

---

# **Sistema de Recomendação de Imagens Baseado em Atenção Visual**

---

**Ernani Viriato de Melo**



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

Uberlândia  
2016



**Ernani Viriato de Melo**

**Sistema de Recomendação de Imagens Baseado  
em Atenção Visual**

Tese de doutorado 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 Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientadora: Denise Guliato

Coorientadora: Sandra de Amo

Uberlândia  
2016

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

---

M528s Melo, Ernani Viriato de, 1981-  
2016 Sistema de recomendação de imagens baseado em atenção  
visual / Ernani Viriato de Melo. - 2016.  
140 f. : il.

Orientadora: Denise Guliato.

Coorientadora: Sandra de Amo.

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

Inclui bibliografia.

1. Computação - Teses.
2. Sistemas de recuperação da  
informação - Teses.
3. Imagens como recursos de informação - Teses.
4. Comércio eletrônico - Teses. I. Guliato, Denise. II. Amo, Sandra de.
- III. Universidade Federal de Uberlândia, Programa de Pós-Graduação em  
Ciência da Computação. IV. Título.

---

CDU: 681.3

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

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da tese intitulada “**Sistema de Recomendação de Imagens Baseado em Atenção Visual**” por **Ernani Viriato de Melo** como parte dos requisitos exigidos para a obtenção do título de **Doutor em Ciência da Computação**.

Uberlândia, 17 de agosto de 2016

Orientadora: Denise  
Profº. Dra. Denise Guliato  
Universidade Federal de Uberlândia

Banca Examinadora:

André Backes  
Prof. Dr. André Ricardo Backes  
Universidade Federal de Uberlândia (UFU)

Maria Camila Nardini Barioni  
Profº. Dra. Maria Camila Nardini Barioni  
Universidade Federal de Uberlândia (UFU)

Altigran Soares da Silva  
Prof. Dr. Altigran Soares da Silva  
Universidade Federal do Amazonas (UFAM)

Paulo Sérgio Silva Rodrigues  
Prof. Dr. Paulo Sérgio Silva Rodrigues  
Fundação Educacional Inaciana Padre Sabóia de Medeiros (FEI)



*Este trabalho é dedicado a minha esposa Marilene e aos meus filhos  
Laura e Otávio.*



---

## Agradecimentos

Inicialmente, gostaria de agradecer a Deus por me guiar, iluminar e propiciar uma vida repleta de conquistas e realizações.

Agradeço às minhas irmãs e principalmente os meus pais, que sempre me motivaram e mostraram a importância do estudo e da ética.

À minha esposa Marilene que cuida da nossa família com muito carinho. Sempre me apoiou e soube lidar com minhas faltas.

Aos meus amigos de laboratório da UFU, Emília, Bruno, Crícia, Daniela e Valquíria pelas conversas, diversão e compartilhamento de conhecimento. Também quero agradecer o meu amigo Henrique que me ajudou muito na viagem à Itália para apresentar este trabalho.

Aos alunos de iniciação científica, Breno, Bruno e Caroline que foram muito importantes na coleta de dados.

Ao secretário da Pós-Graduação, Erisvaldo. Ele me ajudou muito e me deu todo suporte quando eu não podia estar em Uberlândia.

À Pós-Graduação e ao CNPq, que possibilitaram a compra do dispositivo que coleta fixações de olhar e também ajudaram financeiramente nas duas viagens para apresentação do nosso trabalho em conferências internacionais.

Ao Instituto Federal do Triângulo Mineiro - Campus Uberaba que me proporcionou dedicação exclusiva no doutorado por dois anos e horário especial de trabalho nos outros anos restantes.

Agradeço imensamente a professora Dra. Denise Guliato. Trabalhamos juntos por 7 anos, somando mestrado e doutorado, e resumí-la a minha orientadora é muito pouco. A Denise foi e é muito importante para mim não só na condução deste trabalho, mas também como conselheira e amiga.

Agradeço a coorientação da professora Dra. Sandra de Amo. Eu pude contar com o apoio da Sandra nos dois anos iniciais do projeto e este apoio foi fundamental para o desenvolvimento desta tese.



*“As pessoas não sabem o que querem, até mostrarmos a elas.”*  
*(Steve Jobs)*



---

## Resumo

Hoje em dia, a quantidade de usuários que utilizam sites de comércio eletrônico para realizar compras está aumentando muito, principalmente devido à facilidade e rapidez. Muitos sites de comércio eletrônico, diferentemente das lojas físicas, disponibilizam aos seus usuários uma grande variedade de produtos e serviços, e os usuários podem ter muita dificuldade em encontrar produtos de sua preferência. Normalmente, a preferência por um produto pode ser influenciada pela aparência visual da imagem do produto. Neste contexto, os Sistemas de Recomendação de produtos que estão associados a Imagens (SRI) tornaram-se indispensáveis para ajudar os usuários a encontrar produtos que podem ser, eventualmente, agradáveis ou úteis para eles. Geralmente, os SRI usam o comportamento passado dos usuários (cliques, compras, críticas, avaliações, etc.) e/ou atributos de produtos para definirem as preferências dos usuários. Um dos principais desafios enfrentados em SRI é a necessidade de o usuário fornecer algumas informações sobre suas preferências sobre os produtos, a fim de obter novas recomendações do sistema. Infelizmente, os usuários nem sempre estão dispostos a fornecer tais informações de forma explícita. Assim, a fim de lidar com esse desafio, os métodos para obtenção de informações de forma implícita do usuário são desejáveis. Neste trabalho, propõe-se investigar em que medida informações sobre atenção visual do usuário podem ajudar a melhorar a predição de avaliação e consequentemente produzir SRI mais precisos. É também objetivo deste trabalho o desenvolvimento de dois novos métodos, um método baseado em Filtragem Colaborativa (FC) que combina avaliações e dados de atenção visual para representar o comportamento passado dos usuários, e outro método baseado no conteúdo dos itens, que combina atributos textuais, características visuais e dados de atenção visual para compor o perfil dos itens. Os métodos propostos foram avaliados em uma base de imagens de pinturas e uma base de imagens de roupas. Os resultados experimentais mostram que os métodos propostos neste trabalho possuem ganhos significativos em predição de avaliação e precisão na recomendação quando comparados ao estado-da-arte. Vale ressaltar que as técnicas propostas são flexíveis, podendo ser utilizadas em outros cenários que exploram a atenção visual dos itens recomendados.

**Palavras-chave:** Sistemas de Recomendação. Filtragem Colaborativa. Similaridade entre Imagens. Atenção Visual.

---

## Abstract

Nowadays, the amount of users using e-commerce sites for shopping is greatly increasing, mainly due to the easiness and rapidity of this way of consumption. Many e-commerce sites, differently from physical stores, can offer their users a wide range of products and services, and the users can find it very difficult to find products of your preference. Typically, your preference for a product can be influenced by the visual appearance of the product image. In this context, *Image Recommendation Systems* (IRS) have become indispensable to help users to find products that may possibly pleasant or be useful to them. Generally, IRS use past behavior of users (clicks, purchases, reviews, ratings, etc.) and/or attributes of the products to define the preferences of users. One of the main challenges faced by IRS is the need of the user to provide some information about his / her preferences on products in order to get further recommendations from the system. Unfortunately, users are not always willing to provide such information explicitly. So, in order to cope with this challenge, methods for obtaining user's implicit feedback are desirable. In this work, the author propose an investigation to discover to which extent information concerning user visual attention can help improve the rating prediction hence produce more accurate IRS. This work proposes to develop two new methods, a method based on *Collaborative Filtering* (CF) which combines ratings and data visual attention to represent the past behavior of users, and another method based on the content of the items, which combines textual attributes, visual features and visual attention data to compose the profile of the items. The proposed methods were evaluated in a painting dataset and a clothing dataset. The experimental results show significant improvements in rating prediction and precision in recommendation when compared to the state-of-the-art. It is worth mentioning that the proposed techniques are flexible and can be applied in other scenarios that exploits the visual attention of the recommended items.

**Keywords:** Recommendation systems. Collaborative Filtering. Image Similarity. Visual

Attention.

---

## List of illustrations

Figura 1	Exemplo de captura de atenção visual. Cada círculo azul representa uma fixação. O tamanho do círculo representa a duração e o número representa a sequência da fixação. As linhas entre os círculos representam os movimentos dos olhos (sacadas). . . . .	32
Figura 2	Esquema geral da nossa proposta: os usuários avaliam e olham para os produtos; treina-se um modelo de recomendação usando as avaliações, fixações, e movimentos dos olhos; o sistema de recomendação utiliza o modelo para recomendar produtos de acordo com o usuário. . . . .	32
Figura 3	Matriz de utilidade. . . . .	37
Figura 4	Perspectivas das funções de similaridade entre dois usuários $u$ e $v$ . (a) compara o padrão das avaliações dos usuários. (b) calcula o ângulo entre os vetores de avaliações representando os usuários $u$ e $v$ . (c) calcula a distância entre os usuários $u$ e $v$ . Para melhor ilustrar (b) e (c), considerou-se apenas dois itens $i$ e $j$ . . . . .	41
Figura 5	<i>Singular Value Decomposition</i> (SVD). . . . .	45
Figura 6	Visão geral do cálculo do descritor HOG (DALAL; TRIGGS, 2005). . . .	53
Figura 7	Exemplo de um descritor GIST com 512 dimensões. . . . .	54
Figura 8	Processo de extração do descritor LBP: (a) imagem original; (b) imagem em níveis de cinza; (c) cálculo do LBP; (d) imagem gerada após a atualização dos pixels. . . . .	56
Figura 9	Utilizando o descritor de presença de pele proposto em (JAIN, 2011): (a) imagem original; (b) imagem binária onde a cor branca representa a presença de pele. . . . .	57

Figura 10	Rastreadores de olhar montados no desktop. (a) o usuário encaixa a cabeça no rastreador de olhar e não permite movimentos ( <a href="http://www.smivision.com">www.smivision.com</a> ). (b) o rastreador de olhar é um pequeno dispositivo colocado no monitor e o usuário possui uma liberdade limitada de movimentos da cabeça ( <a href="http://www.tobii.com">www.tobii.com</a> ). . . . .	60
Figura 11	Rastreadores de olhar montados na cabeça ( <a href="http://www.tobii.com">www.tobii.com</a> ). (a) pesquisas relacionadas com padrões de compras, embalagens de produtos e disposição dos produtos na loja. (b) comportamento do usuário ao volante. (c) pesquisas nas áreas de esporte e psicologia. . . . .	60
Figura 12	Mapas de Saliência utilizando diferentes modelos computacionais. (a) imagem original; (b)(ITTI; KOCH; NIEBUR, 1998); (c)(ACHANTA et al., 2009); (d)(BRUCE; TSOTSOS, 2005); (e)(GARCIA-DIAZ et al., 2009); (f)(HAREL; KOCH; PERONA, 2006); (g)(JUDD et al., 2009); (h)(MURRAY et al., 2011); (i)(ZHANG; SHEN; FENG, 2008). . . . .	61
Figura 13	Arquitetura do método proposto CFAS. Os retângulos vermelhos são as principais contribuições deste trabalho. . . . .	74
Figura 14	As imagens são segmentadas e rotuladas de acordo com o conjunto $\mathbb{H}$ . (a) nesse exemplo a segmentação é realizada usando uma grade que divide a pessoa em partes do corpo humano e cada parte é rotulada com o respectivo conceito semântico. (b) nesse segundo exemplo a segmentação é obtida dividindo a imagem inteira em uma grade regular e cada célula da grade é rotulada com um conceito semântico relacionado à pintura (1 representa céu, 2 representa oceano e 3 representa um barco). . . . .	75
Figura 15	Representação dos dados da atenção visual de uma imagem de roupa visualizada por dois usuários. (a) imagem segmentada. (b) dois usuários (verde e vermelho) observam a imagem. Cada cor representa um usuário e cada círculo representa uma fixação de olhar e o raio do círculo representa a duração da fixação e o número dentro do círculo representa a sequência das fixações. As linhas entre os círculos representam os movimentos dos olhos (sacadas). (c) representa os dados da atenção visual. O usuário verde percorreu 704 pixels na imagem durante 3,2 segundos e ficou 21% do tempo olhando para o ombro direito, 19% do tempo olhando para o pescoço e 11% do tempo olhando para o pé esquerdo. O segundo usuário (vermelho) percorreu 580 pixels na imagem durante 2,9 segundos e ficou 63% do tempo olhando para o pé esquerdo. . . . .	76

Figura 16	Representação dos dados da atenção visual de uma imagem de pintura visualizada por dois usuários. (a) imagem segmentada. (b) dois usuários (verde e vermelho) observam a imagem. (c) representa os dados da atenção visual. O usuário verde percorreu 362 pixels na imagem durante 3 segundos e olhou 35% do tempo para o céu e 65% do tempo para o barco. O usuário vermelho percorreu 394 pixels na imagem durante 2,4 segundos e olhou 82% do tempo para o céu e 18% do tempo para o oceano. . . . .	76
Figura 17	Arquitetura do método proposto CBAS. Os retângulos vermelhos são as principais contribuições deste trabalho. . . . .	80
Figura 18	Vetor de atributos textuais $T_i$ . . . . .	81
Figura 19	Exemplos de imagens da base de dados UFU-CLOTHING. (a) roupas masculinas e (b) roupas femininas. . . . .	84
Figura 20	O usuário sentou a uma distância de aproximadamente 50 centímetros de uma tela de computador de 22 polegadas. O dispositivo de rastreamento do olhar foi fixado na parte inferior do monitor e a avaliação foi coletada através do teclado. . . . .	85
Figura 21	Partes semânticas do corpo humano. (a) 22 partes de corpo inteiro, (b) 12 partes superiores e (c) 10 partes inferiores. . . . .	86
Figura 22	Exemplos de imagens da base de dados UFU-PAINTINGS. . . . .	88
Figura 23	Segmentação de pinturas: (a) imagem original; (b) Software de segmentação de pinturas. A imagem inteira é dividida em 20 x 20 partes e cada parte é rotulada com um significado semântico. . . . .	89
Figura 24	Gráficos para avaliar o melhor número de vizinhos para os métodos UKB, IKB e CFAS aplicados na <b>UFU-CLOTHING</b> , em termos de RMSE. Medidas de similaridade utilizadas: (a) <i>Pearson Correlation Coefficient</i> (PCC); (b) Cosseno e (c) inversão da distância Euclidiana (Euc). . . . .	98
Figura 25	Gráficos para avaliar o melhor número de vizinhos para os métodos UKB, IKB e CFAS aplicados na <b>UFU-PAINTINGS</b> , em termos de RMSE. Medidas de similaridade utilizadas: (a) <i>Pearson Correlation Coefficient</i> (PCC); (b) Cosseno e (c) inversão da distância Euclidiana (Euc). . . . .	99
Figura 26	Avaliação do número de fatores latentes para cada método comparativo aplicado na UFU-CLOTHING. . . . .	100
Figura 27	Avaliação do número de fatores latentes para cada método comparativo aplicado na UFU-PAINTINGS. . . . .	101

Figura 28	Avaliação do melhor número de vizinhos para os métodos IKT, IKV e CBAS aplicados na UFU-CLOTHING, em termos de RMSE. . . . .	106
Figura 29	Avaliação do melhor número de vizinhos para os métodos IKT, IKV e CBAS aplicados na UFU-PAINTINGS, em termos de RMSE. . . . .	107
Figura 30	Comparação dos métodos CFAS (PCC), CFAS (SVD) e CBAS em termos de RMSE e tamanho da base de fixações: (a) UFU-CLOTHING e (b) UFU-PAINTINGS. . . . .	109
Figura 31	Diferentes divisões de partes do corpo humano. (a) 22 partes; (b) 11 partes; (c) 4 partes; (d) 2 partes. . . . .	110
Figura 32	Pintura rotulada com diferentes conjuntos de rótulos semânticos ( $\mathbb{H}$ ). (a) o conjunto $\mathbb{H}$ possui 41 rótulos de acordo com o Apêndice A. No exemplo, 10 = céu, 13 = árvore, 15 = montanha/pedras, 24 = cavalo, e 29 = leão; (b) o conjunto $\mathbb{H}$ possui 9 rótulos definidos de acordo com a categoria da pintura. No exemplo, 1 = animal e 7 = paisagem. . . .	111
Figura 33	RMSE de acordo com o número de rótulos semânticos ( $ \mathbb{H} $ ). (a) UFU-CLOTHING; (b) UFU-PAINTINGS . . . . .	111
Figura 34	Sistema de Recomendação de Roupas. Tela de parâmetros. . . . .	134
Figura 35	Sistema de Recomendação de Roupas. Tela de entrada de dados. . . .	135
Figura 36	Sistema de Recomendação de Roupas. Execução dos processos de previsão e recomendação. . . . .	136
Figura 37	Sistema de Recomendação de Roupas. Cenário de partida a frio. . . .	137
Figura 38	Sistema de Recomendação de Roupas. Informações sobre os desenvolvedores. . . . .	138
Figura 39	Sistema de Recomendação de Pinturas. Tela de entrada de dados. . . .	140

# **Lista de tabelas**

Tabela 17	Comparação de desempenho em termos de RMSE dos métodos baseados em conteúdo aplicados nas bases UFU-CLOTHING e UFU-PAINTINGS. . . . .	107
Tabela 18	<i>p-values</i> obtidos aos pares pelo <i>sign test</i> nas bases UFU-CLOTHING e UFU-PAINTINGS. . . . .	107
Tabela 19	Avaliação AP@5 e AUC do processo de recomendação dos métodos baseados em conteúdo aplicados na UFU-CLOTHING e UFU-PAINTINGS.	108
Tabela 20	Ganho na AP e AUC do método CBAS quando comparado aos demais métodos baseados em conteúdo. . . . .	108
Tabela 21	Conjunto $\mathbb{H}$ com os 41 rótulos de possíveis significados semânticos para pintura. . . . .	131

---

## **Lista de siglas**

**AEI** Algoritmo Evolutivo Interativo

**AP** *Average Precision*

**AUC** *Area Under the ROC Curve*

**BC** Baseada em Conteúdo

**CBAS** *Content-Based recommender systems with Attentive Similarity*

**CF** *Collaborative Filtering*

**CFAS** *Collaborative Filtering recommender systems with Attentive Similarity*

**FC** Filtragem Colaborativa

**HOG** *Histogram of Gradients*

**IAPS** *Internation Affective Picture System*

**IKB** *Item KNN + Baseline*

**IKSB** *Item KNN + SVD + Baseline*

**IKT** *Item KNN + Textual attributes*

**IKV** *Item KNN + Visual features*

**IRS** *Image Recommendation Systems*

**KNN** *K-Nearest Neighbors*

**KTF** *Korea Telecom Freetel*

**LBP** *Local Binary Pattern histogram*

**NB** *Naive Bayes*

**NDCG** *Normalized Discounted Cumulative Gain*

**NDPM** *Distance-based Performance Measure*

**PCC** *Pearson Correlation Coefficient*

**RMSE** *Root Mean Squared Error*

**RO** Rastreamento do Olhar

**ROI** *Regions of Interest*

**SB** *SVD + Baseline*

**SR** Sistema de Recomendação

**SRBC** Sistema de Recomendação baseado em Conteúdo

**SRI** Sistema de Recomendação de produtos que estão associados a Imagens

**SRP** Sistema de Recomendação de Pinturas

**SRR** Sistema de Recomendação de Roupas

**SVD** *Singular Value Decomposition*

**UKB** *User KNN + Baseline*

**UIB** *UserItemBaseline*

**UFU** Universidade Federal de Uberlândia

---

## List of symbols

- $\mathbb{U}$ : Conjunto de usuários da base de dados do Sistema de Recomendação  
 $\mathbb{I}$ : Conjunto de itens da base de dados do Sistema de Recomendação  
 $\mathbb{R} = \{r_{ui}\}$ : Matriz de utilidade  
 $\hat{r}_{ui}$ : Predição da avaliação do usuário  $u$  sobre o item  $i$   
 $\mathbb{I}(u)$ : Conjunto de itens avaliados pelo usuário  $u$   
 $\mathbb{I}(u, v)$ : Conjunto de itens avaliados pelos usuários  $u$  e  $v$   
 $\mathbb{U}(i)$ : Conjunto de usuários que avaliaram o item  $i$   
 $\mathbb{U}(i, j)$ : Conjunto de usuários que avaliaram os itens  $i$  e  $j$   
 $b_{ui}$ : Predição *baseline* de um usuário  $u$  para um item  $i$   
 $\mu$ : Média global de avaliações  
 $\bar{r}_u$ : Média de avaliações do usuário  $u$   
 $\mathbb{U}(u, i, k)$ :  $k$  usuários vizinhos mais próximos do usuário  $u$  e que avaliaram o item  $i$   
 $\mathbb{I}(u, i, k)$ :  $k$  itens vizinhos mais próximos do item  $i$  que foram avaliados pelo usuário  $u$   
 $\lambda_s$ : Parâmetro de encolhimento (*Shrinkage*)  
 $q_i$ : Vetor de fatores latentes do item  $i$   
 $\mathbb{H}$ : Conjunto de rótulos que se associam a conceitos semânticos  
 $\theta_i$ : Número de usuários que olharam a imagem  $i$   
 $\ell_i$ : Soma total das sacadas de todos os usuários que olharam para a imagem  $i$   
 $\gamma_i$ : Soma da duração, em segundos, de todas as fixações sobre a imagem  $i$   
 $V_i$ : Vetor de atentividade da imagem  $i$  de dimensão  $|\mathbb{H}|$   
 $\mathbb{G}(i)$ : Conjunto de usuários que olharam para a imagem  $i$   
 $\mathbb{G}(u, i)$ : Conjunto de todas as fixações do olhar do usuário  $u$  sobre a imagem  $i$   
 $\mathbb{G}(u, i, t)$ : Conjunto de todas as fixações do olhar do usuário  $u$  sobre o rótulo semântico  $t$  da imagem  $i$   
 $\mathbb{M}(u, i)$ : Conjunto de todas as sacadas do usuário  $u$  sobre a imagem  $i$   
 $\mathbb{S} = \{S_{ij}\}$ : Matriz de similaridades  
 $RS_{ij}$ : Similaridade baseada em avaliações entre dois itens  $i$  e  $j$   
 $AS_{ij}$ : Similaridade atentiva entre dois itens  $i$  e  $j$

$\sigma$ : Peso atribuído à similaridade entre dois vetores de atentividade no cálculo da  $AS_{ij}$

$\lambda_{as}$  Parâmetro de encolhimento da  $AS_{ij}$

$\beta$ : Peso atribuído à similaridade baseada em avaliações no cálculo de  $S_{ij}$  no método CFAS

$\mathbb{A}$ : Conjunto de atributos textuais

$T_i$ : Vetor de atributos textuais da imagem  $i$

$\mathbb{F}$ : Conjunto de vetores de características visuais

$TS_{ij}$ : Similaridade textual entre dois itens  $i$  e  $j$

$FS_{ij}$ : Similaridade visual entre dois itens  $i$  e  $j$

---

# Sumário

<b>1</b>	<b>Introdução</b>	<b>29</b>
1.1	Contextualização . . . . .	29
1.2	Motivação . . . . .	31
1.3	Hipótese e Objetivo . . . . .	32
1.4	Organização da Tese . . . . .	33
<b>2</b>	<b>Fundamentação Teórica</b>	<b>35</b>
2.1	Introdução . . . . .	35
2.2	Sistemas de Recomendação . . . . .	35
2.2.1	Conceitos e Notações . . . . .	36
2.2.2	A Técnica de Filtragem Colaborativa . . . . .	37
2.2.3	A Técnica Baseada em Conteúdo . . . . .	45
2.2.4	Técnica Híbrida . . . . .	48
2.2.5	Problemas em Sistemas de Recomendação . . . . .	48
2.2.6	Métricas para Avaliação de Sistemas de Recomendação . . . . .	50
2.3	Extração de Características Visuais . . . . .	52
2.3.1	Histograma de Gradientes Orientados (HOG) . . . . .	52
2.3.2	GIST . . . . .	54
2.3.3	Histograma de Cor . . . . .	54
2.3.4	Padrões Binários Locais (LBP) . . . . .	55
2.3.5	Descriptor de Pele ( <i>Skin</i> ) . . . . .	56
2.4	Atenção Visual Humana . . . . .	57
2.4.1	Tipos de Rastreadores de Olhar . . . . .	59
2.4.2	Modelos Computacionais de Predição de Atenção Visual . . . . .	61
2.5	Considerações Finais . . . . .	62
<b>3</b>	<b>Trabalhos Correlatos</b>	<b>63</b>
3.1	Sistemas de Recomendação que Combinam Similaridades . . . . .	63

3.2	Sistemas de Recomendação de Imagens . . . . .	66
3.3	Sistemas de Recomendação que Utilizam Rastreadores de Olhar . . . . .	68
3.4	Considerações Finais . . . . .	71
<b>4</b>	<b>Métodos Propostos</b>	<b>73</b>
4.1	Abordagem Baseada em Filtragem Colaborativa (CFAS) . . . . .	73
4.1.1	Processo de Segmentação e Classificação . . . . .	73
4.1.2	Gerenciamento da Atenção Visual e Avaliações . . . . .	74
4.1.3	Processo de Predição . . . . .	77
4.1.4	Processo de Recomendação . . . . .	79
4.2	Abordagem Baseada em Conteúdo (CBAS) . . . . .	79
4.2.1	Gerenciamento da Atenção Visual, Atributos dos Itens e Avaliações	79
4.2.2	Processo de Predição . . . . .	81
4.3	Considerações Finais . . . . .	82
<b>5</b>	<b>Metodologia para Validação dos Métodos Propostos</b>	<b>83</b>
5.1	Base de Dados . . . . .	83
5.1.1	Base UFU-CLOTHING . . . . .	83
5.1.2	Base UFU-PAINTINGS . . . . .	87
5.2	Comparação com Outros Trabalhos . . . . .	90
5.3	Parâmetros de Configuração . . . . .	91
5.4	Avaliação da Significância Estatística . . . . .	93
5.5	Considerações Finais . . . . .	95
<b>6</b>	<b>Experimentos e Análise dos Resultados</b>	<b>97</b>
6.1	Resultados Comparativos entre os Métodos baseados em Filtragem Colaborativa . . . . .	97
6.1.1	Predição de Avaliação . . . . .	97
6.1.2	Recomendação Top-N . . . . .	103
6.1.3	CFAS e o Problema de Partida a Frio . . . . .	103
6.2	Resultados Comparativos entre os Métodos baseados em Conteúdo . . . . .	105
6.2.1	Predição de Avaliação . . . . .	105
6.2.2	Recomendação Top-N . . . . .	108
6.3	Análise da Atenção Visual . . . . .	108
6.3.1	Impacto dos Dados de Atenção Visual nos Métodos Propostos . . . . .	108
6.4	Considerações Finais . . . . .	112
<b>7</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>113</b>
7.1	Principais Contribuições . . . . .	114
7.2	Limitações . . . . .	114
7.3	Trabalhos Futuros . . . . .	115

7.4 Contribuições em Produção Bibliográfica . . . . .	116
<b>Referências</b>	<b>117</b>
<b>Apêndices</b>	<b>129</b>
<b>APÊNDICE A Rótulos com Significados Semânticos para Pintura</b>	<b>131</b>
<b>APÊNDICE B Sistema de Recomendação de Roupas com os Métodos CFAS e CBAS</b>	<b>133</b>
<b>APÊNDICE C Sistema de Recomendação de Pinturas com os Métodos CFAS e CBAS</b>	<b>139</b>



# CAPÍTULO 1

---

## Introdução

### 1.1 Contextualização

A grande quantidade de dados disponíveis em diversos websites, tais como comércio eletrônico, redes sociais, e bibliotecas digitais tem sido uma das causas para que o usuário tenha dificuldades em encontrar conteúdo relevante para ele. Muitas vezes, a busca por um dado nestes websites, tais como vídeos, notícias, imagens, contato de pessoas, produtos de compras, dentre outros é feita por palavras-chave. Por exemplo, em um site de compras de roupas online, os clientes encontram um produto favorito, escrevendo algumas palavras-chave, tais como “short, vestido”. Essa busca pode ser tediosa e frustrante se o site não tem um Sistema de Recomendação (SR) eficiente. Os SR ajudam as pessoas a encontrarem produtos e/ou serviços que correspondam às suas preferências e dão suporte para as pessoas tomarem decisões em vários contextos, tais como: quais produtos comprar (ZHANG; WANG, 2015), qual filme assistir (QIN; MENEZES; SILAGHI, 2010), que música escutar (WANG et al., 2014), que pintura observar (ALBANESE et al., 2013). Neste trabalho, o interesse está em Sistema de Recomendação de produtos que estão associados a Imagens (SRI).

Formalmente, os SR representam o comportamento passado dos usuários (compra, avaliação ou crítica) via uma matriz de utilidade, onde as linhas representam os usuários, as colunas representam os produtos (itens) e a célula  $(u,i)$  contém a avaliação dada pelo usuário  $u$  sobre o item  $i$  (normalmente um número inteiro de 1 a 5 representando estrelas), indicando o interesse do usuário  $u$  pelo item  $i$ . Com isso, o problema de recomendação pode ser interpretado como um problema de estimar as avaliações de um conjunto de itens que ainda não foram avaliados por um dado usuário, e os itens com mais alta avaliação predita são recomendados para este usuário. Em geral, independentemente do tipo de dado que será recomendado (vídeo, imagem, texto, ou áudio), existem três técnicas para o desenvolvimento de um SR (ADOMAVICIUS; TUZHILIN, 2005).

A primeira delas é a técnica baseada em Filtragem Colaborativa (FC) (GE et al., 2011; LIU et al., 2010b; LIU et al., 2010a) que utiliza apenas a matriz de utilidade para representar

as preferências dos usuários e treinar um modelo de recomendação. A técnica FC é muito utilizada, especialmente em grandes sistemas comerciais bem conhecidos, tais como Amazon<sup>1</sup> para a recomendação de produtos de compras, e Netflix<sup>2</sup> para a recomendação de filmes. A principal razão é que essa técnica é baseada em uma estratégia simples e eficaz: o interesse do usuário para um item é calculado considerando as avaliações de outros usuários ou produtos semelhantes.

A segunda técnica para o desenvolvimento de SR é Baseada em Conteúdo (BC) (LOPS; GEMMIS; SEMERARO, 2011). Nessa técnica um conjunto de características (atributos) dos itens previamente avaliados por um usuário é analisado, e a partir dessa análise é construído um modelo ou perfil representando os interesses do usuário. A recomendação de um item consiste na combinação dos atributos do modelo/perfil do usuário com os atributos do item.

Finalmente, a terceira técnica consiste na combinação de diferentes técnicas baseadas em FC e BC produzindo uma solução híbrida (LU et al., 2015). A ideia principal é focar nas vantagens de cada técnica para obter um SR mais preciso.

As três técnicas de desenvolvimento de SR possuem como meta recomendar produtos úteis que irão surpreender e agradar o usuário. Para qualquer tipo de produto/serviço que será recomendado, tais como vídeo, música, roupa, filme, pintura, etc., os SR enfrentam alguns desafios em comum. Como as recomendações são extraídas a partir de dados já existentes, um primeiro desafio é a disponibilidade desses dados. Quanto mais dados relevantes, mais precisa é a recomendação. Em (ADOMAVICIUS; TUZHILIN, 2005) os autores afirmam que em qualquer sistema de recomendação o número de avaliações já obtido é geralmente muito pequeno quando comparado ao número de avaliações necessárias para realizar uma predição de avaliação com boa acurácia. Porém, ao mesmo tempo que é necessário uma grande quantidade de dados para boas recomendações, o desempenho de qualquer SR pode ser comprometido por essa quantidade de dados, isto é, o SR precisa ser escalável em relação ao aumento dos dados com o tempo (SU; KHOSHGOFTAAR, 2009). Outros desafios específicos bastante discutidos na literatura são (LIKA; KOLOMVATSOS; HADJIEFTHYMIADES, 2014): partida a frio do usuário, do inglês *user cold-start* (como recomendar um produto para um usuário que acabou de se cadastrar no sistema?) (BOBADILLA et al., 2012; WONGCHOKPRASITI et al., 2015); e partida a frio de item, do inglês *item cold-start* (como recomendar um item que acabou de ser incluído no banco de dados?) (SAVESKI; MANTRACH, 2014).

Ainda há espaço considerável para melhorar a qualidade da recomendação de produtos, especialmente aqueles em que o aspecto visual do produto é importante para a definição da opinião do usuário. Nesta tese, o foco está no desafio de aumentar e melhorar os dados necessários para uma boa acurácia em SRI sem nenhum esforço adicional por parte do

<sup>1</sup><http://www.amazon.com>.

<sup>2</sup><http://www.netflix.com>

usuário. Este trabalho propõe duas abordagens, uma baseada em FC e outra BC.

## 1.2 Motivação

As imagens são importantes para influenciar as escolhas dos usuários sobre os produtos recomendados. Nota-se que muitos produtos, tais como sapatos, roupas ou pinturas são adquiridos pelos usuários por causa de sua aparência visual. Os principais fatores que influenciam um bom desempenho de um SRI são: a escolha da representação das características das imagens e a obtenção do comportamento e preferências dos usuários (TKALČIČ; BURNIK; KOŠIR, 2010).

Vários SRI propostos na literatura tentam melhorar a representação das imagens agregando alguma informação de acordo com determinada situação ou domínio da aplicação. Em (BOUTEMEDJET; ZIOU, 2008) as imagens são representadas pelas características visuais e situações contextuais. Em (TKALČIČ; BURNIK; KOŠIR, 2010) as imagens são representadas por dados textuais e emotivos, tais como: raiva, alegria, surpresa e tristeza. Em (BARRILERO et al., 2011; KIM; KIM; CHO, 2008) as imagens são descritas apenas por características visuais. A atenção visual humana é uma alternativa promissora para representar imagens e pode ser obtida sem qualquer esforço por parte do usuário (MELO; NOGUEIRA; GULIATO, 2015).

Os dados sobre as preferências dos usuários normalmente são obtidos de forma explícita, tais como avaliações, críticas, etc. Infelizmente, os usuários nem sempre estão dispostos a fornecer tais dados de forma explícita. Assim, a fim de lidar com esse problema, alguns sistemas de recomendação usam *feedback* implícito, que reflete indiretamente a opinião do usuário através de seu comportamento no site (OSTUNI et al., 2013; YU et al., 2013). Tipos de *feedback* implícito incluem: histórico de compras, histórico de navegação, padrões de busca, ou até mesmo o movimento do mouse. Nesta tese, propõe-se a exploração do uso da atenção visual humana para obter *feedback* implícito em domínios onde as imagens associadas aos produtos são informações relevantes. Entender como e onde as pessoas subjetivamente olham enquanto avaliam um produto pode ser uma questão importante para melhorar a qualidade dos SRI. Uma ilustração das ideias propostas está no exemplo motivador a seguir.

*Exemplo motivador:* Vamos considerar um site de compras de roupas online onde os produtos são descritos por suas características, tais como categoria (vestido, blusa, calças, saia), tamanho, tecido (lã, seda, couro, algodão), cor, etc., bem como por uma imagem do produto. Um dispositivo rastreador de olhar é usado para capturar a atenção visual dos usuários durante a navegação nas imagens de roupa. O tempo de fixação num dado ponto da imagem e os movimentos dos olhos são capturados. A Figura 1 ilustra o processo de aquisição da atenção visual. Além disso, os usuários fornecem avaliações sobre os produtos, se assim o desejarem. A partir destes dois tipos de *feedback*, o implícito



Figura 1 – Exemplo de captura de atenção visual. Cada círculo azul representa uma fixação. O tamanho do círculo representa a duração e o número representa a sequência da fixação. As linhas entre os círculos representam os movimentos dos olhos (sacadas).

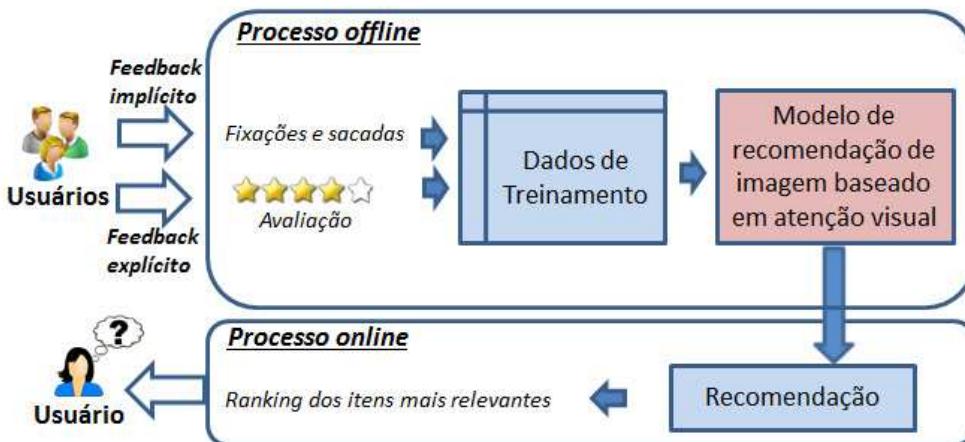


Figura 2 – Esquema geral da nossa proposta: os usuários avaliam e olham para os produtos; treina-se um modelo de recomendação usando as avaliações, fixações, e movimentos dos olhos; o sistema de recomendação utiliza o modelo para recomendar produtos de acordo com o usuário.

