da amostra estão abaixo e 50% dos valores da amostra estão acima desse valor.

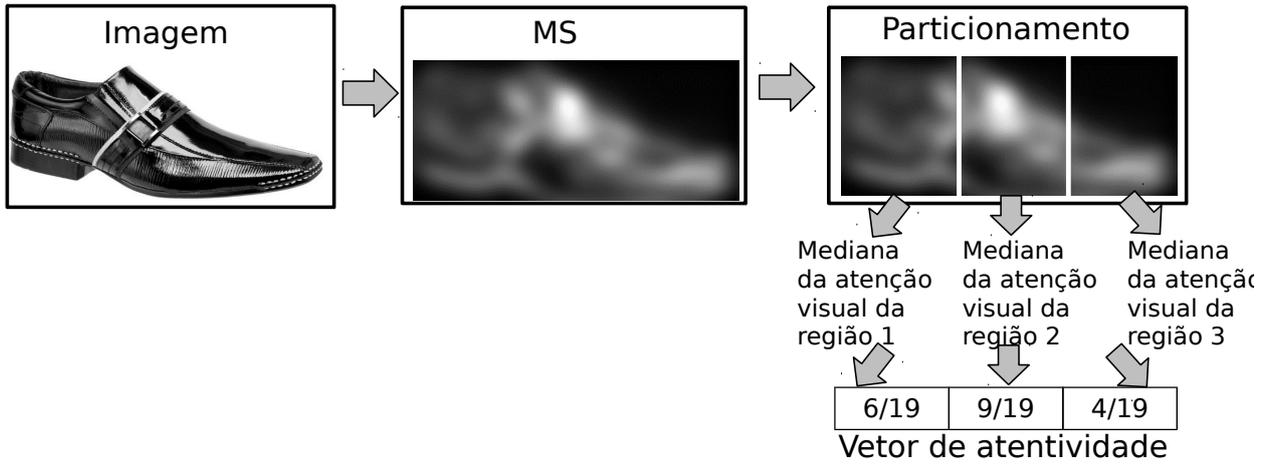


Figura 23 – Exemplo da construção de um vetor de atividade.

Para integrar o vetor de atividade ao sistema de recomendação IKB apresentado na seção 3.3.2.2, este trabalho propõe combinar uma medida de similaridade sim (como o cosseno da Equação 21 ou a correlação de Pearson da Equação 23) com uma nova medida denominada SMS , aqui proposta. Esta medida de similaridade pode ser baseada na intersecção de histogramas (SMEULDERS et al., 2000), divergência de Bregman (BREGMAN, 1967) ou cosseno (GUNAWARDANA; SHANI, 2009), ambas utilizadas no trabalho proposto para indicar o grau de similaridade entre dois vetores de atividade, conforme definido na Equação 32 (baseada na intersecção de histogramas), Equação 33 (baseada na divergência de Bregman) e Equação 34 (baseada na medida de similaridade cosseno).

$$SMS(i, j) = \sum_{t=1}^n \min(V_i[t], V_j[t]) \quad (32)$$

$$SMS(i, j) = \sum_{t=1}^n V_i[t] \log_2\left(\frac{V_i[t]}{V_j[t]}\right) \quad (33)$$

$$SMS(i, j) = \frac{\sum_{t=1}^n V_i[t] \times V_j[t]}{\sqrt{\sum_{t=1}^n (V_i[t])^2} \times \sqrt{\sum_{t=1}^n (V_j[t])^2}} \quad (34)$$

em que, V_i representa o vetor de atividade da imagem i e $|V_i| = |V_j| = n$, em que n é o número de partições do MS e $V_i[t]$ representa o valor da posição t do vetor de atividade V_i .

A medida SMS será utilizada pela nova medida de similaridade $sc(i, j)$, apresentada na Equação 35 que combina $sim(i, j)$ (veja Equações 21 ou 23) com SMS (veja

Equação 32, Equação 33 ou Equação 34). A similaridade entre duas imagens é determinada por $sc(i, j)$ definida como segue:

$$sc(i, j) = (\alpha \cdot (sim(i, j)) + (1 - \alpha) \cdot (SMS(i, j)))^p \quad (35)$$

em que α é um parâmetro predefinido pelo usuário e cujo valor está contido no intervalo $[0, 1]$. O valor de α pondera as medidas $sim(\cdot)$ (que determina a similaridade entre as preferências dos usuários) e $SMS(\cdot)$ (que determina a similaridade atenta das imagens). Valores baixos para α indicam que a atenção visual tem maior importância no cálculo da similaridade entre itens. Outro parâmetro definido pelo usuário é o fator de amplificação p . A utilização desse parâmetro foi baseada no trabalho proposto por (BREESE; HECKERMAN; KADIE, 1998), com o objetivo de favorecer os itens com a similaridade mais elevada e desfavorecer os itens com a similaridade mais baixa.

A predição de avaliação do IKB-MS é obtida substituindo-se $sim(i, j)$ da Equação 27 por $sc(i, j)$, conforme definido na Equação 36.

$$\hat{r}_{ui} = b_{ui} + \frac{\sum_{j \in \mathbb{I}(u, i, k)} sc(i, j)(r_{uj} - b_{uj})}{\sum_{j \in \mathbb{I}(u, i, k)} sc(i, j)} \quad (36)$$

(4) Atualiza a matriz de avaliações com as novas predições:

Com os passos um, dois e três do algoritmo, o IKB-MS realizará as predições com base nas avaliações dos usuários e na atatividade dos itens. Nesta fase todas as avaliações preditas são atualizadas na matriz de avaliações, como representado na Tabela 5.

Tabela 5 – Matriz de avaliações atualizada com as predições.

	item 1	item 2	item 3	item 4	item 5
usuário 1	4	2	5	3	1
usuário 2	1	5	3	1	4
usuário 3	2	3	2	5	3
usuário 4	2	5	4	2	1

4.1.2 Etapa online

Com base nas avaliações realizadas pelos usuários e a similaridade entre itens definida pelas Equações 35 e 36, nesta etapa o usuário receberá uma lista contendo as imagens que possivelmente mais seriam de seu interesse, de acordo com a predição determinada na Equação 35.

4.1.3 Recomendação de itens não avaliados

Um dos maiores problemas ligados à FC é a chamada “partida a frio” (*cold-start*). O problema da partida a frio de itens se deve ao fato de que, quando um item é recém inserido na base de dados, não possui avaliações e, por esse motivo, o sistema não o recomendará.

No entanto, o SR aqui proposto utiliza no cálculo de similaridade, não apenas as informações sobre a preferência dos usuários, mas também a caracterização das imagens via os seus vetores de atentividade. Esse vetor é obtido independentemente da imagem ter sido ou não previamente avaliada. Sendo assim toda imagem possui um vetor de atentividade associado a ela.

O impacto causado pela partida a frio pode ser minimizado, eliminando o termo $sim(.)$ da Equação 35 referente a similaridade entre as avaliações dos usuários, como mostra a Equação 37.

$$sc = SMS(i, j) \quad (37)$$

4.2 Considerações Finais

Neste capítulo foi apresentado um novo método para recomendação de imagens, denominado IKB-MS, que estende o método IKB proposto por (KOREN, 2010).

Para tanto uma nova medida de similaridade foi proposta, referenciada como $sc(.)$, que pondera a similaridade entre as avaliações dos itens e a similaridade atenta entre os itens.

Reduzindo a medida $sc(.)$, considerando apenas a similaridade atenta, o IKB-MS contorna as limitações dos SR baseados em FC no que se refere ao problema de partida a frio de itens.

Experimentos e Análise dos Resultados

Neste capítulo, são apresentados e discutidos os diversos experimentos realizados que comprovam a hipótese de que o uso da atenção visual associado a similaridade de avaliação entre itens melhora a predição em SR de imagens.

5.1 Introdução

Os experimentos contidos neste capítulo visam apresentar o método desenvolvido e a eficácia dos resultados em relação a outros SR. O capítulo está organizado como segue.

Na seção 5.2 são descritos os métodos utilizados para validar a hipótese como bases de dados, critérios de avaliação, algoritmos comparados e parâmetros utilizados. A base de dados descrita na seção 5.2.1 é composta de 6.946 imagens de roupas, desenvolvida em nosso laboratório de pesquisa. Esta base de dados foi utilizada para testar o algoritmo proposto neste trabalho e os algoritmos comumente utilizados na área, tais como, IKB, UKB e SVD, definidos na seção 5.2.3. Para determinar o desempenho dos SR adotou-se as duas métricas RMSE e MAE, definidas na seção 5.2.2. A seção 5.3 ilustra os principais experimentos deste trabalho. Por fim, a seção 5.4 apresenta as considerações finais sobre esse capítulo.