(atenção visual) e o explícito (avaliações), um modelo de recomendação é obtido. Então, este modelo consegue recomendar roupas interessantes para o usuário.

### 1.3 Hipótese e Objetivo

Este trabalho se baseia na hipótese central de que a similaridade entre imagens pode ser melhor representada se a informação de atenção visual for incluída no processo de representação da imagem. Neste trabalho, acredita-se que produtos com regiões atentivas

similares tendem a ser mais similares com relação ao gosto dos usuários.

O objetivo deste trabalho é o desenvolvimento de uma metodologia que inclui informações de atenção visual para melhorar a predição de avaliações nos SRI. Este trabalho propõe duas novas abordagens FC e BC, ambas baseadas na vizinhança entre produtos. Um esquema geral da nossa proposta é ilustrado na Figura 2.

A capacidade de identificar os vizinhos é o ponto mais importante dos métodos baseados em vizinhança e tem um impacto significativo na qualidade da recomendação (RICCI; ROKACH; SHAPIRA, 2011). Em geral, os produtos são representados pelas suas avaliações, fatores latentes ou pelo seu conteúdo. A similaridade entre dois produtos é calculada por uma função de similaridade, tais como cosseno e coeficiente de correlação de Pearson (RICCI; ROKACH; SHAPIRA, 2011). Neste trabalho, propõe-se a adição da informação de atentividade no cálculo de similaridade entre os produtos.

Para alcançar o objetivo geral deste trabalho, vários objetivos específicos devem ser obtidos:

- ❑ Investigar a representação da atenção visual como *feedback* implícito para o treinamento de um SRI.
- ❑ Contribuir com uma nova medida de similaridade entre imagens que considera apenas dados de atenção visual obtidos via um rastreador de olhar.
- ❑ Desenvolver uma nova estratégia baseada em FC que combina a medida de similaridade atentiva proposta com uma medida de similaridade baseada em avaliações.
- ❑ Desenvolver uma estratégia baseada em conteúdo que utilize atributos textuais, características visuais e atentivas para a representação de imagens.

## 1.4 Organização da Tese

O trabalho está organizado como segue.

- ❑ Capítulo 1: apresenta a introdução e a motivação para realização deste trabalho.
- ❑ Capítulo 2: define os conceitos básicos e aplicações de SR. Também são apresentados conceitos sobre extração de características visuais e sobre estratégias de capturar a atenção visual humana.
- ❑ Capítulo 3: mostra uma revisão bibliográfica dos principais trabalhos envolvendo SRI, combinação de similaridades e o uso da atenção visual como *feedback* implícito em SR.
- ❑ Capítulo 4: descreve os dois novos métodos para recomendação de imagens, um baseado em FC e outro BC.

- ❑ Capítulo 5: apresenta a metodologia de avaliação e a construção de duas bases de dados reais para a condução dos experimentos.
- ❑ Capítulo 6: analisa os resultados experimentais obtidos com os métodos propostos e compara os resultados com os principais métodos disponíveis na literatura.
- ❑ Capítulo 7: encerra esta tese apresentando as conclusões, contribuições, limitações encontradas e sugestões de trabalhos futuros.

# CAPÍTULO 2

## Fundamentação Teórica

### 2.1 Introdução

Neste capítulo é apresentado o referencial teórico importante para o entendimento desta tese. Os assuntos abordados são: os conceitos iniciais sobre Sistemas de Recomendação, Extração de Características Visuais e Atenção Visual Humana. A proposta desta tese envolve novas abordagens que combinam atenção visual com técnicas de Filtragem Colaborativa (FC) e baseada em Conteúdo, sendo importante que o leitor tenha uma familiaridade com o tema. Dessa forma esse capítulo foi dividido em quatro seções principais. Na Seção 2.2 são apresentados os conceitos sobre Sistemas de Recomendação e as técnicas de FC, BC e Híbrida. Na Seção 2.3 são apresentados o processo de extração de características visuais e os descritores que foram utilizados nos experimentos do SRI baseado em conteúdo proposto nesta tese. Na Seção 2.4 são discutidos os conceitos relacionados à atenção visual, sua aquisição e representação. Por fim, as considerações finais são apresentadas na Seção 2.5.

### 2.2 Sistemas de Recomendação

Os Sistemas de Recomendação (SR) permitem às empresas gerarem recomendações personalizadas de produtos aos seus clientes. Para gerar as listas de recomendações personalizadas, os SR apoiam-se em algoritmos capazes de explorar as relações entre os usuários e os produtos, a relação entre os usuários ou a relação entre os produtos. Uma recomendação pode ser considerada relevante se o produto oferecido ao usuário for agradável ou útil ao usuário. Quatro exemplos de sistemas de recomendação são:

- ❑ Recomendar artigos de notícias on-line para leitores do jornal, com base em uma previsão de interesses dos leitores.
- ❑ Recomendar aos clientes de uma loja on-line alguns produtos que eles gostariam de comprar, com base no seu passado histórico de compras e/ou pesquisas de produtos.

- Recomendar comunidades em uma rede social, com base na sua rede de amigos.
- Recomendar filmes aos clientes, com base nas avaliações realizadas sobre filmes já assistidos.

Os SR surgem como uma solução importante para empresas com grande quantidade de produtos e serviços, permitindo diminuir o esforço de busca dos usuários. Ao oferecer produtos relevantes ao usuário, as empresas aumentam suas vendas, aumentam a satisfação e fidelidade dos usuários, e compreendem melhor o perfil de seus usuários (RICCI; ROKACH; SHAPIRA, 2011). Os SR despertam tanto o interesse acadêmico quanto empresarial, principalmente depois de 2006, quando a empresa de aluguel de filmes Netflix começou uma competição, denominada *Netflix Prize*. O objetivo da competição era o desenvolvimento de um sistema de recomendação que melhoressse o sistema de recomendação Cinematch da Netflix em 10% em relação à Raiz Quadrada do Erro Quadrático Médio, do inglês *Root Mean Squared Error* (RMSE), em um banco de dados público com cerca de 100 milhões de avaliações de 480 mil usuários sobre 17.770 filmes (BENNETT; LANNING, 2007). Em 21 de setembro de 2009 a equipe “*BellKor’s Pragmatic Chaos*” recebeu o grande prêmio *Netflix Prize*<sup>1</sup> de 1 milhão de dólares.

Nos SR, o interesse dos usuários geralmente é medido por meio de *feedback* explícito, ou seja, por avaliações que indicam as preferências dos usuários. As avaliações mais comuns são: notas de 1 a 5 (1: odiei; 2: não gostei; 3: neutro; 4: gostei; 5: gostei muito), ou a opção *like* de curtiu ou não curtiu. Além do *feedback* explícito, alguns SR usam o *feedback* implícito que reflete indiretamente a opinião do usuário através de seu comportamento no sistema (OSTUNI et al., 2013; NÚÑEZ-VALDÉZ et al., 2012). Tipos de *feedback* implícito incluem histórico de compras, histórico de navegação, os padrões de busca, ou até mesmo os movimentos do mouse.

Em geral, independentemente do tipo de informação que será recomendada (vídeo, imagem, texto, ou áudio), existem três técnicas para o desenvolvimento de um SR (ADOMAVICIUS; TUZHILIN, 2005): Filtragem Colaborativa (FC); Filtragem baseada em Conteúdo; e Híbrida. Os conceitos, notações e cada técnica dos SR serão apresentados a seguir.

### 2.2.1 Conceitos e Notações

Formalmente, um SR trabalha com um conjunto de usuários  $\mathbb{U}$  e um conjunto de itens  $\mathbb{I}$ . O comportamento passado dos usuários (compras, avaliações ou críticas) é representado por uma *matriz de utilidade* esparsa  $\mathbb{R} = \{r_{ui}\}_{1 \leq u \leq |\mathbb{U}|, 1 \leq i \leq |\mathbb{I}|}$ , onde  $u$  representa o  $u$ -ésimo usuário e  $i$  o  $i$ -ésimo produto (ou item).  $|\mathbb{U}|$  é o número total de usuários e  $|\mathbb{I}|$  é o número total de itens. A Figura 3 ilustra uma matriz de utilidade, onde as linhas representam

---

<sup>1</sup><http://www.netflixprize.com/>.

<b>u \ i</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>1</b>	4	2	5	3	*	*
<b>2</b>	*	*	3	*	5	4
<b>3</b>	3	2	4	*	*	3
<b>4</b>	4	3	3	*	2	1
<b>5</b>	4	5	3	*	4	1
<b>6</b>	*	*	*	*	3	*

Figura 3 – Matriz de utilidade.

os usuários, colunas representam produtos e a célula  $(u, i)$  contém a avaliação dada pelo usuário  $u$  sobre o produto  $i$  indicando o nível de preferência. As células com “\*” indicam que o valor da avaliação não foi informado pelo usuário.

O problema de recomendação pode ser interpretado como um problema de estimar avaliações para um conjunto de produtos que ainda não foram avaliados por um dado usuário, situação esta representada por “\*” na Figura 3. O SR realiza a predição dessas avaliações desconhecidas e recomenda os produtos com os maiores valores de avaliação predita para o usuário. A predição da célula  $(u, i)$  é denotada como  $\hat{r}_{ui}$ . O conjunto de itens avaliados pelo usuário  $u$  é denotado por  $\mathbb{I}(u)$  e o conjunto de usuários que avaliaram o item  $i$  é denotado por  $\mathbb{U}(i)$ .

### 2.2.2 A Técnica de Filtragem Colaborativa

A técnica Filtragem Colaborativa (FC) é muito popular, bastante utilizada e se baseia apenas no comportamento dos usuários, ou seja, na matriz de utilidade. O pressuposto fundamental por trás desta técnica é que, intuitivamente, se os usuários concordam sobre a qualidade ou relevância de alguns itens, então eles provavelmente vão concordar sobre outros itens.

A técnica de FC pode ser dividida em duas categorias (RICCI; ROKACH; SHAPIRA, 2011): métodos baseados em vizinhança (ou também chamados de métodos baseados em memória) e métodos baseados em modelos. Os métodos baseados em vizinhança focam no cálculo da similaridade entre usuários ou entre produtos (itens). Para predizer a avaliação de um produto  $i$  por um usuário  $u$ , uma abordagem usuário-usuário encontra outros usuários similares a  $u$  e utiliza suas avaliações sobre o item  $i$  para realizar a predição (BREESE; HECKERMAN; KADIE, 1998; HERLOCKER et al., 1999), enquanto uma abordagem item-item usa as avaliações do usuário  $u$  sobre os produtos mais similares ao produto  $i$  para realizar a predição (DESHPANDE; KARYPIS, 2004; SARWAR et al., 2001). A abordagem item-item se tornou mais popular por ter maior escalabilidade e melhor acurácia em várias situações (BELL; KOREN, 2007; SARWAR et al., 2001; TAKÁCS et al., 2007). Os métodos

baseados em modelos usam um algoritmo de aprendizado de máquina para construir um modelo de recomendação (PATEREK, 2007; KOREN, 2008; GEMULLA et al., 2011).

### 2.2.2.1 Métodos *Baseline*

Antes de discutir realmente os algoritmos de Filtragem Colaborativa, discutiremos os métodos de predição *baseline*. Os métodos *baseline* consideram a tendência dos usuários em avaliar os itens abaixo ou acima da média e a tendência dos itens de serem avaliados abaixo ou acima da média (RICCI; ROKACH; SHAPIRA, 2011). Estes métodos são úteis para o estabelecimento de linhas de base, pré-processamento e normalização de dados para a utilização em algoritmos mais sofisticados. Estes algoritmos também podem ser úteis para novos usuários que não possuem avaliações. Denota-se  $b_{ui}$  a predição *baseline* de um usuário  $u$  para um item  $i$ .

O método *baseline* mais simples é prever a avaliação média de todas as avaliações do sistema:  $b_{ui} = \mu$ , onde  $\mu$  é a avaliação média global para todo  $u \in \mathbb{U}$  e  $i \in \mathbb{I}$ . Também pode ser utilizada a avaliação média sobre todos os itens avaliados pelo usuário  $u$  ( $b_{ui} = \bar{r}_u$ ) ou a avaliação média de todos os usuários sobre o item  $i$  ( $b_{ui} = \bar{r}_i$ ).

A predição pode ser melhorada combinando a tendência do usuário em avaliar os itens abaixo ou acima da média e a tendência do item de ser avaliado abaixo ou acima da média, denominada *UserItemBaseline* (UIB), conforme Equação 1.

$$\hat{r}_{ui} = b_{ui} = \mu + b_u + b_i, \quad (1)$$

onde  $b_u$  e  $b_i$  indicam os desvios sobre a média de avaliação global  $\mu$  do usuário  $u$  e do item  $i$ , respectivamente. Eles podem ser definidos simplesmente da seguinte forma (BELL; KOREN, 2007; HERLOCKER; KONSTAN; RIEDL, 2002; POTTER, 2008):

$$b_u = \frac{1}{|\mathbb{I}(u)|} \sum_{i \in \mathbb{I}(u)} (r_{ui} - \mu) \quad (2)$$

$$b_i = \frac{1}{|\mathbb{U}(i)|} \sum_{u \in \mathbb{U}(i)} (r_{ui} - \mu - b_u) \quad (3)$$

Diante de amostras dispersas, ou seja, alguns usuários ou itens com poucas avaliações, o método UIB pode ser mais regularizado com a adição dos parâmetros  $\lambda_u$  e  $\lambda_i$  (KOREN, 2010):

$$b_u = \frac{1}{|\mathbb{I}(u)| + \lambda_u} \sum_{i \in \mathbb{I}(u)} (r_{ui} - \mu) \quad (4)$$

$$b_i = \frac{1}{|\mathbb{U}(i)| + \lambda_i} \sum_{u \in \mathbb{U}(i)} (r_{ui} - \mu - b_u) \quad (5)$$

Este ajuste faz com que a predição  $b_{ui}$  fique mais perto da média global quando o usuário  $u$  ou o item  $i$  tiver poucas avaliações. Se o usuário  $u$  é novo, portanto, não tem

avaliações, o seu desvio  $b_u$  é definido como zero. Do mesmo modo, se o item  $i$  é novo, o seu desvio  $b_i$  também é definido como zero.

### 2.2.2.2 Filtragem Colaborativa Usuário-Usuário

Filtragem Colaborativa Usuário-Usuário, também conhecida como Filtragem Colaborativa - *K-Nearest Neighbors* (KNN), foi o primeiro método baseado em FC implementado na literatura, com o intuito de recomendar artigos (RESNICK et al., 1994). A ideia central é o SR recomendar para um usuário ativo (a quem se deseja recomendar) os itens que outros usuários, com gostos similares a ele, gostaram no passado. A similaridade entre dois usuários é calculada baseada na similaridade do histórico de avaliação/compra dos usuários. A hipótese desta abordagem é que a nota do usuário ativo  $u$  para um item  $i$  é provavelmente similar à nota recebida por outro usuário  $v$ , se  $u$  e  $v$  avaliaram itens no passado de forma semelhante.

A maneira mais comum de realizar a predição  $\hat{r}_{ui}$  da avaliação de um usuário  $u$  sobre um item  $i$  é calculando a média ponderada das avaliações do conjunto  $\mathbb{U}(u, i, k)$ , que são os  $k$  usuários vizinhos mais próximos do usuário  $u$  e que avaliaram o item  $i$ , conforme a Equação 6.

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in \mathbb{U}(u, i, k), v \neq u} (r_{vi} - \bar{r}_v) \cdot s_{uv}}{\sum_{v \in \mathbb{U}(u, i, k), v \neq u} |s_{uv}|}, \quad (6)$$

onde  $s_{uv}$  é a similaridade entre os usuários  $u$  e  $v$ .  $\bar{r}_u$  e  $\bar{r}_v$  são a média de avaliações do usuário  $u$  e  $v$  respectivamente, conforme Equação 7.

$$\bar{r}_u = \frac{1}{|\mathbb{I}(u)|} \sum_{i \in \mathbb{I}(u)} r_{ui} \quad (7)$$

O modelo *User KNN + Baseline* (UKB) adapta o cálculo da predição usando o predictor *baseline* (KOREN, 2010), conforme Equação 8.

$$\hat{r}_{ui} = b_{ui} + \frac{\sum_{v \in \mathbb{U}(u, i, k), v \neq u} (r_{vi} - b_{vi}) \cdot s_{uv}}{\sum_{v \in \mathbb{U}(u, i, k), v \neq u} |s_{uv}|}, \quad (8)$$

Um detalhe muito importante nesta abordagem é a escolha da função de similaridade entre usuários. Diversas funções de similaridade foram propostas e avaliadas na literatura.

- *Pearson Correlation Coefficient* (PCC): Esta é a função mais popular e amplamente utilizada. A função calcula a correlação estatística de Pearson entre as avaliações comuns a dois usuários para determinar a similaridade entre ambos (Figura 4 (a)). A correlação entre os usuários  $u$  e  $v$  é calculada da seguinte forma:

$$s_{uv} = \frac{\sum_{i \in \mathbb{I}(u, v)} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in \mathbb{I}(u, v)} (r_{ui} - \bar{r}_u)^2} \cdot \sqrt{\sum_{i \in \mathbb{I}(u, v)} (r_{vi} - \bar{r}_v)^2}} \quad (9)$$

onde  $\mathbb{I}(u, v)$  é o conjunto de itens avaliados por ambos usuários  $u$  e  $v$ . O valor de similaridade retornado pelo PCC se encontra no intervalo de  $[-1, 1]$ . Valores próximos a 1 indicam que os dois usuários tendem a concordar, e valores próximos a -1 indicam que os dois usuários tendem a discordar um do outro. Valores iguais a 0 indicam que o histórico de avaliações dos dois usuários não são correlacionados.

Para melhorar o cálculo de similaridade, Koren e Bell (KOREN; BELL, 2011) modificaram o PCC substituindo o valor da média de avaliações pela predição *baseline*, conforme Equação 10.

$$s_{uv} = \frac{\sum_{i \in \mathbb{I}(u,v)} (r_{ui} - b_{ui})(r_{vi} - b_{vi})}{\sqrt{\sum_{i \in \mathbb{I}(u,v)} (r_{ui} - b_{ui})^2} \cdot \sqrt{\sum_{i \in \mathbb{I}(u,v)} (r_{vi} - b_{vi})^2}} \quad (10)$$

- *Cosseno:* Os usuários são representados como vetores e a similaridade é medida pela função cosseno entre dois vetores de avaliações (Figura 4 (b)). A similaridade cosseno é dada na Equação 11.

$$s_{uv} = \frac{\sum_{i \in \mathbb{I}} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in \mathbb{I}} (r_{ui})^2} \cdot \sqrt{\sum_{i \in \mathbb{I}} (r_{vi})^2}} \quad (11)$$

onde avaliações desconhecidas são consideradas iguais a zero. O valor da similaridade cosseno está no intervalo de  $[0, 1]$ . Valores próximos a 1 indicam que os dois usuários tendem a concordar, e valores próximos a 0 indicam que os dois usuários tendem a discordar um do outro.

- *Baseado na distância Euclidiana:* Na inversão da distância Euclidiana utiliza-se a perspectiva da magnitude das avaliações, e quanto maior a distância mais similares são os usuários (Figura 4 (c)), é dada na Equação 12.

$$s_{uv} = \frac{1}{1 + \sqrt{\sum_{i \in \mathbb{I}(u,v)} (r_{ui} - r_{vi})^2}} \quad (12)$$

O valor da similaridade da inversão da distância Euclidiana está no intervalo de  $[0, 1]$ . Valores próximos a 1 indicam que os dois usuários tendem a concordar, e valores próximos a 0 indicam que os dois usuários tendem a discordar um do outro.

Como exemplo, vamos predizer as recomendações para o usuário **1** da Figura 3. O primeiro passo é encontrar quais são os usuários que possuem gosto semelhante ao usuário **1**. A Tabela 1 apresenta a similaridade entre o usuário **1** e os demais usuários. Note que de acordo com a perspectiva da função de similaridade o resultado da similaridade tem algumas diferenças, por exemplo, o usuário **1** é mais similar ao usuário **3** se considerarmos o PCC ou a distância Euclidiana invertida, ou é mais similar ao usuário **4** se considerarmos a função cosseno. Caso dois usuários não tenham avaliado itens em comum, a similaridade cosseno é igual a zero e não é possível calcular o PCC e a distância Euclidiana invertida.

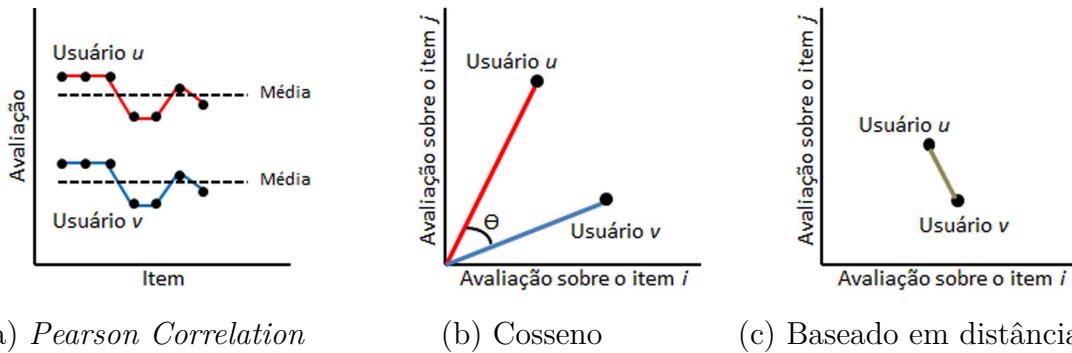


Figura 4 – Perspectivas das funções de similaridade entre dois usuários  $u$  e  $v$ . (a) compara o padrão das avaliações dos usuários. (b) calcula o ângulo entre os vetores de avaliações representando os usuários  $u$  e  $v$ . (c) calcula a distância entre os usuários  $u$  e  $v$ . Para melhor ilustrar (b) e (c), considerou-se apenas dois itens  $i$  e  $j$ .

Tabela 1 – Similaridade entre o usuário 1 e os demais usuários da Figura 3.

$s_{uv}$	Função de similaridade		
	Pearson (Eq. 9)	Cosseno	Distância Euclidiana invertida
$s_{12}$	0	0,28	0,33
$s_{13}$	0,98	0,79	0,41
$s_{14}$	0,19	0,80	0,31
$s_{15}$	-0,98	0,68	0,21
$s_{16}$	*	0	*

Tabela 2 – Predição da avaliação do usuário 1 sobre os itens 5 e 6 da Figura 3.

item	Predição de avaliação ( $\hat{r}_{ui}$ )		
	Pearson (Eq. 9)	Cosseno	Distância Euclidiana invertida
5	2,90	2,90	2,90
6	3,24	2,70	2,82

Para predizer a nota a um determinado item que o usuário 1 não avaliou ainda, podemos utilizar a Equação 6. A Tabela 2 mostra o resultado da predição sobre os itens 5 e 6 usando apenas os dois vizinhos mais próximos ( $k = 2$ ), ou seja, os usuários 3 e 4. Note que a escolha da função de similaridade é muito importante, pois a mesma será usada como peso no cálculo da predição de avaliação.

### 2.2.2.3 Filtragem Colaborativa Item-Item

Filtragem Colaborativa Usuário-Usuário sofre com o problema de escalabilidade a medida que o número de usuários cresce no sistema. Encontrar os vizinhos mais próximos a um usuário pode ser muito custoso. Para atender os grandes websites de comércio eletrônico, alternativas mais escaláveis são necessárias. A abordagem Filtragem Colaborativa Item-Item é uma solução interessante e bastante popular, pois na maioria das vezes é mais escalável e possui melhor acurácia na predição de avaliação (BELL; KOREN, 2007;

SARWAR et al., 2001; TAKÁCS et al., 2007).

A ideia da abordagem Item-Item é que ao invés de usar semelhanças entre os usuários, considera-se semelhanças entre itens. Se dois itens tendem a ter avaliações próximas dos mesmos usuários, então eles são semelhantes e os usuários deverão ter preferências similares para itens similares. A abordagem Item-Item oferece melhor desempenho quando comparado a Usuário-Usuário nas situações em que o número de usuários é superior ao número de itens.

A predição baseada em Item-Item é análoga à predição baseada em Usuário-Usuário, a diferença é que a predição é feita utilizando o conjunto  $\mathbb{I}(u, i, k)$ , que são os  $k$  itens vizinhos mais próximos do item  $i$  que foram avaliados pelo usuário  $u$ . A maneira mais comum de realizar a predição  $\hat{r}_{ui}$  é dada na Equação 13.

$$\hat{r}_{ui} = \frac{\sum_{j \in \mathbb{I}(u, i, k), j \neq i} r_{uj} \cdot s_{ij}}{\sum_{j \in \mathbb{I}(u, i, k), j \neq i} |s_{ij}|}, \quad (13)$$

onde  $s_{ij}$  é a similaridade entre os itens  $i$  e  $j$ .

O modelo *Item KNN + Baseline* (IKB) adapta o cálculo da predição usando o preditor *baseline* (KOREN, 2010), conforme Equação 14.

$$\hat{r}_{ui} = b_{ui} + \frac{\sum_{j \in \mathbb{I}(u, i, k), j \neq i} (r_{uj} - b_{uj}) \cdot s_{ij}}{\sum_{j \in \mathbb{I}(u, i, k), j \neq i} |s_{ij}|}, \quad (14)$$

Assim como na abordagem Usuário-Usuário, existe uma variedade de funções para calcular a similaridade entre itens. As funções mais comuns são o PCC (KOREN; BELL, 2011) (Equação 15) e a Cosseno (Equação 16).

$$s_{ij} = \frac{\sum_{u \in \mathbb{U}(i, j)} (r_{ui} - b_{ui})(r_{uj} - b_{uj})}{\sqrt{\sum_{u \in \mathbb{U}(i, j)} (r_{ui} - b_{ui})^2} \cdot \sqrt{\sum_{u \in \mathbb{U}(i, j)} (r_{uj} - b_{uj})^2}} \quad (15)$$

onde  $\mathbb{U}(i, j)$  é o conjunto de usuários que avaliaram os itens  $i$  e  $j$  simultaneamente.

$$s_{ij} = \frac{\sum_{u \in \mathbb{U}} r_{ui} \cdot r_{uj}}{\sqrt{\sum_{u \in \mathbb{U}} (r_{ui})^2} \cdot \sqrt{\sum_{u \in \mathbb{U}} (r_{uj})^2}} \quad (16)$$

Sarwar et al. (SARWAR et al., 2001) recomenda o uso da medida de similaridade cosseno ajustada, onde as avaliações são traduzidas via a dedução de média dos usuários antes de calcular a similaridade cosseno, conforme Equação 17.

$$s_{ij} = \frac{\sum_{u \in \mathbb{U}(i, j)} (r_{ui} - \bar{r}_u) \cdot (r_{uj} - \bar{r}_u)}{\sqrt{\sum_{u \in \mathbb{U}(i, j)} (r_{ui} - \bar{r}_u)^2} \cdot \sqrt{\sum_{u \in \mathbb{U}(i, j)} (r_{uj} - \bar{r}_u)^2}} \quad (17)$$

Como a predição de avaliação depende muito dos pesos da similaridade  $s_{ij}$ , em alguns conjuntos de dados, a transformação desses pesos pode ser interessante (MELVILLE; SINDHWANI, 2010). A seguir são apresentadas duas estratégias para modificar  $s_{ij}$ , denotadas por  $s'_{ij}$ . Neste caso, a similaridade  $s_{ij}$  é substituída por  $s'_{ij}$  no cálculo da predição de avaliação  $\hat{r}_{ui}$ .

- *Ponderação de Importância (Significance Weighting)*: É comum para muitos pares de itens que o número de usuários que avaliaram ambos os itens seja pequeno. Com isso, a função de similaridade pode produzir uma predição de semelhança ruim. Uma abordagem para resolver este problema é multiplicar o valor da similaridade por um fator de ponderação de importância, também denominado *shrunk correlation*, que desvaloriza as correlações com poucas avaliações em comum (KOREN; BELL, 2011). O fator de ponderação de importância é dado na Equação 18.

$$s'_{ij} = \frac{|\mathbb{U}(i, j)| - 1}{|\mathbb{U}(i, j)| - 1 + \lambda_s} s_{ij} \quad (18)$$

onde  $\lambda_s$  é um parâmetro de encolhimento definido pelo usuário.  $|\mathbb{U}(i, j)|$  é a quantidade de usuários que avaliaram ambos os itens  $i$  e  $j$ .

- *Amplificação (Case Amplification)*: A fim de favorecer os itens com a similaridade mais elevada e desfavorecer os itens com a similaridade mais baixa, Breese et al. (BREESE; HECKERMAN; KADIE, 1998) introduziu o parâmetro de amplificação que transforma os valores de similaridade originais conforme Equação 19.

$$s'_{ij} = s_{ij} \cdot |s_{ij}|^{p-1} \quad (19)$$

onde  $p$  é o fator de amplificação.

*Exemplo:* Este exemplo compara o impacto da aplicação do fator de amplificação na predição da avaliação  $\hat{r}_{15}$  do usuário **1** sobre o item **5** da Figura 3. O resultado da similaridade entre o item **5** e os demais itens é dado na Tabela 3.

Tabela 3 – Similaridade entre o item **5** e os demais itens da Figura 3 após a aplicação do fator de amplificação  $p$  da Equação 19.

$s_{ij}$	$p = 1$	$p = 2$
$s_{51}$	0,5	0,25
$s_{52}$	0,8	0,64
$s_{53}$	0,3	0,09

Considerando as avaliações do usuário **1** sobre os itens **1**, **2** e **3** da Figura 3 ( $r_{11} = 4$ ,  $r_{12} = 2$  e  $r_{13} = 5$ ), e o fator de amplificação  $p = 1$  da Equação 19, a predição de avaliação  $\hat{r}_{15}$  do usuário **1** sobre o item **5** é calculada da seguinte forma:  $\hat{r}_{15} = ((0,5 * 4) + (0,8 * 2) + (0,3 * 5)) / (0,5 + 0,8 + 0,3) = 3,18$ . Com um fator de amplificação  $p = 2$ , a similaridade entre os itens **5** e **2** será favorecida em relação as demais similaridades, como demonstrado no cálculo da predição de avaliação  $\hat{r}_{15} = ((0,25 * 4) + (0,64 * 2) + (0,09 * 5)) / (0,25 + 0,64 + 0,09) = 2,78$ .

Métodos baseados em vizinhança se tornaram muito populares, porque eles são intuitivos e relativamente simples de implementar. Eles também oferecem duas importantes propriedades:

- *Explicável*: É muito importante explicar como as recomendações automatizadas são realizadas. Os usuários esperam um sistema com clareza nas suas previsões, em vez de apresentar recomendações "caixa preta". As explicações não só enriquecem a experiência do usuário, mas também incentivam os usuários a interagir com o sistema, corrigir impressões erradas e melhorar a precisão em longo prazo.
- *Novas avaliações*: Modelos baseados em vizinhança podem fornecer recomendações atualizadas imediatamente depois que os usuários entram com novas avaliações. Não existe a necessidade de treinar modelos e estimar parâmetros.

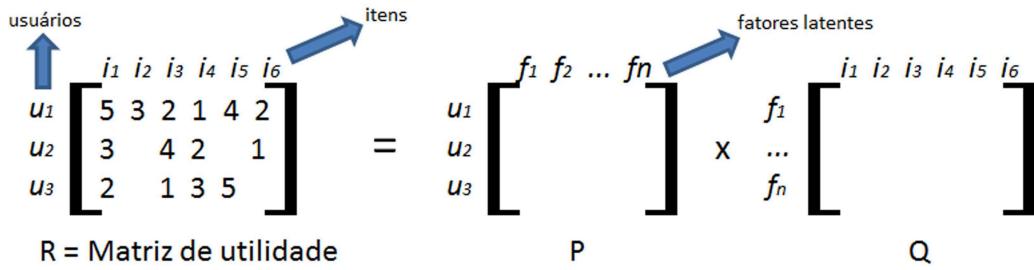
Entretanto, as avaliações usuário-item são representadas por vetores de alta dimensão. Na abordagem Usuário-Usuário, cada usuário é representado por um vetor de dimensão  $|I|$  e na abordagem Item-Item, cada item é representado por um vetor de dimensão  $|U|$ . A maneira mais utilizada para reduzir a dimensionalidade é modelar os itens e usuários utilizando a fatoração de matrizes.

#### 2.2.2.4 Filtragem Colaborativa Baseada em Modelos de Fatoração de Matrizes

Vários algoritmos de recomendação são baseados em modelos de fatores latentes. A maioria deles é baseada na fatoração da matriz de utilidade utilizando a técnica *Singular Value Decomposition* (SVD) (KOREN; BELL; VOLINSKY, 2009). 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  $q_i \in \mathbb{R}^M$ , onde  $M$  é o número de fatores latentes. A predição de avaliação  $\hat{r}_{ui}$  é obtida pelo produto escalar dos vetores  $p_u$  e  $q_i$ , conforme Equação 20.

$$\hat{r}_{ui} = p_u^T \cdot q_i \quad (20)$$

Este tem sido um dos modelos mais populares na FC. As avaliações são profundamente influenciadas por um conjunto de fatores que são muito específicos de acordo com o domínio (por exemplo, tipo do decote e tamanho da manga para roupas; intensidade da cor e textura para pinturas). 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  $i$  possui os fatores latentes. Para cada usuário  $u$ , os elementos de  $p_u$  medem o grau

Figura 5 – *Singular Value Decomposition* (SVD).

de interesse que o usuário  $u$  tem em itens com alta correspondência de fatores latentes. O processo de aprendizado para obter a matriz  $P$  de vetores de fatores latentes de todos usuários e a matriz  $Q$  de vetores de fatores latentes de todos os itens é um problema de otimização, onde dada uma matriz de utilidade esparsa  $R$  deve-se encontrar as matrizes  $P$  e  $Q$  tal que  $P.Q$  é uma matriz muito próxima de  $R$ , como ilustrado na Figura 5 e descrito em (KOREN; BELL, 2011). A estratégia e o momento da aprendizagem podem mudar de acordo com o SR.

O modelo *SVD + Baseline* (SB) combina os modelos SVD e *Baseline* (MENON; ELKAN, 2010). A predição de avaliação  $\hat{r}_{ui}$  é a soma do preditor *Baseline* ( $\hat{r}_{ui}^{(1)}$ , de acordo com a Equação 1) com o preditor SVD ( $\hat{r}_{ui}^{(2)}$ , de acordo com a Equação 20).

$$\hat{r}_{ui} = \hat{r}_{ui}^{(1)} + \hat{r}_{ui}^{(2)} \quad (21)$$

Os valores de  $b_u$ ,  $b_i$ ,  $p_u$  e  $q_i$  devem ser aprendidos juntos, como demonstrados no trabalho descrito em (KOREN; BELL, 2011).

Pode-se também combinar fatores latentes com vizinhança. O modelo *Item KNN + SVD + Baseline* (IKSB) substitui o valor de  $s_{ij}$  da Equação 14 pelo valor  $c_{ij}$  (o inverso da distância Euclidiana normalizada entre dois vetores de fatores latentes de itens) calculado como mostra a Equação 22 (JAHRER; TÖSCHER, 2012).

$$c_{ij} = \left( \frac{\sum_{m=1}^M (q_{im} - q_{jm})^2}{\sqrt{\sum_{m=1}^M q_{im}^2} \cdot \sqrt{\sum_{m=1}^M q_{jm}^2}} \right)^{-2}, \quad (22)$$

onde  $M$  é o número de fatores latentes;  $q_i$  e  $q_j$  são vetores de fatores latentes dos itens  $i$  e  $j$ , respectivamente.

### 2.2.3 A Técnica Baseada em Conteúdo

O Sistema de Recomendação baseado em Conteúdo (SRBC) tem suas raízes na recuperação da informação. Os SRBC analisam um conjunto de descrições de itens previamente avaliados pelo usuário e constroem um modelo ou perfil dos interesses do usuário com base nas características dos itens. O processo básico consiste em combinar o perfil do

usuário com as características dos itens para recomendar novos itens interessantes. Se o perfil reflete bem as preferências do usuário, o SRBC vai ter grande eficácia no processo de recomendação. Os detalhes dos SRBC diferem de acordo com a representação dos itens e a construção das preferências dos usuários.

### 2.2.3.1 Representação dos itens

Os itens que podem ser recomendados para o usuário são normalmente armazenados em uma tabela de uma base de dados. A Tabela 4 mostra um exemplo simples da representação de três roupas. O nome das colunas são atributos, também chamados de propriedades ou características. Cada linha representa um item e cada célula um valor para cada atributo.

Tabela 4 – Uma base de dados de roupas

Item	Atributos			
	Gênero	Categoria	Tamanho	Tecido
1	Masculino	Calça	Médio	Jeans
2	Feminino	Saia	Pequeno	Couro
3	Feminino	Blusa	Grande	Lã

Cada item é descrito pelo mesmo conjunto de atributos estruturados e os valores dos atributos são conhecidos. Com isso, muitos algoritmos de aprendizagem de máquina e/ou recuperação da informação podem ser utilizados para aprender um perfil de usuário.

Uma das técnicas usadas em diversos domínios de aplicação pela abordagem por conteúdo é a utilização do modelo de espaço vetorial para a representação do conteúdo e a medida *Term-Frequency* (TF). Nesta técnica, um item  $i$  é denominado documento e um conjunto de documentos  $\mathbb{D} = d_1, d_2, \dots, d_N$  contém um conjunto de termos (possíveis valores dos atributos dos itens)  $\mathbb{T} = t_1, t_2, \dots, t_n$ . Cada documento  $d_i$  pode ser representado por um vetor do espaço n-dimensional de pesos dos termos de  $\mathbb{T}$ , tal que  $d_i = w_{1i}, w_{2i}, \dots, w_{ni}$ , sendo  $w_{ki}$  o peso do termo  $t_k$  no documento  $d_i$ .

A frequência inversa (*Inverted Document Frequency* - IDF) dos itens (documentos) é calculada conforme Equação 23.

$$IDF(t_k) = \log \frac{N}{n_k}, \quad (23)$$

onde  $N$  é o total de itens e  $n_k$  é a quantidade de documentos onde o termo  $t_k$  aparece. O *Term Frequency - Inverted Document Frequency* (TF-IDF) é dada na Equação 24.

$$TF - IDF(t_k, d_i) = TF(t_k, d_i).IDF(t_k), \quad (24)$$

onde  $TF(t_k, d_i)$  é a frequência do termo  $t_k$  no documento  $d_i$ . Agora é possível mostrar o cálculo do peso  $w_{ki}$  do termo  $t_k$  no documento  $d_i$ , conforme Equação 25.

$$w_{ki} = \frac{TF - IDF(t_k, d_i)}{\sqrt{\sum_{s=1}^{|T|}(TF - IDF(t_s, d_i))^2}} \quad (25)$$

### 2.2.3.2 Perfil do usuário

Várias técnicas podem ser usadas para aprender um perfil de interesses do usuário. Em geral, o perfil do usuário pode ser: (1) uma personalização do usuário, onde os usuários fornecem manualmente os seus próprios interesses e preferências; (2) um modelo que descreve os tipos de itens que são interessantes para o usuário; (3) um histórico das interações do usuário (por exemplo, compras ou avaliações passadas) (PAZZANI; BILLSUS, 2007). Então os itens são comparados com o perfil do usuário para determinar quais itens serão recomendados.

- *Perfil baseado na personalização do usuário.* As preferências do usuário são disponíveis em uma forma mais simples e direta. No entanto, este perfil possui algumas desvantagens: o processo requer um esforço do usuário para informar suas preferências manualmente e os usuários nem sempre estão dispostos a fornecer tais informações de forma explícita; os usuários podem alterar os seus interesses e não informar tais alterações ao sistema; é difícil informar ao sistema a ordem de preferência sobre os atributos dos itens (BRUSILOVSKY; KOBZA; NEJDL, 2007).
- *Perfil baseado em modelo.* Várias técnicas podem ser utilizadas para modelar as preferências do usuário, tais como, redes neurais artificiais (ZHANG; CALLAN; MINKA, 2002), mineração de regras de associação (AMO et al., 2012; COHEN, 1995), métodos probabilísticos (MOONEY; ROY, 2000; LEWIS; RINGUETTE, 1994; MCCALLUM; NIGAM et al., 1998), classificadores lineares (LEWIS et al., 1996), e árvores de decisão (KIM et al., 2001; PAZZANI; BILLSUS, 1997). Normalmente, estas técnicas são projetadas para trabalhar com dados estruturados de baixa dimensionalidade, e podem não ser interessantes para lidar com dados de alta dimensionalidade, tais como características visuais de imagens.
- *Perfil baseado no histórico de interações dos usuários.* Este perfil é representado por todo o conteúdo dos itens já avaliados e armazenados pelo usuário. Neste caso, um algoritmo de vizinhança KNN é usado para a predição de avaliação dos itens levando em consideração o comportamento passado do usuário e a similaridade entre os itens (PAZZANI; BILLSUS, 2007).

Em um algoritmo KNN, os  $k$  itens mais similares são calculados usando diferentes medidas de similaridade, tais como similaridade Cosseno, distância Euclidiana, etc. Um exemplo usando a similaridade Cosseno e o TF-IDF é dado na Equação 26

$$s_{ij} = \frac{\sum_{k \in \mathbb{T}} w_{ki} \cdot w_{kj}}{\sqrt{\sum_{k \in \mathbb{T}} (w_{ki})^2} \cdot \sqrt{\sum_{k \in \mathbb{T}} (w_{kj})^2}}, \quad (26)$$

onde  $w_{ki}$  é o peso do termo  $t_k$  no item  $i$ .

## 2.2.4 Técnica Híbrida

Com o intuito de preservar características importantes e diminuir/eliminar problemas das abordagens por Conteúdo e Filtragem Colaborativa, pode ser interessante gerar sistemas híbridos que combinem dois ou mais métodos de recomendação. O sistema híbrido mais comum é combinar uma abordagem Filtragem Colaborativa com outra técnica para evitar o problema de partida a frio. Segundo (BURKE, 2007), os sistemas híbridos podem ser classificados da seguinte forma:

- *Por peso.* A recomendação do sistema é dada pela combinação ponderada de vários métodos de recomendação.
- *Por troca.* O sistema usa algum critério para alternar entre as técnicas de recomendação que será utilizada.
- *Por união (misto).* Recomendações de vários métodos são apresentadas ao usuário ao mesmo tempo.
- *Combinação de características.* Um único algoritmo com características de diferentes sistemas gera uma recomendação única.
- *Cascata.* Um sistema refina as recomendações de outro sistema, reordenando os itens recomendados. Nesta técnica, inicialmente utiliza-se uma técnica de recomendação para produzir um ranking de candidatos e uma segunda técnica refina a recomendação entre o conjunto de candidatos, reordenando os itens candidatos.
- *Melhora de característica.* A saída de um sistema é a entrada de outro, usando os valores das recomendações e outros fatores para gerar a nova recomendação.
- *Meta-nível.* O modelo aprendido por um sistema é utilizado em outro.

## 2.2.5 Problemas em Sistemas de Recomendação

Nesta seção serão enumerados os principais problemas/desafios que as abordagens apresentadas encontram para prover boas recomendações. Estes problemas são objeto

de muitas pesquisas que buscam maneiras de minimizá-los ou até eliminá-los (RICCI; ROKACH; SHAPIRA, 2011).

- **Novo usuário (Partida a frio de usuário):** Para receber recomendações, tanto na abordagem de Filtragem Colaborativa como na Baseada em Conteúdo, os usuários deverão ter avaliado um número suficiente de itens. Só a partir daí o Sistema de Recomendação conseguirá aprender as preferências dos usuários e então conseguir recomendar itens. Com pouca informação o sistema não é capaz de fornecer recomendações confiáveis (BARJASTEH et al., 2015; GRAUS; WILLEMSSEN, 2015).
- **Novo item (Partida a frio de item):** Este é um problema típico apenas dos métodos FC. Novos itens são adicionados diariamente nos sistemas. Até que um novo item seja avaliado por um expressivo número de usuários, os métodos FC não conseguirão recomendá-lo (AHARON et al., 2015; SHARMA et al., 2015). Uma das vantagens mais importantes da abordagem por conteúdo em relação à FC é não sofrer do problema de novos itens, ou seja, quando um novo item é inserido no sistema, facilmente as suas características podem ser descritas e comparadas com outros itens, o que possibilita que ele seja recomendado aos usuários imediatamente.
- **Esparsidade dos dados:** É comum que os sistemas cresçam e tenham muitos usuários e muitos itens. Na prática, um usuário avalia somente uma pequena parte dos itens, resultando numa matriz de utilidade esparsa (BARGAH; MISHRA, 2016; NIU et al., 2016). O sucesso de um Sistema de Recomendação de Filtragem Colaborativa, em específico, depende da disponibilidade de uma massa crítica de usuários. Da mesma forma, a abordagem Baseada em Conteúdo sofre com os poucos dados disponíveis. Se um usuário não interage com o sistema e não disponibiliza suas preferências de forma explícita ou implícita, dificilmente o Sistema de Recomendação conseguirá calcular boas recomendações.
- **Escalabilidade:** Em muitos casos, o número de usuários e de itens pode chegar à ordem dos milhões. Na Filtragem Colaborativa por vizinhança, é necessário o cálculo da similaridade entre todos os usuários ou entre todos os itens, além de predizer a avaliação para milhões de itens não avaliados para cada usuário ativo. Da mesma forma, a técnica baseada em Conteúdo precisa ser escalável para a quantidade de características que um item pode possuir, assim como a quantidade de itens que precisam ser comparados ao perfil do usuário para serem recomendados. Alguns métodos tentam reduzir este problema usando técnicas de agrupamento para reduzir o número do cálculo de similaridades ou técnicas baseadas em modelo, como o SVD, para reduzir a dimensionalidade da representação dos dados (YI; LEE; JUNG, 2015).
- **Superespecialização:** Um usuário com preferências muito limitadas (poucas ou muitas na mesma categoria de itens) sofre o problema da pouca variabilidade de

sugestões. O SRBC não recomenda novidades e sugere apenas itens que são semelhantes ao perfil do usuário. Para resolver este problema, é possível introduzir alguma aleatoriedade no sistema, de forma a surgirem outras recomendações diferentes, mas que façam sentido (SANTOS, 2015).

- **Análise de Conteúdo Limitado:** A descrição dos itens de forma eficiente e completa é uma das principais características para o sucesso de um SRBC. A atribuição de características aos itens pode ser insuficiente para definir aspectos distintos dos itens que se tornam necessários para instigar o interesse do usuário (LOPS; GEMMIS; SEMERARO, 2011; SANTOS, 2015).

## 2.2.6 Métricas para Avaliação de Sistemas de Recomendação

Segundo os autores em (HERLOCKER et al., 2004), a avaliação de SR é muito difícil por duas razões principais. Primeiro, diferentes algoritmos podem ser melhores ou piores em diferentes conjuntos de dados. A segunda razão é que o objetivo para o qual a avaliação é realizada pode diferir. As métricas podem ser organizadas em duas categorias: *métricas de acurácia na predição* e *métricas de acurácia no ranqueamento*.

### 2.2.6.1 Métricas de Acurácia na Predição

Nos SR o intuito da métrica de acurácia na predição é verificar o quanto as avaliações preditas pelo sistema estão próximas das verdadeiras avaliações realizadas pelo usuário. A acurácia da predição é tipicamente independente da interface do usuário e pode ser medida em um experimento offline.

Para verificar a acurácia das predições, as avaliações conhecidas são divididas em: i) conjunto de treinamento, que representa as avaliações realizadas pelos usuários  $r_{ui}$ ; e ii) conjunto de teste  $\mathbb{T}$ . O sistema gera predições de avaliação  $\hat{r}_{ui}$  para o conjunto de teste  $\mathbb{T}$  (BURKE, 2002). As métricas mais utilizadas são o Erro Médio Absoluto, do inglês *Mean Absolute Error* (MAE) e a Raiz Quadrada do Erro Quadrático Médio, do inglês *Root Mean Squared Error* (RMSE).

A métrica MAE mede o desvio médio absoluto entre a avaliação predita e a verdadeira avaliação do usuário, conforme Equação 27 (SHARDANAND; MAES, 1995).

$$MAE = \sqrt{\frac{1}{|\mathbb{T}|} \sum_{(u,i) \in \mathbb{T}} (\hat{r}_{ui} - r_{ui})^2} \quad (27)$$

A métrica RMSE é uma variação da métrica MAE, conforme Equação 28 (HERLOC-KER et al., 2004). Comparado à MAE, a métrica RMSE penaliza grandes erros (SHANI;

GUNAWARDANA, 2013). Por exemplo, um erro com valor 1 aumenta em 1 no somatório, já um erro de valor 2 aumenta em 4 no somatório.

$$RMSE = \sqrt{\frac{1}{|\mathbb{T}|} \sum_{(u,i) \in \mathbb{T}} (\hat{r}_{ui} - r_{ui})^2} \quad (28)$$

### 2.2.6.2 Métricas de Acurácia no Ranqueamento

As métricas de acurácia no ranqueamento medem a habilidade do SR em apresentar uma listagem (Top-N), em que os itens estão ordenados de acordo com a preferência do usuário.

Na tarefa de recomendação Top-N, o sistema recomenda para o usuário os  $N$  primeiros itens mais relevantes para ele, proporcionando 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. As métricas mais populares para esse cenário são Precisão e Revocação.

Para calcular a Precisão e a Revocação, as avaliações são convertidas para uma escala binária (1-relevante, 0-não relevante). 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 é considerado não relevante. Precisão é definida como a razão entre o número dos itens relevantes recomendados ao usuário e o número total de itens recomendados. Dado um conjunto  $\mathbb{T}$  de usuários de teste, a Média de Precisão, do inglês *Average Precision* (AP), é calculada como mostra a Equação 29.

$$AP@N = \frac{\sum_{u \in \mathbb{T}} \frac{|itens\ relevantes\ recomendados|}{|Nitens\ recomendados|}}{|\mathbb{T}|} \quad (29)$$

A Revocação (ou sensitividade) é definida como a razão entre os itens relevantes recomendados e o número total de itens relevantes disponíveis ( $Recall = \frac{|itens\ relevantes\ recomendados|}{|total\ de\ itens\ relevantes|}$ ). A medida de Especificidade mede a capacidade de se predizer um item como não relevante de forma correta ( $Especificidadde = \frac{|itens\ preditos\ como\ não\ relevantes|}{|total\ de\ itens\ não\ relevantes|}$ ). A relação entre as medidas de Revocação e Especificidade pode ser representada através da curva ROC (*Receiver Operating Characteristics*). A área abaixo da curva ROC é chamada de *Area Under the ROC Curve* (AUC) (SHANI; GUNAWARDANA, 2013). Quanto mais próximo de 1 o valor de AUC, melhor é o desempenho do SR.

Algumas alternativas, tais como *Normalized Discounted Cumulative Gain* (NDCG) e *Distance-based Performance Measure* (NDPM), consideram a importância da ordem dos itens na lista (Top-N) no cálculo da acurácia do ranqueamento. Estas métricas consideram os itens no topo como mais importantes e os itens na parte inferior como menos importantes. O valor retornado pelas métricas é entre 0 e 1 independentemente do intervalo dos dados (avaliações). Uma pesquisa abrangente sobre métricas para avaliar SR pode ser encontrada no artigo (SHANI; GUNAWARDANA, 2013).

## 2.3 Extração de Características Visuais

Em um SRI baseado em conteúdo, o processo de extração de características de imagens corresponde à obtenção automática de informações visuais, tais como cor, textura, forma, orientação, entre outros. Estas características são então utilizadas para descrever as imagens. Os extratores de características (descritores) são responsáveis por extrair tais características as quais são organizadas em vetores de características. Tais vetores são utilizados para o cálculo de similaridade entre as imagens. Geralmente, os vetores de características representam uma descrição compacta da imagem, tornando mais prática a sua análise do que trabalhar diretamente sobre os pixels da imagem. A seguir são apresentados os descritores de características utilizados nos experimentos do Capítulo 6.

### 2.3.1 Histograma de Gradientes Orientados (HOG)

O Histograma de Gradientes Orientados, do inglês *Histogram of Gradients* (HOG), proposto em (DALAL; TRIGGS, 2005), tem o intuito de representar a aparência e a forma local de objetos sob mudanças de iluminação e cor da imagem. O HOG é obtido a partir de uma grade densa e sobreposta de blocos com histogramas das orientações dos gradientes de uma imagem. Sua construção pode ser resumida em quatro etapas: cálculo de orientação e magnitude das arestas na imagem; divisão da imagem em blocos de células; cálculo do histograma de gradientes por células, agrupados em blocos; concatenação destes histogramas. Um vetor descritor será então a concatenação dos vários histogramas.

A primeira etapa do processo consiste em calcular os gradientes de primeira ordem da imagem. O cálculo dos gradientes captura contornos, silhuetas e um pouco de textura da imagem. As máscaras especiais usadas para este cálculo são dadas por:  $[-1 \ 0 \ 1]$  para o cálculo do gradiente na horizontal e  $[-1 \ 0 \ 1]^T$  para o cálculo do gradiente na vertical. O gradiente é calculado em cada canal de cor do modelo RGB, e o gradiente final, para cada pixel, é o gradiente com a maior magnitude entre os canais de cor.

A segunda etapa consiste na criação dos histogramas de orientação do gradiente, que tem como objetivo produzir uma codificação que seja sensível ao conteúdo local da imagem e ao mesmo tempo invariante a pequenas variações de pose e aparência. A imagem, ou parte da imagem dependendo da aplicação, é dividida em pequenas estruturas, denominadas *células*. Estas células possuem dimensões  $N \times N$  pixels, normalmente utiliza-se células de tamanho 8x8 pixels. A orientação de gradiente em cada um dos pixels da célula é usado para criar um histograma da orientação dos gradientes da célula. A orientação é classificada em uma das 9 divisões (*bins*), igualmente espaçados entre 0 e 180 graus. A contribuição de cada pixel para os histogramas é ponderada pelo valor da magnitude do gradiente na posição do pixel.

A terceira etapa consiste na normalização local dos histogramas em regiões formadas por um conjunto de células, denominadas *blocos*. A normalização introduz um aumento na

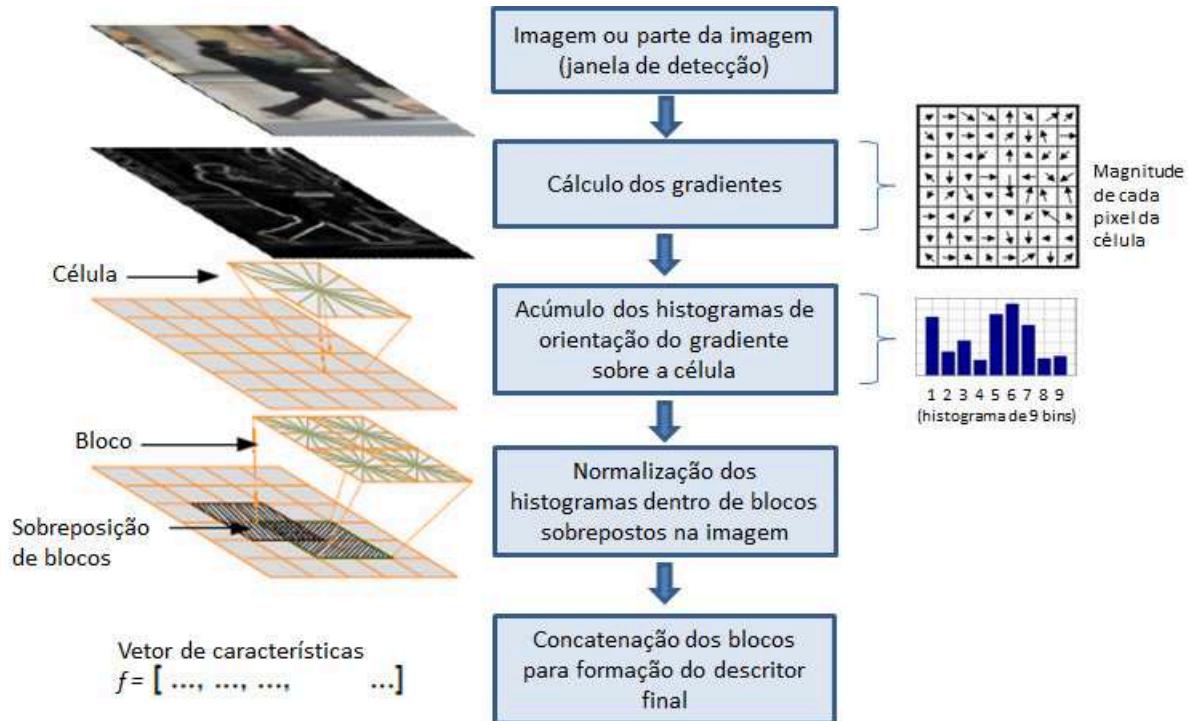


Figura 6 – Visão geral do cálculo do descritor HOG (DALAL; TRIGGS, 2005).

invariância à iluminação, sombreamento e contraste dos contornos. Cada bloco é formado por um conjunto de  $2 \times 2$  células. Antes da formação dos histogramas nas células, uma máscara Gaussiana do tamanho do bloco é aplicada sobre a magnitude do gradiente dos pixels do bloco. Os blocos possuem uma sobreposição de tamanho igual a uma célula, fazendo com que cada célula pertença a 4 blocos diferentes, com a exceção das células das bordas e dos cantos da imagem, que pertencem, respectivamente, a dois blocos e a um bloco. Isso faz com que cada célula seja normalizada em relação a diferentes blocos, aparecendo assim mais de uma vez no descritor final.

A quarta etapa consiste em concatenar todos os blocos presentes na imagem após a normalização, num único vetor, criando assim o descritor final. Por exemplo, se uma imagem tem  $128 \times 64$  pixels, e se utilizarmos células de  $8 \times 8$  pixels, então teremos  $16 \times 8$  células e  $15 \times 7$  blocos. Cada bloco é formado por  $2 \times 2$  células, sendo que cada célula possui um histograma de 9 bins. Logo, o descritor final (HOG) possui uma dimensão igual a  $15 * 7 * 2 * 2 * 9 = 3780$ .

A ideia básica por traz da utilização do descritor HOG é a de que o contorno e aparência locais de objetos pode muitas vezes ser bem caracterizada pela distribuição local das orientações do gradiente da imagem, mesmo sem o conhecimento exato da posição do gradiente. A Figura 6 ilustra todo o processo de construção do descritor HOG.

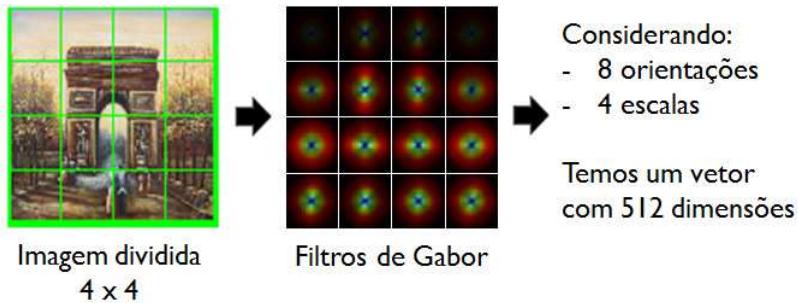


Figura 7 – Exemplo de um descritor GIST com 512 dimensões.

### 2.3.2 GIST

Um descritor global que se destaca na caracterização de cenas e recuperação de imagens é o GIST (o nome é um acrônimo de essência de cena, do inglês, *gist of the scene* (OLIVA; TORRALBA, 2006)). Originalmente, a ideia desse descritor é prover uma descrição em baixa dimensão de um cenário e que não necessite de segmentação. Esse descritor é invariante a transformações de luminância, obscuridade, escala, entre outras. No entanto, não é invariante a operações de translação, rotação e oclusão (DOUZE et al., 2009). Os autores propuseram transmitir um sumário estrutural que é significativo o suficiente para permitir a caracterização da imagem, denominado *Envelope Espacial*.

O descritor GIST leva em consideração o fato de que uma cena é um arranjo em um espaço tridimensional e a representa através de propriedades que funcionam como descritores do espaço que ela ocupa, as dimensões perceptuais (naturalidade, abertura, aspereza, expansão, e robustez). Os autores observaram que as dimensões perceptuais capturam a maioria das estruturas tridimensionais das cenas do mundo real. Além disso, cenas que compartilham a mesma categoria semântica tendem a se aglomerar dentro da mesma região de um espaço multidimensional em que os eixos são as propriedades perceptuais, ou seja, possuem dimensões perceptuais similares. O conjunto de dimensões espaciais perceptuais forma o envelope espacial de uma cena.

Uma cena pode ser representada por uma combinação de características globais. Uma opção de representação é a combinação ponderada da saída de bancos de filtros multiescalares e multiorientados, como os filtros de Gabor. O descritor GIST divide a imagem em  $N \times N$  células e aplica os filtros de Gabor em cada uma das células para diferentes [orientações] e [escalas]. O formato do descritor GIST é um vetor com [escalas] \* [orientações] \*  $N \times N$  dimensões. A Figura 7 mostra o esquema do descritor GIST.

### 2.3.3 Histograma de Cor

A cor é uma das características mais utilizadas para representar uma imagem no processo de comparação de imagens (LU, 1999). Dentre os principais motivos, pode-se destacar: os descritores são facilmente entendidos e fáceis de serem implementados; a

informação de cor está presente na grande maioria dos domínios de imagens; a informação de cor pode ser processada de forma automática; e os resultados obtidos pela utilização da informação de cor, normalmente, são satisfatórios (LU, 1999).

Os modelos de representação de cores mais conhecidos são o RGB (*red, green, blue*), que mapeia cada cor nos seus componentes vermelho, verde e azul e o HSV (*hue, saturation, value*), que mapeia mais fielmente a percepção humana com o apelo intuitivo das designações matiz (cor), sombra (saturação) e tonalidade (valor) (LU, 1999). A representação das cores de uma imagem mais comum é o histograma.

Os histogramas representam a quantização dos espaços de cores pela contagem do número de pixels que cada cor quantizada possui na imagem. Os histogramas podem ser classificados como: globais, baseados em partições regulares e baseados em regiões (KIMURA et al., 2011). Os histogramas globais consideram toda a imagem no processo de quantização dos pixels. Os histogramas são invariantes à translação e rotação das imagens, entretanto, os histogramas sofrem pela falta de informação espacial dos pixels na imagem. Para diminuir este problema, pode-se utilizar histogramas de cor locais: abordagem baseada em partições regulares ou abordagem baseada em regiões. Na abordagem baseada em partições regulares, a imagem é dividida em células de tamanho fixo, extraindo o histograma de cor de cada célula separadamente. Na abordagem baseada em regiões, a imagem é segmentada em regiões utilizando um algoritmo de segmentação. A quantidade de regiões, tamanho, localização e forma podem variar de uma imagem para outra.

Nos experimentos descritos no Capítulo 6 serão utilizados histogramas de cor globais (para a base de dados de pinturas) e histogramas de cor locais (para a base de dados de roupas).

### 2.3.4 Padrões Binários Locais (LBP)

O descriptor *Local Binary Pattern histogram* (LBP) proposto em (OJALA; PIETIKÄINEN; MÄENPÄÄ, 2002) é bastante utilizado para descrever textura. Este descriptor baseia-se na conjuntura que padrões binários locais e a região de vizinhança de um pixel são características fundamentais na textura da imagem. Dessa forma, o histograma formado a partir destas características é uma boa maneira para se representar a textura.

No descriptor LBP, a intensidade  $v_c$  de cada pixel de uma imagem é substituída por um vetor binário  $b$ , determinado pela comparação entre a intensidade do pixel e as intensidades dos pixels vizinhos. Em cada comparação, se o valor do pixel central for menor que o seu vizinho, ele é marcado com 1, caso contrário é marcado com 0. Os valores obtidos para cada vizinho são concatenados e o número binário gerado é convertido na base decimal para substituir o valor central  $v_c$ . A Figura 8 ilustra este processo, para uma vizinhança (máscara) de 3 x 3 pixels, e apresenta o resultado gerado por esta transformação.

A imagem gerada é dividida em regiões, habitualmente organizadas em uma malha regular e, em seguida, para cada região isoladamente, calcula-se o histograma normalizado

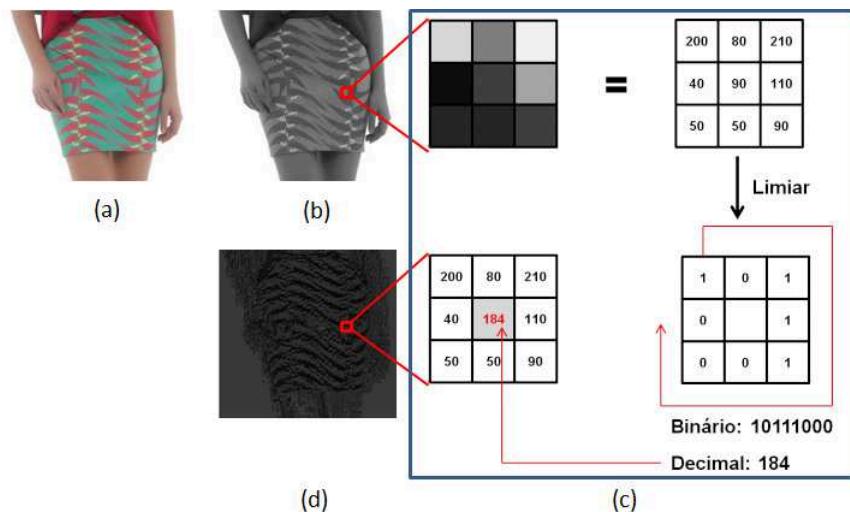


Figura 8 – Processo de extração do descritor LBP: (a) imagem original; (b) imagem em níveis de cinza; (c) cálculo do LBP; (d) imagem gerada após a atualização dos pixels.

de intensidade dos pixels, e os valores resultantes são concatenados em um único vetor.

### 2.3.5 Descritor de Pele (*Skin*)

Existem diversos descritores de pele descritos na literatura, alguns são utilizados apenas para a identificação da presença da face humana, da mão ou de qualquer outra parte do corpo humano (JAIN, 2011; SONG et al., 2011), e outros são utilizados para análise da pele com diversas finalidades, tais como, detecção de problemas dermatológicos (BOSMAN; PETKOV; JONKMAN, 2010; LASKARIS et al., 2010; KAUR; AGGARWAL, 2013), detecção de pornografia (LIU et al., 2008), etc.

Alguns descritores de presença de pele são baseados no tom da pele (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). A cor da pele em uma imagem é sensível a vários fatores, tais como, iluminação variável, ajustes de câmera, tipos de pele humana de etnia diferente, e também ao ambiente externo (penteado, óculos e cores de fundo), tornando o problema desafiador. Uma pesquisa abrangente de técnicas de detecção de tons de pele pode ser encontrada nos artigos (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007; TERRILLON et al., 2003).

O descritor de pele utilizado nos experimentos do Capítulo 6 é baseado no tom da pele e foi proposto em (JAIN, 2011). Inicialmente, a imagem é convertida para o modelo YCbCr, onde o espaço de cor RGB é linearmente transformado em Y (luminância) e dois componentes de diferença-azul Cb e diferença-vermelho Cr. Este espaço de cor é motivado pelo fato de que o sistema visual humano tem consideravelmente menos acuidade espacial para informação de cor do que para o brilho e é usado principalmente para a transmissão reduzida de dados para a televisão e vídeo. A conversão de RGB para YCbCr é descrita

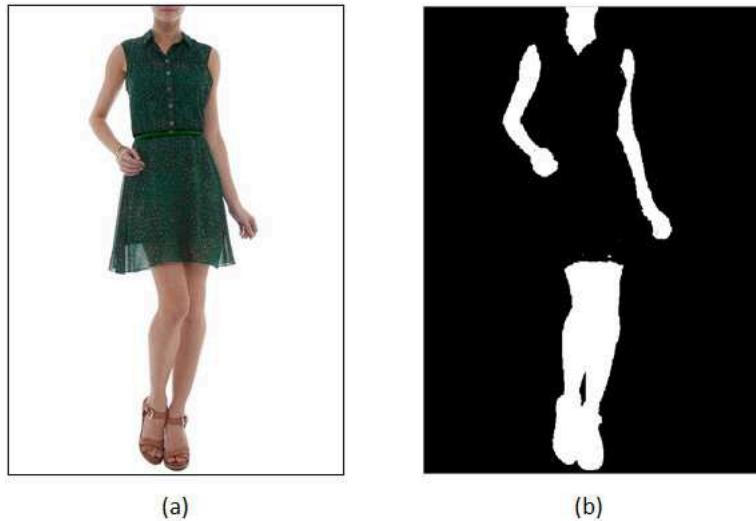


Figura 9 – Utilizando o descritor de presença de pele proposto em (JAIN, 2011): (a) imagem original; (b) imagem binária onde a cor branca representa a presença de pele.

na Equação 30.

$$\begin{aligned}
 Y &= 16 + 1/256 * (65,738R + 129,057G + 25,064B), \\
 Cb &= 128 + 1/256 * (-37,945R - 74,494G + 112,439B), \\
 Cr &= 128 + 1/256 * (112,439R - 94,154G - 18,285B),
 \end{aligned} \tag{30}$$

onde os valores de R, G, B e Y são tipicamente entre 0 e 255, e os valores dos componentes Cb e Cr estão no intervalo de 16 a 240.

O descritor baseado no tom de pele proposto em (JAIN, 2011) descarta o canal de luminância Y e define a região de pele dentro da distribuição Cb e Cr descrita na Equação 31. A Figura 9 mostra um exemplo de detecção de presença de pele em uma imagem utilizando este descritor.

$$77 \leq Cb \leq 127, 133 \leq Cr \leq 173 \tag{31}$$

## 2.4 Atenção Visual Humana

Os seres humanos se deparam com uma enorme quantidade de informações visuais. No entanto, esta quantidade é muito maior que toda a informação que o cérebro é capaz de processar e assimilar (DUCHOWSKI, 2007). Por meio de movimentos rápidos dos olhos, conhecidos como sacadas (*saccades*), o cérebro prioriza as informações relevantes e recebe somente parte da informação visual a cada instante (DUCHOWSKI, 2007).

A atenção visual é a habilidade que o sistema visual humano utiliza para selecionar e processar somente as regiões mais importantes, chamadas de regiões de interesse (do

inglês *Regions of Interest* (ROI)), de uma cena visual e ignorar o restante das informações. A atenção visual pode ser obtida por meio de dispositivos de rastreamento do olhar (*eye tracker*).

Segundo a empresa Tobii (TOBII, 2015), líder no ramo de atenção visual, o termo Rastreamento do Olhar (RO), do inglês *eye tracking*, apresenta dois significados distintos. Ele pode ser usado para representar técnicas que são capazes de determinar o ponto que está sendo visto pelo usuário, ou seja, onde o usuário está olhando. Por outro lado, este termo também pode ser usado para designar apenas as técnicas capazes de localizar a posição do olho do usuário em relação a sua cabeça. Outro termo em inglês *eye gaze tracking* também é bastante utilizado para representar a primeira definição e evitar a ambiguidade. Nesta tese, adota-se a primeira definição para o termo RO.

Pesquisas relacionadas com o RO iniciaram no fim do século XIX. Rayner (RAYNER, 1998) dividiu o estudo dos movimentos dos olhos em três eras. Na primeira era, que aconteceu entre 1879 a 1920, surgiram os primeiros dispositivos capazes de registrar os movimentos dos olhos de forma automática. Estes dispositivos eram invasivos, exigiam contato direto com os olhos dos participantes, impossibilitando qualquer movimento de cabeça. Os dispositivos eram compostos por lentes de contato com um furo na região da pupila (centro da visão). Estas lentes eram conectadas em outros dispositivos que registravam com grande dificuldade os movimentos dos olhos (TOBII, 2015). Naquela época, o estudo estava relacionado com problemas oftalmológicos.

A segunda era aconteceu entre as décadas de 1930 e 1950. Destacaram-se a criação do primeiro dispositivo montado na cabeça do participante, que permitiu maior liberdade de movimentação aos participantes, e os primeiros estudos de usabilidade foram com os movimentos dos olhos de um piloto durante o pouso de uma aeronave (JACOB; KARN, 2003). Também foram iniciados estudos relacionados com distúrbios de leitura e percepção de cenas.

Após alguns anos sem grandes inovações na área, iniciou-se a terceira era na metade da década de 1970. Esta era foi marcada pelo surgimento dos microcomputadores que possibilitou o RO em tempo real, grandes melhorias de coleta e gravação dos movimentos dos olhos, aumento na precisão dos dados, e maior liberdade de movimentação da cabeça.

Atualmente o RO é usado para responder uma infinita variedade de questões relacionadas a diversas áreas, tais como: testes de usabilidade, interação homem-computador, psicologia, publicidade e recuperação da informação.

- **Testes de usabilidade:** O RO pode ser usado para testar a usabilidade de websites e software (LENZNER; KACZMIREK; GALESIC, 2014; ALLEY et al., 2014; CARENINI et al., 2014). O RO também é usado para estudar a interação do usuário com dispositivos móveis, jogos, televisão interativa, interfaces de mapas digitais e outros dispositivos físicos. A atenção visual é capaz de fornecer informações adicionais

que o participante do teste não consegue relatar ao pesquisador. O RO mostra as reações imediatas dos usuários e a distribuição de sua atenção em uma interface.

- **Interação Homen-Computador:** O RO pode ser usado com um dispositivo de entrada e não somente como uma ferramenta para coletar dados (FUJII et al., 2013; HUANG; LO; SHI, 2013; DEBELJAK; OCEPEK; ZUPAN, 2012). Os movimentos dos olhos são registrados e utilizados em tempo real como informações de entrada e controle para o computador. Este recurso pode ser utilizado principalmente para usuários com deficiência.
- **Psicologia:** O RO ajuda os pesquisadores a entender aspectos cognitivos, percepção e habilidades de pessoas desde a infância até a idade adulta (OLÁH et al., 2014; DAVIDSE et al., 2014). Antes mesmo das crianças aprenderem a fala, o RO pode fornecer informações detalhadas sobre como elas percebem o mundo.
- **Embalagens e publicidade:** A maioria das decisões de compra, em um supermercado por exemplo, são feitas em uma fração de segundo. Os produtos devem ser visualmente atraentes e captar rapidamente a atenção dos clientes. O RO é muito poderoso para avaliar e melhorar o conteúdo e a distribuição das informações em: embalagens de produtos, publicidade de sites, banner e outdoor (POPA et al., 2015; CIAN; KRISHNA; ELDER, 2014; VELAZQUEZ; PASCH, 2014).
- **Recuperação da informação:** Dado um conjunto de informações na tela de um computador, o usuário pode selecionar objetos ou regiões que são de seu interesse. A atenção visual nestes objetos e/ou regiões pode ajudar o sistema a entender as preferências do usuário e melhorar a recuperação da informação (JUNG et al., 2013; LIANG et al., 2010; ZHANG; SHEN; FENG, 2008).