5.2 Configurações dos Experimentos

Nesta seção são descritas as medidas de avaliação, os parâmetros utilizados, a base de dados e os trabalhos com os quais a proposta dessa dissertação será comparada.

5.2.1 Base de Dados

A avaliação do método proposto nesse trabalho foi conduzida utilizando a base de dados UFU-CLOTHING Database disponível no endereço: <http://www.lincci.ufu.br/>, denominada UFU-CLOTHING. A base de dados é composta por 6.946 imagens de rou-

pas coletadas de vários *sites* de compras brasileiros, como Privalia¹, Posthaus², Oqvestir³ e Netshoes⁴, usando palavras-chave como camiseta, terno, vestido, etc. A resolução das imagens é de 500x700 *pixels*. A Figura 24 apresenta algumas imagens da UFU-CLOTHING Database. As imagens foram organizadas de acordo com o gênero (masculino e feminino), classe do corpo (completo, superior e inferior) e categoria (calça, terno, camiseta, etc). A Tabela 6 apresenta uma consolidação dessas informações.



Figura 24 – Exemplos de imagens da base de dados UFU-CLOTHING. (a) roupas masculinas e (b) roupas femininas

Tabela 6 – Número de imagens separadas pelo gênero, classe do corpo e categoria.

Masculino								
Completo	Superior				Inferior			Total
Terno	Camisa	Camiseta Polo	Camiseta	Jaqueta	Bermuda	Calça	Shorts	
185	180	336	586	372	187	150	258	2.254
Feminino								
Completo	Macacão	Superior			Inferior			Total
Vestido	Macacão	Camisa	Camiseta	Jaqueta	Saia	Calça	Shorts	
783	98	1.316	325	192	510	737	731	4.692

Para se obter as avaliações, o conjunto de dados foi separado de acordo com o sexo: masculino e feminino. Cada usuário do sexo masculino avaliou 300 itens (210 itens são comuns a todos os usuários do sexo masculino e 90 itens foram escolhidos de forma aleatória). A partir destes 300 itens, o usuário avaliou 20 ternos, 30 camisas, 50 camisetas polo, 60 camisetas, 40 jaquetas, 30 bermudas, 30 calças e 40 shorts. Cada usuário do sexo feminino também avaliou 300 itens (210 itens são comuns a todos os usuários do

¹ <http://br.privalia.com>

² <http://www.posthaus.com.br>

³ <http://www.oqvestir.com.br>

⁴ <http://www.netshoes.com.br>

sexo feminino e 90 itens foram escolhidos aleatoriamente). Os itens avaliados incluem 20 macacões, 40 vestidos, 60 camisas, 30 camisetas, 30 jaquetas, 40 calças, 40 saias e 40 shorts.

Um total de 245 voluntários, divididos em 122 homens e 123 mulheres com idades entre 18 e 60 anos, avaliaram os itens do banco de dados. A ordem de exibição das roupas foi aleatória e o usuário avaliou cada roupa de acordo com sua preferência: 0 (nada a dizer); 1 (odiei); 2 (não gostei); 3 (neutro); 4 (gostei); 5 (gostei muito). O tempo de apresentação de cada roupa dependia do tempo necessário para o usuário efetuar sua avaliação. Cada participação durou, em média, 15 minutos. Atualmente, a base de dados é composta por 73.414 avaliações. A esparsidade dos dados atinge aproximadamente 95%.

Cada imagem da base de dados pertence a uma classe: superior, inferior ou completo. Essas imagens são particionadas de forma automática de acordo com a classe pertencente. Se a imagem for da classe superior, serão 12 partições superiores do corpo, tais como pescoço, ombro direito, ombro esquerdo, etc. Se a imagem for da classe inferior, serão 10 partições inferiores, tais como joelho direito, joelho esquerdo, etc. Por fim, se for uma imagem da classe completo, será particionada em 22 partições. A Figura 25 mostra alguns exemplos de imagens particionadas. Isso foi possível pois, geralmente, em *sites* de compras *online* de roupas, as imagens contém modelos posando profissionalmente de maneira similar. Então, cada imagem pode ser facilmente particionada em partições do corpo humano.

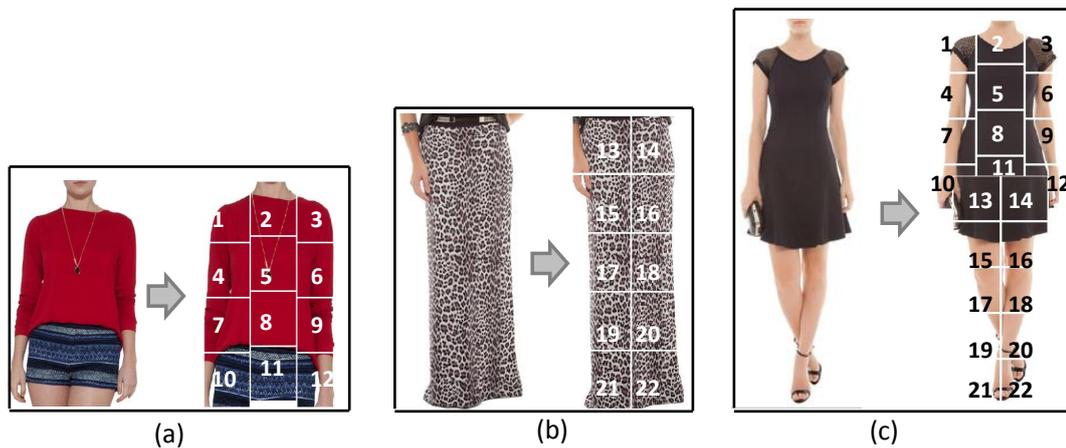


Figura 25 – **Imagens divididas em 12, 10 e 22 partições respectivamente.** (a) superior, (b) inferior e (c) completo. No caso de (a) e (b) o vetor de atenção também tem 22 posições mas são preenchidas as 12 e 10 primeiras posições, respectivamente sendo que as posições restantes ficarão com zero.

O vetor de atenção criado para este experimento contém 22 posições, cada uma correspondendo a uma região, como ilustrado na Figura 26.

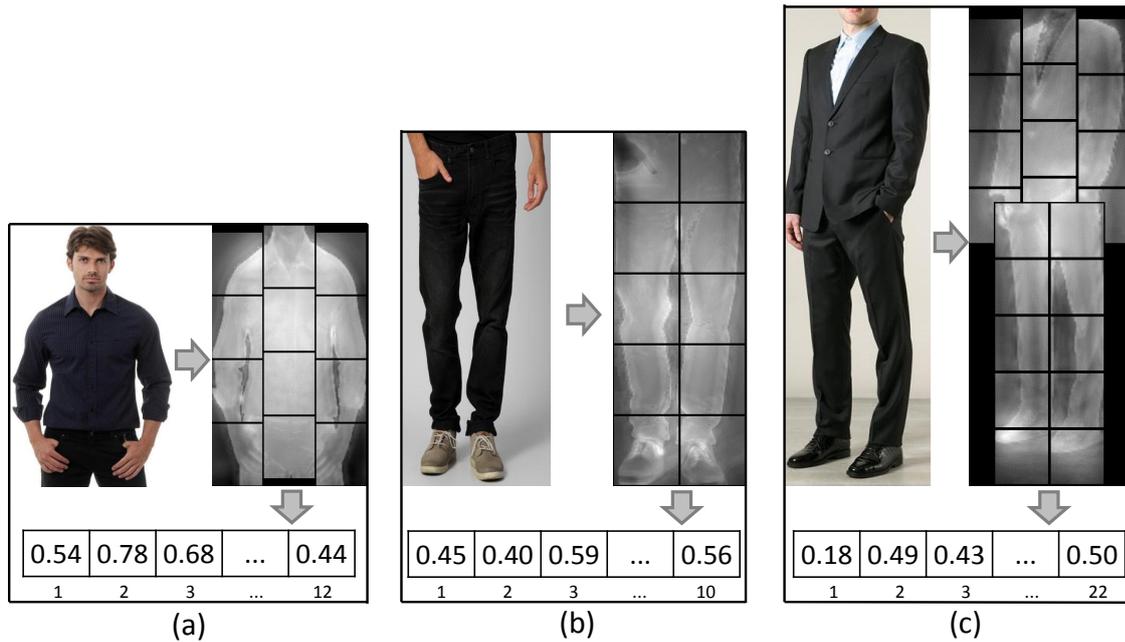


Figura 26 – Vetores de atentividade correspondendo aos MS extraídos de imagens exemplo. (a) 12 partições do MS, (b) 10 partições do MS e (c) 22 partições do MS.