#### 2.4.1 Tipos de Rastreadores de Olhar

Os primeiros Rastreadores de Olhar (ROs) eram compostos por máquinas semelhantes aos equipamentos utilizados atualmente em consultórios oftalmológicos. O usuário posiciona a cabeça no equipamento e não pode se mover. Atualmente, os ROs são baseados em câmeras de vídeo e possuem algoritmos que detectam a posição dos olhos e a orientação da cabeça do usuário. Com isso, os ROs são mais confortáveis e permitem experimentos mais duradouros.

Embora as tecnologias atuais dispensem restrições de movimentos da cabeça, ainda existem alguns ROs que exigem que as pessoas fiquem com as cabeças fixas, por conta do baixo custo do equipamento ou de sua alta precisão. A Figura 10 mostra dois exemplos de dispositivos montados em um desktop, um com a cabeça do usuário fixa e outro com liberdade limitada para o usuário.



Figura 10 – Rastreadores de olhar montados no desktop. (a) o usuário encaixa a cabeça no rastreador de olhar e não permite movimentos ([www.smivision.com](http://www.smivision.com)). (b) o rastreador de olhar é um pequeno dispositivo colocado no monitor e o usuário possui uma liberdade limitada de movimentos da cabeça ([www.tobii.com](http://www.tobii.com)).



Figura 11 – Rastreadores de olhar montados na cabeça ([www.tobii.com](http://www.tobii.com)). (a) pesquisas relacionadas com padrões de compras, embalagens de produtos e disposição dos produtos na loja. (b) comportamento do usuário ao volante. (c) pesquisas nas áreas de esporte e psicologia.

Uma categoria de ROs, denominada dispositivos montados na cabeça, se destaca pela capacidade de obter dados do comportamento do usuário em qualquer ambiente do mundo real, tais como, lojas, restaurantes e no trânsito. A Figura 11 mostra três diferentes aplicações destes ROs. No primeiro exemplo, o usuário está no supermercado e os pesquisadores analisam seus padrões de compra e a eficiência das embalagens. No segundo exemplo, os pesquisadores analisam o que chama a atenção do usuário ao dirigir um carro. No terceiro exemplo, a atenção visual mostra a velocidade de raciocínio juntamente com a tomada de decisão do jogador de basquete.

Infelizmente, os dispositivos de rastreamento do olhar comerciais são muito caros. Alguns pesquisadores, portanto, têm se voltado para o uso de câmeras de web comuns como dispositivos para obter imagens dos olhos (KIM; RAMAKRISHNA, 1999; DARRELL et al., 2002; RUDDARRAJU et al., 2003; LIN et al., 2006; SU; SU; CHEN, 2005; GORODNICHY, 2006) e um algoritmo de rastreamento do olhar, como por exemplo o projeto Opengazer desenvolvido pela *Gatsby Foundation* e *Samsung* (ZIELINSKI, 2015). Outra alternativa é a predição da atenção humana em uma imagem por meio de modelos computacionais

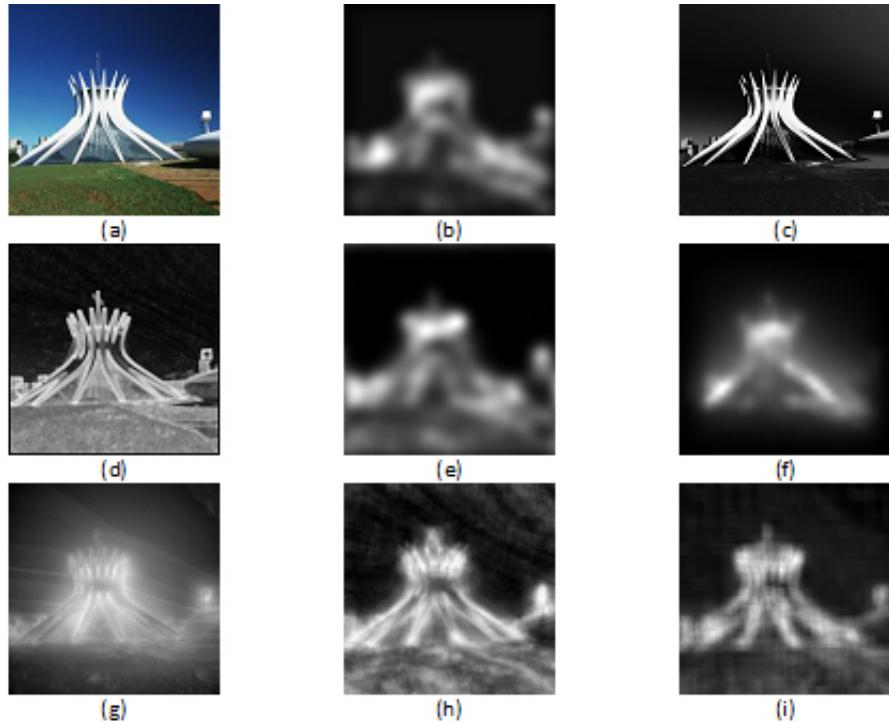


Figura 12 – Mapas de Saliência utilizando diferentes modelos computacionais. (a) imagem original; (b)(ITTI; KOCH; NIEBUR, 1998); (c)(ACHANTA et al., 2009); (d)(BRUCE; TSOTSOS, 2005); (e)(GARCIA-DIAZ et al., 2009); (f)(HAREL; KOCH; PERONA, 2006); (g)(JUDD et al., 2009); (h)(MURRAY et al., 2011); (i)(ZHANG; SHEN; FENG, 2008).

(BORJI; ITTI, 2013).

#### 2.4.2 Modelos Computacionais de Predição de Atenção Visual

Vários modelos computacionais têm sido propostos para tentar predizer a atenção visual em uma cena sem a necessidade de ROs (BORJI; ITTI, 2013). Estes modelos têm aplicabilidade em diversas áreas: recuperação da informação, assistência à navegação, controle de robôs, sistemas de vigilância, segmentação e detecção de objetos, reconhecimento e compreensão de cena, projetos de publicidade (MELO; AMO; GULIATO, 2014; GASPARINI; CORCHS; SCHETTINI, 2007; LI; ZHAO, 2011; HATA; TOYOURA; MAO, 2012).

Dada uma cena, os modelos computacionais de atenção visual produzem um mapa de saliência que indica a conspicuidade de cada pixel da imagem, ou seja, a probabilidade de uma região atrair a atenção de humanos. O mapa de saliência é visualizado como uma imagem em níveis de cinza, em que o brilho de um pixel é proporcional à sua saliência. A Figura 12 apresenta o resultado da aplicação de diferentes modelos de atenção visual a uma imagem original.

A modelagem da atenção visual tem sido uma área de pesquisa muito ativa desde a década de 1990. A predição da atenção visual é influenciada por dois principais fatores: *bottom-up* e *top-down*. A atenção visual *bottom-up* é guiada por meio de elementos da

cena visual que se destacam dos demais a partir de características primitivas, tais como, descontinuidade de intensidade em diferentes escalas, cor e orientações, sem qualquer informação contextual. A atenção visual *top-down* é guiada por meio de informações do córtex cerebral, ou seja, busca por regiões de interesse a partir de características de alto nível, tais como, conhecimento e expectativa do indivíduo. Uma pesquisa abrangente sobre diferentes modelos de predição de atenção visual pode ser encontrada no artigo (BORJI; ITTI, 2013).

## 2.5 Considerações Finais

Neste capítulo foram apresentados diversos conceitos e propriedades preliminares relacionados a sistemas de recomendação. Foram enfatizados os principais trabalhos envolvendo formalismos para a representação dos dados (itens, usuários e avaliações), cálculos de similaridade e a predição de avaliação. Tanto os conceitos apresentados quanto os trabalhos abordados são essencialmente importantes, pois serão utilizados e referenciados em diversos momentos neste trabalho.

Tendo em vista que os métodos propostos neste trabalho utilizam atenção visual, fez-se importante abordar os principais conceitos e mecanismos para a extração das fixações dos olhares e dos movimentos dos olhos. Como os dispositivos de rastreamento do olhar comerciais são muito caros, discutimos também maneiras alternativas com custo baixo ou zero para a extração da atenção visual, porém com menor eficácia.

Neste trabalho daremos ênfase ao uso de informações capturadas por dispositivos de rastreamento do olhar, mas a nossa abordagem é facilmente ajustável para o uso de modelos computacionais de atenção visual.

# CAPÍTULO 3

---

## Trabalhos Correlatos

O foco principal desta tese é o desenvolvimento de métodos para SRI em que as informações de atentividade são combinadas no cálculo de similaridade entre produtos. Este capítulo apresenta os trabalhos mais relevantes que abordam temas relacionados com esta tese: i) trabalhos baseados em vizinhança que combinam similaridades; ii) trabalhos que tem por objetivo a recomendação de imagens; iii) trabalhos que usam atenção visual como *feedback* implícito em SR.

### 3.1 Sistemas de Recomendação que Combinam Similaridades

Nos SR baseados em vizinhança, a capacidade de identificar vizinhos é a parte mais significativa para a melhoria da qualidade na recomendação. Os valores de similaridade desempenham dois papéis nestes SR: i) permitem a seleção de vizinhos confiáveis, cujas avaliações são utilizadas na predição, e ii) fornecem os meios de ponderar a importância destes vizinhos na predição.

A agregação de duas ou mais similaridades pode ser realizada utilizando funções com propriedades e comportamentos diferentes. As funções lineares são as mais comuns, entretanto as funções não-lineares também podem ser utilizadas pois em determinadas situações permitem refinar o processo de predição de avaliações, criando predições relevantes e mais precisas. O objetivo das funções de agregação é combinar similaridades (entradas) que normalmente são interpretadas como graus de preferência, força de evidência, ou apoio a uma hipótese, e assim por diante (BELIAKOV; CALVO; JAMES, 2011).

Nesta tese, são consideradas funções de agregação definidas sobre o intervalo  $f : [0, 1]^n \rightarrow [0, 1]$ , entretanto, outras opções poderiam ser utilizadas. O valor de  $n$  é a quantidade de entradas e o valor de entrada 0 é interpretado como nenhuma associação entre os itens comparados, ou nenhuma evidência de similaridade entre os itens, e naturalmente, uma agregação de 0s tem como resultado 0. Da mesma forma, o valor 1 é interpretado

como total associação entre os itens, a mais forte evidência de similaridade entre os itens, e uma agregação em que todas as entradas são iguais a 1 produz como resultado o valor 1.

As funções de agregação também exigem monotonicidade em cada argumento, ou seja, um aumento em qualquer entrada não pode resultar em uma diminuição no resultado final da agregação.

**Definição 1. Função de Agregação.** *Uma função de agregação é uma função de  $n$  argumentos com  $n > 1$ , que mapeia  $n$ -dimensões para o intervalo unitário  $f : [0, 1]^n \rightarrow [0, 1]$ , com as propriedades*

- (i)  $f(\underbrace{0, 0, \dots, 0}_{n\text{-vezes}}) = 0$  e  $f(\underbrace{1, 1, \dots, 1}_{n\text{-vezes}}) = 1$ .
- (ii)  $x \leq y$  implica  $f(x) \leq f(y)$  para todos  $x, y \in [0, 1]^n$ .

A função de agregação é classificada de acordo com o seu comportamento global em relação às entradas (CALVO et al., 2002). A escolha da função de agregação depende das características das medidas de similaridade sendo usadas como entrada e quanto sensível ou ampla a saída é desejada.

**Definição 2. Classes.** *Uma função de agregação é classificada como:*

- (i) *Média.* Se é limitada por  $\min(x) \leq f(x) \leq \max(x)$ ;
- (ii) *Conjuntiva.* Se é limitada por  $f(x) \geq \min(x)$ ;
- (iii) *Disjuntiva.* Se é limitada por  $f(x) \leq \max(x)$ ;
- (iv) *Misturada.* Caso contrário.

Ao agregar medidas de similaridades usando a classe Média, garante-se que a similaridade entre dois itens é representada pela tendência central das similaridades. Por outro lado, dependendo das características das entradas, a classe Misturada pode ser mais atraente para o SR (BELIAKOV; CALVO; JAMES, 2011).

As funções de agregação da classe Média são as mais utilizadas na combinação de similaridades em SR. Dado um vetor de pesos  $w = (w_1, \dots, w_n)$ , onde  $\sum w_j = 1$  e  $w_j \geq 0 \forall j$ . Os casos especiais mais utilizados da classe Média são (BELIAKOV; CALVO; JAMES, 2011):

□ *Médias aritméticas:*

$$WAM_w = \sum_{j=1}^n w_j \cdot x_j \quad (32)$$

□ *Médias geométricas:*

$$G_w = \prod_{j=1}^n x_j^{w_j} \quad (33)$$

□ *Médias harmônicas:*

$$H_w = \left( \sum_{j=1}^n \frac{w_j}{x_j} \right)^{-1} \quad (34)$$

□ *Médias de potência:*

$$M_{w,[r]} = \left( \sum_{j=1}^n w_j \cdot x_j^r \right)^{\frac{1}{r}} \quad (35)$$

em que  $x_j$  é o valor de entrada da medida de similaridade  $j$  ponderada pelo peso  $w_j$ .

Normalmente, o peso atribuído a uma dada entrada indica a importância daquela entrada na função de agregação. As funções  $WAM_w, G_w, H_w$  e  $M_{w,[r]}$  são idempotentes, estritamente monótonas e homogêneas.

**Definição 3. Propriedades.** Uma função de agregação é:

- (i) *Idempotente.* Para cada  $t \in [0, 1]$  a saída de  $f(t, t, \dots, t) = t$ ;
- (ii) *Estritamente monótona.* Se  $x \leq y$  e  $x \neq y$  implica  $f(x) \leq f(y)$ ;
- (iii) *Homogeneidade.* Para todo  $\lambda \in [0, 1]$  e para todo  $x = (x_1, \dots, x_n)$ ,  $f(\lambda x_1, \dots, \lambda x_n) = \lambda f(x)$ .

No trabalho apresentado em (STANESCU; NAGAR; CARAGEA, 2013), os autores definem uma função de agregação  $WAM_w$  para combinar duas similaridades. A primeira similaridade considera um conjunto de filmes com *tags* que representam tópicos e a segunda similaridade considera as avaliações sobre os filmes. Em (MOBASHER; JIN; ZHOU, 2004) os autores também usam uma  $WAM_w$  para combinar duas medidas de similaridades. A primeira considera relações entre conceitos representados pelo website e a segunda medida de similaridade considera as avaliações sobre os itens. Em (LI; KIM, 2003) os autores usam um algoritmo de clusterização para agrupar itens. A similaridade entre dois itens é uma combinação  $WAM_w$  de duas similaridades baseadas em avaliações. A primeira similaridade considera todos os itens da base de dados e a segunda similaridade considera apenas os itens do grupo que o item (cuja avaliação será predita) foi alocado no processo de clusterização. Esta abordagem é interessante para diminuir o problema de partida a frio. Em (DONG et al., 2015) os autores também definem uma combinação linear entre duas similaridades. A primeira medida similaridade considera características voltadas ao sentimento dos usuários com relação aos itens, e a segunda similaridade considera as avaliações.

No trabalho descrito em (CHOI; SUH, 2013), os autores utilizam uma combinação geométrica  $G_w$  para calcular a similaridade entre dois usuários. A combinação é formada pela agregação de duas medidas de similaridades baseadas em avaliações, a primeira é calculada pela similaridade entre itens e a segunda é calculada pela similaridade entre usuários. A ideia é minimizar o problema de partida a frio. Em (BODHIT; AMIN, 2013) os autores também descrevem uma combinação geométrica utilizando três medidas de similaridades entre usuários: Proximidade, Impacto e Popularidade. O fator Proximidade é baseado na diferença aritmética simples entre duas avaliações, mas considera ainda se as duas avaliações estão de acordo ou não, penalizando avaliações em desacordo. Neste caso, se duas avaliações estão no mesmo lado de uma determinada escala de avaliação obtida

pela mediana, então as avaliações são consideradas estarem de acordo. O fator Impacto considera o quanto fortemente um item é o preferido ou destacado pelos usuários. Quando um item é fortemente preferido ou não, pode-se considerar que a preferência do usuário pelo produto foi expressa claramente, e, portanto, maior credibilidade pode ser dada para o resultado da similaridade. O fator Popularidade dá maior importância para avaliações que estão mais distantes da avaliação média do item.

A combinação de similaridade pode ser usada com diversas características e de diversas formas. Como as imagens podem ser representadas por várias características, a combinação de similaridade pode ser uma boa alternativa para melhorar SRI baseados em vizinhança.

## **3.2 Sistemas de Recomendação de Imagens**

A diferença entre os termos “Recomendação de imagens” e “Recuperação de imagens” não está bem definida na literatura. Vários trabalhos, como (CHAO et al., 2009; IWATA; WANATABE; SAWADA, 2011; HUANG; LO; SHI, 2013) são descritos como “Recomendação de imagens”. Entretanto, estes trabalhos, na verdade, apresentam estratégias para recuperar uma lista ordenada de imagens semelhantes a uma imagem de consulta, considerando apenas a semelhança entre as características da imagem e ignorando o comportamento dos usuários.

Nesta tese, entendemos como recomendação de imagens a tarefa que considera o comportamento passado dos usuários (compras, navegação na internet, avaliações) para personalizar a recuperação de produtos representados por imagens. A recomendação de tais imagens pode ser aplicada em diferentes situações. Os SRI baseados em conteúdo focam em determinar características que descrevem as imagens e medidas de similaridade para identificar a similaridade entre as imagens. Os SRI baseados em FC usam apenas o comportamento passado dos usuários, representado pela matriz de utilidade, e não requerem as características das imagens.

Widisinghe et al. (WIDISINGHE et al., 2010) descrevem uma nova abordagem FC, denominada picSEEK, que considera o contexto para a recomendação de imagens. O usuário informa o contexto desejado, por exemplo, “estudante de medicina” e então o sistema recomenda imagens. O método possui três etapas: i) calcular contextos similares ao informado pelo usuário; ii) calcular a similaridade entre o usuário atual e os demais usuários considerando apenas imagens selecionadas (feedback implícito apenas) no contexto semelhante ao informado pelo usuário atual; iii) utilizar um algoritmo KNN para realizar a predição do ranking de imagens para então recomendá-las. A limitação deste trabalho é a necessidade dos produtos serem categorizados de acordo com o contexto e a necessidade do usuário informar o contexto para poder receber a recomendação.

Os autores em (TKALČIČ; BURNIK; KOŠIR, 2010) investigam a influência da afetividade

no desempenho da recomendação de imagens. Eles propuseram modelar a preferência do usuário em relação aos estados emotivos (raiva, alegria, surpresa, aversão, medo, neutro e tristeza). A hipótese é que uma imagem (por exemplo, as pinturas famosas *Scream* de Edvard Munch e *Poppies Blooming* de Claude Monet) pode provocar diferentes estados emotivos nos usuários. O perfil da imagem é representado por dados genéricos, como gênero e tempo médio de observação, e metadados afetivos que indicam uma estatística das emoções que a imagem induz nos usuários. O perfil do usuário pode ser modelado e aprendido de diferentes maneiras, dependendo da máquina de aprendizagem utilizada (árvore de decisão, máquina de vetores de suporte, etc.). Os dados de treinamento são avaliações anteriores e os estados emotivos induzidos pelas imagens observadas pelo usuário. Com isso, o modelo classifica os itens que ainda não foram avaliados pelo usuário a fim de ordená-los para a recomendação.

Barrilero et al. (BARRILERO et al., 2011) apresentam uma abordagem que usa apenas características visuais (luminância, crominância e textura) para descrever uma imagem. Um algoritmo KNN com distância Mahalanobis (MAESSCHALCK; JOUAN-RIMBAUD; MAS-SART, 2000) é usado para a predição da avaliação das imagens. Um diferencial dessa abordagem é que os dados podem ser alocados dentro de uma rede nas nuvens.

Kim et al. (KIM; KIM; CHO, 2008) propuseram uma solução híbrida para a recomendação de imagens. Uma abordagem FC é utilizada para obter os itens vizinhos mais próximos e a partir dos vizinhos obtidos, uma abordagem por conteúdo é utilizada para prever a importância das imagens para o usuário. As imagens são representadas pelo descriptor Momentos de cor (STRICKER; ORENGO, 1995). A principal contribuição do trabalho é que o processo de recomendação é realizado em uma arquitetura distribuída *peer-to-peer*, diferentemente do processo tradicional que normalmente é realizado em único servidor central.

Albanese et al. (ALBANESE et al., 2013) propuseram uma nova abordagem para a recomendação em sistemas de navegação multimídia. A técnica utilizada para entender as preferências dos usuários é baseada na técnica *PageRank*. Em um sistema de navegação multimídia, se um usuário seleciona um item  $i$  imediatamente após um item  $j$ , é implícito afirmar que existe uma relação entre os itens  $i$  e  $j$  e eles podem ser semelhantes. Os autores modelam o comportamento de cada usuário utilizando um grafo direcionado, onde os nós representam os itens e as arestas representam as relações entre os itens. O comportamento geral de todos os usuários também é representado por um grafo. Uma função qualquer pode ser definida para calcular a similaridade entre os itens, considerando características de baixo nível. O cálculo da recomendação personalizada é realizado com uma combinação linear do comportamento de navegação do usuário, da comunidade (global) e da similaridade intrínseca entre os itens. Uma contribuição importante deste trabalho é endereçar o problema de “novo item”, em que os autores consideram apenas as características de baixo nível da imagem, e o problema de “novo usuário”, em que desconsideram

apenas o comportamento da comunidade no processo de recomendação. Como estudo de caso, os autores criaram um sistema que ajuda os usuários a navegar na galeria de pinturas Uffizi, Florença, Itália. Descritores MPEG-7 (MANJUNATH; SALEMBIER; SIKORA, 2002) foram utilizados para caracterizar imagens.

Boutemedjet e Ziou (BOUTEMEDJET; ZIOU, 2008) propuseram um modelo híbrido, denominado CBIS, que representa as preferências do usuário usando um método probabilístico que combina avaliações, conteúdo da imagem (representado por dados textuais e visuais) e situações contextuais. Os experimentos foram conduzidos usando uma base de dados com 4775 imagens de fotografias com 87 tipos de palavras-chave. As características visuais utilizadas foram: correlograma de cor, histograma de borda e um histograma de 100 pontos chaves obtido pelo agrupamento de SIFTs da imagem usando a técnica de *bag-of-visual-words* (LOWE, 2004).

Os trabalhos descritos nesta seção apresentam limitações em diferentes situações. Em (WIDISINGHE et al., 2010; BOUTEMEDJET; ZIOU, 2008) a limitação está na necessidade dos produtos serem categorizados de acordo com o contexto e a necessidade do usuário informar o contexto para poder receber a recomendação. O trabalho descrito (TKALČIČ; BURNIK; KOŠIR, 2010) é destinado apenas a produtos que são fortemente relacionados a emoção. Em (BARRILERO et al., 2011; KIM; KIM; CHO, 2008) a descrição das imagens é pobre, com apenas alguns descritores visuais, prejudicando a recomendação de imagens. Por fim, o trabalho apresentado em (ALBANESE et al., 2013) é limitado a sistemas de navegação.

Em SRI, a atenção visual humana é uma característica importante para descrever imagens. A grande vantagem é que a atenção visual pode ser obtida sem qualquer esforço por parte do usuário e por parte do administrador do sistema. Em nosso melhor conhecimento, não existem trabalhos que descrevem produtos de acordo com sua atentividade. Nesta tese está sendo proposto o uso da atenção visual para descrever os produtos de forma implícita via rastreador de olhar. Um resumo dos trabalhos relacionados com SRI é apresentado na Tabela 5.

### 3.3 Sistemas de Recomendação que Utilizam Rastreadores de Olhar

Como os usuários são geralmente pouco interessados em fornecer informações aos SR (*feedback explícito*), alguns trabalhos usam apenas o *feedback* implícito para recomendar itens aos usuários. Os feedbacks implícitos mais populares são o histórico de compras e o clique de dados. O uso da atentividade pode ser uma alternativa importante e não intrusiva para obter um retorno do usuário.

Os autores em (JUNG et al., 2013) descrevem um Algoritmo Evolutivo Interativo (AEI), onde as informações do olhar e das sacadas são obtidas via um rastreador de olhar e usadas

Tabela 5 – Comparaçāo dos trabalhos relacionados com recomendação de imagens.

<b>Trab. Relacio- nado</b>	<b>Foco</b>	<b>Tipo SR</b>	<b>Abordagem</b>	<b>Entrada</b>	<b>Estudo de Caso</b>
(WIDI-SINGHE et al., 2010)	Contexto	SRFC	Método KNN: similaridade entre contextos e similaridade usuário-usuário	contexto e usuário	de 10 a 100 imagens no domínio da medicina e 1000 imagens ao todo
(TKALČIČ; BURNIK; KOŠIR, 2010)	Emoção	SRBC	Vários podem ser aplicados: SVM, árvore de decisão, etc.	usuário	70 imagens da International Affective Picture System (IAPS) e 14 usuários
(BARRILERO et al., 2011)	C. visuais (luminância, crominância e textura) e o SR em rede	SRBC	Método KNN: similaridade entre itens	usuário	200 pinturas da galeria de arte Ciudad de la pintura e 60 usuários
(KIM; KIM; CHO, 2008)	C. visuais (momento de cor HSV) e arquitetura distribuída	Híbrido	Método KNN: Etapa 1: similaridade BC dos itens; Etapa 2: FC usuário-usuário sobre o resultado da Etapa1	usuário	1000 imagens do Korea Telecom Freetel (KTF) e 1000 usuarios
(ALBA-NESE et al., 2013)	Técnica PageRank (navegação)	Híbrido	Modelagem utilizando grafo. Combinação da similaridade FC usuário-usuário e similaridade BC dos itens	usuário	474 pinturas da galeria Uffizi e 47 usuários
(BOUTE-MEDJET; ZIOU, 2008)	Contexto	Híbrido	Modelo probabilístico. Etapa 1: similaridade BC (atributos textuais, visuais com SIFT e bag-of-features); Etapa 2: FC utilizando contexto	contexto (palavras-chave) e usuário	4.775 imagens da Universidade de Washington com 87 palavras-chave e 24 usuários

para ordenar automaticamente um conjunto de soluções que são exibidas na tela, ou seja, as informações do olhar refletem o gosto das pessoas. A preferência humana é medida com base no comprimento do olhar e na quantidade de fixações e essas informações são usadas como valor na função objetivo do AEI. O método pode ser resumido nos seguintes passos: i) inicializar um conjunto de soluções N, denominada soluções pais; ii) avaliar as soluções de acordo com uma função objetivo (através das informações do olhar); iii) realizar a mutação entre as melhores soluções e gerar a prole; iv) a nova geração será exibida novamente ao usuário; assim por diante.

No trabalho (XU; JIANG; LAU, 2008), os autores propõem um algoritmo que pode retornar uma recomendação de conteúdo on-line personalizado de acordo com o comportamento passado do usuário. O comportamento passado do usuário é obtido por meio do tempo de atenção em documentos, imagens e vídeos combinado com características dos objetos. O processo de recomendação possui duas etapas: i) o usuário faz uma consulta e um motor de busca é utilizado para buscar os primeiros 300 registros. O motor de busca indicado pelos autores é o Google Web Search para documentos, o Google Image Search para imagens, e o YouTube para vídeos; ii) o algoritmo prediz o nível de atentividade dos itens retornados pelo motor de busca, baseando-se no conteúdo e no tempo do olhar em outros itens. O nível de atentividade predito é usado para ranquear a recomendação de itens, ou seja, a atentividade é usada para representar a preferência de um usuário por um produto.

O tempo da atenção, ou também chamado tempo de leitura, é um tipo de feedback implícito bem recente. A relação entre a sua confiabilidade ainda é duvidosa devido ao fato que nem sempre o que chama a atenção da pessoa é o que ela mais gosta. Kelly e Belkin (KELLY; BELKIN, 2001; KELLY; BELKIN, 2004) alegam que não existe confiança entre o interesse por um documento e o tempo de fixação. Entretanto, outra opinião é dada por Kubat et al. (KUBAT; TAPIA et al., 2007) comprovando que quanto mais tempo o usuário gasta na leitura de um documento, maior a importância do mesmo. Os trabalhos descritos anteriormente (JUNG et al., 2013; XU; JIANG; LAU, 2008), que usam a atenção visual como *feedback* implícito, pressupõem que o tempo de atenção está relacionado com a preferência do usuário.

Nesta tese, aborda-se a atenção visual de maneira diferente. Usa-se a atenção visual para caracterizar a imagem e ajudar a calcular a similaridade entre imagens, ou seja, se as regiões de duas imagens (com o mesmo significado semântico) tiverem a atenção visual semelhante, então as duas imagens tem grande probabilidade de serem semelhantes. Além disso, o usuário que receberá a recomendação não necessariamente precisa de um rastreador de olhar, diferentemente das abordagens apresentadas nesta subseção, pois alguns usuários com rastreadores de olhar já são suficientes para caracterizar as imagens.

### 3.4 Considerações Finais

Neste capítulo forma abordados os principais trabalhos envolvendo combinação de similaridades, recomendação de imagens e o uso da atenção visual como *feedback* implícito. Em geral, independentemente do tipo de informação que será recomendado (vídeo, imagem, texto, áudio), as três principais técnicas de sistemas de recomendação podem ser utilizadas: Filtragem Colaborativa, baseada em Conteúdo e Híbrida.

É difícil encontrar novidades em abordagens FC para imagens em específico, já que as abordagens FC usam apenas a matriz de utilidade para realizar a predição. O trabalho descrito em (WIDISINGHE et al., 2010) usa o contexto como informação de entrada para o SR baseado em FC realizar a recomendação.

Entretanto, existem vários trabalhos com foco em imagens e que usam ou a técnica baseada em conteúdo ou a abordagem híbrida. Em geral, os trabalhos tentam enriquecer a representação das imagens, adicionando informações como emoção, contexto, características visuais, e palavras-chave para melhorar a recomendação de imagens.

A hipótese central desta tese é que a atenção visual pode ser uma característica importante para descrever imagens. Não foram encontrados trabalhos que abordem esta hipótese em nenhuma técnica de recomendação.



# CAPÍTULO 4

---

## Métodos Propostos

Neste capítulo serão apresentados dois novos métodos propostos nesta tese para o desenvolvimento de SRI que utilizam informações da atenção visual obtidas via rastreador de olhar: *Collaborative Filtering recommender systems with Attentive Similarity* (CFAS) e *Content-Based recommender systems with Attentive Similarity* (CBAS). Os métodos são baseados na vizinhança entre itens. Nesta tese os termos itens e imagens serão usados de forma indistinta.

### 4.1 Abordagem Baseada em Filtragem Colaborativa (CFAS)

A arquitetura do método CFAS é sumarizada na Figura 13. Especificamente, o método CFAS é dividido em quatro componentes principais: Processo de segmentação e classificação; Gerenciamento da atenção visual e avaliações; Processo de predição; e o Processo de recomendação.

#### 4.1.1 Processo de Segmentação e Classificação

Neste processo assume-se que uma coleção de imagens possui um conjunto de rótulos que se associam a conceitos semânticos, denominado conjunto  $\mathbb{H}$ . O conteúdo e a cardinalidade de  $\mathbb{H}$  depende do domínio da aplicação. Cada imagem da coleção é então segmentada em partes e cada parte deve ser rotulada (classificada) de acordo com o conjunto  $\mathbb{H}$ . Dois exemplos diferentes de segmentação são ilustrados na Figura 14. Na Figura 14 (a) o domínio de aplicação é “vestuário” e o conjunto  $\mathbb{H}$  representa as partes do corpo humano, ou seja,  $\mathbb{H} = \{ombro\ direito, pescoço, ombro\ esquerdo, joelho\ direito, joelho\ esquerdo, \dots\}$ , e na Figura 14 (b) o domínio de aplicação é “pintura” e o conjunto de rótulos  $\mathbb{H}$  representa paisagens, objetos, animais, pessoas, construções, entre outros.

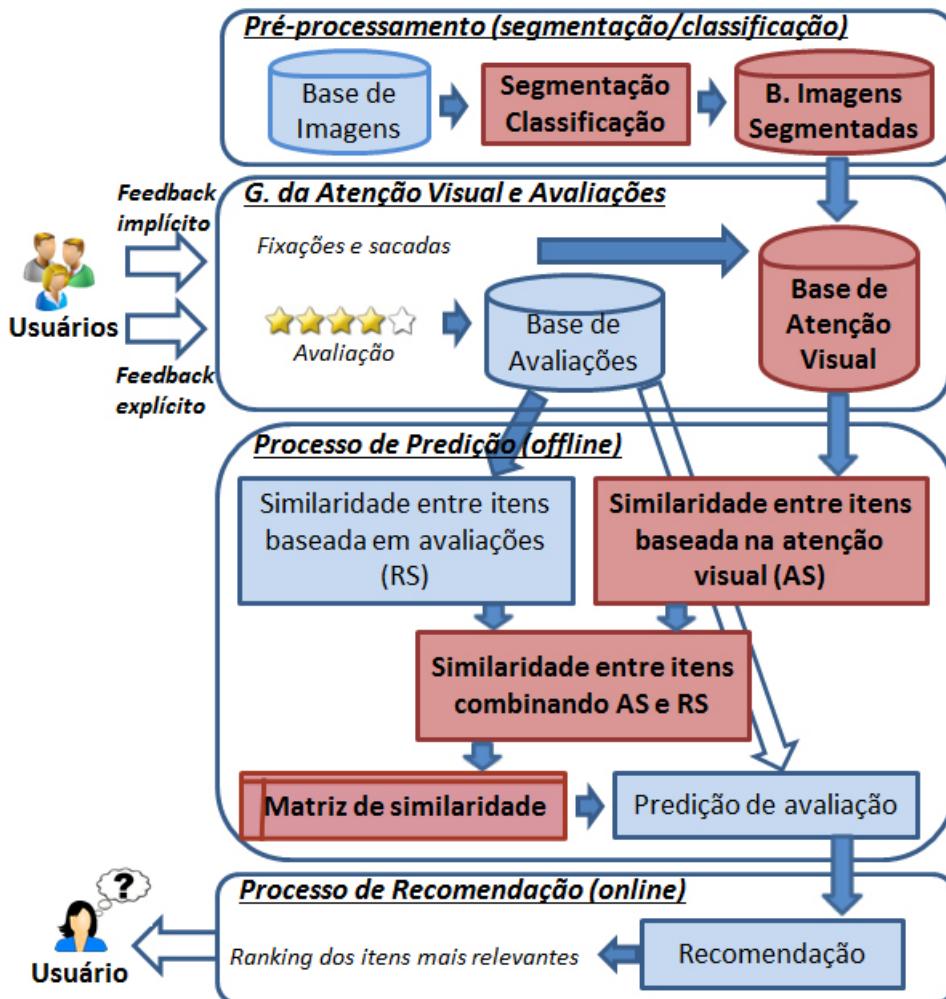


Figura 13 – Arquitetura do método proposto CFAS. Os retângulos vermelhos são as principais contribuições deste trabalho.

#### 4.1.2 Gerenciamento da Atenção Visual e Avaliações

Este componente contém duas bases de dados. A primeira base, denominada base de avaliações, armazena a matriz de utilidade. A representação da matriz de utilidade foi definida e detalhada na Subseção 2.2.1. A segunda base, denominada base de atenção visual, armazena as fixações e os movimentos dos olhos dos usuários (informação fornecida implicitamente pelo usuário e atualizada em tempo real). A representação formal da atenção visual é obtida pela junção da base de imagens segmentadas e rotuladas e da coleta de fixações e movimentos dos olhos.

*Processo de coleta de fixações e movimentos dos olhos:* quando um usuário navega sobre itens (imagens) de um sistema usando um computador com dispositivo de rastreamento do olhar, os dados da atenção visual são capturados e armazenados, veja Figuras 15 (b) e 16 (b). Cada imagem  $i$  é então descrita por quatro atributos de atenção visual:  $[\theta_i, \ell_i, \gamma_i, V_i]$ , em que  $\theta_i$  é o número de usuários que olharam para a imagem  $i$ ;  $\ell_i$  é a soma total do percurso, em número de pixels, de todos os usuários que olharam para a imagem

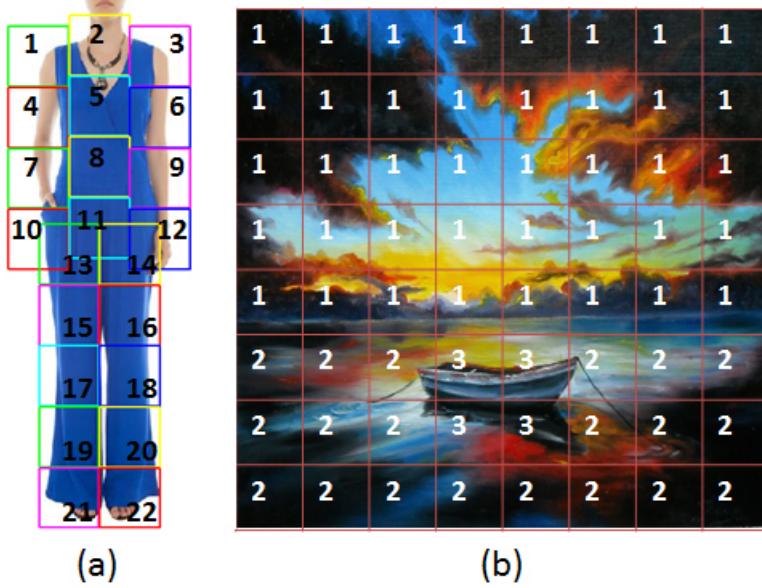


Figura 14 – As imagens são segmentadas e rotuladas de acordo com o conjunto  $\mathbb{H}$ . (a) nesse exemplo a segmentação é realizada usando uma grade que divide a pessoa em partes do corpo humano e cada parte é rotulada com o respectivo conceito semântico. (b) nesse segundo exemplo a segmentação é obtida dividindo a imagem inteira em uma grade regular e cada célula da grade é rotulada com um conceito semântico relacionado à pintura (1 representa céu, 2 representa oceano e 3 representa um barco).