5.2.2 Critérios de Avaliação

Os experimentos foram conduzidos aplicando a técnica de validação cruzada (*k-fold-cross-validation*). Nestes experimentos, foi utilizado $k = 10$, ou seja, dividiu-se aleatoriamente a base de dados de avaliações em 10 conjuntos D_1, \dots, D_k de tamanhos iguais (ou aproximados). Posteriormente, k iterações de treinamento e de validação são executadas, de tal forma que, em cada iteração i , um conjunto diferente D_i é usado para validação, enquanto os conjuntos restantes são usados para aprendizagem.

Para avaliar o método IKB-MS no cenário de partida a frio de itens, foram selecionados aleatoriamente 100 itens. Para cada item i foram criados dois conjuntos de dados: teste e treinamento. O conjunto de teste é composto somente por avaliações sobre o item i . O conjunto de treinamento é composto somente pela informação da atenção visual de i . O experimento foi executado 100 vezes, uma vez para cada item previamente escolhido.

Para avaliar a tarefa de predição, este trabalho adotou duas métricas populares usadas para mensurar a eficácia dos SR: *Root Mean Squared Error* (RMSE) (ARMSTRONG; COLLOPY, 1992) e *Mean Absolute Error* (MAE) (SHARDANAND; MAES, 1995), ambas descritas na seção 3.4.1. Essas métricas verificam o quanto as avaliações previstas pelo sistema estão próximas das verdadeiras avaliações realizadas pelo usuário.

Para a tarefa de recomendação Top-N, os resultados são apresentados em termos do Precisão Média - *Average Precision* (AP) (seção 3.4.2), o qual é calculado com base nas

medidas de precisão e revocação. A AP mede a habilidade do SR em apresentar uma lista de itens, em que os itens estão ordenados de acordo com a preferência do usuário em ordem decrescente. Nos experimentos, adotaram-se como relevantes as avaliações maiores ou iguais a 4 (estrelas) e não-relevantes, as avaliações menores que 4 (estrelas).

Por fim, para comprovar estatisticamente se há diferenças significantes nos resultados de eficácia dos algoritmos, aplicou-se o teste de sinal apresentado na seção 3.4.3.

5.2.3 Algoritmos Comparados

Neste trabalho, o método proposto foi comparado com os principais SR comumente utilizados na área, disponíveis publicamente no *software* MyMediaLite (GANTNER et al., 2011):

1. *Random*: método que prevê aleatoriamente uma avaliação desconhecida, por exemplo, \hat{r}_{ui} é um número aleatório entre 1 e 5.
2. *Global Average (GA)*: método que prediz uma avaliação desconhecida usando a média global de todas as avaliações.
3. *User Average (UA)*: método que usa o valor da média das avaliações do usuário para predizer uma avaliação desconhecida do usuário correspondente.
4. *Item Average (IA)*: método que usa a média de avaliações do item para predizer uma avaliação desconhecida do item correspondente.
5. *UserItemBaseline (UIB)*: método que usa a média global μ , mais o *bias* para predição do usuário e do item (KOREN, 2010)
6. *SVD*: modelo tradicional de fatoração de matrizes.
7. *SVD++*: método que considera o *feedback* implícito dos usuários para a predição (KOREN, 2008).
8. *ItemKNN+Baseline (IKB)*: método que prediz uma avaliação desconhecida calculando a média ponderada das avaliações dos itens vizinhos. Esse método é o baseline da abordagem baseada em FC (KOREN, 2010).

5.2.4 Configurações dos algoritmos utilizados

Nos experimentos desta pesquisa, os MS foram extraídos utilizando três modelos de atenção visual diferentes: (ITTI; KOCH; NIEBUR, 1998) (HAREL; KOCH; PERONA, 2006) e (JUDD et al., 2009). Enquanto que os dois primeiros modelos foram utilizados nos experimentos respeitando a proposta original, o modelo proposto por (JUDD et al., 2009) sofreu alterações com relação ao número de características utilizadas para gerar o MS.

Ao invés de se utilizar 33 características como o modelo original, optou-se por utilizar 30 características, devido à base de dados empregada nos experimentos não conter imagens de face, pessoas ou carros.

Os parâmetros dos métodos disponíveis no MyMediaLite foram escolhidos empiricamente. A maioria dos parâmetros com a configuração padrão do MyMediaLite obtiveram melhores resultados. Para o cálculo da similaridade SMS descrita na Equação 35, o fator de amplificação adotado foi $p = 3$, baseando-se em (BREESE; HECKERMAN; KADIE, 1998). Já o valor de α que determina a importância de $sim(i, j)$ e $SMS(i, j)$, e o parâmetro k que representa o número de vizinhos mais próximos, proporcionam um maior impacto nos resultados e por esse motivo serão discutidos na próxima seção.

5.3 Experimentos

Analisou-se o método IKB-MS usando três diferentes MCAV, sendo cada variação do método denominada como: IKB-MS-I (IKB + (ITTI; KOCH; NIEBUR, 1998)), IKB-MS-H (IKB + (HAREL; KOCH; PERONA, 2006)), IKB-MS-J (IKB + (JUDD et al., 2009)), cada uma dessas variações foram testadas em termos de: escolha da medida de similaridade usada na SMS , do valor de α usado na medida $sc(i, j)$, o melhor número de partições para a base de dados utilizada e o tamanho da vizinhança para a predição, cada um desses experimentos será melhor apresentado no decorrer desta seção. Também foi comparado o método IKB-MS com os algoritmos apresentados na seção 5.2.3 em termos de precisão na predição de avaliação (utilizando as medidas de similaridade cosseno e Pearson), na partida a frio de itens e recomendação Top-N. Por fim, um teste estatístico é realizado, para se comparar a significância dos ganhos obtidos pelo IKB-MS. Lembrando que quanto menor o RMSE e o MAE mais eficaz é o método.

5.3.1 Escolha da medida de similaridade usada na SMS

Antes do algoritmo proposto ser comparado com os demais métodos da literatura, um experimento foi realizado com o intuito de determinar qual medida de similaridade (intersecção de histogramas, cosseno ou KL) utilizada na SMS tem uma melhor eficácia em relação as demais. Os resultados podem ser comparados na Figura 27, e vão de acordo com o esperado, pois quanto mais zeros no vetor de atividade pior será o resultado da KL e dependendo da imagem o vetor de atividade de tamanho 22 pode ter até 12 zeros. Já a cosseno que determina a diferença entre o ângulo dos vetores de atividade também não é medida ideal para representar o valor real da similaridade. Com base nos resultados obtidos por esse experimento a medida de similaridade a ser utilizada na SMS nos demais experimentos será a intersecção de histogramas.

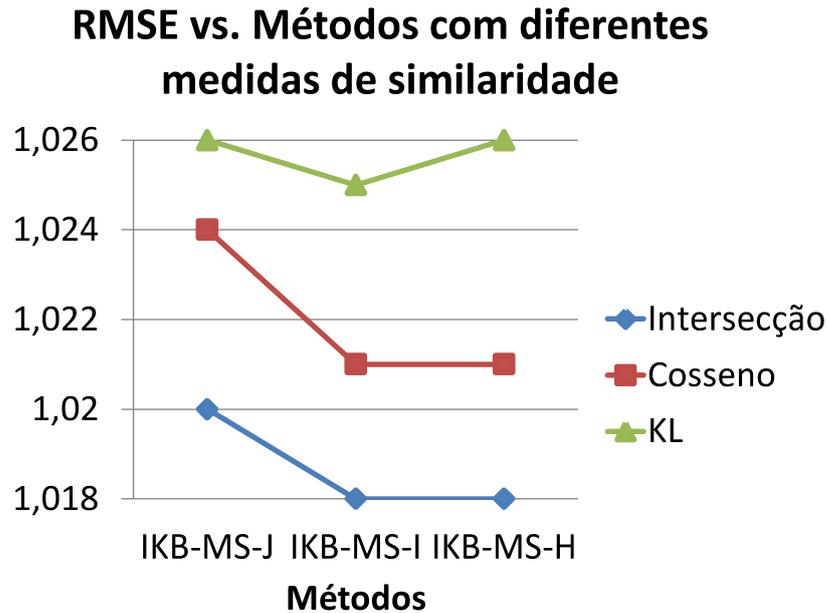


Figura 27 – Comparando as diferentes medidas de similaridades utilizadas na SMS.

5.3.2 Escolha do α usado na medida $sc(i, j)$

Com a medida de similaridade a ser usada na *SMS* definida, outra questão a ser analisada é o valor de α da medida $sc(i, j)$ (Equação 35). Para tal, foi realizado um experimento variando o valor de α de 0.1 à 0.9. Os resultados obtidos estão representados na Figura 28, observando a figura notamos que $\alpha = 0.8$ proporcionou os melhores resultados e por esse motivo para os próximos experimentos o valor de α utilizado será 0.8.

5.3.3 Escolha do número de partições

Para ser possível determinar o quanto o particionamento das imagens interfere no resultado, realizou-se um experimento variando o número de partições e por consequência o tamanho do vetor de atenção também sofre alteração. Com isso as imagens de roupa da base de dados foram dividida em 2, 4, 11 ou 22 partições e os resultados obtidos estão na Figura 29, analisando esses resultados foi observado que 22 partições consegue representar melhor a imagem e por esse motivo, para os próximos experimentos o vetor de atenção terá tamanho 22, correspondendo a imagem particionada.

5.3.4 Escolha do tamanho de vizinhança para a predição

Para que as comparações entre os métodos IKB e IKB-MS fossem justas, foram feitos experimentos para se obter o tamanho de vizinhança mais adequado.

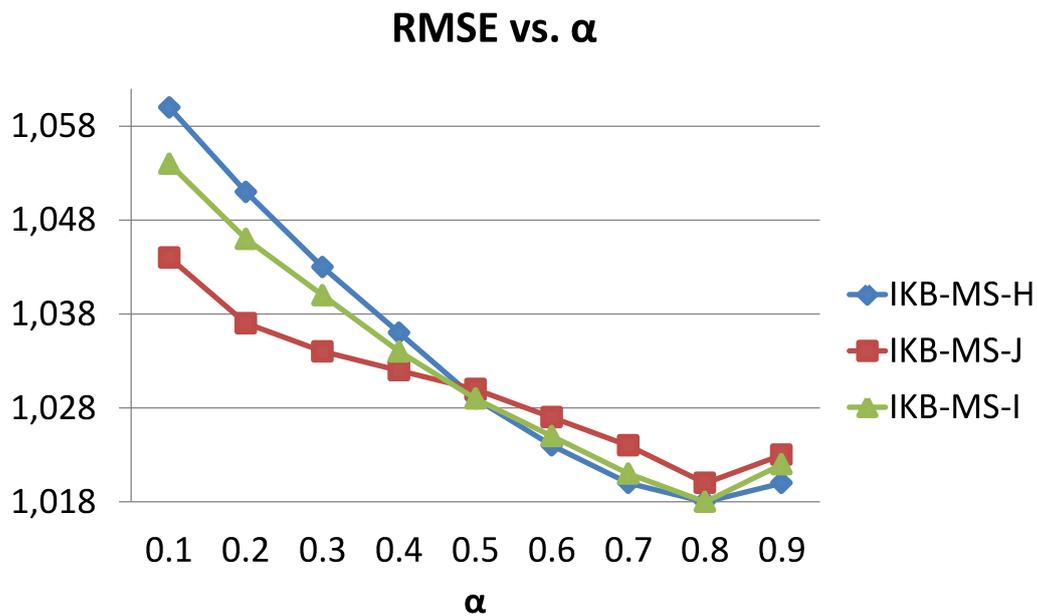


Figura 28 – Comparando diferentes valores de α .

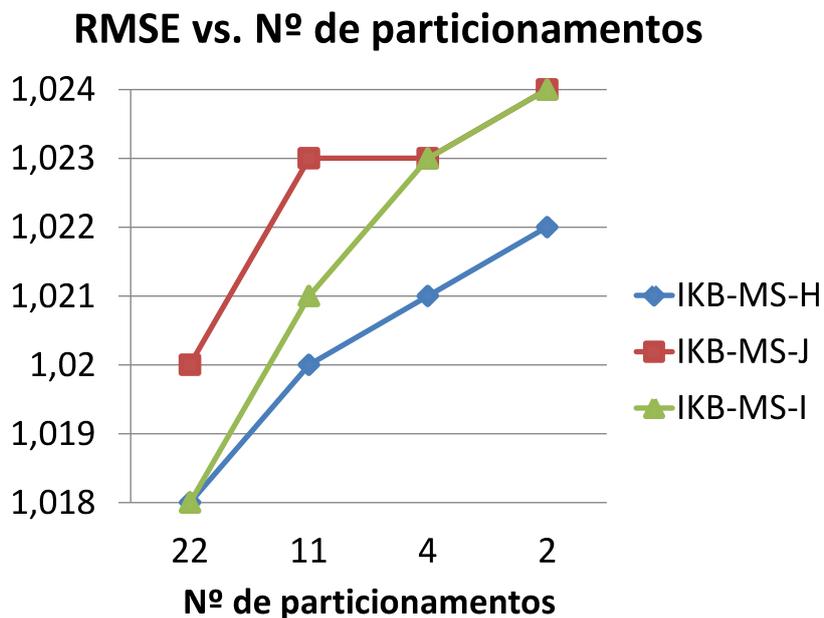


Figura 29 – Comparando diferentes números de partições das imagens.

O tamanho de vizinhança é determinante para o resultado de predição usando IKB-MS. As medidas de similaridades utilizadas para ponderar a predição de ambos os métodos nesse experimento foram o cosseno (Equação 21) e a correlação de Pearson (Equação 23).

A Figura 30 apresenta a comparação entre o IKB e o IKB-MS com a medida de similaridade cosseno variando a vizinhança k entre 10 e 50. Quanto mais baixo o RMSE

melhor é a precisão de predição.

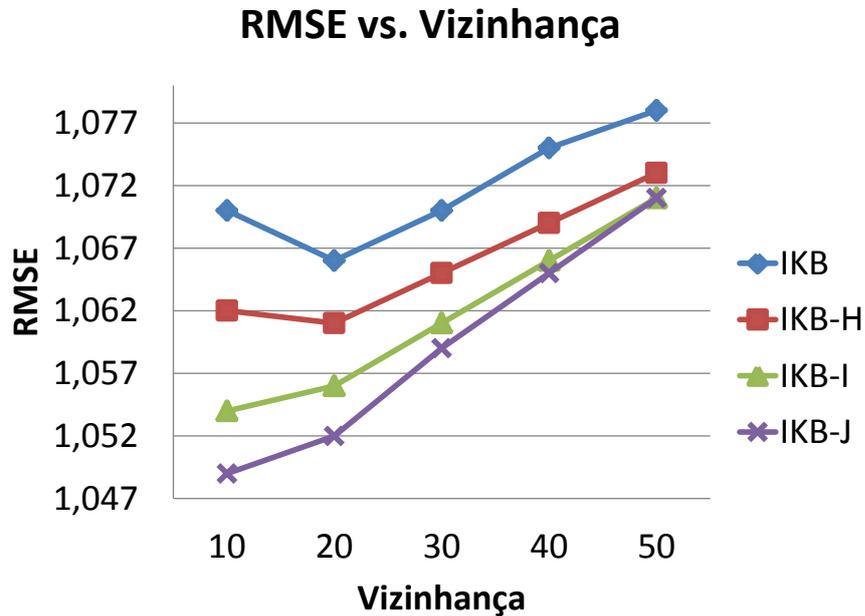


Figura 30 – Comparando IKB com IKB-MS e ponderando a predição com o cosseno. O RMSE é mostrado com vários tamanhos de vizinhança.

A vizinhança foi variada da mesma forma em um segundo experimento, mas agora utilizando a correlação de Pearson para ponderar a predição. A Figura 31 apresenta a comparação entre o IKB e o IKB-MS usando os três MCAV diferentes, em termos do RMSE. Conforme a vizinhança vai sendo alterada, o IKB-MS mantém bons resultados, enquanto que os resultados do IKB vão ficando cada vez piores.

Note que em ambos os casos a eficácia do IKB-MS é dependente do MCAV sendo utilizado. No segundo experimento tanto o IKB quanto o IKB-MS obtiveram os melhores resultados com $k = 20$.

Com base nesses resultados os próximos experimentos, utilizarão a correlação de Pearson para ponderar a predição, com o tamanho da vizinhança definido para $k = 20$.

Uma vez determinada a medida de similaridade e o tamanho a vizinhança novos experimentos foram executados para comparar o IKB-MS com os demais trabalhos competitivos listados na seção 5.2.3. A Tabela 7 apresenta um resumo dos resultados obtidos em termos do RMSE e MAE. Foram conduzidos experimentos para comparar o IKB-MS com métodos de recomendação que utilizam fatores latentes como é o caso do SVD e SVD++. O número de fatores latentes usado foi determinado considerando o melhor resultado obtido pelo algoritmo.