$i$ ;  $\gamma_i$  é a soma da duração, em segundos, de todas as fixações sobre a imagem  $i$ ; e  $V_i$  é o vetor de atentividade de dimensão igual a  $|\mathbb{H}|$ . Cada posição do vetor de atentividade  $V_i$  é relacionada com um rótulo  $t$  de uma imagem  $i$ . Os valores de  $\ell_i$ ,  $\gamma_i$ , e cada dimensão  $V_i[t]$  da imagem  $i$  são obtidos de acordo com as Equações 36, 37, e 38 respectivamente.

$$\ell_i = \sum_{u \in \mathbb{G}(i)} \left( \sum_{m \in \mathbb{M}(u,i)} l_m \right), \quad (36)$$

$$\gamma_i = \sum_{u \in \mathbb{G}(i)} \left( \sum_{g \in \mathbb{G}(u,i)} d_g \right), \quad (37)$$

$$V_i[t] = \sum_{u \in \mathbb{G}(i)} \left( \frac{\sum_{g \in \mathbb{G}(u,i,t)} d_g}{\sum_{g \in \mathbb{G}(u,i)} d_g} \right), \quad (38)$$

onde o conjunto  $\mathbb{G}(i)$  contém os usuários que olharam para a imagem  $i$ ,  $\mathbb{M}(u,i)$  é o conjunto de todas as sacadas do usuário  $u$  sobre a imagem  $i$ ,  $l_m$  é o tamanho da sacada  $m$ ,  $\mathbb{G}(u,i)$  é o conjunto de todos os dados de fixações do olhar do usuário  $u$  sobre a imagem  $i$ ,  $\mathbb{G}(u,i,t)$  é o conjunto de todos os dados das fixações do olhar do usuário  $u$  sobre o rótulo  $t$  da imagem  $i$ , e  $d_g$  é a duração, em segundos, da fixação  $g$ . A representação dos dados de atenção visual de uma imagem de roupa visualizada por dois usuários é ilustrada na Figura 15 e de uma imagem de pintura é ilustrada na Figura 16.

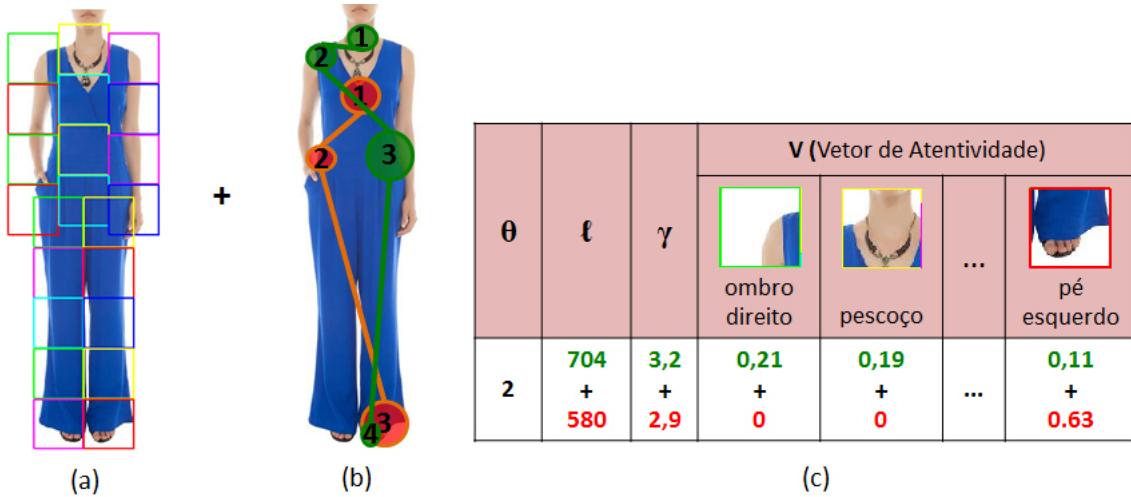


Figura 15 – Representação dos dados da atenção visual de uma imagem de roupa visualizada por dois usuários. (a) imagem segmentada. (b) dois usuários (verde e vermelho) observam a imagem. Cada cor representa um usuário e cada círculo representa uma fixação de olhar e o raio do círculo representa a duração da fixação e o número dentro do círculo representa a sequência das fixações. As linhas entre os círculos representam os movimentos dos olhos (sacadas). (c) representa os dados da atenção visual. O usuário verde percorreu 704 pixels na imagem durante 3,2 segundos e ficou 21% do tempo olhando para o ombro direito, 19% do tempo olhando para o pescoço e 11% do tempo olhando para o pé esquerdo. O segundo usuário (vermelho) percorreu 580 pixels na imagem durante 2,9 segundos e ficou 63% do tempo olhando para o pé esquerdo.

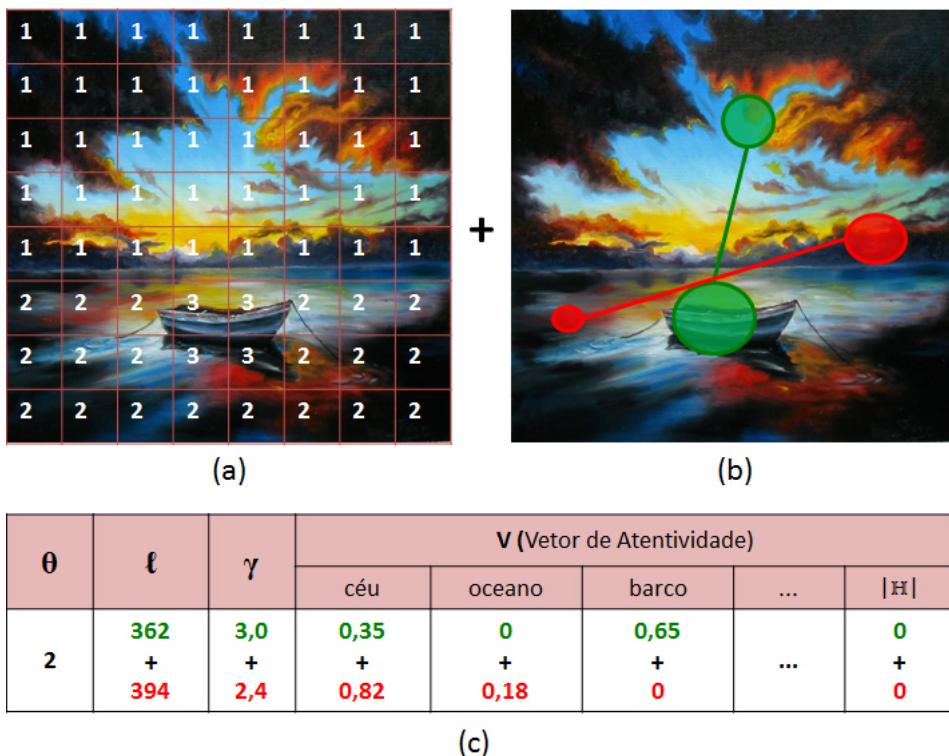


Figura 16 – Representação dos dados da atenção visual de uma imagem de pintura visualizada por dois usuários. (a) imagem segmentada. (b) dois usuários (verde e vermelho) observam a imagem. (c) representa os dados da atenção visual. O usuário verde percorreu 362 pixels na imagem durante 3 segundos e olhou 35% do tempo para o céu e 65% do tempo para o barco. O usuário vermelho percorreu 394 pixels na imagem durante 2,4 segundos e olhou 82% do tempo para o céu e 18% do tempo para o oceano.

### 4.1.3 Processo de Predição

O processo de predição ocorre de maneira *offline* e tem como objetivo principal a predição de avaliações desconhecidas da matriz de utilidade. Este processo ocorre quando o usuário atualiza suas avaliações, quando um novo item é inserido na base de dados, ou quando informações sobre o item são atualizadas. Este processo é dividido em duas etapas principais: cálculo da similaridade entre itens e a predição de avaliação.

O cálculo de similaridade entre itens é muito importante para uma acurácia alta em sistemas de recomendação baseados em vizinhança. A similaridade entre todos os itens é representada por uma matriz de similaridade  $\mathbb{S} = \{S_{ij}\}_{1 \leq i \leq |\mathbb{I}|, 1 \leq j \leq |\mathbb{I}|}$ . A similaridade  $S_{ij}$  entre dois itens  $i$  e  $j$  é calculada pela agregação de duas similaridades: similaridade atentiva ( $AS_{ij}$ ) e similaridade baseada em avaliações ( $RS_{ij}$ ).

#### 4.1.3.1 A medida de similaridade atentiva ( $AS_{ij}$ )

A similaridade atentiva ( $AS_{ij}$ ) entre duas imagens  $i$  e  $j$  é dada por uma função de agregação  $f_{AS} : [0, 1]^2 \rightarrow [0, 1]$  que considera dois termos. O primeiro termo ( $simV(V_i, V_j)$ ) considera a similaridade entre dois vetores de atentividade ( $V_i$  e  $V_j$ ) e o segundo termo ( $siml(\ell_i, \ell_j)$ ) considera a similaridade entre os comprimentos das sacadas ( $\ell_i$  e  $\ell_j$ ), como definido na Equação 39.

$$AS_{i,j} = f_{AS}(simV(V_i, V_j), siml(\ell_i, \ell_j)) \quad (39)$$

Os vetores de atentividade  $V_i$  e  $V_j$  são histogramas de atentividade, onde cada *bin* representa um rótulo semântico e o valor atribuído a um rótulo representa o quanto este rótulo é atentivo. Dividindo estes vetores pelo número de usuários que olharam as imagens  $i$  e  $j$  ( $\theta_i$  e  $\theta_j$  respectivamente), os vetores serão normalizados. Para calcular a similaridade  $simV(V_i, V_j)$  entre dois vetores pode-se utilizar diversas funções de distância, tais como distância Euclidiana, Mahalanobis, intersecção de histograma (SWAIN; BALLARD, 1991), etc. Tanto a similaridade entre os vetores de atentividade  $simV(V_i, V_j)$  quanto a similaridade entre os comprimentos das sacadas ( $siml(\ell_i, \ell_j)$ ) devem estar no intervalo  $[0, 1]$ . A similaridade  $siml(\ell_i, \ell_j)$  pode ser calculada também usando diferentes funções de distância, desde que seja normalizada. O valor da similaridade atentiva ( $AS_{ij}$ ) também está no intervalo  $[0, 1]$ , onde 0 significa que as imagens  $i$  e  $j$  são totalmente diferentes e 1 significa que as imagens  $i$  e  $j$  são totalmente similares do ponto de vista da atenção visual. Se  $i$  ou  $j$  não possuem dados de atenção visual, o cálculo de  $AS_{ij}$  é inviável.

#### 4.1.3.2 A medida de similaridade baseada em avaliações ( $RS_{ij}$ )

Na nossa estratégia, a similaridade  $RS_{ij}$  entre dois itens  $i$  e  $j$  pode ser calculada usando qualquer uma das funções de similaridade entre itens baseada em avaliações, como descrito

na Subseção 2.2.2.3. Para a FC Item-Item, as mais comuns são o PPC (Equação 10) e a Cosseno (Equação 16).

Uma estratégia para reduzir a dimensionalidade é o uso de fatoração de matrizes. A similaridade  $RS_{ij}$  pode ser calculada usando vetores de fatores latentes de itens, como demonstrado na Equação 22.

#### 4.1.3.3 Ponderação de importância para a similaridade atentiva

Em uma abordagem Filtragem Colaborativa, o sucesso no cálculo de similaridade  $RS_{ij}$  depende das avaliações dos usuários do sistema. Uma maneira de tentar ser mais justo e dar maior importância para similaridades entre itens com maior quantidade de avaliações, é utilizar o fator de ponderação de importância, como descrito na Equação 18. Do mesmo modo, a similaridade atentiva pode ser prejudicada se um dos itens tiver poucas visualizações. Definimos então uma estratégia para modificar o valor de  $AS_{ij}$ , denotada  $AS'_{ij}$ , utilizando um fator de ponderação de importância, conforme Equação 40. Neste caso, a similaridade atentiva  $AS_{ij}$  é substituída por  $AS'_{ij}$  no cálculo da similaridade  $S_{ij}$ .

$$AS'_{ij} = \frac{|\mathbb{G}(i, j)| - 1}{|\mathbb{G}(i, j)| - 1 + \lambda_{as}} AS_{ij} \quad (40)$$

em que  $\lambda_{as}$  é um parâmetro de encolhimento definido pelo usuário,  $|\mathbb{G}(i, j)|$  é a quantidade de usuários que possuem fixações do olhar sobre ambos os itens  $i$  e  $j$ .

#### 4.1.3.4 A medida de similaridade entre itens ( $S_{ij}$ )

A medida de similaridade  $S_{ij}$  proposta neste trabalho entre dois itens  $i$  e  $j$  é obtida por uma função de agregação  $f_{CF} : [0, 1]^2 \rightarrow [0, 1]$  que combina a similaridade atentiva ( $AS_{ij}$ ) e a similaridade baseada em avaliações ( $RS_{ij}$ ), conforme Equação 41.

$$S_{i,j} = f_{CF} (RS_{ij}, AS_{ij}) \quad (41)$$

#### 4.1.3.5 Predição de avaliação

Quando um usuário atualiza suas avaliações, o sistema precisa recalcular a predição de avaliações desconhecidas (os asteriscos da matriz de utilidade, veja Figura 3) para este usuário. Então, os pares  $(u, i)$  para cada  $r_{ui}$  desconhecido são separados em um conjunto  $\mathbb{Z} = \{(u, i) | r_{ui} \text{ é desconhecido}\}$ .

Depois de identificado o conjunto  $\mathbb{Z}$ , para cada par  $(u, i) \in \mathbb{Z}$ , calcula-se a predição de avaliação  $\hat{r}_{ui}$  utilizando o modelo *Item KNN + Baseline* (KOREN, 2010), descrito na Equação 14.

O problema de partida a frio de item é muito comum em SR baseados em Filtragem Colaborativa. Uma solução muito utilizada para este problema é a utilização de métodos

BC. Entretanto, nem sempre é viável um sistema de recomendação possuir atributos dos itens, seja por dificuldades de manutenção ou espaço em disco. O método CFAS aborda parcialmente o problema de partida a frio de item, onde novos itens que ainda não foram avaliados, mas que já foram observados por alguns usuários podem ser recomendados. A predição de avaliação  $\hat{r}_{ui}$  para um novo item  $i$  na Equação 14 pode ser reduzida para a Equação 42.

$$\hat{r}_{ui} = \mu + b_u + \frac{\sum_{j \in \mathbb{I}(u,i,k)} AS_{ij} (r_{uj} - b_{uj})}{\sum_{j \in \mathbb{I}(u,i,k)} AS_{ij}} \quad (42)$$

#### 4.1.4 Processo de Recomendação

Todas as predições de avaliação são armazenadas em um conjunto  $\mathbb{X} = \{(u, i) | \hat{r}_{ui} \text{ é predita}\}$ .

O processo de recomendação ocorre *online*. Dado um usuário  $u$ , o sistema carrega o conjunto  $\mathbb{X}$  e recomenda os itens com as avaliações mais altas para o usuário  $u$ .

## 4.2 Abordagem Baseada em Conteúdo (CBAS)

A arquitetura do método CBAS é sumarizada na Figura 17. O método CBAS é dividido em quatro componentes principais: Processo de segmentação e classificação; Gerenciamento da atenção visual, atributos dos itens e avaliações; Processo de predição; e o Processo de recomendação. Os processos de segmentação/classificação e recomendação são idênticos aos do método CFAS descrito nas subseções 4.1.1 e 4.1.4 respectivamente.

### 4.2.1 Gerenciamento da Atenção Visual, Atributos dos Itens e Avaliações

Este componente contém três bases de dados: base de avaliações, base de atenção visual, e a base de atributos dos itens. As duas primeiras bases são as mesmas discutidas no método CFAS, entretanto a base de avaliações é utilizada apenas no momento da predição e não no cálculo de similaridade entre itens. A base de atributos dos itens armazena o conteúdo dos itens: atributos textuais (que podem ser obtidos manualmente ou automaticamente) e características visuais (extraídas automaticamente das imagens).

A representação formal dos atributos textuais e características visuais será descrita a seguir.

- *Atributos textuais:* seja  $\mathbb{A} = \{A_1, A_2, \dots, A_{n_A}\}$  o conjunto de atributos textuais e  $n_A$  o número de atributos textuais. Cada atributo textual  $A_p$  possui um número variável de conceitos ou valores, denotado por  $|A_p|$ . Então, o atributo  $A_p$  é representado por um vetor e cada dimensão deste vetor é associada a um conceito deste atributo. Se o

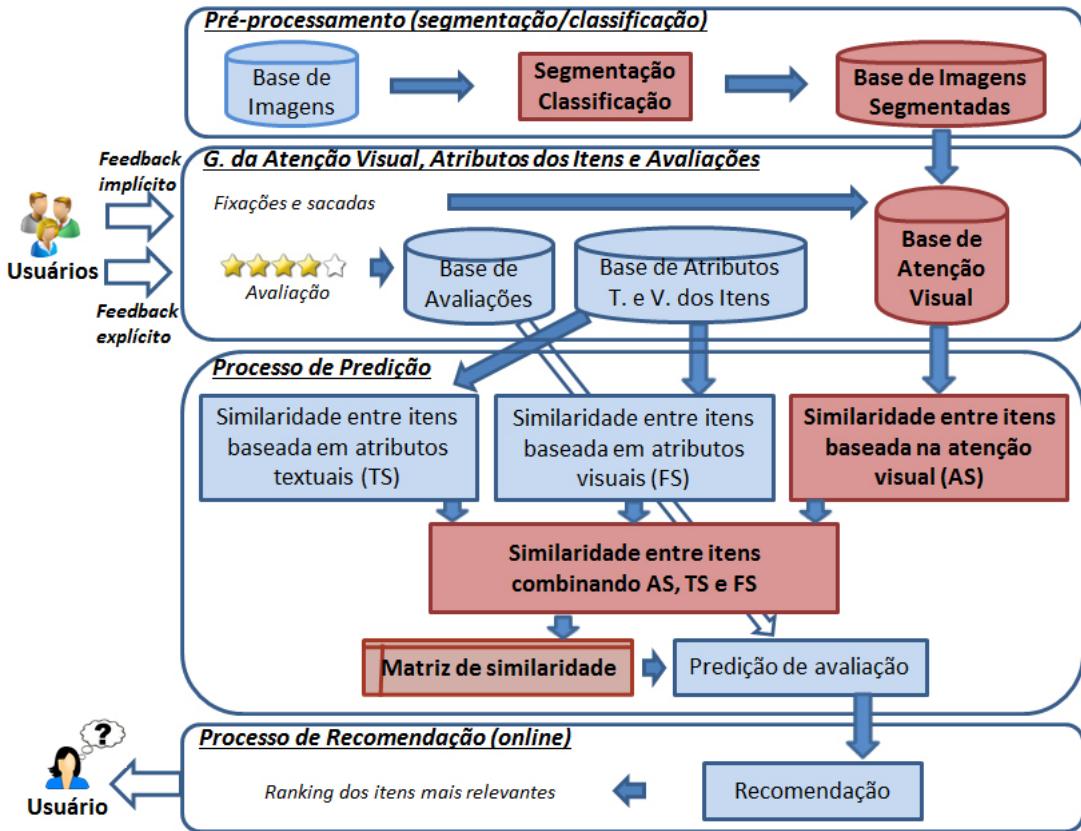


Figura 17 – Arquitetura do método proposto CBAS. Os retângulos vermelhos são as principais contribuições deste trabalho.

$k$ -ésimo conceito ocorre em  $A_p$ , então  $A_p[k] = 1$ , caso contrário  $A_p[k] = 0$ . A imagem  $i$  é representada pela concatenação dos vetores  $A_1, A_2, \dots, A_{n_A}$ . Esta concatenação produz um vetor binário, denotado  $T_i$ , com tamanho  $|T_i| = \sum_{p=1}^{n_A} |A_p|$ .

Como exemplo, considere que a imagem  $i$ , apresentada na Figura 14 (a), seja modelada pelo conjunto  $\mathbb{A}$  contendo apenas três atributos textuais  $\mathbb{A} = \{\text{Manga}, \text{Botão}, \text{Colarinho}\}$ , sendo que:

- *Manga* ( $A_1$ ) possui os possíveis valores (conceitos): *sem manga; manga curta; manga longa*.
- *Botão* ( $A_2$ ) possui os possíveis valores (conceitos): *com botão; sem botão*.
- *Colarinho* ( $A_3$ ) possui os possíveis valores (conceitos): *tomara que caia, gola V, redondo fechado, redondo aberto, gola de camisa, mula manca*.

O vetor de atributos textuais  $T_i$  da imagem  $i$  é demonstrado na Figura 18. Nesse exemplo a roupa é sem manga, não tem botão e possui gola V.

- *Características visuais:* as características visuais de uma imagem  $i$  são representadas por um conjunto de vetores de características  $\mathbb{F}_i = \{F_1, F_2, \dots, F_{n_F}\}$ , onde  $F_q$  representa uma característica visual e  $n_F$  é o número de características visuais. Exemplos de características visuais foram descritos na Seção 2.3.

A1 (Manga)			A2 (Botão)		A3 (Colarinho)					
Sem manga	manga curta	manga longa	com botão	sem botão	tomara que caia	gola V	redondo fechado	redondo aberto	gola de camisa	mula manca
$T_i$	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>

Figura 18 – Vetor de atributos textuais  $T_i$ .

## 4.2.2 Processo de Predição

O processo de predição ocorre do mesmo modo do método CFAS, com exceção do cálculo de similaridade entre itens. A similaridade  $S_{ij}$  entre dois itens  $i$  e  $j$  é calculada combinando três similaridades: similaridade atentiva ( $AS_{ij}$ ), descrita na subseção 4.1.3, similaridade textual ( $TS_{ij}$ ), e similaridade visual ( $FS_{ij}$ ).

### 4.2.2.1 A medida de similaridade textual ( $TS_{ij}$ )

Sejam  $T_i$  e  $T_j$  os vetores de atributos textuais dos itens  $i$  e  $j$ , respectivamente. A similaridade textual é calculada usando a função cosseno, como mostra a Equação 43.

$$TS_{ij} = \frac{\sum_{k=1}^n T_i[k] \cdot T_j[k]}{\sqrt{\sum_{k=1}^n T_i[k]^2} \cdot \sqrt{\sum_{k=1}^n T_j[k]^2}} \quad (43)$$

onde  $n = |T_i| = |T_j|$ . O valor da similaridade textual ( $TS_{ij}$ ) está no intervalo  $[0,1]$ , onde 0 significa que as imagens  $i$  e  $j$  são totalmente diferentes e 1 significa que as imagens  $i$  e  $j$  são totalmente similares considerando apenas os atributos textuais.

### 4.2.2.2 A medida de similaridade visual ( $FS_{ij}$ )

Considerando que na comparação entre imagens: i) pode-se utilizar diferentes funções de distância para diferentes características visuais; ii) dependendo do domínio da aplicação, uma característica visual pode ser mais importante do que outras. O cálculo da similaridade visual ( $FS_{ij}$ ) entre duas imagens  $i$  e  $j$  é dada pela combinação ponderada das similaridades de cada característica visual. Seja  $\mathbb{W} = \{w_1, \dots, w_{n_F}\}$  o conjunto de pesos de acordo com as características visuais. A similaridade visual ( $FS_{ij}$ ) é descrita na Equação 44.

$$FS_{ij} = \frac{\sum_{q=1}^{n_F} w_q \cdot s f_q(i, j)}{\sum_{q=1}^{n_F} w_q} \quad (44)$$

onde  $n_F$  é o número de características visuais e  $sf_q(i, j)$  é a similaridade entre a característica visual  $F_q$  das imagens  $i$  e  $j$ . Note que a similaridade  $sf_q(i, j)$  deve estar definida no intervalo  $[0,1]$ , onde 0 significa completamente dissimilar e 1 totalmente similar.

#### 4.2.2.3 A medida de similaridade entre itens ( $S_{ij}$ )

A similaridade  $S_{ij}$  entre duas imagens  $i$  e  $j$  é definida pela função de agregação  $f_{CB} : [0, 1]^3 \rightarrow [0, 1]$  que combina as três similaridades:  $AS_{ij}$ ,  $TS_{ij}$  e  $FS_{ij}$ .

$$S_{ij} = f_{CB}(TS_{ij}, FS_{ij}, AS_{ij}) \quad (45)$$

#### 4.2.2.4 Predição de avaliação

A predição de avaliação  $\hat{r}_{ui}$  também é calculada utilizando o *Item KNN + Baseline* (KOREN, 2010), descrito na Equação 14.

### 4.3 Considerações Finais

Neste capítulo foram abordadas as estratégias para adicionar a atenção visual em sistemas de recomendação de imagens. Os métodos propostos são interessantes para vários tipos de situação e tem como objetivo melhorar a predição de avaliação, e consequentemente, melhorar o processo de recomendação.

Para sistemas que não possuem informações do conteúdo dos itens, nossa abordagem baseada em Filtragem Colaborativa (CFAS) minimiza o problema de *cold-start* de item. Já para grandes bases de dados, sugerimos combinar a similaridade entre vetores de fatores latentes dos itens com a similaridade atentiva.

Para sistemas que possuem dados textuais e visuais dos itens, o método CBAS foi proposto. Este método é ideal principalmente para eliminar o problema de *cold-start* de item e para lidar com situações adversas, onde em uma mesma base de dados podemos ter imagens com muitas características textuais e poucas visuais e vice-versa.

Um aspecto importante para o sucesso de nossas abordagens é a necessidade da segmentação de imagem em partes rotuladas com conceitos semânticos. A segmentação e classificação automática é uma tarefa muito desafiadora, fortemente dependente da aplicação e muito importante para a compreensão da imagem. Existem vários trabalhos nesta área (ATHANASIADIS et al., 2007; MESNIL et al., 2014; ZHENG et al., 2014), e apesar de muito importante para os métodos propostos, a segmentação de imagem não será discutida nesta tese.

# CAPÍTULO 5

---

## Metodologia para Validação dos Métodos Propostos

Neste capítulo são apresentados e discutidos aspectos importantes quanto a validação dos métodos propostos. Foram necessárias a construção de duas bases de dados reais e a utilização de métricas para avaliar o processo de predição de avaliação e recomendação. Por fim, definiu-se com quais trabalhos da literatura os métodos propostos serão comparados.

### 5.1 Base de Dados

Para validar a metodologia proposta, foram construídas duas bases de dados: a primeira, denominada UFU-CLOTHING(MELO; NOGUEIRA; GULIATO, 2015), é uma base de roupas e foi construída em sua totalidade nesta tese pelo nosso grupo de pesquisa do laboratório LINCCI<sup>1</sup>; a segunda base de dados, denominada UFU-PAINTINGS(FELÍCIO et al., 2016), é uma base de pinturas e foi construída em parceria com o grupo de pesquisa do laboratório LSI<sup>2</sup>, ou seja, os alunos do LSI coletaram as imagens, os atributos textuais e as fixações de olhares e os alunos do LINCCI extraíram as características visuais e implementaram a segmentação nas imagens.

#### 5.1.1 Base UFU-CLOTHING

A base de roupas UFU-CLOTHING inclui imagens avaliadas por usuários, atributos textuais e visuais, e dados de atenção visual (fixações do olhar e sacadas). De acordo com o nosso melhor conhecimento, não existe na literatura uma base de dados pública que une todas estas informações.

<sup>1</sup>LINCCI: Laboratório Integrado de Computação Científica ([www.lincci.ufu.br](http://www.lincci.ufu.br))

<sup>2</sup>LSI: Laboratório de Sistemas de Informação ([www.lsi.ufu.br](http://www.lsi.ufu.br))

Tabela 6 – Número de imagens separadas por gênero, classe do corpo e categoria. Juntas são 2.254 roupas masculinas e 4.692 roupas femininas.

Masculino								
Inteiro	Parte superior				Parte inferior			
Terno	Camisa	Polo shirt	Camiseta	Casaco	Bermuda	Calça	Shorts	
185	180	336	586	372	187	150	258	
Feminino								
Inteiro	Parte superior				Parte inferior			
Vestido	Macacão	Camisa	Camiseta	Casaco	Saia	Calça	Shorts	
783	98	1316	325	192	510	737	731	



(a)



(b)

Figura 19 – Exemplos de imagens da base de dados UFU-CLOTHING. (a) roupas masculinas e (b) roupas femininas.

Inicialmente, foram coletadas 6.946 imagens de roupas de vários websites brasileiros de compras online, tais como br.privalia.com, www.posthaus.com.br, www.oqvestir.com.br e www.netshoes.com.br, usando palavras-chave como “camiseta”, “vestido”, etc. A resolução das imagens é de 500 x 700 pixels. As imagens foram organizadas de acordo com o gênero (masculino ou feminino), classe do corpo (inteiro, parte de cima, parte de baixo), e categoria (camisa, calça, etc.), como mostra a Tabela 6. O fundo das imagens é limpo, evitando o desvio da atenção. A Figura 19 mostra alguns exemplos da base de dados UFU-CLOTHING.

Para obter os dados de avaliação e atenção visual, dois conjuntos de roupas foram separados de acordo com o gênero: masculino e feminino. Cada usuário masculino avaliou 300 itens (210 itens são comuns a todos os usuários masculinos e 90 itens são escolhidos aleatoriamente). Dentre os 300 itens, estão incluídos: 20 ternos, 30 camisas, 50 camisas



Figura 20 – O usuário sentou a uma distância de aproximadamente 50 centímetros de uma tela de computador de 22 polegadas. O dispositivo de rastreamento do olhar foi fixado na parte inferior do monitor e a avaliação foi coletada através do teclado.

polo, 60 camisetas, 40 casacos, 30 calças, e 40 shorts. Cada usuário feminino também avaliou 300 itens (210 itens comuns a todos os usuários femininos e 90 itens escolhidos aleatoriamente). Os itens avaliados incluem 20 macacões, 40 vestidos, 60 camisas, 30 camisetas, 30 casacos, 40 calças, 40 saias, e 40 shorts.

Um total de 245 usuários, sendo 122 homens e 123 mulheres, avaliaram os itens da base de dados UFU-CLOTHING. Os usuários tinham idade entre 18 e 60 anos e eram estudantes ou funcionários da Universidade Federal de Uberlândia (UFU). Todos os usuários sentaram a uma distância de aproximadamente 50 centímetros de um monitor de 22 polegadas com uma resolução de 1680 x 1050 pixels, como mostra a Figura 20.

O rastreador de olhar não intrusivo Tobii X2-60 foi usado para coletar os movimentos dos olhos a uma frequência de 60 Hz. Para cada usuário, foi realizada a calibração do dispositivo Tobii X2-60 sob uma grade de nove pontos de calibração para minimizar os erros da coleta. O rastreador de olhar extrai o movimento dos olhos, o local e a duração das fixações.

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 UFU-CLOTHING é composta de 469.071 fixações de olhar e 73.414 avaliações. A esparsidade dos dados atinge aproximadamente 95 %.

### 5.1.1.1 Segmentação e classificação

As imagens são modelos humanos posando da mesma maneira, facilitando a segmentação das imagens em partes do corpo humano. Nesta tese, foi desenvolvido um algoritmo de segmentação baseado em partição considerando a posição de cada parte do corpo humano na imagem, possibilitando a segmentação e classificação automática das imagens.

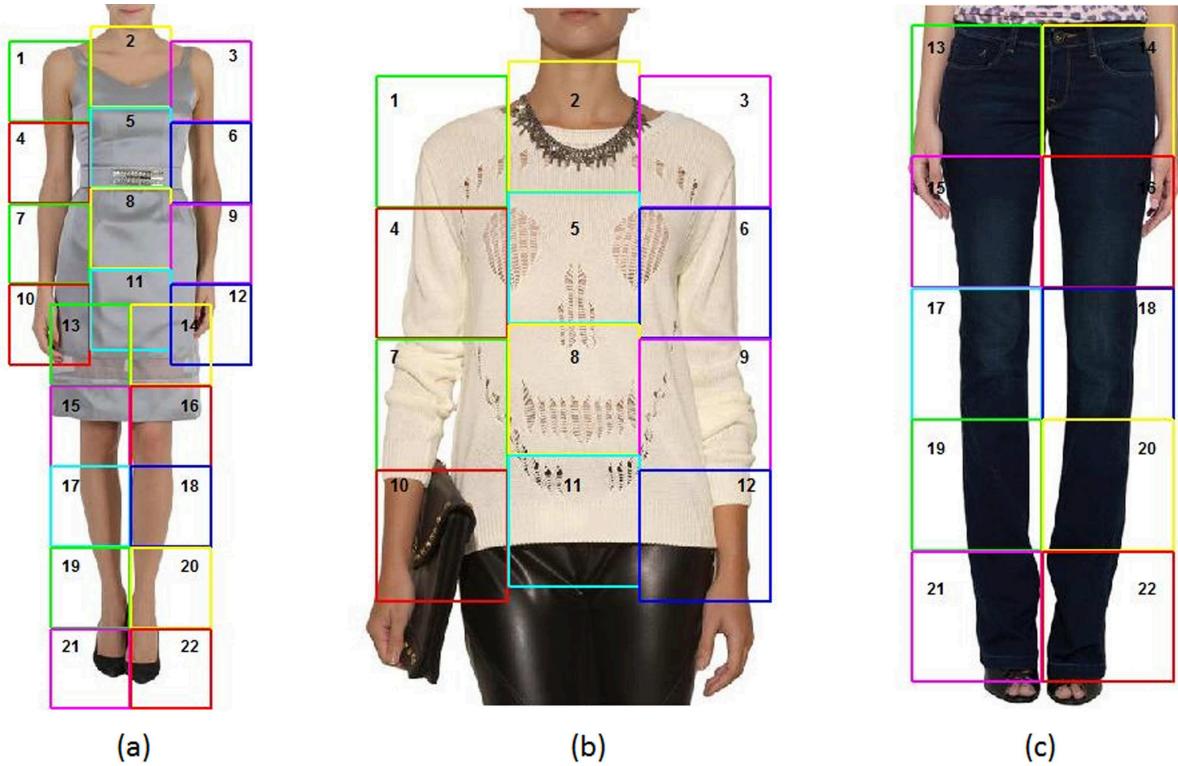


Figura 21 – Partes semânticas do corpo humano. (a) 22 partes de corpo inteiro, (b) 12 partes superiores e (c) 10 partes inferiores.

Por uma questão de eficiência e baseando-se na literatura (LIU et al., 2012; FU et al., 2013), as imagens foram segmentadas em 22 partes e cada parte foi classificada de acordo com a parte do corpo humano, sendo 12 partes superiores do corpo, tais como pescoço, ombro direito, ombro esquerdo, etc., e 10 partes inferiores, tais como joelho direito, joelho esquerdo, etc., como mostra a Figura 21. Alguns trabalhos sobre recuperação de roupas também segmentam pessoas em partes do corpo humano para comparar imagens (LIU et al., 2012; FU et al., 2013). Nestes trabalhos foram usadas mais de 22 partes, pois eles trabalharam com sobreposição de partes para contornar o problema da discrepância entre as poses humanas. No nosso caso, não foi necessária a sobreposição, pois os modelos estavam em poses praticamente idênticas.

### 5.1.1.2 Atributos textuais e características visuais

Como mostra a Tabela 7, o conjunto de atributos textuais específicos de roupas ( $\mathbb{A}$ ) foi construído de acordo com estudo empírico sobre catálogo de roupas e inspirado no trabalho apresentado em (LIU et al., 2012) que descreve um sistema de recuperação de roupas. Os atributos textuais definidos estão resumidos em três classes: *atributos globais*, *parte superior do corpo* e *parte inferior do corpo*. Os valores dos atributos foram atribuídos manualmente às 6.946 imagens de roupas.

Para descrever o conteúdo visual de uma imagem, cada parte do corpo humano é carac-

Tabela 7 – Atributos específicos de roupas: 12 atributos e 60 valores de atributos.