Analisando os resultados, é importante notar que, em todos os casos, a proposta deste trabalho alcançou melhores resultados do que os SR comumente utilizados na área, variando de 2.94% (IKB) a 71.21% (Random) de diferença. Esses ganhos em termos de

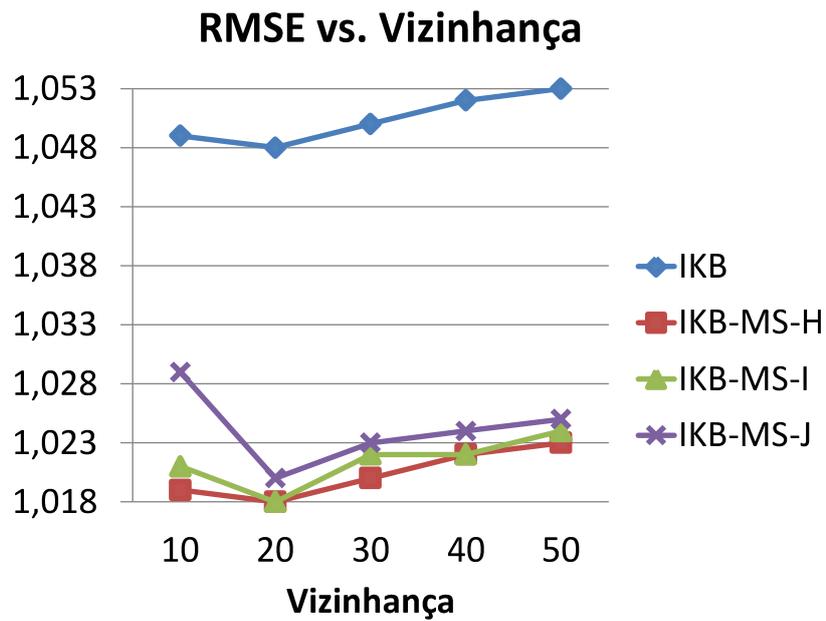


Figura 31 – Comparando IKB com IKB-MS e ponderando a predição com a correlação de Pearson. O RMSE é mostrado com vários tamanhos de vizinhaça.

Tabela 7 – Comparação da eficácia dos métodos em termos do RMSE e MAE.

Métodos	RMSE	MAE	Parâmetros
1. Random	1,743	1,428	
2. GA	1,292	1,113	
3. UA	1,205	1,019	
4. IA	1,248	1,026	
5. UIB	1,114	0,921	Nº de vizinhos (K) = 20
6. SVD	1,105	0,863	Fatores latentes = 10
7. SVD++	1,107	0,910	Fatores latentes = 25
8. IKB	1,048	0,838	Nº de vizinhos (K) = 20
9. IKB-MS-J	1,020	0,813	Nº de vizinhos (K) = 20
10. IKB-MS-I	1,018	0,810	Nº de vizinhos (K) = 20
11. IKB-MS-H	1,018	0,808	Nº de vizinhos (K) = 20

eficácia se devem ao fato de que, a caracterização das imagens da base de dados tende a complementar as avaliações dessas imagens, ou seja, cada item da base passa a ser melhor representado, o que possibilita uma melhor recomendação.

5.3.5 Minimizando a partida a frio de itens

Em um cenário em que o SR não possui atributos dos itens, seja por dificuldades de manutenção ou espaço em disco, este trabalho propõe minimizar o problema de partida a

frio utilizando a atenção visual para identificar possíveis vizinhos semelhantes ao novo item i . Uma situação, em que um novo item i não foi avaliado, é possível calcular a similaridade SMS entre o item i e os demais itens, conforme a Equação 36 apresentada no capítulo 4. Consequentemente, é possível prever as avaliações sobre o item i utilizando o método IKB-MS.

Para avaliar a eficácia da proposta deste trabalho, os métodos comumente utilizados na área diretamente aplicados a este problema foram selecionados: *Random*, *Global Average* (GA), *User Average* (UA) e *UserItemBaseline* (UIB). Os métodos *Item Average* (IA) e *ItemKNN+Baseline* (IKB) não podem ser utilizados neste experimento, pois tanto a média das avaliações dos itens quanto a média ponderada, avaliações dos itens vizinhos são necessárias para realizar a predição de uma avaliação desconhecida. Assim como IA e IKB, os métodos baseados em fatores latentes (LATHAUWER; MOOR; VANDEWALLE, 2000; BILLSUS; PAZZANI, 1998) também não foram utilizados, uma vez que os fatores latentes de novos itens são desconhecidos.

A Tabela 8 apresenta os resultados dos métodos sendo avaliados em termos de RMSE e MAE. O melhor resultado alcançado entre os trabalhos correlatos foi obtido pelo método UIB com $RMSE = 1,212$ e pelo UA com $MAE = 1,025$. A melhor configuração do IKB-MS foi obtida pelo IKB-MS-J que alcançou um $RMSE = 1,174$ e $MAE = 0,983$, apresentando um ganho, em termos do RMSE de 3,23% e em termos do MAE de 4,27% com relação aos melhores resultados dos concorrentes.

Tabela 8 – Comparação da eficácia em termos do RMSE e MAE para o problema de partida a frio.

Métodos	RMSE	MAE	Parâmetros
1. Random	1,723	1,405	
2. GA	1,281	1,100	
3. UA	1,214	1,025	
4. UIB	1,212	1,027	
5. IKB-MS-H	1,191	0,995	Nº de vizinhos (K) = 20
4. IKB-MS-I	1,180	0,988	Nº de vizinhos (K) = 20
3. IKB-MS-J	1,174	0,983	Nº de vizinhos (K) = 20

Com os resultados apresentados até este momento é possível notar que há um ganho de eficácia do IKB-MS em relação aos algoritmos comumente utilizados na área, principalmente quando novos itens são inseridos na base de dados. Diferentemente dos métodos IKB, SVD, SVD++ e IA, o IKB-MS consegue gerar recomendações mesmo quando o item possui poucas ou nenhuma avaliação.

5.3.6 Recomendação Top-N

Na prática, é desejável que os SR recomendem apenas um pequeno conjunto de itens relevantes para o usuário.

Diferentes itens podem ser relevantes para diferentes usuários. Em uma tarefa de recomendação Top-N, o sistema recomenda ao usuário os primeiros N itens mais relevantes para ele/ela. A Tabela 9 mostra a média de precisão (AP) da recomendação top-1 (AP@1) até top-5 (AP@5). Em todos os casos o IKB-MS, nas suas diferentes versões, alcançou os melhores resultados quando comparado aos métodos competidores, com ganhos variando entre 5,39% (AP@1) a 3,72% (AP@5).

Tabela 9 – Média de precisão na recomendação Top-N.

	AP@1	AP@2	AP@3	AP@4	AP@5
1. Randon	0,348	0,356	0,351	0,349	0,350
2. GA	0,356	0,352	0,350	0,351	0,354
3. UA	0,356	0,352	0,350	0,351	0,354
4. IA	0,556	0,551	0,553	0,556	0,549
5. UIB	0,639	0,609	0,593	0,574	0,560
6. SVD	0,669	0,639	0,624	0,603	0,585
7. SVD++	0,636	0,605	0,590	0,567	0,551
8. IKB	0,724	0,690	0,664	0,636	0,617
9. IKB-MS-H	0,758	0,719	0,689	0,665	0,641
10. IKB-MS-I	0,760	0,719	0,689	0,667	0,643
11. IKB-MS-J	0,763	0,719	0,686	0,663	0,640

Com a recomendação Top-N foi possível avaliar o quão eficaz estão as recomendações se o intuito for mostrar os N itens mais relevantes, estratégia comum nos *sites* de recomendação de itens.

Para ficar mais clara a forma como o SR se comportaria em um Top-N, uma ilustração é apresentada na Figura 32. Esta figura nada mais é do que o Top-N executado pelo método proposto, em que, o retângulo 1 (um) representa o treinamento realizado pelo sistema usando o MCAV de (HAREL; KOCH; PERONA, 2006), ou seja, é a etapa offline descrita no algoritmo 1. Já no retângulo 2 (dois), é necessário escolher um determinado usuário entre os 245 existentes para que se possa gerar a recomendação. A ordem da lista contendo os itens avaliados por esse usuário é opcional, e tem o intuito de mostrar quais itens forma melhor/pior avaliados por aquele usuário e posteriormente confrontá-los com os itens recomendados. Essa recomendação é realizada de forma online e instantânea após o botão de recomendação ser pressionado. Por fim o retângulo 3 (três) representa o Top-5, contendo os 5 itens que possivelmente o usuário 100 teria interesse.

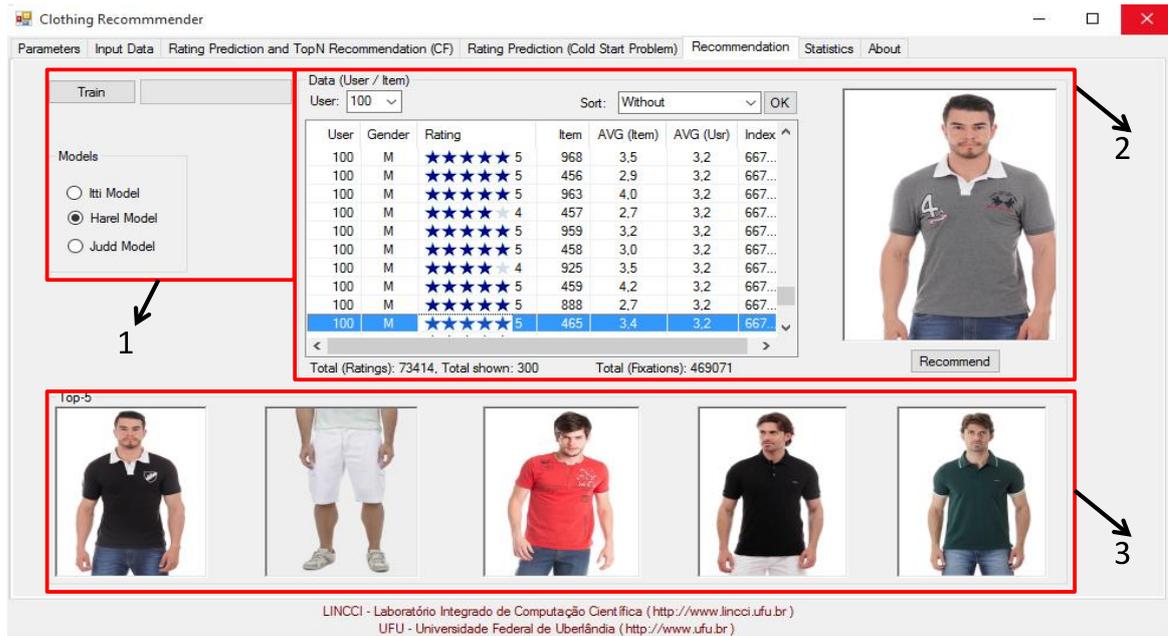


Figura 32 – Top-N executado pelo IKB-MS.

5.3.7 Teste Estatístico

A superioridade dos resultados alcançados pelo IKB-MS é confirmada estatisticamente com 95% de confiança, como mostrado na Tabela 10. A hipótese nula (H_0) é de que o IKB-MS tem eficácia menor do que os demais métodos comumente utilizados na área e a hipótese alternativa H_1 é de que IKB-MS é melhor do que os métodos comumente utilizados. Os resultados obtidos mostram que ao comparar o método IKB-MS contra os métodos da literatura, o p -valor é menor que 0.05 e portanto a probabilidade de H_0 ser verdadeira é menor do que 0.05. Com base nesse valor, é possível rejeitar H_0 e aceitar H_1 .

5.4 Considerações Finais

Nesse capítulo, experimentos foram realizados comparando a abordagem proposta contra outros algoritmos comumente utilizados na área. Esses experimentos mostraram a eficácia da nova medida de similaridade (descrita neste trabalho) no SR. Um destaque para o algoritmo IKB-MS é o tratamento de novos itens inseridos na base de dados e, com isso, proporcionando um ganho de 4,27% em relação aos algoritmos comumente utilizados na área. O próximo capítulo conclui esta dissertação revisando os objetivos que foram impostos a este trabalho, além de algumas considerações e propostas para trabalhos futuros.

Tabela 10 – p -valores da hipótese nula, que os sistemas são igualmente eficazes.

	1	2	3	4	5	6	7	8	9	10	11
1. Randon	-										
2. GA	**	-									
3. IA	**	**	-	.69							
4. UA	**	**	.35	-							
5. UIB	**	**	**	**	-	.73	.73				
6. SVD	**	**	**	**	.30	-	.26				
7. SVD++	**	**	**	**	.30	.77	-				
8. IKB	**	**	**	**	**	**	**	-			
9. IKB-MS-J	**	**	**	**	**	**	**	**	-		
10. IKB-MS-I	**	**	**	**	**	**	**	**	**	-	.44
11. IKB-MS-H	**	**	**	**	**	**	**	**	**	*	-

vazio: p -valor > 0.9 *: p -valor < 0.05 **: p -valor < 0.001

Conclusão

6.1 Conclusões

Neste trabalho, foi proposto um novo método para recomendação, denominado IKB-MS com o objetivo de melhorar a qualidade das recomendações de imagens. Para isso, informações sobre a atenção visual humana foram associadas ao método de recomendação baseado em FC, conhecido como Item KNN + Baseline (IKB). Ressalta-se que a atenção visual pode ser modelada por qualquer MCAV. Neste trabalho, o IKB-MS foi avaliado usando três MCAV diferentes, propostos por (ITTI; KOCH; NIEBUR, 1998) (HAREL; KOCH; PERONA, 2006) e (JUDD et al., 2009).

Os experimentos foram conduzidos usando uma base de dados de roupas construída para esse propósito e a eficácia do método foi medida em termos do RMSE e MAE, validações por meio de testes estatísticos também ajudaram a comprovar a hipótese de que o uso da atenção visual humana em SR melhora o resultado de predição.

A análise dos resultados mostrou que o IKB-MS, independente do MCAV utilizado, apresentou os melhores resultados, quando comparado com métodos competitivos. A proposta apresentada também contorna o problema de partida a frio de itens, uma vez que, embora um novo item não tenha avaliações de usuários mas, ainda assim o seu MS pode ser computado, permitindo que o item possa ser recomendado. A seguir, são apresentadas as principais contribuições atingidas com essa pesquisa.

6.2 Limitações encontradas

- Parametrização: A escolha do melhor valor para p , α e k pode variar conforme a base de dados utilizada, necessitando de diversos testes para se obter a melhor configuração.
- Particionamento: É extremamente dependente da aplicação e em varias delas o processo pode ter de ser feito de forma manual.

- ❑ Escalabilidade: Pode vir a ser uma limitação caso a base de dados seja muito grande, devido a necessidade de armazenar a matriz de utilidade na memória.

6.3 Principais Contribuições

Alicerçado pelos resultados dos experimentos realizados, as principais contribuições alcançadas no desenvolvimento do trabalho são:

- ❑ Integração das informações obtidas via MCAV ao SR baseado em FC utilizando a medida de similaridade proposta, que pondera a atenção visual e as avaliações dos usuários.
- ❑ Minimização do problema de partida a frio dos itens na filtragem colaborativa.

6.4 Trabalhos Futuros

Embora os resultados dos experimentos mostrem o bom desempenho do novo método desenvolvido, superando algoritmos da literatura, a pesquisa desenvolvida abre espaço para novos questionamentos e possíveis linhas para a continuação do trabalho, por exemplo:

- ❑ Incorporação de informações de conteúdo aos itens com o objetivo de melhorar ainda mais o problema da partida a frio de itens.
- ❑ Aumentar a base de roupas.
- ❑ Testar o IKB-MS com diferentes estratégias de particionamento.
- ❑ Testar em outras bases de dados.

6.5 Contribuições em Produção Bibliográfica

Desde o início desta pesquisa, contribuições e trabalhos foram gerados:

- ❑ Artigo completo intitulado “IKB-MS: Uma Abordagem Filtragem Colaborativa com Atenção Visual para Sistemas de Recomendação de Roupas” foi aceito no 21st Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), em Julho de 2015 o congresso acontecerá na cidade de Manaus, no Brasil. Em Outubro, o congresso tem uma qualificação B3. Esse artigo apresenta os resultados obtidos com a integração dos MCAV de (ITTI; KOCH; NIEBUR, 1998; HAREL; KOCH; PERONA, 2006; JUDD et al., 2009) ao IKB.

- ❑ Artigo completo intitulado “A Clothing Recommendation System Improved by Visual Attention” foi aceito no IEEE International Conference on Tools with Artificial Intelligence (ICTAI) em Julho de 2015. O congresso acontecerá na cidade de Vietri Sul Mare na Itália, apresentando um qualis A2.
- ❑ No início desta pesquisa, participamos de um artigo completo intitulado “Image Retrieval via Generalized I-Divergence in the Bag-of-Visual-Words Framework”, apresentado no 21st IEEE International Conference on Electronics Circuits and Systems (IEEE-ICECS-2014) em Dezembro de 2014 na cidade de Marseille na França. O congresso tem uma qualificação B1. Nesse artigo, foram apresentados os resultados variando a medida de similaridade para a recomendação de imagens e o MCAV utilizado foi de (ITTI; KOCH; NIEBUR, 1998).