Atributos	Valores
<b>Global</b>	
Classe do corpo	Inteiro; Superior; Inferior
Categoria (Masculino)	Terno; Camisa; Camisa Polo; Camiseta; Casaco; Bermuda; Calça; Shorts
Categoria (Feminino)	Macacão; Vestido; Camisa; Camiseta; Casaco; Calça; Saia; Shorts
Cor predominante	Multi-cor; Vermelho; Laranja; Amarelo; Verde; Azul; Roxo; Preto; Branco; Cinza; Marrom; Rosa; Bege
Intensidade da cor	Claro; Médio; Escuro
Padrão	Liso; Listrado vertical; Listrado horizontal; Xadrez; Com desenho; Floral; Bolado; Estampado
Forma	Solto; Justo; Largo
<b>Parte superior</b>	
Manga	Sem manga; Manga curta; Manga longa
Colarinho	Tomara que caia; Gola-V; Redondo fechado; Redondo aberto; Gola de camisa; Mula manca
Botão	sim/não
Zíper	sim/não
<b>Parte inferior</b>	
Tamanho	Pequeno; Médio; Grande

terizada por cinco características visuais ( $n_F = 5$ ): HOG proposto em (FELZENZWALB et al., 2010) que utiliza 31 *bins* de orientação, *HSV Color Histogram*, LBP (PIETIKÄINEN, 2005), GIST (OLIVA; TORRALBA, 2006), e *skin descriptor* (JAIN, 2011)). A dimensão destes vetores descritores para cada parte do corpo humano é: 279, 72, 256, 512 e 16, respectivamente. Estes descritores são muito utilizados em sistemas de recuperação de roupas (LIU et al., 2012; FU et al., 2013). Cada característica visual  $F_q$  da imagem  $i$ , tal que  $F_q \in \mathbb{F}_i$ , é obtida pela concatenação da  $q$ -ésima característica visual de cada parte do corpo humano. O conteúdo visual da imagem  $i$  é então descrita por um conjunto  $\mathbb{F}_i = \{F_{HOG}, F_{HSV}, F_{LBP}, F_{GIST}, F_{SKIN}\}$  de  $n_F$  elementos, cada um representando uma característica visual.

### 5.1.2 Base UFU-PAINTINGS

A base de dados UFU-PAINTINGS é composta por 605 imagens de pinturas coletadas do website [pintura.aut.org](http://pintura.aut.org). A resolução das imagens é variada. A Figura 22 ilustra alguns exemplos da base de dados UFU-PAINTINGS.

Um total de 194 usuários avaliaram 200 imagens aleatoriamente. Os usuários eram estudantes da UFU e a atenção visual do usuário foi coletada do mesmo modo que na UFU-CLOTHING. Atualmente, a base de dados UFU-PAINTINGS é composta de 444.780 fixações de olhar e 38.742 avaliações. A esparsidade dos dados atinge aproximadamente



Figura 22 – Exemplos de imagens da base de dados UFU-PAINTINGS.

67 %.

### 5.1.2.1 Segmentação e classificação

A base de dados UFU-PAINTINGS contém várias pinturas diversificadas e produzidas entre os séculos XIV e XXI. Cada imagem desta base pertence a apenas uma das 9 categorias de pinturas: Animal, Arquitetura, Arte abstrata, Mitologia, Natureza morta, Nudismo, Paisagem, Pessoa e Religiosa. Em cada categoria, pode-se encontrar diversas partes da imagem com significados semânticos diferentes.

Nesta tese, foi desenvolvido um software que divide a imagem em uma grade de 20 x 20 partes de tamanhos iguais. O usuário, utilizando o mouse, rotula (classifica) manualmente cada parte da imagem com um significado semântico. O conjunto  $\mathbb{H}$ , definido na Seção 4.2.1, é composto por 41 rótulos de possíveis significados semânticos. O Apêndice A apresenta os 41 rótulos possíveis para a base de dados UFU-PAINTINGS. A Figura 23 ilustra uma tela do software desenvolvido para segmentação e rotulação de uma imagem.

### 5.1.2.2 Atributos textuais e características visuais

Tanto as imagens das pinturas quanto os atributos textuais foram retirados do website [pintura.aut.org](http://pintura.aut.org). O conjunto de atributos textuais ( $\mathbb{A}$ ) para a base de dados UFU-PAINTINGS é apresentado na Tabela 8.

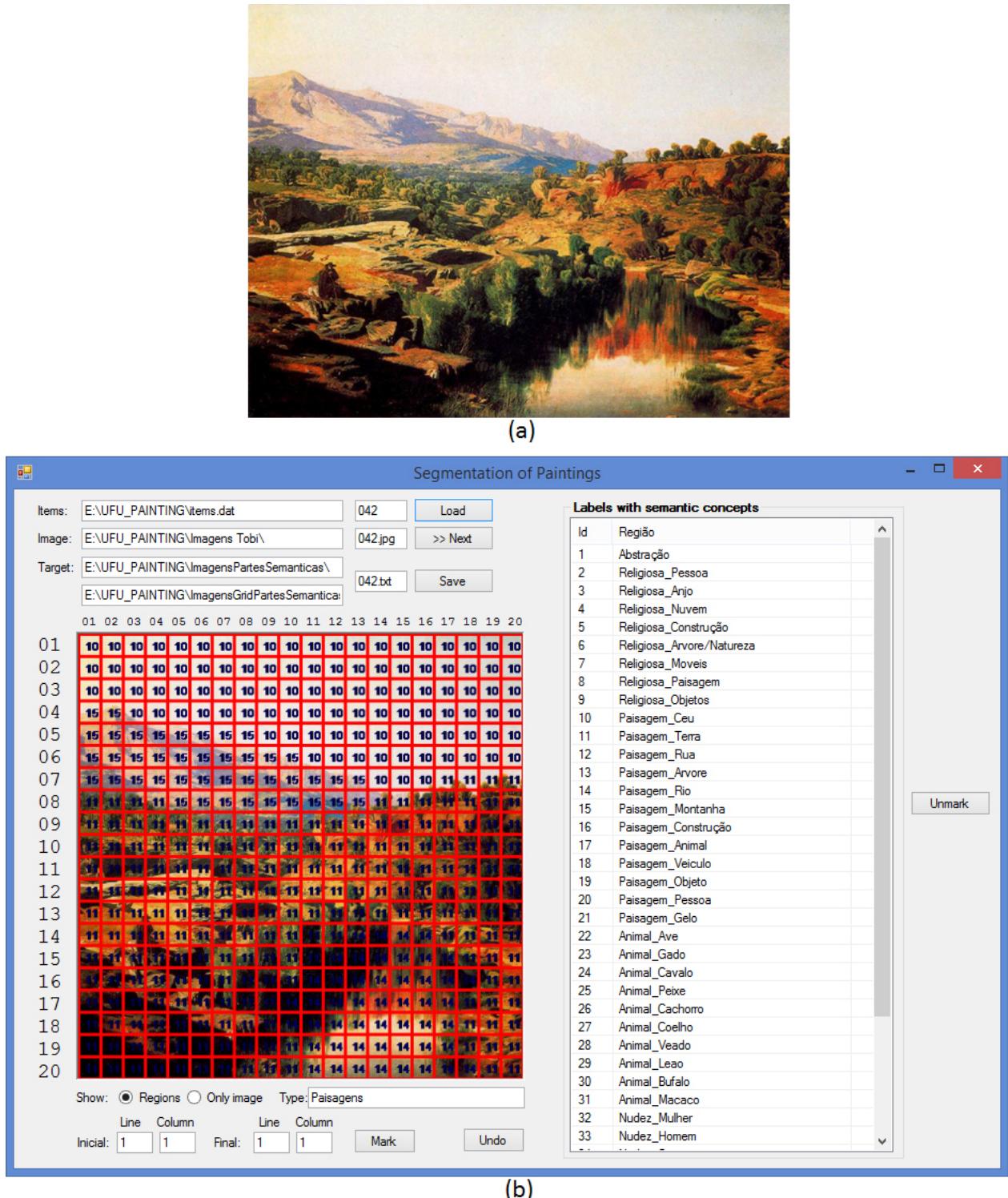


Figura 23 – Segmentação de pinturas: (a) imagem original; (b) Software de segmentação de pinturas. A imagem inteira é dividida em 20 x 20 partes e cada parte é rotulada com um significado semântico.

Tabela 8 – Atributos específicos de pinturas.

Atributos	Valores
Categoria	Animal, Arquitetura, Arte abstrata, Mitologia, Natureza morta, Nudismo, Paisagem, Pessoa e Religiosa
Pintor	280 pintores (valores) diferentes.
Época	Séculos XIV, XV, XVI, XVII, XVIII, XIX, XX e XXI
Movimento	Barroco, Classicismo, Neoclassicismo, Romantismo, Cubismo, Espacialismo, Expressionismo, Fauvismo, Gótico, Renascimento, Impressionismo, Modernismo, Realismo, Simbolismo e Surrealismo.
País	Alemanha, Austria, Espanha, EUA, França, Holanda, Inglaterra, Itália e Rússia.

A comparação do conteúdo visual de imagens de pinturas é um problema extremamente difícil, influenciada pelas diferenças semânticas, contextuais, cognitivas e presença de fortes gradientes locais ocasionados pelas pinceladas (SHRIVASTAVA et al., 2011). Para descrever o conteúdo visual das pinturas, a imagem inteira é caracterizada por três descriptores de características: HOG e GIST, baseados nos trabalhos de recuperação de imagens baseada em conteúdo descritos em (SHRIVASTAVA et al., 2011; MELO; AMO; GULIATO, 2014); e Histograma de Cor, baseado no artigo (ZHANG; HUANG, 2014). A dimensão destes vetores descriptores é: 16399 para o HOG, 512 para o GIST e 72 para o Histograma de Cor.

## 5.2 Comparação com Outros Trabalhos

Para mostrar a eficiência da nossa metodologia, as nossas abordagens são comparadas com métodos bem conhecidos do estado-da-arte. Os métodos comparados são divididos em dois grupos: Métodos baseados em Filtragem Colaborativa e Métodos baseados em Conteúdo. Todos estes métodos utilizados estão disponíveis publicamente no *framework* MyMediaLite (GANTNER et al., 2011).

*Métodos baseados em Filtragem Colaborativa:*

1. *Random*: este método prediz aleatoriamente uma avaliação desconhecida, isto é,  $\hat{r}_{ui}$  = número aleatório entre 1 e 5.
2. *Global Average*: este método prediz uma avaliação desconhecida usando a média global de todas as avaliações, isto é,  $\hat{r}_{ui} = \mu$ .
3. *User Average*: este método usa o valor da média das avaliações do usuário para predizer uma avaliação desconhecida do usuário correspondente.
4. *Item Average*: este método usa a média de avaliações do item para predizer uma avaliação desconhecida do item correspondente.

5. *UIB*: descrito na Seção 2.2.2.1.
6. *UserKNN+Baseline* (*UKB*): descrito na Seção 2.2.2.2.
7. *ItemKNN+Baseline* (*IKB*): descrito na Seção 2.2.2.3.
8. *SVD*: descrito na Seção 2.2.2.4.
9. *SVD+Baseline* (*SB*): descrito na Seção 2.2.2.4, também chamado *Biased MF*, é amplamente utilizado como *baseline* em sistemas de recomendação.
10. *SVD++*: este é um modelo de fatoração de matrizes (KOREN, 2008) que considera o *feedback implícito* dos usuários para a predição. Este modelo também é muito utilizado como *baseline*.
11. *ItemKNN+SVD+Baseline* (*IKS*B): descrito na Subseção 2.2.2.4. Nesta tese, este método foi implementado sob o *framework* MyMediaLite, pois o mesmo não está disponível neste *framework*.

*Métodos baseados em Conteúdo:*

1. *Naive Bayes* (*NB*): este é um método probabilístico baseado em redes bayesianas (GANTNER et al., 2011). Neste método, somente os atributos textuais são usados.
2. *Item KNN + Textual attributes* (*IKT*): este método realiza a predição de avaliação, como descrito na Equação 14, usando a medida de similaridade textual  $TS_{ij}$ , como definida na Equação 43.
3. *Item KNN + Visual features* (*IKV*): este método realiza a predição de avaliação, como descrito na Equação 14, usando a medida de similaridade visual  $FS_{ij}$ , como definida na Equação 44.

Os métodos propostos nesta tese CFAS e CBAS foram desenvolvidos utilizando os recursos do *framework* MyMediaLite (GANTNER et al., 2011). Os sistemas recomendadores de roupas e de pinturas desenvolvidos nesta tese estão ilustrados nos Apêndices B e C, respectivamente.

## 5.3 Parâmetros de Configuração

Os experimentos foram conduzidos aplicando a técnica de validação cruzada (*k-fold-cross-validation*). Utilizou-se  $k = 10$ , 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 validação são executadas de tal forma que em cada iteração  $i$  um diferente conjunto  $D_i$  é usado para validação enquanto os conjuntos restantes são usados para aprendizagem.

Para avaliar o processo de predição de avaliação e o processo de recomendação foram utilizadas as métricas mais populares: RMSE, AP e AUC. No cálculo de AP e AUC foram considerados itens avaliados com 4 ou 5 estrelas como relevantes para o usuário, e itens avaliados com 1, 2 ou 3 estrelas como não relevantes.

Para avaliar o método CFAS baseada em Filtragem Colaborativa com o cenário de *cold-start* de item, selecionou-se aleatoriamente 100 itens com uma razoável quantidade de visualizações para garantir que os itens foram visualmente inspecionados. Para cada item  $i$  criou-se 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  $i$  e o restante da base de dados. O experimento foi executado 100 vezes, uma vez para cada item previamente escolhido.

Os parâmetros dos métodos disponíveis no MyMediaLite foram configurados com valores indicados na literatura correlata como sendo os mais adequados, ou seja, adotou-se a configuração padrão (*default*) do MyMediaLite (GANTNER et al., 2011). Estes parâmetros estão resumidos na Tabela 9.

Para configurar os parâmetros que fazem parte apenas dos métodos propostos foram realizados diversos experimentos, e de acordo com os resultados obtidos adotou-se:

- ❑ a intersecção para o cálculo de similaridade entre vetores de atentividade e similaridade entre os comprimentos das sacadas, conforme Equações 46 e 47, onde  $n = |V_i| = |V_j| = |\mathbb{H}|$  é o número de rótulos semânticos e  $\sum_{t=1}^{|\mathbb{H}|} \left( \frac{V_i[t]}{\theta_i} \right) = 1$ ;

$$simV(V_i, V_j) = \sum_{t=1}^n \min \left( \frac{V_i[t]}{\theta_i}, \frac{V_j[t]}{\theta_j} \right) \quad (46)$$

$$sim\ell(\ell_i, \ell_j) = \frac{\min \left( \frac{\ell_i}{\theta_i}, \frac{\ell_j}{\theta_j} \right)}{\max \left( \frac{\ell_i}{\theta_i}, \frac{\ell_j}{\theta_j} \right)} \quad (47)$$

- ❑ a função linear de agregação, conforme Equação 48, para calcular a similaridade atentiva, onde  $\sigma$  é um parâmetro especificado pelo usuário, cujo valor varia entre 0 a 1. O primeiro termo é importante para comparar a atentividade de acordo com o número de fixações das partes rotuladas semanticamente e o segundo termo é importante para compreender a intensidade das sacadas em uma imagem, por exemplo, se o tamanho de um vestido pode chamar a atenção de um usuário. O valor do parâmetro  $\sigma$  depende do contexto da aplicação. Por exemplo, se o tamanho das sacadas não é importante para a definição de opinião do usuário, o valor de  $\sigma$  pode ser alto, considerando mais a atentividade das regiões. Foi utilizado o  $\sigma = 0,8$  para a base UFU-CLOTHING e o  $\sigma = 0,9$  para a base UFU-PAINTINGS (o tamanho das sacadas não é uma informação tão relevante para o domínio de pinturas);

$$AS_{i,j} = \sigma (simV(V_i, V_j)) + (1 - \sigma) (sim\ell(\ell_i, \ell_j)) \quad (48)$$

- ❑ o parâmetro de encolhimento  $\lambda_{as}$ , definido na Equação 40, igual a 25;
- ❑ a função linear de agregação, conforme Equação 49, para calcular a similaridade combinada  $S_{ij}$  do método CFAS, onde o valor de  $\beta$  está no intervalo [0,1] e é um parâmetro especificado pelo usuário para controlar o peso da similaridade baseada em avaliações. Adotou-se  $\beta = 0,75$ .

$$S_{i,j} = \beta.(RS_{ij}) + (1 - \beta).(AS_{ij}) \quad (49)$$

- ❑ a função geométrica de agregação, conforme Equação 50, para calcular a similaridade combinada  $S_{ij}$  do método CBAS.

$$S_{i,j} = (TS_{ij})^{n_T/\eta} \cdot (FS_{ij})^{n_F/\eta} \cdot (AS_{ij})^{\alpha/\eta}, \quad (50)$$

onde  $n_T = \sum_{k=1}^{|T_i|} T_i[k]$  é a quantidade de características textuais da imagem  $i$  e representa o grau de amplificação da similaridade textual.  $n_F$  é o número de características visuais e representa o grau de amplificação da similaridade visual, e  $\alpha = \gamma_i/\theta_i$  (se  $\theta_i > 0$ ) ou  $\alpha = 0$  (se  $\theta_i = 0$ ). O valor de  $\alpha$  é a média de tempo, em segundos, dos usuários que olharam para a imagem  $i$  e representa o grau de amplificação da atenção visual.  $\eta$  é a soma dos amplificadores, ou seja,  $\eta = n_T + n_F + \alpha$ . À medida que os valores destes amplificadores aumentam, mais importante é o respectivo termo no resultado final. Se o item  $i$  não tem características textuais,  $n_T$  é igual a zero e consequentemente  $TS_{ij}$  será ignorado no cálculo de similaridade. O mesmo se aplica para  $FS_{ij}$  e  $AS_{ij}$ . Note também que  $S_{ij}$  é diferente de  $S_{ji}$ . A combinação geométrica descrita na Equação 50 é interessante, pois não necessita de parâmetros especificados pelo usuário, como por exemplo, a ponderação de similaridades. Além disso, de acordo com o domínio da aplicação e o comportamento dos dados, os amplificadores (veja Equação 19) são atribuídos automaticamente.

- ❑ o vetor de pesos do cálculo de similaridade visual  $\mathbb{W} = \{0.2, 0.2, 0.2, 0.2, 0.2\}$  referente às características Histograma de Cor, HOG, LBP, GIST e SKIN para a base UFU-CLOTHING e  $\mathbb{W} = \{0.4, 0.3, 0.3\}$  referente às características Histograma de Cor, HOG e GIST para a base UFU-PAINTINGS.

Os parâmetros com maior impacto nos resultados serão discutidos no Capítulo 6. Os parâmetros são: medida de similaridade baseada em avaliações; número de vizinhos mais próximos e número de fatores latentes.

## 5.4 Avaliação da Significância Estatística

Segundo (DEMŠAR, 2006), a eficiência de um SR pode ser avaliada usando testes estatísticos por meio do *sign test*. Nesta tese, a avaliação do SR foi conduzida definindo

Tabela 9 – Valores dos parâmetros usados nos experimentos.

Parâmetro	Descrição	Método	De-fault*	Valor usado
Iterações B	Número de iterações para o aprendizado de $b_u$ e $b_i$ .	Todos que usam <i>Baseline</i> .	10	10
$\lambda_u$	Parâmetro regularizador para o cálculo de $b_u$ , definido na Equação 4.	Todos que usam linha de base e vizinhança.	15	15
		Todos que usam linha de base e SVD.	0,015	0,015
$\lambda_i$	Parâmetro regularizador para o cálculo de $b_i$ , definido na Equação 5.	Todos que usam linha de base e vizinhança.	10	10
		Todos que usam linha de base e SVD.	0,015	0,015
$\lambda_s$	<i>Shrinkage</i> : parâmetro de encolhimento, definido na Equação 18.	Todos os métodos KNN.	25	25
Iterações LF	Número de iterações para o aprendizado dos fatores latentes.	Métodos SVD.	30	30
Amplificação	Parâmetro que favorece os vizinhos com similaridade elevada, definido na Equação 19.	Métodos baseados em vizinhança: UKB, IKB e CFAS.	-	4
		Métodos IKS-B e CFAS-SVD (combina fatores latentes e similaridade atentiva).	-	2

os dados como descrito por Shani and Gunawardana (SHANI; GUNAWARDANA, 2013), que usa o *sign test* para a tarefa de predição de avaliação em uma configuração pareada usando o mesmo conjunto de teste. Calcula-se o RMSE médio de cada usuário. Para comparar dois métodos  $A$  e  $B$ , calcula-se o número de usuários cuja média de RMSE é menor em  $A$  do que em  $B$ , denotado por  $m_A$ , e o número de usuários cuja média de RMSE é menor em  $B$  do que em  $A$ , denotado por  $m_B$ . O nível de significância ou *p-value* é obtido de acordo com a Equação 51.

$$p = (0,5)^n \sum_{i=m_A}^n \frac{n!}{i!(n-i)!}, \quad (51)$$

em que  $n = m_A + m_B$ . Quando o *p-value* é menor que um valor pré-definido (normalmente, 0,05), rejeita-se a hipótese nula de que o método  $A$  não é realmente melhor do que o método  $B$  com uma confiança de  $(1 - p) * 100\%$ .

## 5.5 Considerações Finais

Neste capítulo foram apresentadas as bases de dados reais utilizadas para validar a estratégia proposta, o *sign test* que é utilizado para comparar dois métodos com uma mesma base de dados e validar a significância estatística dos resultados obtidos na tarefa de predição de avaliação, uma descrição dos métodos a serem comparados e seus parâmetros. Além disso, também descreveu-se como foi realizada a coleta dos dados para compor as bases de dados UFU-CLOTHING e UFU-PAINTINGS.



# CAPÍTULO 6

---

## Experimentos e Análise dos Resultados

Neste capítulo são apresentados os resultados obtidos a partir da validação da metodologia proposta. Os resultados obtidos foram divididos em dois grupos: Métodos baseados em Filtragem Colaborativa e Métodos baseados em Conteúdo. Para cada grupo, foram executados os experimentos para as tarefas de predição de avaliação e recomendação Top-N, que mostraram a superioridade dos métodos propostos.

### 6.1 Resultados Comparativos entre os Métodos baseados em Filtragem Colaborativa

Nesta Seção, o método CFAS é analisado e comparado com os demais métodos baseados em Filtragem Colaborativa listados na Seção 5.2.

#### 6.1.1 Predição de Avaliação

Os parâmetros com maior impacto nos resultados dos métodos baseados em vizinhança são: i) o número  $k$  de vizinhos mais próximos; ii) a medida de similaridade baseada em avaliações. Para os métodos baseados em fatoração de matrizes o principal parâmetro é o número de fatores latentes. Os experimentos foram realizados alterando os valores destes parâmetros para a escolha do melhor valor de cada parâmetro para cada método comparativo. Estes parâmetros foram utilizados para conduzir o melhor resultado de cada método.

##### 6.1.1.1 Métodos baseados em vizinhança

Nesta subseção são feitas comparações entre o CFAS e os métodos UKB e IKB. Todos usam os parâmetros: i) o número  $k$  de vizinhos mais próximos; ii) a medida de similaridade baseada em avaliações. Foram executados os experimentos variando o número de vizinhos ( $k$ ) de 10 a 50 e o RMSE foi calculado para cada método. Estes experimentos

foram conduzidos usando três medidas de similaridade baseada em avaliações: *Pearson Correlation Coefficient* (PCC), Cosseno e a Euclidiana invertida (Euc). A Figura 24 ilustra os resultados aplicados na base UFU-CLOTHING e a Figura 25 ilustra os resultados aplicados na base UFU-PAINTINGS.

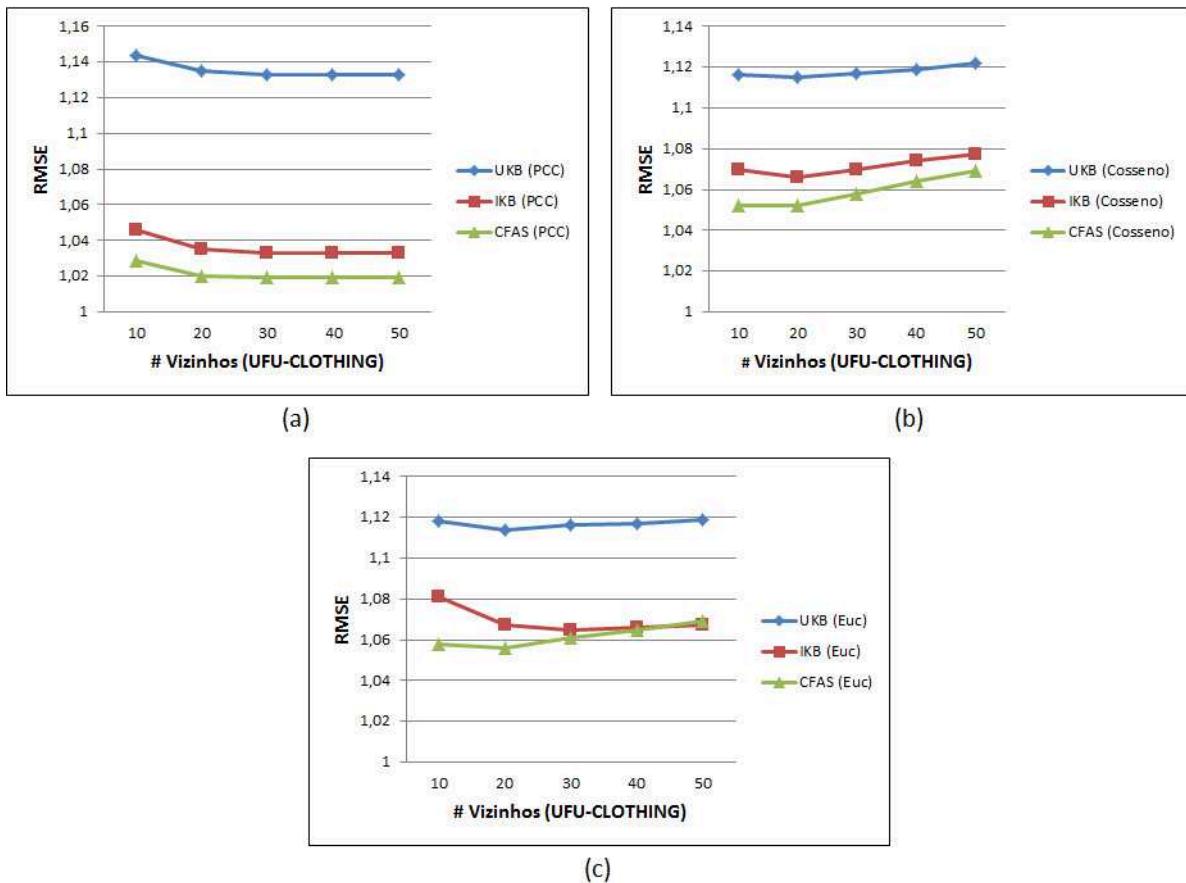


Figura 24 – Gráficos para avaliar o melhor número de vizinhos para os métodos UKB, IKB e CFAS aplicados na **UFU-CLOTHING**, em termos de RMSE. Medidas de similaridade utilizadas: (a) *Pearson Correlation Coefficient* (PCC); (b) Cosseno e (c) inversão da distância Euclidiana (Euc).

□ **Base de dados UFU-CLOTHING:** Usando a medida de similaridade PCC, os melhores resultados usando a UFU-CLOTHING (Figura 24(a)) foram obtidos com  $k = 30$ . O RMSE dos métodos UKB, IKB e CFAS foram 1,133, 1,033 e 1,019 respectivamente. O ganho do método proposto CFAS em relação ao UKB foi de 10,05% a 10,13% e o ganho em relação ao IKB foi de 1,35% a 1,62%.

Com a medida Cosseno (Figura 24(b)), os melhores resultados foram obtidos com  $k = 20$ , com RMSE=1,115 para o UKB, RMSE=1,066 para o IKB e RMSE=1,052 para o CFAS. O ganho do CFAS em relação ao UKB foi de 5,28% a 5,73% e o ganho em relação ao IKB foi de 0,74% a 1,68%.

Com a distância Euclidiana (Euc) (Figura 24(c)), o melhor resultado do UKB foi RMSE = 1,114 com o  $k = 20$ , do IKB foi RMSE = 1,065 com  $k = 30$  e do CFAS

foi  $\text{RMSE} = 1,056$  com  $k = 20$ . Nestes casos, o ganho do CFAS em relação ao UKB foi de 5,2% e em relação ao IKB foi de 0,85%. Note que o método CFAS é superior em todos os casos, com exceção no  $k = 50$ .

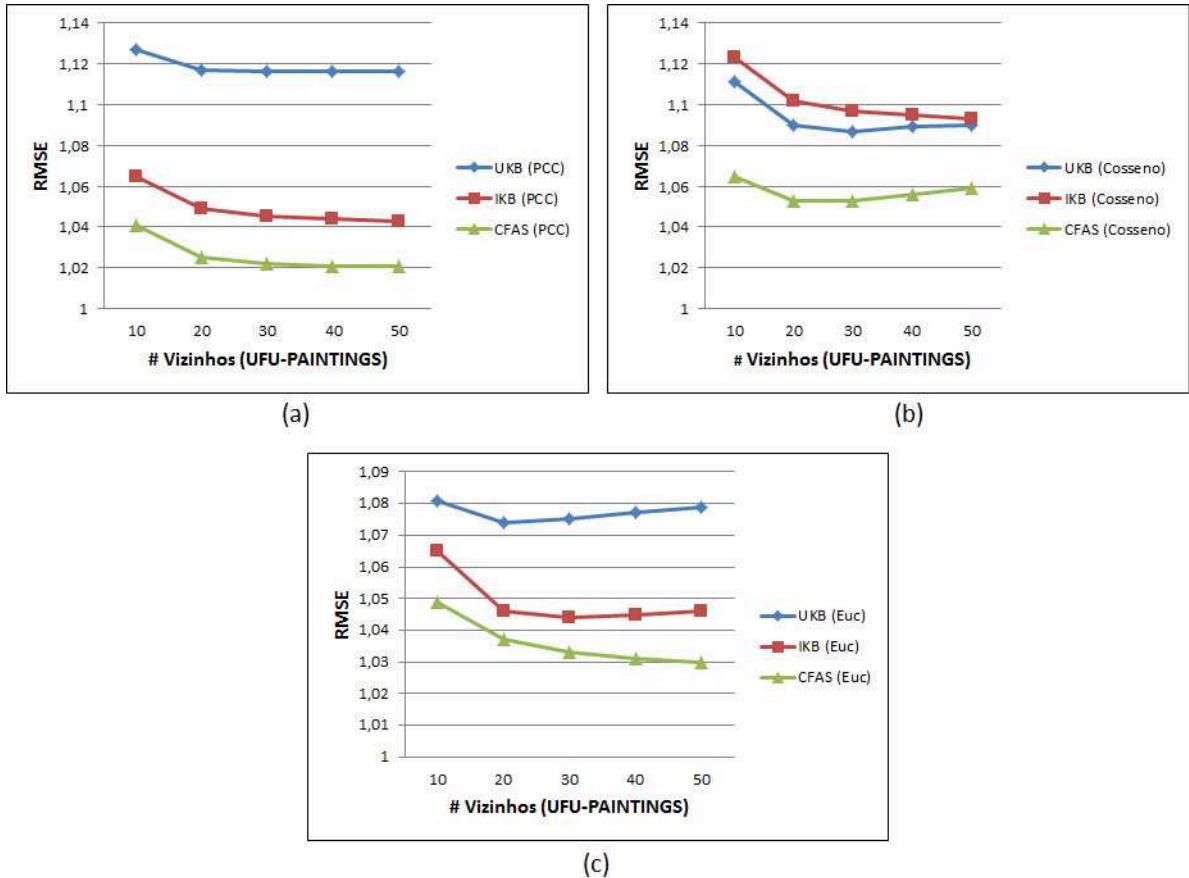


Figura 25 – Gráficos para avaliar o melhor número de vizinhos para os métodos UKB, IKB e CFAS aplicados na **UFU-PAINTINGS**, em termos de RMSE. Medidas de similaridade utilizadas: (a) *Pearson Correlation Coefficient* (PCC); (b) Cosseno e (c) inversão da distância Euclidiana (Euc).

□ **Base de dados UFU-PAINTINGS:** Usando a medida de similaridade PCC (Figura 25(a)), os melhores resultados foram obtidos com  $k = 40$ . O RMSE dos métodos UKB, IKB e CFAS foram 1,116, 1,044 e 1,021 respectivamente. O ganho do CFAS em relação ao UKB foi de 7,63% a 8,51% e o ganho em relação ao IKB foi de 2,10% a 2,28%.

Usando a medida Cosseno (Figura 25(b)), os melhores resultados foram obtidos com  $k = 30$  para os métodos UKB e CFAS e  $k = 50$  para o método IKB , com  $\text{RMSE}=1,087$  para o UKB,  $\text{RMSE}=1,093$  para o IKB e  $\text{RMSE}=1,053$  para o CFAS. O ganho do CFAS em relação ao UKB foi de 2,84% a 4,14% e o ganho em relação ao IKB foi de 3,11% a 5,16%.

Usando a distância Euclidiana (Figura 25(c)), o melhor resultado do UKB foi  $\text{RMSE} = 1,074$  com o  $k = 20$ , do IKB foi  $\text{RMSE} = 1,044$  com  $k = 30$  e do CFAS foi  $\text{RMSE}$

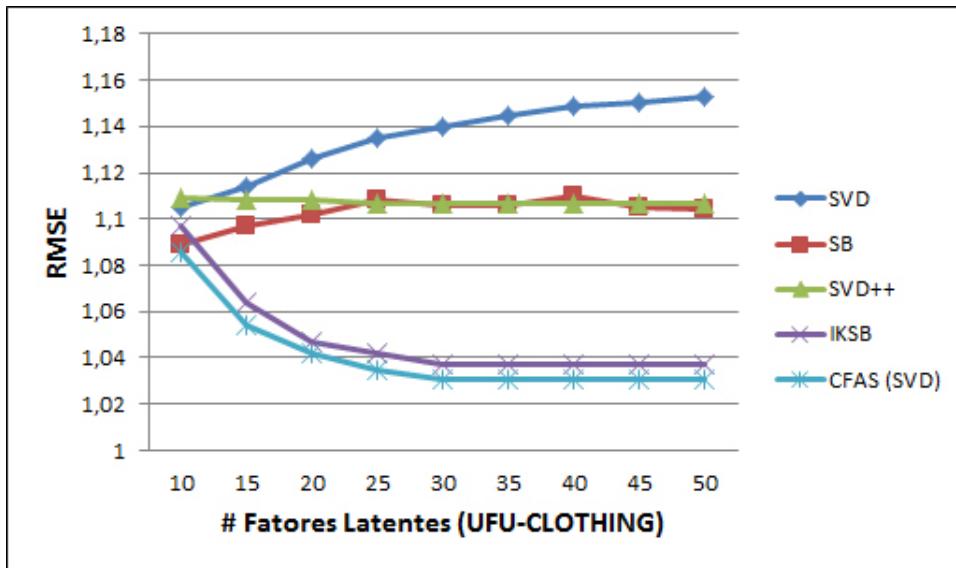


Figura 26 – Avaliação do número de fatores latentes para cada método comparativo aplicado na UFU-CLOTHING.

$= 1,030$  com  $k = 50$ . O ganho do CFAS em relação ao UKB foi de 2,96% a 4,54% e em relação ao IKB foi de 0,86% a 1,52%.

Independentemente da medida de similaridade, o método CFAS demonstrou superioridade em relação aos demais métodos. Analisando o método CFAS em relação às medidas de similaridade (veja Figuras 24 e 25), percebe-se que os resultados obtidos usando a medida de similaridade PCC em ambas as bases de dados são muito superiores em todas as variações de vizinhança.

#### 6.1.1.2 Métodos baseados em fatoração de matrizes

O método proposto CFAS pode combinar a similaridade entre fatores latentes com a similaridade atentiva, como definido na Equação 22. Denota-se esta combinação como CFAS (SVD). Os métodos SVD, SB, SVD++ e o IKS requerem o número de fatores latentes como parâmetro. Os experimentos foram realizados variando o número de fatores latentes entre 10 e 50. As Figuras 26 e 27 mostram um gráfico em termos de número de fatores latentes versus RMSE aplicados na UFU-CLOTHING e UFU-PAINTINGS respectivamente.

- **Base de dados UFU-CLOTHING:** Os melhores resultados obtidos na UFU-CLOTHING, Figura 26, foram: RMSE=1,031 para o CFAS (SVD), RMSE=1,037 para o IKS, RMSE=1,107 para o SVD++, RMSE=1,089 para o SB e RMSE=1,105 para o SVD. O ganho do método proposto CFAS (SVD) em relação aos métodos IKS, SVD++, SB e SVD foi de 0,6%, 6,9%, 5,3% e 6,7% respectivamente.