Referências

- ADOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. **Knowledge and Data Engineering, IEEE Transactions on, IEEE**, v. 17, n. 6, p. 734–749, 2005.
- ANSARI, A.; ESSEGAIER, S.; KOHLI, R. Internet recommendation systems. **Journal of Marketing research**, American Marketing Association, v. 37, n. 3, p. 363–375, 2000.
- ARMSTRONG, J. S.; COLLOPY, F. Error measures for generalizing about forecasting methods: Empirical comparisons. **International journal of forecasting**, Elsevier, v. 8, n. 1, p. 69–80, 1992.
- ARYA, S. Nearest neighbor searching and applications. University of Maryland at College Park, 1996.
- AZEVEDO-MARQUES, P. M. Diagnóstico auxiliado por computador na radiologia. **Radiologia Brasileira**, SciELO Brasil, v. 34, 2001.
- BACON, W. F.; EGETH, H. E. Overriding stimulus-driven attentional capture. **Perception & psychophysics**, Springer, v. 55, n. 5, p. 485–496, 1994.
- BALABANOVIĆ, M.; SHOHAM, Y. Fab: content-based, collaborative recommendation. **Communications of the ACM**, ACM, v. 40, n. 3, p. 66–72, 1997.
- BALOIAN, N. et al. A model for a collaborative recommender system for multimedia learning material. In: **Groupware: Design, Implementation, and Use**. [S.l.]: Springer, 2004. p. 281–288.
- BILLSUS, D.; PAZZANI, M. J. Learning collaborative information filters. In: **ICML**. [S.l.: s.n.], 1998. v. 98, p. 46–54.
- BORJI, A.; AHMADABADI, M. N.; ARAABI, B. N. Cost-sensitive learning of top-down modulation for attentional control. **Machine Vision and Applications**, Springer, v. 22, n. 1, p. 61–76, 2011.
- BREESE, J. S.; HECKERMAN, D.; KADIE, C. Empirical analysis of predictive algorithms for collaborative filtering. In: MORGAN KAUFMANN PUBLISHERS INC. **Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence**. [S.l.], 1998. p. 43–52.

- BREGMAN, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. **USSR computational mathematics and mathematical physics**, Elsevier, v. 7, n. 3, p. 200–217, 1967.
- BURKE, R. Hybrid recommender systems: Survey and experiments. **User modeling and user-adapted interaction**, Springer, v. 12, n. 4, p. 331–370, 2002.
- BURT, P. J.; ADELSON, E. H. The laplacian pyramid as a compact image code. **Communications, IEEE Transactions on**, IEEE, v. 31, n. 4, p. 532–540, 1983.
- CAZELLA, S. C. **Aplicando a Relevância da Opinião de usuários em Sistema de Recomendação para Pesquisadores**. Tese (Doutorado) — Instituto de Informática da Universidade Federal do Rio Grande do Sul., 2006.
- COTTER, P.; SMYTH, B. Ptv: Intelligent personalised tv guides. In: **AAAI/IAAI**. [S.l.: s.n.], 2000. p. 957–964.
- COVER, T.; HART, P. Nearest neighbor classifiers. **IEEE Transactions on Computers**, p. 23–11, 1974.
- DAHLEN, B. J. et al. Jump-starting movielens: User benefits of starting a collaborative filtering system with "dead data". **University of Minnesota TR**, v. 98, p. 017, 1998.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **The Journal of Machine Learning Research**, JMLR. org, v. 7, p. 1–30, 2006.
- FELZENSZWALB, P.; MCALLESTER, D.; RAMANAN, D. A discriminatively trained, multiscale, deformable part model. In: IEEE. **Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on**. [S.l.], 2008. p. 1–8.
- FERNÁNDEZ, Y. B. **Propuesta metodológica para el razonamiento semántico en sistemas de recomendación personalizada y automática. Aplicación al caso de contenidos audiovisuales**. Tese (Doutorado) — Ph. D. thesis Universidade de Vigo, 2007.
- FERNÁNDEZ, Y. B. et al. Avatar: an improved solution for personalized tv based on semantic inference. **Consumer Electronics, IEEE Transactions on**, IEEE, v. 52, n. 1, p. 223–231, 2006.
- FISCHER, B.; WEBER, H. Express saccades and visual attention. **Behavioral and Brain Sciences**, Cambridge Univ Press, v. 16, n. 03, p. 553–567, 1993.
- GABOR, D. Theory of communication. part 1: The analysis of information. **Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering**, IET, v. 93, n. 26, p. 429–441, 1946.
- GANTNER, Z. et al. Mymedialite: A free recommender system library. In: **ACM. Proceedings of the fifth ACM conference on Recommender systems**. [S.l.], 2011. p. 305–308.
- GOLDBERG, D. et al. Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, ACM, v. 35, n. 12, p. 61–70, 1992.

- GOLDBERG, K. et al. Eigentaste: A constant time collaborative filtering algorithm. **Information Retrieval**, Springer, v. 4, n. 2, p. 133–151, 2001.
- GREENSPAN, H. et al. Overcomplete steerable pyramid filters and rotation invariance. In: IEEE. **Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on**. [S.l.], 1994. p. 222–228.
- GUNAWARDANA, A.; SHANI, G. A survey of accuracy evaluation metrics of recommendation tasks. **The Journal of Machine Learning Research**, JMLR. org, v. 10, p. 2935–2962, 2009.
- GUO, C.; ZHANG, L. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. **Image Processing, IEEE Transactions on**, IEEE, v. 19, n. 1, p. 185–198, 2010.
- HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**, University of Illinois at Urbana-Champaign. [S.l.]: Elsevier, 2006.
- HAREL, J.; KOCH, C.; PERONA, P. Graph-based visual saliency. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2006. p. 545–552.
- HERLOCKER, J. L.; KONSTAN, J. A.; RIEDL, J. Explaining collaborative filtering recommendations. In: ACM. **Proceedings of the 2000 ACM conference on Computer supported cooperative work**. [S.l.], 2000. p. 241–250.
- HERLOCKER, J. L. et al. Evaluating collaborative filtering recommender systems. **ACM Transactions on Information Systems (TOIS)**, ACM, v. 22, n. 1, p. 5–53, 2004.
- HILL, W. et al. Recommending and evaluating choices in a virtual community of use. In: ACM PRESS/ADDISON-WESLEY PUBLISHING CO. **Proceedings of the SIGCHI conference on Human factors in computing systems**. [S.l.], 1995. p. 194–201.
- HU, X.; ZHU, W.; LI, Q. Hers: A hybrid clothes recommender system based on user ratings and product features. **arXiv preprint arXiv:1411.6754**, 2014.
- ITTI, L. **Models of bottom-up and top-down visual attention**. Tese (Doutorado) — California Institute of Technology, 2000.
- ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE Computer Society, v. 20, n. 11, p. 1254–1259, 1998.
- JUDD, T. et al. Learning to predict where humans look. In: IEEE. **Computer Vision, 2009 IEEE 12th international conference on**. [S.l.], 2009. p. 2106–2113.
- KOCH, C. Shifts in selective visual attention: Towards the underlying neural circuitry. **Human Neurobiology**, v. 4, p. 219–227, 1985.
- KOREN, Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: ACM. **Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2008. p. 426–434.

- _____. Factor in the neighbors: Scalable and accurate collaborative filtering. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM, v. 4, n. 1, p. 1, 2010.
- KOREN, Y.; BELL, R. Advances in collaborative filtering. In: **Recommender systems handbook**. [S.l.]: Springer, 2011. p. 145–186.
- KRULWICH, B.; BURKEY, C. Learning user information interests through extraction of semantically significant phrases. In: AAAI PRESS MENLO PARK. **Proceedings of the AAAI spring symposium on machine learning in information access**. [S.l.], 1996. p. 100–112.
- LAAR, P. V. D.; HESKES, T.; GIELEN, S. Task-dependent learning of attention. **Neural networks**, Elsevier, v. 10, n. 6, p. 981–992, 1997.
- LANG, K. Newsweeder: Learning to filter netnews. In: **Proceedings of the 12th international conference on machine learning**. [S.l.: s.n.], 1995. p. 331–339.
- LATHAUWER, L. D.; MOOR, B. D.; VANDEWALLE, J. A multilinear singular value decomposition. **SIAM journal on Matrix Analysis and Applications**, SIAM, v. 21, n. 4, p. 1253–1278, 2000.
- LEMOS, F. D. et al. Towards a context-aware photo recommender system. In: **Context-Aware Recommender System Workshops**. [S.l.: s.n.], 2012.
- LINDEN, G.; SMITH, B.; YORK, J. Amazon. com recommendations: Item-to-item collaborative filtering. **Internet Computing, IEEE**, IEEE, v. 7, n. 1, p. 76–80, 2003.
- LIU, H.; YU, L. Toward integrating feature selection algorithms for classification and clustering. **Knowledge and Data Engineering, IEEE Transactions on**, IEEE, v. 17, n. 4, p. 491–502, 2005.
- LOPERA, M. C. L. et al. **Multi-resolution graph-based representation for analysis of large histo-pathological images**. Tese (Doutorado) — Universidad Nacional de Colombia, 2012.
- LUO, J.; SINGHAL, A. On measuring low-level saliency in photographic images. In: **IEEE. cvpr**. [S.l.], 2000. p. 1084.
- MA, Y.-F. et al. A generic framework of user attention model and its application in video summarization. **Multimedia, IEEE Transactions on**, IEEE, v. 7, n. 5, p. 907–919, 2005.
- MARLIN, B. M. Modeling user rating profiles for collaborative filtering. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2003. p. None.
- MCCALLUM, A. K. **Reinforcement learning with selective perception and hidden state**. Tese (Doutorado) — University of Rochester, 1996.
- MELO, E. V.; AMO, S. de; GULIATO, D. Cross-domain image matching improved by visual attention. **Journal of WSCG (Plzen, Print)**, Václav Skala-UNION Agency, v. 22, p. 65–72, 2014.