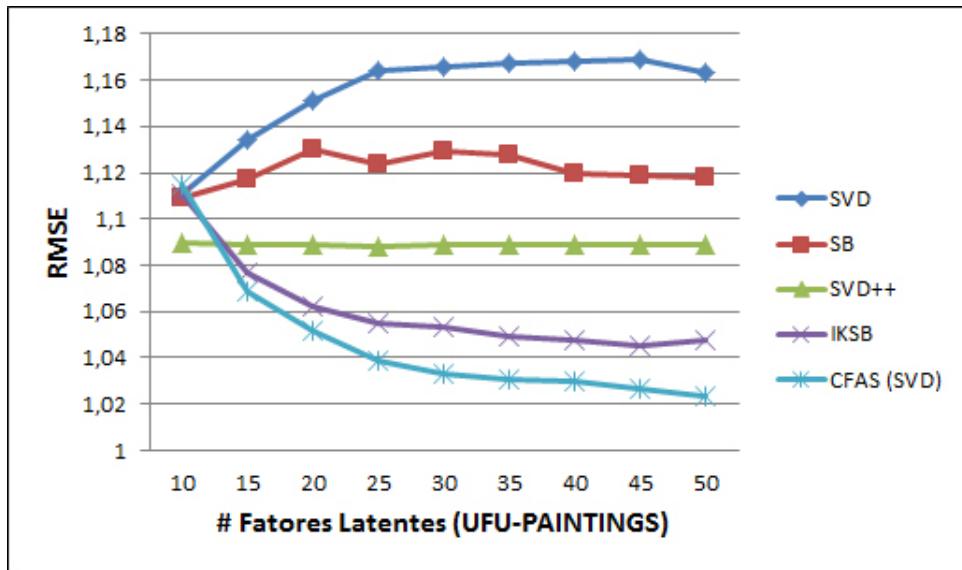


Figura 27 – Avaliação do número de fatores latentes para cada método comparativo aplicado na UFU-PAINTINGS.

□ **Base de dados UFU-PAINTING:** Os melhores resultados obtidos na UFU-PAINTINGS, Figura 27, foram: RMSE=1,023 para o CFAS (SVD), RMSE=1,045 para o IKS, RMSE=1,089 para o SVD++, RMSE=1,109 para o SB e RMSE=1,111 para o SVD. O ganho do CFAS (SVD) em relação aos métodos IKS, SVD++, SB e SVD foi de 2,1%, 6,1%, 7,8% e 7,9% respectivamente.

As Figuras 26 e 27 demonstram que em ambas as bases de dados, o método proposto CFAS (SVD) foi superior em todos os casos.

#### 6.1.1.3 Análise estatística de todos os métodos baseados em FC

A Tabela 10 resume os melhores resultados obtidos nos experimentos para os métodos baseados em vizinhança, métodos baseados em fatores latentes e métodos que não possuem os parâmetros mencionados nesta Subseção 6.1.1, são eles *Random*, *Global Average*, *User Average*, *Item Average* e *UIB*. Os resultados são apresentados em termos de RMSE juntamente com o valor do parâmetro (se usado) para cada base de dados.

Na tarefa de predição de avaliação, o ganho do método proposto em termos de RMSE aparentemente é pequeno, entretanto estes ganhos são muito significativos na área de sistemas de recomendação. A superioridade do método CFAS com relação aos demais métodos do estado-da-arte foi confirmada pelo cálculo da significância estatística. O método estatístico utilizado foi o *sign test* e os resultados são apresentados na Tabela 11 (UFU-CLOTHING) e na Tabela 12 (UFU-PAINTINGS). Pode-se perceber que o método CFAS (PCC) apresenta um *p – value* menor que 0,05 quando comparado com os demais métodos. A proposta com fatores latentes, denominada CFAS (SVD), foi superior estatisticamente em relação a todos os demais métodos baseados em fatores latentes.

Tabela 10 – Comparação de desempenho em termos de RMSE dos métodos baseados em Filtragem Colaborativa aplicados nas bases UFU-CLOTHING e UFU-PAINTINGS.

Métodos	UFU-CLOTHING		UFU-PAINTINGS	
	RMSE	Parâmetros	RMSE	Parâmetros
1. Random	1,741		1,752	
2. Global Average	1,292		1,306	
3. User Average	1,205		1,184	
4. Item Average	1,248		1,234	
5. UIB	1,114		1,100	
6. UKB	1,114	Viz: 20; Sim: Euc	1,074	Viz: 20; Sim: Euc
7. IKB	1,033	Viz: 30; Sim: PCC	1,044	Viz: 40; Sim: PCC
8. SVD	1,105	F.L.: 10	1,111	F.L.: 10
9. SB	1,089	F.L.: 10	1,109	F.L.: 10
10. SVD++	1,107	F.L.: 30	1,089	F.L.: 15
11. IKS	1,037	Viz: 30; F.L.: 30	1,045	Viz: 40; F.L.: 45
<b>12. CFAS</b>	<b>1,019</b>	Viz: 30; Sim: PCC	<b>1,021</b>	Viz: 40; Sim: PCC
<b>13. CFAS (SVD)</b>	<b>1,031</b>	Viz: 30; F.L.: 30	<b>1,023</b>	Viz: 40; F.L.: 50

**Legenda:**

Viz: número de vizinhos mais próximos; F.L.: número de fatores latentes.

Tabela 11 – *p-values* obtidos aos pares pelo *sign test* na UFU-CLOTHING.

Métodos	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Random	-												
2. Global Average	**	-											
3. User Average	**	**	-	,35									
4. Item Average	**	**	,69	-									
5. UIB	**	**	**	**	-	,26		,92		,81			
6. UKB	**	**	**	**	,77	-					,92		
7. IKB	**	**	**	**	**	**	-	**	**	**	,15		,92
8. SVD	**	**	**	**	,10	**		-			*		
9. SB	**	**	**	**	**	**		*	-		**		
10. SVD++	**	**	**	**	,22	,10		,96			-		
11. IKS	**	**	**	**	**	**	,87	**	**	**	-		
<b>12. CFAS (PCC)</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>-</b>	<b>**</b>							
<b>13. CFAS (SVD)</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>,10</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>**</b>	<b>-</b>	

**Legenda:**

vazio  $\Rightarrow p\text{-value} > 0,9$ ,

\*  $\Rightarrow 0,001 \leq p\text{-value} < 0,05$  (significância estatística),

\*\*  $\Rightarrow p\text{-value} < 0,001$  (significância estatística).

Tabela 12 –  $p$ -values obtidos aos pares pelo *sign test* na UFU-PAINTINGS.

Métodos	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Random	-												
2. Global Average	**	-											
3. User Average	**	**	-	,07									
4. Item Average	**	**	,94	-									
5. UIB	**	**	**	**	-			,55	0,38				
6. UKB	**	**	**	**	**	-		**	**	**			
7. IKB	**	**	**	**	**	**	-	**	**	**	,33		
8. SVD	**	**	**	**	,50			-	0,50	0,84			
9. SB	**	**	**	**	0,66			0,55	-				
10. SVD++	**	**	**	**	**			,19	0,05	-			
11. IKS <sub>B</sub>	**	**	**	**	**	**	,71	**	**	**	-		
12. CFAS (PCC)	**	**	**	**	**	**	**	**	**	**	**	-	**
13. CFAS (SVD)	**	**	**	**	**	**	**	**	**	**	**	-	

**Legenda:**

vazio  $\Rightarrow p$ -value  $> 0,9$ ,

\*  $\Rightarrow 0,001 \leq p$ -value  $< 0,05$  (significância estatística),

\*\*  $\Rightarrow p$ -value  $< 0,001$  (significância estatística).

### 6.1.2 Recomendação Top-N

Na tarefa de recomendação Top-N, o sistema recomenda ao usuário os primeiros  $N$  itens mais relevantes para ele. Na prática é desejável que os SR recomendem apenas um pequeno conjunto de itens relevantes para o usuário. Para evitar problemas com eficiência e precisão,  $N$  deve ser escolhido com cuidado (RICCI; ROKACH; SHAPIRA, 2011). Um valor de  $N$  pequeno é interessante para manter uma precisão mais coerente em bases de testes pequenas. A Tabela 13 apresenta a AP para Top-5 e AUC dos métodos FC aplicados nas bases UFU-CLOTHING e UFU-PAINTINGS. Os parâmetros utilizados foram os mesmos dos experimentos apresentados na Tabela 10.

Note que em todos os casos, o CFAS obteve melhores resultados quando comparado com os outros métodos. Os ganhos obtidos do CFAS (PCC) e do CFAS (SVD) em relação aos demais métodos estão representados nas Tabelas 14 e 15, respectivamente.

### 6.1.3 CFAS e o Problema de Partida a Frio

O problema de novos itens é um grande desafio em sistemas de recomendação, especialmente quando eles são baseados em Filtragem Colaborativa. Se um novo item  $i$  ocorre em um método FC baseado em vizinhança, não é possível calcular a similaridade entre o item  $i$  e os demais itens. Consequentemente, não é possível calcular a predição de avaliação, como mostram as Equações 13 e 14. Em métodos FC baseados em fatores latentes, não é possível calcular o vetor de fatores latentes  $q_i$ , o que torna impossível predizer a avaliação, conforme Equação 20.

Tabela 13 – Avaliação AP@5 e AUC do processo de recomendação dos métodos FC aplicados na UFU-CLOTHING e UFU-PAINTINGS.

Métodos	UFU-CLOTHING		UFU-PAINTINGS	
	AP@5	AUC	AP@5	AUC
1. Random	0,354	0,501	0,421	0,500
2. Global Average	0,351	0,499	0,416	0,500
3. User Average	0,351	0,499	0,416	0,500
4. Item Average	0,549	0,687	0,603	0,699
5. UIB	0,560	0,695	0,601	0,701
6. UKB	0,568	0,704	0,613	0,716
7. IKB	0,627	0,743	0,631	0,736
8. SVD	0,589	0,718	0,611	0,709
9. SB	0,598	0,724	0,610	0,715
10. SVD++	0,557	0,689	0,603	0,702
11. IKS <sub>B</sub>	0,628	0,746	0,630	0,739
<b>12. CFAS (PCC)</b>	<b>0,644</b>	<b>0,757</b>	<b>0,651</b>	<b>0,760</b>
<b>13. CFAS (SVD)</b>	<b>0,633</b>	<b>0,751</b>	<b>0,644</b>	<b>0,754</b>

Tabela 14 – Ganho na AP e AUC do método CFAS (PCC) quando comparado aos demais métodos FC.

Métodos	UFU-CLOTHING		UFU-PAINTINGS	
	AP@5	AUC	AP@5	AUC
1. Random	82%	51%	55%	52%
2. Global Average	83%	52%	56%	52%
3. User Average	83%	52%	56%	52%
4. Item Average	17%	10%	08%	09%
5. UIB	15%	09%	08%	08%
6. UKB	13%	08%	06%	06%
7. IKB	03%	02%	03%	03%
8. SVD	09%	05%	07%	07%
9. SB	08%	05%	07%	06%
10. SVD++	16%	10%	08%	08%
11. IKS <sub>B</sub>	03%	01%	03%	03%

Tabela 15 – Ganho na AP e AUC do método CFAS (SVD) quando comparado aos demais métodos FC.

Métodos	UFU-CLOTHING		UFU-PAINTINGS	
	AP@5	AUC	AP@5	AUC
1. Random	79%	50%	53%	51%
2. Global Average	80%	51%	55%	51%
3. User Average	80%	51%	55%	51%
4. Item Average	15%	09%	07%	08%
5. UIB	13%	08%	07%	08%
6. UKB	11%	07%	05%	05%
7. IKB	01%	01%	02%	02%
8. SVD	07%	05%	05%	06%
9. SB	06%	04%	06%	05%
10. SVD++	14%	09%	07%	07%
11. IKS <sub>B</sub>	01%	01%	02%	02%

Tabela 16 – Comparaçāo em termos de RMSE para o problema de partida a frio de item aplicados nas bases UFU-CLOTHING e UFU-PAINTINGS.

Métodos	UFU-CLOTHING		UFU-PAINTINGS	
	RMSE	Parâmetros	RMSE	Parâmetros
1. Random	1,723		1,740	
2. Global Average	1,281		1,304	
3. User Average	1,214		1,209	
4. UserItemBaseline	1,212		1,207	
<b>5. CFAS</b>	<b>1.159</b>	Vizinhos: 30	<b>1.172</b>	Vizinhos: 40

Em um cenário onde o sistema de recomendação não possui atributos dos itens (o SR não pode ser BC), seja por dificuldades de manutenção ou espaço em disco, propõe-se superar este problema utilizando a atenção visual para identificar possíveis vizinhos semelhantes a um novo item  $i$ . Em uma situação onde um novo item  $i$  não foi avaliado, porém já foi visualizado por alguns usuários, é possível calcular a similaridade atentiva entre o item  $i$  e os demais itens e, consequentemente, predizer avaliações sobre o item  $i$  utilizando o método CFAS, conforme Equação 42.

Para avaliar a eficiência desta estratégia, foram selecionados os métodos baseados em FC do estado-da-arte que podem ser aplicados a este problema: *Random*; *Global Average*; *User Average*; e *UserItemBaseline*, conforme definido na Equação 1 e reduzido a  $\mu + b_u$ , pois  $b_i$  é desconhecido.

A Tabela 16 descreve os resultados e mostra que o método proposto CFAS obteve melhor acurácia na tarefa de predição de avaliação. O melhor método do estado-da-arte foi o *UserItemBaseline* com RMSE = 1,212 para UFU-CLOTHING e RMSE = 1,207 para UFU-PAINTINGS, enquanto o CFAS obteve um RMSE = 1,159 para UFU-CLOTHING e RMSE = 1,172 para UFU-PAINTINGS, ou seja, um ganho de 4,4% e 3% respectivamente.

## 6.2 Resultados Comparativos entre os Métodos baseados em Conteúdo

Nesta seção, o método proposto CBAS é analisado e comparado com os demais métodos baseados em conteúdo listados na Seção 5.2.

### 6.2.1 Predição de Avaliação

Nesta seção o método CBAS será comparado com o método estatístico *Naive Bayes* (*NB*) e os métodos baseados em vizinhança IKT e IKV. O método *Item KNN + Textual attributes* (*IKT*) utiliza a medida de similaridade textual definida na Equação 43 e o método *Item KNN + Visual features* (*IKV*) utiliza a medida de similaridade visual definida na Equação 44.

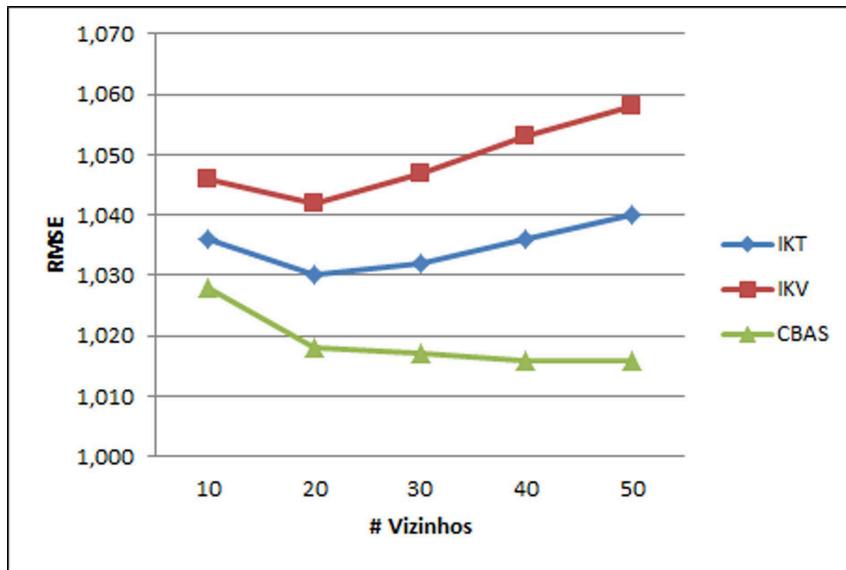


Figura 28 – Avaliação do melhor número de vizinhos para os métodos IKT, IKV e CBAS aplicados na UFU-CLOTHING, em termos de RMSE.

#### 6.2.1.1 Métodos baseados em vizinhança

Nos métodos IKT, IKV e CBAS o parâmetro principal para o desempenho do método é o número  $k$  de vizinhos mais próximos. Os experimentos foram realizados variando o número de vizinhos de 10 a 50 e o RMSE foi calculado para cada método.

- **Base de dados UFU-CLOTHING:** Aplicando o CBAS na UFU-CLOTHING (Figura 28), o melhor resultado foi  $\text{RMSE} = 1,016$  com  $k = 40$ . Os melhores resultados obtidos pelos métodos IKT e IKV foram  $\text{RMSE} = 1,030$  e  $\text{RMSE} = 1,042$  com  $k = 20$ . O ganho do CBAS em relação ao IKT e IKV foi de 0,8% a 2,3% e 1,7% a 4% respectivamente.
- **Base de dados UFU-PAINTINGS:** Aplicando o CBAS na UFU-PAINTINGS (Figura 29), o melhor resultado foi  $\text{RMSE} = 1,021$  com  $k = 20$ . Os melhores resultados obtidos pelos métodos IKT e IKV foram  $\text{RMSE} = 1,050$  com  $k = 30$  e  $\text{RMSE} = 1,089$  com  $k = 40$  respectivamente. O ganho do CBAS em relação ao IKT foi de 1,5% a 3% e em relação ao IKV foi de 4,8% a 8%.

As Figuras 28 e 29 demonstram que em ambas as bases de dados, o método CBAS foi superior em todos os casos.

#### 6.2.1.2 Análise estatística de todos os métodos baseados em conteúdo

A Tabela 17 resume os melhores resultados obtidos nos experimentos para os métodos baseados em vizinhança e para o método NB.

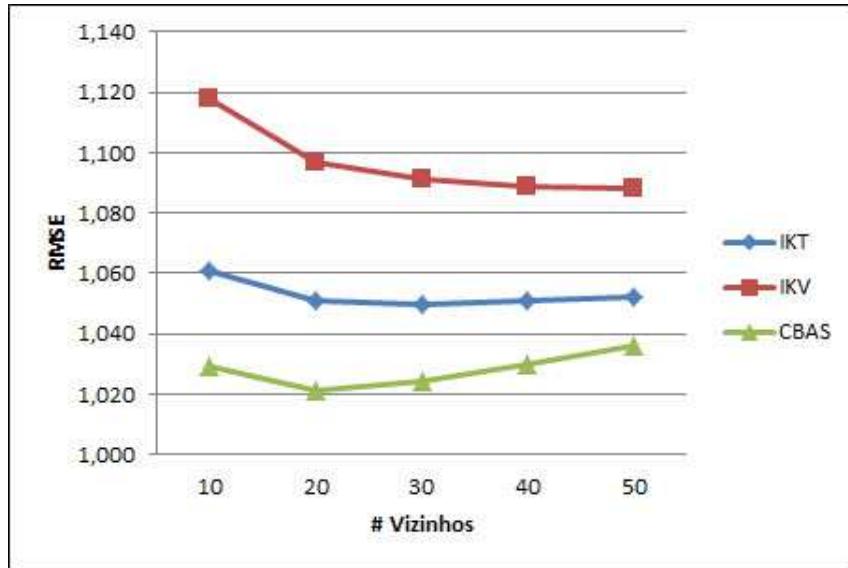


Figura 29 – Avaliação do melhor número de vizinhos para os métodos IKT, IKV e CBAS aplicados na UFU-PAINTINGS, em termos de RMSE.

Tabela 17 – Comparaçāo de desempenho em termos de RMSE dos métodos baseados em conteúdo aplicados nas bases UFU-CLOTHING e UFU-PAINTINGS.

Métodos	UFU-CLOTHING		UFU-PAINTINGS	
	RMSE	Parâmetros	RMSE	Parâmetros
<i>Naive Bayes</i> (NB)	1,332		1,210	
2. IKT	1,030	Vizinhos: 20	1,050	Vizinhos: 30
3. IKV	1,042	Vizinhos: 20	1,089	Vizinhos: 40
<b>4. CBAS</b>	<b>1,016</b>	Vizinhos: 40	<b>1,021</b>	Vizinhos: 20

Tabela 18 – *p-values* obtidos aos pares pelo *sign test* nas bases UFU-CLOTHING e UFU-PAINTINGS.

Métodos	UFU-CLOTHING				UFU-PAINTINGS			
	1	2	3	4	1	2	3	4
1. NB	-				-			
2. IKT	**	-	**		**	-	**	
3. IKV	**	-			**	-		
<b>4. CBAS</b>	**	**	**	-	**	**	**	-

**Legenda:**

**vazio** ⇒  $p\text{-value} > 0,9$ ,

**\*\*** ⇒  $p\text{-value} < 0,001$  (significância estatística).

A superioridade do método CBAS em relação ao métodos IKT, IKV e NB foi confirmada usando também o *sign test*. O CBAS apresentou o  $p\text{-value}$  menor que 0,05 quando comparado com os demais métodos, conforme Tabela 18.

Tabela 19 – Avaliação AP@5 e AUC do processo de recomendação dos métodos baseados em conteúdo aplicados na UFU-CLOTHING e UFU-PAINTINGS.

<b>Métodos</b>	<b>UFU-CLOTHING</b>		<b>UFU-PAINTINGS</b>	
	<b>AP@5</b>	<b>AUC</b>	<b>AP@5</b>	<b>AUC</b>
1. NaiveBayes (NB)	0,495	0,644	0,572	0,677
2. IKT	0,624	0,753	0,630	0,735
3. IKV	0,618	0,747	0,605	0,705
<b>4. CBAS</b>	<b>0,646</b>	<b>0,766</b>	<b>0,654</b>	<b>0,757</b>

Tabela 20 – Ganhos na AP e AUC do método CBAS quando comparado aos demais métodos baseados em conteúdo.

<b>Métodos</b>	<b>UFU-CLOTHING</b>		<b>UFU-PAINTINGS</b>	
	<b>AP@5</b>	<b>AUC</b>	<b>AP@5</b>	<b>AUC</b>
1. NaiveBayes (NB)	31%	19%	14%	12%
2. IKT	04%	02%	04%	03%
3. IKV	05%	03%	08%	07%

### 6.2.2 Recomendação Top-N

A Tabela 19 apresenta a AP para Top-5 e AUC dos métodos BC aplicados nas bases UFU-CLOTHING e UFU-PAINTINGS. Note que em ambas as bases, o CBAS obteve melhores resultados quando comparado aos demais métodos. A Tabela 20 apresenta os ganhos obtidos do método CBAS em relação aos demais métodos NB, IKT e IKV.

## 6.3 Análise da Atenção Visual

Nesta seção, avalia-se como a atenção visual e os rótulos semânticos afetam a predição de avaliação dos métodos CFAS e CBAS.

### 6.3.1 Impacto dos Dados de Atenção Visual nos Métodos Propostos

As bases de dados UFU-CLOTHING e UFU-PAINTINGS possuem 469.071 e 444.780 fixações de olhar, respectivamente. Neste experimento, analisa-se a relação entre a quantidade de dados de atenção visual com o desempenho dos métodos propostos CFAS e CBAS.

A Figura 30 mostra o resultado dos métodos CFAS (PCC), CFAS (SVD) e CBAS em termos de predição de avaliação. Neste experimento, decrementa-se aleatoriamente e gradativamente a quantidade de fixações de olhares no conjunto de treinamento para verificar a influência da atenção visual sobre o desempenho dos métodos. Em ambas as bases de dados e ambos os métodos, percebe-se que ao diminuir o número de fixações de olhares o desempenho dos métodos também diminui. Isto se deve ao fato de que imagens

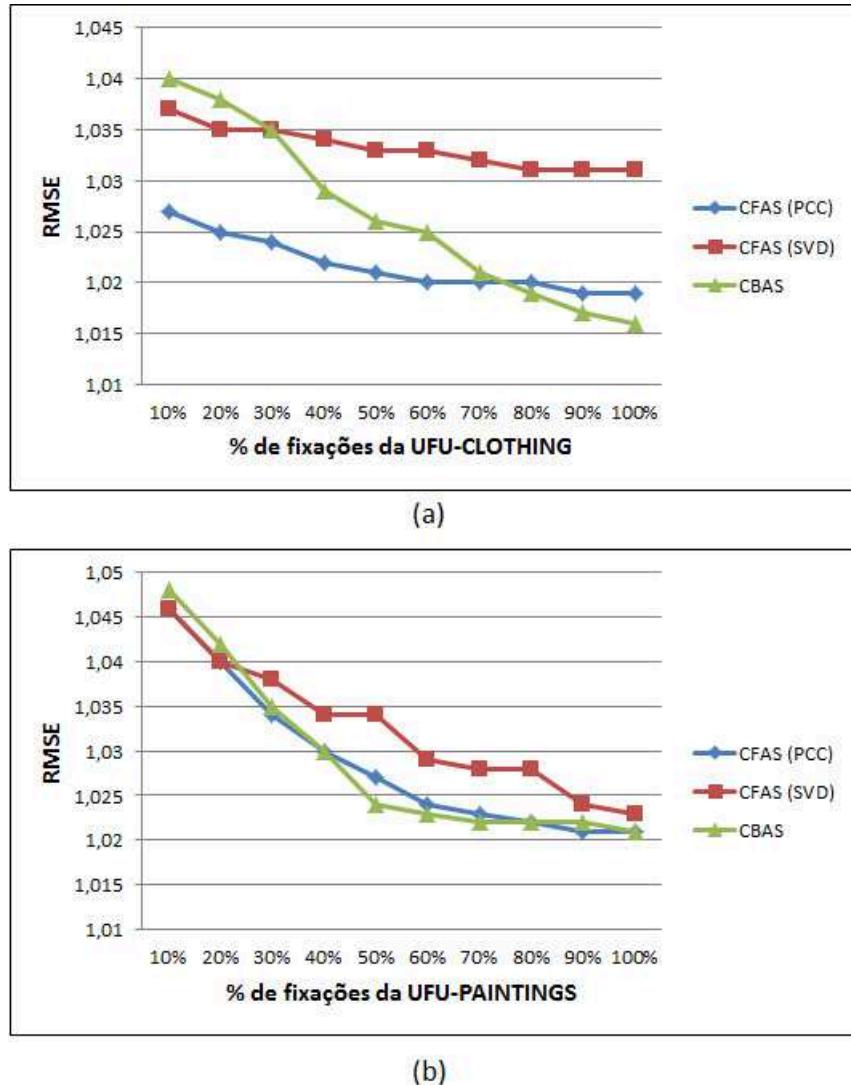


Figura 30 – Comparação dos métodos CFAS (PCC), CFAS (SVD) e CBAS em termos de RMSE e tamanho da base de fixações: (a) UFU-CLOTHING e (b) UFU-PAINTINGS.

observadas por uma quantidade muito pequena de usuários prejudicam a representatividade da similaridade atentiva. Entretanto, acredita-se que depois de certa quantidade de observações sobre um item, o mesmo já terá sua atentividade bem descrita e não precisará de mais observações.

Para avaliar o impacto da representatividade dos rótulos semânticos, realizou-se outro experimento alterando o conjunto  $\mathbb{H}$  de rótulos semânticos.

**□ Base de dados UFU-CLOTHING:** Para a base UFU-CLOTHING, o corpo humano foi dividido em 22, 11, 4 e 2 partes, conforme Figura 31. Os resultados comprovam que dividir as pessoas em 22 partes do corpo humano tornam o cálculo da similaridade atentiva mais detalhista e com uma acurácia melhor na predição de avaliação quando comparado com um número menor de partes do corpo. A Figura 33 (a) mostra que o desempenho dos métodos cai de acordo com a diminuição do

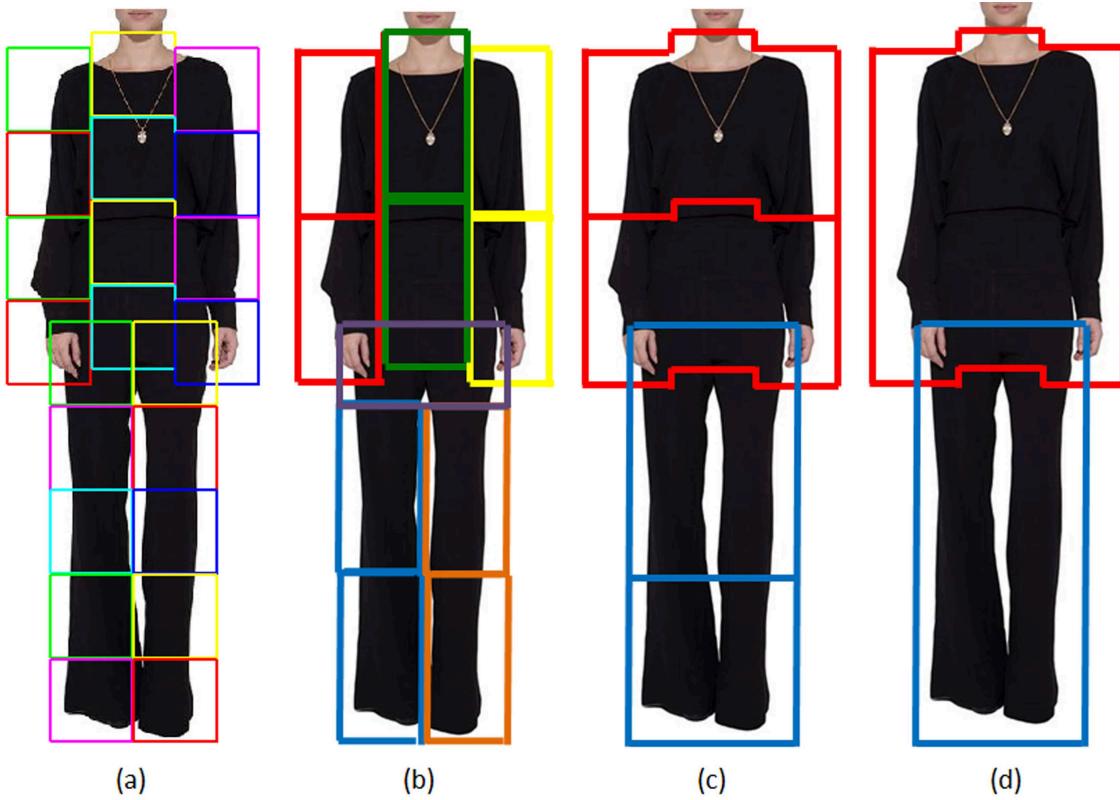


Figura 31 – Diferentes divisões de partes do corpo humano. (a) 22 partes; (b) 11 partes; (c) 4 partes; (d) 2 partes.

número de rótulos semânticos e aumento das regiões que representam as partes do corpo humano.

- **Base de dados UFU-PAINTINGS:** Para a base UFU-PAINTINGS, utilizou-se a categoria da pintura para definir os rótulos semânticos. Com isso o experimento contou com duas representações: 41 rótulos semânticos (originais) e 9 rótulos semânticos (Animal, Arquitetura, Arte abstrata, Mitologia, Natureza morta, Nudismo, Paisagem, Pessoa e Religiosa). Um exemplo de como uma pintura é rotulada é dado na Figura 32.

A Figura 33 (b) mostra que a representatividade maior, com 41 rótulos, obteve uma acurácia melhor na predição de avaliação.

Em ambas as bases, os experimentos com maior quantidade de rótulos semânticos obteve melhores resultados. Entretanto, este fato não comprova que quanto mais rótulos semânticos, melhor a acurácia para qualquer aplicação. No caso de roupas, por exemplo, podemos exceder os detalhes da atentividade dividindo o corpo humano em mais de 22 partes, prejudicando o desempenho dos métodos propostos.



Figura 32 – Pintura rotulada com diferentes conjuntos de rótulos semânticos ( $\mathbb{H}$ ). (a) o conjunto  $\mathbb{H}$  possui 41 rótulos de acordo com o Apêndice A. No exemplo, 10 = céu, 13 = árvore, 15 = montanha/pedras, 24 = cavalo, e 29 = leão; (b) o conjunto  $\mathbb{H}$  possui 9 rótulos definidos de acordo com a categoria da pintura. No exemplo, 1 = animal e 7 = paisagem.

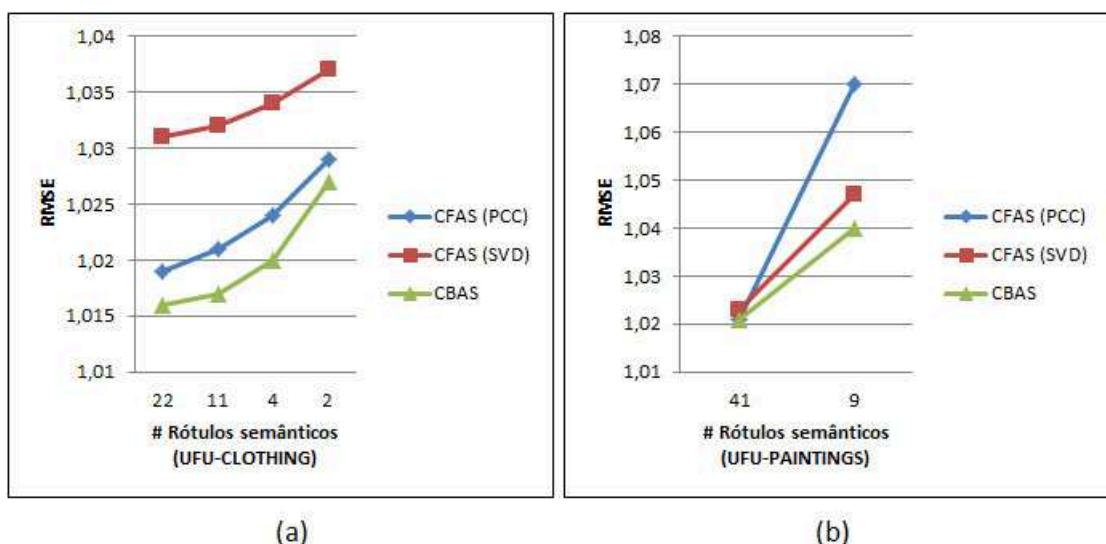


Figura 33 – RMSE de acordo com o número de rótulos semânticos ( $|H|$ ). (a) UFU-CLOTHING; (b) UFU-PAINTINGS

## 6.4 Considerações Finais

Este capítulo apresentou um conjunto de experimentos usando os métodos CFAS, CBAS e os principais métodos da literatura para SR. Os experimentos foram divididos de acordo com a técnica de recomendação: FC e BC. Também foi analisado o impacto do uso da atenção visual nos métodos propostos.

Os experimentos mostraram que o uso de informações sobre a atenção visual humana é potencialmente vantajoso em SRI. Em todos os experimentos, os métodos CFAS, CBAS obtiveram os melhores resultados quando comparados com os métodos tradicionais da literatura. O próximo capítulo apresenta as principais contribuições deste trabalho, discute as limitações dos métodos propostos e os trabalhos futuros a serem realizados.

## CAPÍTULO 7

# Conclusões e Trabalhos Futuros

Este trabalho apresentou duas novas abordagens para o desenvolvimento de Sistemas de Recomendação de Imagens usando a atenção visual humana, como *feedback implícito*. O principal objetivo da proposta apresentada é aumentar e melhorar a representação dos itens permitindo bons valores de acurácia na predição de avaliações desconhecidas.

As duas novas abordagens propostas neste trabalho são: uma baseada em Filtragem Colaborativa e outra baseada em Conteúdo. O primeiro método baseado em FC é o CFAS que combina qualquer função de similaridade baseada em avaliações usando a matriz de utilidade e a similaridade atentiva. No entanto, esta alternativa não é escalável. Uma alternativa escalável para grandes quantidades de usuários, itens e avaliações é o uso do método CFAS (SVD) que combina a similaridade entre fatores latentes e a similaridade atentiva. O método baseado em Conteúdo, denominado CBAS, combina as similaridades textual, visual e atentiva. A grande vantagem deste método é a possibilidade de ponderar cada similaridade de acordo com a disponibilidade do conteúdo dos itens, evitando parâmetros de ponderação.

Os experimentos foram conduzidos de acordo com o processo de predição de avaliação e recomendação. Os resultados mostraram que os métodos propostos alcançaram os melhores resultados quando comparados com os métodos clássicos da literatura. A adição da atenção visual na representação de itens mostrou ser interessante para melhorar a eficiência do método baseado em vizinhança item-item. Também foram realizados experimentos que analisam a relação dos rótulos semânticos e número de fixações de olhares com o desempenho dos métodos propostos. Foi verificado que é muito importante que o item seja bastante observado para uma boa caracterização atentiva. Acredita-se que depois de certa quantidade de observações sobre o item, não existe a necessidade de coletar novas fixações de olhares sobre este item. Outro ponto importante é que o usuário não precisa do rastreador de olhar para receber recomendações, mas o SRI proposto precisa de avaliações de alguns usuários com o rastreador de olhar para poder compor a representação atentiva dos itens.

## 7.1 Principais Contribuições

Este trabalho apresentou duas novas abordagens para sistemas de recomendação de imagens. As principais diferenças dos métodos propostos em relação aos métodos da literatura são:

- ❑ uma nova maneira de descrever o conteúdo da imagem em partes semânticas atentivas através do comportamento dos usuários ao olhar para a imagem;
- ❑ Uma nova medida de similaridade, denominada similaridade atentiva, que calcula a semelhança entre duas imagens de acordo com a atenção visual;
- ❑ Uma nova maneira de combinar uma medida de similaridade baseada em avaliações com a medida de similaridade atentiva em um método FC, denominado CFAS. A similaridade baseada em avaliações pode ser calculada tanto pela matriz de utilidade quanto por fatores latentes;
- ❑ Uma nova maneira de combinar a similaridade entre atributos textuais, visuais e atenção visual em um método BC, denominado CBAS.

## 7.2 Limitações

Algumas das limitações deste trabalho são descritas a seguir.