- MELVILLE, P.; SINDHWANI, V. Recommender systems. In: **Encyclopedia of machine learning**. [S.l.]: Springer, 2010. p. 829–838.
- MILANESE, R. **Detecting salient regions in an image: from biological evidence to computer implementation**. Tese (Doutorado) — University of Geneva, 1993.
- MONGELO, A. I. et al. Validação de método baseado em visão computacional para automação de contagem de viabilidade de leveduras em indústrias alcooleiras. **Outubro, Bento Gonçalves, RS. Citado na**, p. 3, 2011.
- NIEBUR, E.; KOCH, C. Control of selective visual attention: Modeling the "where" pathway. **Advances in neural information processing systems**, MORGAN KAUFMANN PUBLISHERS, p. 802–808, 1996.
- O'DONOVAN, J.; SMYTH, B. Trust in recommender systems. In: ACM. **Proceedings of the 10th international conference on Intelligent user interfaces**. [S.l.], 2005. p. 167–174.
- OLIVA, A.; TORRALBA, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. **International journal of computer vision**, Springer, v. 42, n. 3, p. 145–175, 2001.
- PAIVA, J. P.; YEPES, I. Uso de visão computacional para controle autônomo de câmera embarcada em veículo aéreo não tripulado. **Anais do Computer on the Beach**, p. 409–410, 2014.
- PRIVITERA, C. M.; STARK, L. W. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 22, n. 9, p. 970–982, 2000.
- RAJASHEKAR, U. et al. Gaffe: A gaze-attentive fixation finding engine. **Image Processing, IEEE Transactions on**, IEEE, v. 17, n. 4, p. 564–573, 2008.
- RAO, R. P. et al. Eye movements in iconic visual search. **Vision research**, Elsevier, v. 42, n. 11, p. 1447–1463, 2002.
- RESNICK, P. et al. Grouplens: an open architecture for collaborative filtering of netnews. In: ACM. **Proceedings of the 1994 ACM conference on Computer supported cooperative work**. [S.l.], 1994. p. 175–186.
- RUSSELL, B. C. et al. Labelme: a database and web-based tool for image annotation. **International journal of computer vision**, Springer, v. 77, n. 1-3, p. 157–173, 2008.
- SANCHEZ, F. et al. Social and content hybrid image recommender system for mobile social networks. **Mobile Networks and Applications**, Springer, v. 17, n. 6, p. 782–795, 2012.
- SARWAR, B. et al. Analysis of recommendation algorithms for e-commerce. In: ACM. **Proceedings of the 2nd ACM conference on Electronic commerce**. [S.l.], 2000. p. 158–167.
- _____. Item-based collaborative filtering recommendation algorithms. In: ACM. **Proceedings of the 10th international conference on World Wide Web**. [S.l.], 2001. p. 285–295.

- SCHAFER, J. B. et al. Collaborative filtering recommender systems. In: **The adaptive web**. [S.l.]: Springer, 2007. p. 291–324.
- SCHWARTZ, E. L.; GREVE, D. N.; BONMASSAR, G. Space-variant active vision: definition, overview and examples. **Neural Networks**, Elsevier, v. 8, n. 7, p. 1297–1308, 1995.
- SHANI, G.; GUNAWARDANA, A. Tutorial on application-oriented evaluation of recommendation systems. **AI Communications**, IOS Press, Nieuwe Hemweg 6 B Amsterdam 1013 BG Netherlands, v. 26, n. 2, p. 225–236, 2013.
- SHARDANAND, U.; MAES, P. Social information filtering: algorithms for automating “word of mouth”. In: ACM PRESS/ADDISON-WESLEY PUBLISHING CO. **Proceedings of the SIGCHI conference on Human factors in computing systems**. [S.l.], 1995. p. 210–217.
- SI, L.; JIN, R. Flexible mixture model for collaborative filtering. In: **ICML**. [S.l.: s.n.], 2003. v. 3, p. 704–711.
- SIKLOSSY, I. Mimicking the visual pathway. In: . [S.l.]: University Edinburgh - Division of Informatics, Edinburgh, 2005.
- SIMON, C. P.; BLUME, L.; DOERING, C. I. **Matemática para economistas**. [S.l.]: Bookman, 2004.
- SMEULDERS, A. W. et al. Content-based image retrieval at the end of the early years. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 22, n. 12, p. 1349–1380, 2000.
- SMYTH, B.; COTTER, P. Personalized electronic program guides for digital tv. **Ai Magazine**, v. 22, n. 2, p. 89, 2001.
- SU, X.; KHOSHGOFTAAR, T. M. A survey of collaborative filtering techniques. **Advances in artificial intelligence**, Hindawi Publishing Corp., v. 2009, p. 4, 2009.
- SZELISKI, R. **Computer vision: algorithms and applications**. [S.l.]: Springer Science & Business Media, 2010.
- TERVEEN, L. et al. Phoaks: A system for sharing recommendations. **Communications of the ACM**, ACM, v. 40, n. 3, p. 59–62, 1997.
- THEEUWES, J. Perceptual selectivity for color and form. **Perception & Psychophysics**, Springer, v. 51, n. 6, p. 599–606, 1992.
- TORRALBA, A. Modeling global scene factors in attention. **JOSA A**, Optical Society of America, v. 20, n. 7, p. 1407–1418, 2003.
- TORTORA, G. J.; DERRICKSON, B. **Corpo humano: fundamentos de anatomia e fisiologia**. [S.l.]: ArtMed, 2012.
- TSOTSOS, J. K. Analyzing vision at the complexity level. **Behavioral and brain sciences**, Cambridge Univ Press, v. 13, n. 03, p. 423–445, 1990.

- UHR, L. Layered "recognition cone" networks that preprocess, classify, and describe. **IEEE Transactions on Computers**, IEEE Computer Society, Los Alamitos, CA, USA, v. 21, n. 7, p. 758–768, 1972. ISSN 0018-9340.
- ULLMAN, S.; POGGIO, T. **Vision: A computational investigation into the human representation and processing of visual information**. [S.l.]: MIT Press, 2010.
- VIOLA, P.; JONES, M. J. Robust real-time face detection. **International journal of computer vision**, Springer, v. 57, n. 2, p. 137–154, 2004.
- VOZALIS, E.; MARGARITIS, K. G. Analysis of recommender systems algorithms. In: **The 6th Hellenic European Conference on Computer Mathematics & its Applications**. [S.l.: s.n.], 2003. p. 732–745.
- WANG, W. et al. Simulating human saccadic scanpaths on natural images. In: **IEEE. Computer vision and pattern recognition (cvpr), 2011 ieee conference on**. [S.l.], 2011. p. 441–448.
- WASFI, A. M. A. Collecting user access patterns for building user profiles and collaborative filtering. In: **ACM. Proceedings of the 4th international conference on Intelligent user interfaces**. [S.l.], 1998. p. 57–64.
- WIDISINGHE, A. et al. picseek: Collaborative filtering for context-based image recommendation. In: **IEEE. Information and Automation for Sustainability (ICIAFs), 2010 5th International Conference on**. [S.l.], 2010. p. 225–232.
- WITKIN, A. P. Scale-space filtering. In: **Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 2**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1983. (IJCAI'83), p. 1019–1022. Disponível em: <<http://dl.acm.org/citation.cfm?id=1623516.1623607>>.
- WOLFE, J. M.; HOROWITZ, T. S. What attributes guide the deployment of visual attention and how do they do it? **Nature Reviews Neuroscience**, Nature Publishing Group, v. 5, n. 6, p. 495–501, 2004.
- XIAO-SONG, H. et al. A probabilistic clothes recommender based on clothes features. In: **IEEE. Management Science & Engineering (ICMSE), 2014 International Conference on**. [S.l.], 2014. p. 76–81.
- ZHENG, N. et al. **Computer vision and pattern recognition**. Taylor & Francis, 2007.
- ZIMMERMAN, J. et al. Tv personalization system. In: **Personalized Digital Television**. [S.l.]: Springer, 2004. p. 27–51.