- ❑ **Parametrização:** Além dos parâmetros convencionais dos métodos baseados em vizinhança, devemos configurar o parâmetro  $\sigma$  do cálculo da similaridade atentiva para definir a proporção entre a diferença de fixações e a diferença dos movimentos dos olhos. Outro parâmetro é o de encolhimento  $\lambda_{as}$  da similaridade atentiva. Para o método CFAS, temos o parâmetro  $\beta$  para ponderar as similaridades atentiva e baseada em avaliações. Em geral, os valores dos parâmetros podem variar de acordo com o domínio da aplicação e a base de dados. Encontrar os melhores valores de parâmetros não é uma tarefa fácil e depende fortemente dos experimentos.
- ❑ **Segmentação:** A metodologia apresentada neste trabalho é baseada na segmentação das imagens de acordo com um conjunto de rótulos semânticos. Dependendo do domínio da aplicação, esta tarefa se torna muito desafiadora e complexa de ser realizada. Outro ponto importante e difícil de mensurar é a definição do conjunto de rótulos semânticos que representam as imagens da base de dados. Uma boa segmentação é fundamental para a acurácia da similaridade atentiva.
- ❑ **Escalabilidade:** Na metodologia proposta neste trabalho, a medida que novos dados de atenção visual são coletados, a similaridade atentiva deve ser recalculada. Com o aumento do número de itens, o tempo de processamento também aumenta

exponencialmente. Uma alternativa para grandes bases de dados é o agrupamento de itens de acordo com sua atentividade, e o encerramento de coleta de dados visuais de um determinado item quando o mesmo já atingiu certa quantidade de fixações de olhar e/ou tempo de fixações.

- **Rastreador de Olhar:** Os métodos propostos nesta tese requerem um dispositivo de rastreamento do olhar que atualmente é extremamente caro. Entretanto, espera-se que em poucos anos, celulares, notebooks, TVs terão dispositivos de rastreamento do olhar embarcados favorecendo a utilização dos métodos propostos. Outra solução a curto prazo e de baixo custo é a utilização de modelos computacionais para a predição da atenção visual.

### 7.3 Trabalhos Futuros

Durante o desenvolvimento desta pesquisa, algumas questões foram levantadas para melhorar o desempenho e validação da metodologia proposta. Entretanto, estas questões não foram investigadas no escopo desta pesquisa, constituindo assim, assuntos para trabalhos futuros. A seguir são listadas as questões levantadas que proporcionarão trabalhos futuros:

- Abordagem FC usuário-usuário: os métodos propostos nesta tese são baseados na similaridade entre itens. Um aspecto que poderá ser analisado é a similaridade entre usuários de acordo com a percepção visual, podendo assim criar novos métodos FC usuário-usuário baseado em atenção visual. Este trabalho futuro é interessante para tentar minimizar o problema de partida a frio de usuário.
- Abordagem híbrida: os SR híbridos combinam métodos FC e BC tentando minimizar o problema de cada método, por exemplo, minimizar o problema de partida a frio da abordagem FC e minimizar o problema de superespecialização da abordagem por conteúdo. Uma questão levantada nesta pesquisa é: se ao combinar métodos tradicionais da literatura com as abordagens baseadas em atenção visual propostas nesta tese consegue-se melhorar a acurácia na predição? A ideia principal da nova abordagem híbrida é utilizar as informações disponíveis dos itens (dados de atenção visual, avaliações, conteúdo textual e visual) para que o SR possa tomar a decisão de qual método utilizar para cada predição de avaliação.
- Mapas de Saliência: Para eliminar o problema do alto custo do Rastreador de Olhar, uma solução é a utilização de modelos computacionais para a predição da atenção visual. As abordagens propostas nesta tese podem se adequar obtendo a atenção visual das imagens através de mapas de saliência e não por meio das fixações de olhares.

## 7.4 Contribuições em Produção Bibliográfica

No que tange a produção científica relacionada com esta tese, foram publicados artigos em eventos internacionais e nacionais, listados a seguir:

1. MELO, E. V.; AMO, S. ; GULIATO, D. Cross-domain image matching improved by visual attention. In: International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG), 2014, Plzen, Czech Republic. Journal of WSCG, v. 22. n. 2, p. 65-72, 2014. (Ciência da Computação: Qualis B1)
2. MELO, E. V.; NOGUEIRA, E. A.; GULIATO, D. Content-based filtering enhanced by human visual attention applied to clothing recommendation. In: IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), 2015, Vietri Sul Mare, Italy. p. 644-651. (Ciência da Computação: Qualis A2)
3. NOGUEIRA, E. A. ; MELO, E. V. ; FARIA, E. R. ; GULIATO, D. IKB-MS: A Collaborative Filtering Approach Associated with Human Visual Attention for Clothing Recommendation. In: ACM 21st Brazilian Symposium on Multimedia and the Web (WebMedia), 2015, Manaus, Brazil. p. 149-156. (Ciência da Computação: Qualis B3)

O artigo listado abaixo foi submetido a um periódico e está em avaliação:

1. MELO, E. V. ; GULIATO, D. A Collaborative Filtering Approach for Image Recommendation through use of Eye Gaze Tracking. Knowledge-Based Systems. (Ciência da Computação: Qualis A2)

---

## Referências

- ACHANTA, R. et al. Frequency-tuned salient region detection. In: IEEE. **Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on.** [S.l.], 2009. p. 1597–1604.
- 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.
- AHARON, M. et al. Excuseme: Asking users to help in item cold-start recommendations. In: ACM. **Proceedings of the 9th ACM Conference on Recommender Systems.** [S.l.], 2015. p. 83–90.
- ALBANESE, M. et al. A multimedia recommender system. **ACM Transactions on Internet Technology (TOIT)**, ACM, v. 13, n. 1, p. 3, 2013.
- ALLEY, S. et al. Do personally tailored videos in a web-based physical activity intervention lead to higher attention and recall?—an eye-tracking study. **Frontiers in public health**, Frontiers Media SA, v. 2, 2014.
- AMO, S. D. et al. **Mining contextual preference rules for building user profiles.** [S.l.]: Springer, 2012.
- ATHANASIADIS, T. et al. Semantic image segmentation and object labeling. **Circuits and Systems for Video Technology, IEEE Transactions on**, IEEE, v. 17, n. 3, p. 298–312, 2007.
- BARGAH, P.; MISHRA, N. Solving sparsity problem in movie based recommendation system. **Advances in Image and Video Processing**, v. 4, n. 2, p. 71, 2016.
- BARJASTEH, I. et al. Cold-start item and user recommendation with decoupled completion and transduction. In: ACM. **Proceedings of the 9th ACM Conference on Recommender Systems.** [S.l.], 2015. p. 91–98.
- BARRILERO, M. et al. In-network content based image recommendation system for content-aware networks. In: IEEE. **Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on.** [S.l.], 2011. p. 115–120.
- BELIAKOV, G.; CALVO, T.; JAMES, S. Aggregation of preferences in recommender systems. In: **Recommender Systems Handbook.** [S.l.]: Springer, 2011. p. 705–734.

- BELL, R. M.; KOREN, Y. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In: IEEE. **Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on.** [S.l.], 2007. p. 43–52.
- BENNETT, J.; LANNING, S. The netflix prize. In: **Proceedings of KDD cup and workshop.** [S.l.: s.n.], 2007. v. 2007, p. 35.
- BOBADILLA, J. et al. A collaborative filtering approach to mitigate the new user cold start problem. **Knowledge-Based Systems**, Elsevier, v. 26, p. 225–238, 2012.
- BODHIT, A.; AMIN, K. Possible solutions of new user or item cold-start problem. **International Journal of Mathematics**, v. 1, n. 3, 2013.
- BORJI, A.; ITTI, L. State-of-the-art in visual attention modeling. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 35, n. 1, p. 185–207, 2013.
- BOSMAN, H. H.; PETKOV, N.; JONKMAN, M. F. Comparison of color representations for content-based image retrieval in dermatology. **Skin Research and Technology**, Wiley Online Library, v. 16, n. 1, p. 109–113, 2010.
- BOUTEMEDJET, S.; ZIOU, D. A graphical model for context-aware visual content recommendation. **Multimedia, IEEE Transactions on**, IEEE, v. 10, n. 1, p. 52–62, 2008.
- 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.
- BRUCE, N.; TSOTSOS, J. Saliency based on information maximization. In: **Advances in neural information processing systems.** [S.l.: s.n.], 2005. p. 155–162.
- BRUSILOVSKY, P.; KOBSA, A.; NEJDL, W. **The adaptive web: methods and strategies of web personalization.** [S.l.]: Springer Science & Business Media, 2007.
- BURKE, R. Hybrid recommender systems: Survey and experiments. **User modeling and user-adapted interaction**, Springer, v. 12, n. 4, p. 331–370, 2002.
- \_\_\_\_\_. Hybrid web recommender systems. In: **The adaptive web.** [S.l.]: Springer, 2007. p. 377–408.
- CALVO, T. et al. Aggregation operators: properties, classes and construction methods. In: **Aggregation operators.** [S.l.]: Springer, 2002. p. 3–104.
- CARENINI, G. et al. Highlighting interventions and user differences: Informing adaptive information visualization support. In: ACM. **Proceedings of the 32nd annual ACM conference on Human factors in computing systems.** [S.l.], 2014. p. 1835–1844.
- CHAO, X. et al. A framework for robust feature selection for real-time fashion style recommendation. In: ACM. **Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics.** [S.l.], 2009. p. 35–42.

- CHOI, K.; SUH, Y. A new similarity function for selecting neighbors for each target item in collaborative filtering. **Knowledge-Based Systems**, Elsevier, v. 37, p. 146–153, 2013.
- CIAN, L.; KRISHNA, A.; ELDER, R. S. This logo moves me: Dynamic imagery from static images. **Journal of Marketing Research**, American Marketing Association, v. 51, n. 2, p. 184–197, 2014.
- COHEN, W. W. Fast effective rule induction. In: **Proceedings of the twelfth international conference on machine learning**. [S.l.: s.n.], 1995. p. 115–123.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: **IEEE Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on**. [S.l.], 2005. v. 1, p. 886–893.
- DARRELL, T. et al. Exploring vision-based interfaces: How to use your head in dual pointing tasks. 2002.
- DAVIDSE, N. J. et al. A twin-case study of developmental number sense impairment. **Cognitive neuropsychology**, Taylor & Francis, v. 31, n. 3, p. 221–236, 2014.
- DEBELJAK, M.; OCEPEK, J.; ZUPAN, A. **Eye Controlled Human Computer Interaction for Severely Motor Disabled Children**. [S.l.]: Springer, 2012.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **The Journal of Machine Learning Research**, JMLR.org, v. 7, p. 1–30, 2006.
- DESHPANDE, M.; KARYPIS, G. Item-based top-n recommendation algorithms. **ACM Transactions on Information Systems (TOIS)**, ACM, v. 22, n. 1, p. 143–177, 2004.
- DONG, R. et al. Combining similarity and sentiment in opinion mining for product recommendation. **Journal of Intelligent Information Systems**, Springer, p. 1–28, 2015.
- DOUZE, M. et al. Evaluation of gist descriptors for web-scale image search. In: **ACM Proceedings of the ACM International Conference on Image and Video Retrieval**. [S.l.], 2009. p. 19.
- DUCHOWSKI, A. **Eye tracking methodology: Theory and practice**. [S.l.]: Springer Science & Business Media, 2007.
- FELÍCIO, C. Z. et al. Visual perception similarities to improve the quality of user cold start recommendations. In: **Advances in Artificial Intelligence: 29th Canadian Conference on Artificial Intelligence, Victoria, BC, Canada, May**. [S.l.: s.n.], 2016. p. 96–101.
- FELZENSZWALB, P. F. et al. Object detection with discriminatively trained part-based models. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 32, n. 9, p. 1627–1645, 2010.
- FU, J. et al. Efficient clothing retrieval with semantic-preserving visual phrases. In: **Computer Vision–ACCV 2012**. [S.l.]: Springer, 2013. p. 420–431.

- FUJII, K. et al. Gaze contingent cartesian control of a robotic arm for laparoscopic surgery. In: IEEE. **Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on**. [S.l.], 2013. p. 3582–3589.
- 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.
- GARCIA-DIAZ, A. et al. Decorrelation and distinctiveness provide with human-like saliency. In: SPRINGER. **Advanced concepts for intelligent vision systems**. [S.l.], 2009. p. 343–354.
- GASPARINI, F.; CORCHS, S.; SCHETTINI, R. Low-quality image enhancement using visual attention. **Optical Engineering**, International Society for Optics and Photonics, v. 46, n. 4, p. 040502–040502, 2007.
- GE, Y. et al. Collaborative filtering with collective training. In: ACM. **Proceedings of the fifth ACM conference on Recommender systems**. [S.l.], 2011. p. 281–284.
- GEMULLA, R. et al. Large-scale matrix factorization with distributed stochastic gradient descent. In: ACM. **Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2011. p. 69–77.
- GORODNICHY, D. Perceptual cursor-a solution to the broken loop problem in vision-based hands-free computer control devices. 2006.
- GRAUS, M. P.; WILLEMSSEN, M. C. Improving the user experience during cold start through choice-based preference elicitation. In: ACM. **Proceedings of the 9th ACM Conference on Recommender Systems**. [S.l.], 2015. p. 273–276.
- 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.
- HATA, M.; TOYOURA, M.; MAO, X. Automatic generation of accentuated pencil drawing with saliency map and lic. **The Visual Computer**, Springer, v. 28, n. 6-8, p. 657–668, 2012.
- HERLOCKER, J.; KONSTAN, J. A.; RIEDL, J. An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. **Information Retrieval**, Springer, v. 5, n. 4, p. 287–310, 2002.
- HERLOCKER, J. L. et al. An algorithmic framework for performing collaborative filtering. In: ACM. **Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval**. [S.l.], 1999. p. 230–237.
- \_\_\_\_\_. Evaluating collaborative filtering recommender systems. **ACM Transactions on Information Systems (TOIS)**, ACM, v. 22, n. 1, p. 5–53, 2004.
- HUANG, B.; LO, A. H.; SHI, B. E. Integrating eeg information improves performance of gaze based cursor control. In: IEEE. **Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on**. [S.l.], 2013. p. 415–418.

- 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.
- IWATA, T.; WANATABE, S.; SAWADA, H. Fashion coordinates recommender system using photographs from fashion magazines. In: **IJCAI Proceedings-International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 2011. v. 22, n. 3, p. 2262.
- JACOB, R.; KARN, K. S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. **Mind**, Citeseer, v. 2, n. 3, p. 4, 2003.
- JAHRER, M.; TÖSCHER, A. Collaborative filtering ensemble. In: **KDD Cup**. [S.l.: s.n.], 2012. p. 61–74.
- JAIN, G. **Emotion Recognition from Eye Region Signals using Local Binary Patterns**. Tese (Doutorado) — University of Toronto, 2011.
- 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.
- JUNG, J. et al. Evolutionary programming based recommendation system for online shopping. In: **IEEE. Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific**. [S.l.], 2013. p. 1–4.
- KAKUMANU, P.; MAKROGIANNIS, S.; BOURBAKIS, N. A survey of skin-color modeling and detection methods. **Pattern Recognition**, Elsevier, v. 40, n. 3, p. 1106–1122, 2007.
- KAUR, S.; AGGARWAL, D. Image content based retrieval system using cosine similarity for skin disease images. **Advances in Computer Science: an International Journal**, v. 2, n. 4, p. 89–95, 2013.
- KELLY, D.; BELKIN, N. J. Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback. In: **ACM. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval**. [S.l.], 2001. p. 408–409.
- \_\_\_\_\_. Display time as implicit feedback: understanding task effects. In: **ACM. Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval**. [S.l.], 2004. p. 377–384.
- KIM, J. K.; KIM, H. K.; CHO, Y. H. A user-oriented contents recommendation system in peer-to-peer architecture. **Expert Systems with Applications**, Elsevier, v. 34, n. 1, p. 300–312, 2008.
- KIM, J. W. et al. Application of decision-tree induction techniques to personalized advertisements on internet storefronts. **International Journal of Electronic Commerce**, Taylor & Francis, v. 5, n. 3, p. 45–62, 2001.
- KIM, K.-N.; RAMAKRISHNA, R. Vision-based eye-gaze tracking for human computer interface. In: **IEEE. Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on**. [S.l.], 1999. v. 2, p. 324–329.

KIMURA, P. A. et al. Evaluating retrieval effectiveness of descriptors for searching in large image databases. **Journal of Information and Data Management**, v. 2, n. 3, p. 305, 2011.

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.

KOREN, Y.; BELL, R.; VOLINSKY, C. Matrix factorization techniques for recommender systems. **Computer**, Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, 17 th Fl New York NY 10016-5997 United States, v. 42, n. 8, p. 30–37, 2009.

KUBAT, M.; TAPIA, M. et al. Time spent on a web page is sufficient to infer a user's interest. In: ACTA PRESS. **Iasted European Conference on Proceedings of the Iasted European Conference: internet and multimedia systems and applications**. [S.l.], 2007. p. 41–46.

LASKARIS, N. et al. Fuzzy description of skin lesions. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **SPIE Medical Imaging**. [S.l.], 2010. p. 762717–762717.

LENZNER, T.; KACZMIREK, L.; GALESIC, M. Left feels right a usability study on the position of answer boxes in web surveys. **Social Science Computer Review**, SAGE Publications, v. 32, n. 6, p. 743–764, 2014.

LEWIS, D. D.; RINGUETTE, M. A comparison of two learning algorithms for text categorization. In: **Third annual symposium on document analysis and information retrieval**. [S.l.: s.n.], 1994. v. 33, p. 81–93.

LEWIS, D. D. et al. Training algorithms for linear text classifiers. In: ACM. **Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval**. [S.l.], 1996. p. 298–306.

LI, Q.; KIM, B. M. An approach for combining content-based and collaborative filters. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the sixth international workshop on Information retrieval with Asian languages-Volume 11**. [S.l.], 2003. p. 17–24.

LI, S.; ZHAO, M. Image inpainting with salient structure completion and texture propagation. **Pattern Recognition Letters**, Elsevier, v. 32, n. 9, p. 1256–1266, 2011.

LIANG, Z. et al. Content-based image retrieval using a combination of visual features and eye tracking data. In: ACM. **Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications**. [S.l.], 2010. p. 41–44.

- LIKA, B.; KOLOMVATSOS, K.; HADJIEFTHYMIADES, S. Facing the cold start problem in recommender systems. **Expert Systems with Applications**, Elsevier, v. 41, n. 4, p. 2065–2073, 2014.
- LIN, Y.-P. et al. Webcam mouse using face and eye tracking in various illumination environments. In: IEEE. **Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the**. [S.l.], 2006. p. 3738–3741.
- LIU, B.-b. et al. Pornographic images detection based on cbir and skin analysis. In: IEEE. **Fourth international conference on semantics, knowledge and grid**. [S.l.], 2008. p. 487–488.
- LIU, C. et al. Distributed nonnegative matrix factorization for web-scale dyadic data analysis on mapreduce. In: ACM. **Proceedings of the 19th international conference on world wide web**. [S.l.], 2010. p. 681–690.
- LIU, N. N. et al. Unifying explicit and implicit feedback for collaborative filtering. In: ACM. **Proceedings of the 19th ACM international conference on Information and knowledge management**. [S.l.], 2010. p. 1445–1448.
- LIU, S. et al. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In: IEEE. **Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on**. [S.l.], 2012. p. 3330–3337.
- LOPS, P.; GEMMIS, M. D.; SEMERARO, G. Content-based recommender systems: State of the art and trends. In: **Recommender systems handbook**. [S.l.]: Springer, 2011. p. 73–105.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, Springer, v. 60, n. 2, p. 91–110, 2004.
- LU, G. **Multimedia database management systems**. [S.l.]: Artech House Boston, 1999.
- LU, J. et al. Recommender system application developments: a survey. **Decision Support Systems**, Elsevier, v. 74, p. 12–32, 2015.
- MAESSCHALCK, R. D.; JOUAN-RIMBAUD, D.; MASSART, D. L. The mahalanobis distance. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 50, n. 1, p. 1–18, 2000.
- MANJUNATH, B. S.; SALEMPIER, P.; SIKORA, T. **Introduction to MPEG-7: multimedia content description interface**. [S.l.]: John Wiley & Sons, 2002.
- MCCALLUM, A.; NIGAM, K. et al. A comparison of event models for naive bayes text classification. In: CITESEER. **AAAI-98 workshop on learning for text categorization**. [S.l.], 1998. v. 752, p. 41–48.
- MELO, E. V.; AMO, S.; GULIATO, D. Cross-domain image matching improved by visual attention. **Journal of WSCG**, v. 22, n. 2, p. 65–72, 2014.

- MELO, E. V.; NOGUEIRA, E. A.; GULIATO, D. Content-based filtering enhanced by human visual attention applied to clothing recommendation. In: **IEEE. Tools with Artificial Intelligence (ICTAI), 2015 IEEE 27th International Conference on**. [S.l.], 2015. p. 644–651.
- MELVILLE, P.; SINDHWANI, V. Recommender systems. In: **Encyclopedia of machine learning**. [S.l.]: Springer, 2010. p. 829–838.
- MENON, A. K.; ELKAN, C. A log-linear model with latent features for dyadic prediction. In: **IEEE. Data Mining (ICDM), 2010 IEEE 10th International Conference on**. [S.l.], 2010. p. 364–373.
- MESNIL, G. et al. Learning semantic representations of objects and their parts. **Machine learning**, Springer, v. 94, n. 2, p. 281–301, 2014.
- MOBASHER, B.; JIN, X.; ZHOU, Y. Semantically enhanced collaborative filtering on the web. In: **Web Mining: From Web to Semantic Web**. [S.l.]: Springer, 2004. p. 57–76.
- MOONEY, R. J.; ROY, L. Content-based book recommending using learning for text categorization. In: **ACM. Proceedings of the fifth ACM conference on Digital libraries**. [S.l.], 2000. p. 195–204.
- MURRAY, N. et al. Saliency estimation using a non-parametric low-level vision model. In: **IEEE. Computer vision and pattern recognition (cvpr), 2011 ieee conference on**. [S.l.], 2011. p. 433–440.
- NIU, J. et al. Fuir: Fusing user and item information to deal with data sparsity by using side information in recommendation systems. **Journal of Network and Computer Applications**, Elsevier, v. 70, p. 41–50, 2016.
- NÚÑEZ-VALDÉZ, E. R. et al. Implicit feedback techniques on recommender systems applied to electronic books. **Computers in Human Behavior**, Elsevier, v. 28, n. 4, p. 1186–1193, 2012.
- OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 24, n. 7, p. 971–987, 2002.
- OLÁH, K. et al. Social category formation is induced by cues of sharing knowledge in young children. **PloS one**, Public Library of Science, v. 9, n. 7, p. e101680, 2014.
- OLIVA, A.; TORRALBA, A. Building the gist of a scene: The role of global image features in recognition. **Progress in brain research**, Elsevier, v. 155, p. 23–36, 2006.
- OSTUNI, V. C. et al. Top-n recommendations from implicit feedback leveraging linked open data. In: **ACM. Proceedings of the 7th ACM conference on Recommender systems**. [S.l.], 2013. p. 85–92.
- PATEREK, A. Improving regularized singular value decomposition for collaborative filtering. In: **Proceedings of KDD cup and workshop**. [S.l.: s.n.], 2007. v. 2007, p. 5–8.

- PAZZANI, M.; BILLSUS, D. Learning and revising user profiles: The identification of interesting web sites. **Machine learning**, Springer, v. 27, n. 3, p. 313–331, 1997.
- PAZZANI, M. J.; BILLSUS, D. Content-based recommendation systems. In: **The adaptive web**. [S.l.]: Springer, 2007. p. 325–341.
- PIETIKÄINEN, M. Image analysis with local binary patterns. In: **Image Analysis**. [S.l.]: Springer, 2005. p. 115–118.
- POPA, L. et al. Reading beyond the glance: eye tracking in neurosciences. **Neurological Sciences**, Springer, p. 1–6, 2015.
- POTTER, G. Putting the collaborator back into collaborative filtering. In: **ACM. Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition**. [S.l.], 2008. p. 3.
- QIN, S.; MENEZES, R.; SILAGHI, M. A recommender system for youtube based on its network of reviewers. In: **IEEE. Social Computing (SocialCom), 2010 IEEE Second International Conference on**. [S.l.], 2010. p. 323–328.
- RAYNER, K. Eye movements in reading and information processing: 20 years of research. **Psychological bulletin**, American Psychological Association, v. 124, n. 3, p. 372, 1998.
- 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.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. **Introduction to recommender systems handbook**. [S.l.]: Springer, 2011.
- RUDDARRAJU, R. et al. Perceptual user interfaces using vision-based eye tracking. In: **ACM. Proceedings of the 5th international conference on Multimodal interfaces**. [S.l.], 2003. p. 227–233.
- SANTOS, A. Menk dos. A hybrid recommendation system based on human curiosity. In: **ACM. Proceedings of the 9th ACM Conference on Recommender Systems**. [S.l.], 2015. p. 367–370.
- SARWAR, B. et al. Item-based collaborative filtering recommendation algorithms. In: **ACM. Proceedings of the 10th international conference on World Wide Web**. [S.l.], 2001. p. 285–295.
- SAVESKI, M.; MANTRACH, A. Item cold-start recommendations: learning local collective embeddings. In: **ACM. Proceedings of the 8th ACM Conference on Recommender Systems**. [S.l.], 2014. p. 89–96.
- SHANI, G.; GUNAWARDANA, A. Tutorial on application-oriented evaluation of recommendation systems. **AI Communications**, IOS Press, 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.

- SHARMA, M. et al. Feature-based factorized bilinear similarity model for cold-start top-n item recommendation. In: SIAM. **Proceedings of the 15th SIAM International Conference on Data Mining, SDM**. [S.l.], 2015. v. 15, p. 190–198.
- SHRIVASTAVA, A. et al. Data-driven visual similarity for cross-domain image matching. In: ACM. **ACM Transactions on Graphics (TOG)**. [S.l.], 2011. v. 30, n. 6, p. 154.
- SONG, Z. et al. Predicting occupation via human clothing and contexts. In: IEEE. **Computer Vision (ICCV), 2011 IEEE International Conference on**. [S.l.], 2011. p. 1084–1091.
- STANESCU, A.; NAGAR, S.; CARAGEA, D. A hybrid recommender system: User profiling from keywords and ratings. In: IEEE. **Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on**. [S.l.], 2013. v. 1, p. 73–80.
- STRICKER, M. A.; ORENGO, M. Similarity of color images. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology**. [S.l.], 1995. p. 381–392.
- SU, M.-C.; SU, S.-Y.; CHEN, G.-D. A low-cost vision-based human-computer interface for people with severe disabilities. **Biomedical Engineering: Applications, Basis and Communications**, World Scientific, v. 17, n. 06, p. 284–292, 2005.
- SU, X.; KHOSHGOFTAAR, T. M. A survey of collaborative filtering techniques. **Advances in artificial intelligence**, Hindawi Publishing Corp., v. 2009, p. 4, 2009.
- SWAIN, M. J.; BALLARD, D. H. Color indexing. **International Journal of Computer Vision**, Springer, v. 7, n. 1, p. 11–32, 1991.
- TAKÁCS, G. et al. Major components of the gravity recommendation system. **ACM SIGKDD Explorations Newsletter**, ACM, v. 9, n. 2, p. 80–83, 2007.
- TERRILLON, J.-C. et al. Analysis of a large set of color spaces for skin pixel detection in color images. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Quality Control by Artificial Vision**. [S.l.], 2003. p. 433–446.
- TKALČIČ, M.; BURNIK, U.; KOŠIR, A. Using affective parameters in a content-based recommender system for images. **User Modeling and User-Adapted Interaction**, Springer, v. 20, n. 4, p. 279–311, 2010.
- TOBII. **Tobii Technology**. 2015. [Http://www.tobii.com](http://www.tobii.com). Accessed: 2015-05-27.
- VELAZQUEZ, C. E.; PASCH, K. E. Attention to food and beverage advertisements as measured by eye-tracking technology and the food preferences and choices of youth. **Journal of the Academy of Nutrition and Dietetics**, Elsevier, v. 114, n. 4, p. 578–582, 2014.
- WANG, M. et al. Context-aware music recommendation with serendipity using semantic relations. In: **Semantic Technology**. [S.l.]: Springer, 2014. p. 17–32.
- 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.

- WONGCHOKPRASITTI, C. et al. User model in a box: Cross-system user model transfer for resolving cold start problems. In: **User Modeling, Adaptation and Personalization**. [S.l.]: Springer, 2015. p. 289–301.
- XU, S.; JIANG, H.; LAU, F. Personalized online document, image and video recommendation via commodity eye-tracking. In: ACM. **Proceedings of the 2008 ACM conference on Recommender systems**. [S.l.], 2008. p. 83–90.
- YI, M.-Y.; LEE, O.-J.; JUNG, J. J. Mbti-based collaborative recommendation system: A case study of webtoon contents. In: SPRINGER. **International Conference on Context-Aware Systems and Applications**. [S.l.], 2015. p. 101–110.
- YU, X. et al. Recommendation in heterogeneous information networks with implicit user feedback. In: ACM. **Proceedings of the 7th ACM conference on Recommender systems**. [S.l.], 2013. p. 347–350.
- ZHANG, C.; HUANG, L. Content-based image retrieval using multiple features. **CIT. Journal of Computing and Information Technology**, SRCE-Sveučilišni računski centar, v. 22, n. LISS 2013, p. 1–10, 2014.
- ZHANG, J.; SHEN, L.; FENG, D. D. A personalized image retrieval based on visual perception. **Journal of Electronics (China)**, Springer, v. 25, n. 1, p. 129–133, 2008.
- ZHANG, X.; WANG, H. Study on recommender systems for business-to-business electronic commerce. **Communications of the IIMA**, v. 5, n. 4, p. 8, 2015.
- ZHANG, Y.; CALLAN, J.; MINKA, T. Novelty and redundancy detection in adaptive filtering. In: ACM. **Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval**. [S.l.], 2002. p. 81–88.
- ZHENG, S. et al. Dense semantic image segmentation with objects and attributes. In: IEEE. **Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on**. [S.l.], 2014. p. 3214–3221.
- ZIELINSKI, P. **Opengazer: open-source gaze tracker for ordinary webcams (software)**, Samsung and The Gatsby Charitable Foundation. 2015. <Http://www.inference.phy.cam.ac.uk/opengazer/>. Accessed: 2015-05-27.



## Apêndices



# APÊNDICE A

## Rótulos com Significados Semânticos para Pintura

Tabela 21 – Conjunto H com os 41 rótulos de possíveis significados semânticos para pintura.

<b>Id</b>	<b>Rótulo</b>	<b>Id</b>	<b>Rótulo</b>
01	Abstração	22	Animal_Ave
02	Religiosa_Pessoa	23	Animal_Gado
03	Religiosa_Anjo	24	Animal_Cavalo
04	Religiosa_Nuvem	25	Animal_Peixe
05	Religiosa_Construção	26	Animal_Cachorro
06	Religiosa_Arvore/Vegetação	27	Animal_Coelho
07	Religiosa_Moveis	28	Animal_Veado
08	Religiosa_Paisagem	29	Animal_Leão
09	Religiosa_Objetos	30	Animal_Búfalo
10	Paisagem_Ceu	31	Animal_Macaco
11	Paisagem_Terra	32	Nudez_Mulher
12	Paisagem_Rua	33	Nudez_Homem
13	Paisagem_Arvore	34	Nudez_Criança
14	Paisagem_Rio/Oceano	35	Retrato_Mulher
15	Paisagem_Montanha	36	Retrato_Homem
16	Paisagem_Construção	37	Retrato_Criança
17	Paisagem_Animal	38	NaturezaMorta_Fruta
18	Paisagem_Veiculo/Barco	39	NaturezaMorta_Objeto
19	Paisagem_Objetos	40	NaturezaMorta_Quitanda/Pão
20	Paisagem_Pessoa	41	NaturezaMorta_Planta
21	Paisagem_Gelo		



## APÊNDICE B

---

# **Sistema de Recomendação de Roupas com os Métodos CFAS e CBAS**

O Sistema de Recomendação de Roupas (SRR) desenvolvido nesta tese é dividido em cinco abas principais: *Parameters*, *Input Data*, *Rating Prediction and TopN Recommendation*, *Cold-Start Problem* e *About*.

A aba *Parameters* é ilustrada na Figura 34. Nesta tela configura-se os principais parâmetros dos métodos CFAS e CBAS e dos métodos comparativos.

A aba *Input Data* é ilustrada na Figura 35. Nesta tela o sistema carrega os dados de entrada (usuários, itens, similaridade visual entre itens, avaliações e dados de atenção visual). Além disso, esta tela mostra as imagens com as fixações, e dados estatísticos sobre os itens e sobre os usuários. Através dessa tela, é possível utilizar filtros para visualizar informações específicas, por exemplo, avaliações somente de um determinado usuário, ou de apenas um item.

A aba *Rating Prediction and TopN Recommendation* é ilustrada na Figura 36. Nesta tela configuram-se quais métodos (FC e BC) vão ser executados e comparados, e a quantidade de *Folds* do *Cross-Validation*. Após a execução dos métodos, a tela exibe o resultado (em termos de RMSE e MAE) do processo de predição de avaliação, o resultado (em termos de precisão, revocação, e AUC) do processo de recomendação, e o resultado estatístico do teste de sinal.

A aba *Cold-Start Problem* é ilustrada na Figura 37. Nesta tela configura-se o cenário de teste para o problema de partida a frio tanto de usuários quanto de itens. O resultado é exibido em termos de RMSE e MAE.

A aba *About* exibe informações sobre os desenvolvedores do software, como mostra a Figura 38.

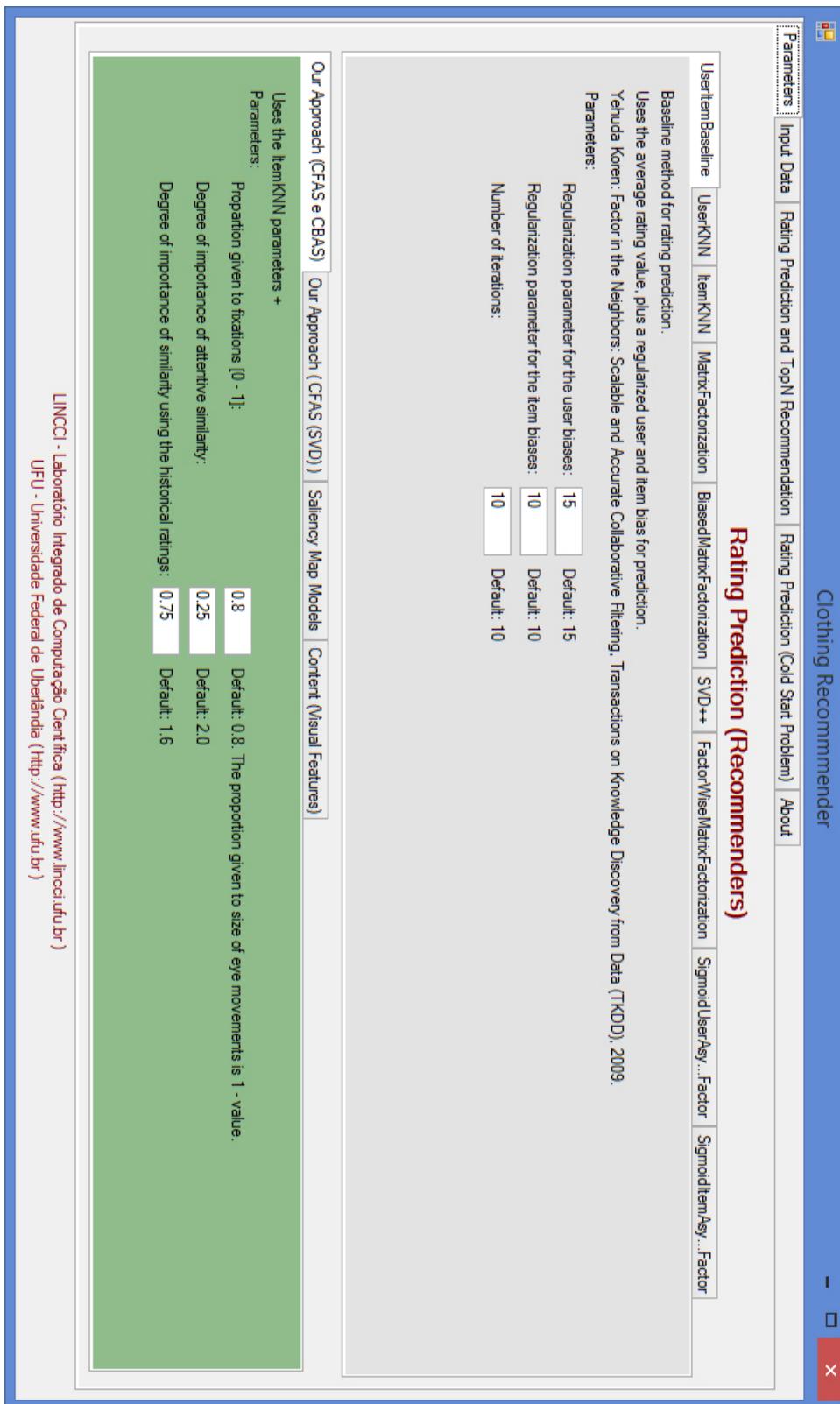


Figura 34 – Sistema de Recomendação de Roupas. Tela de parâmetros.

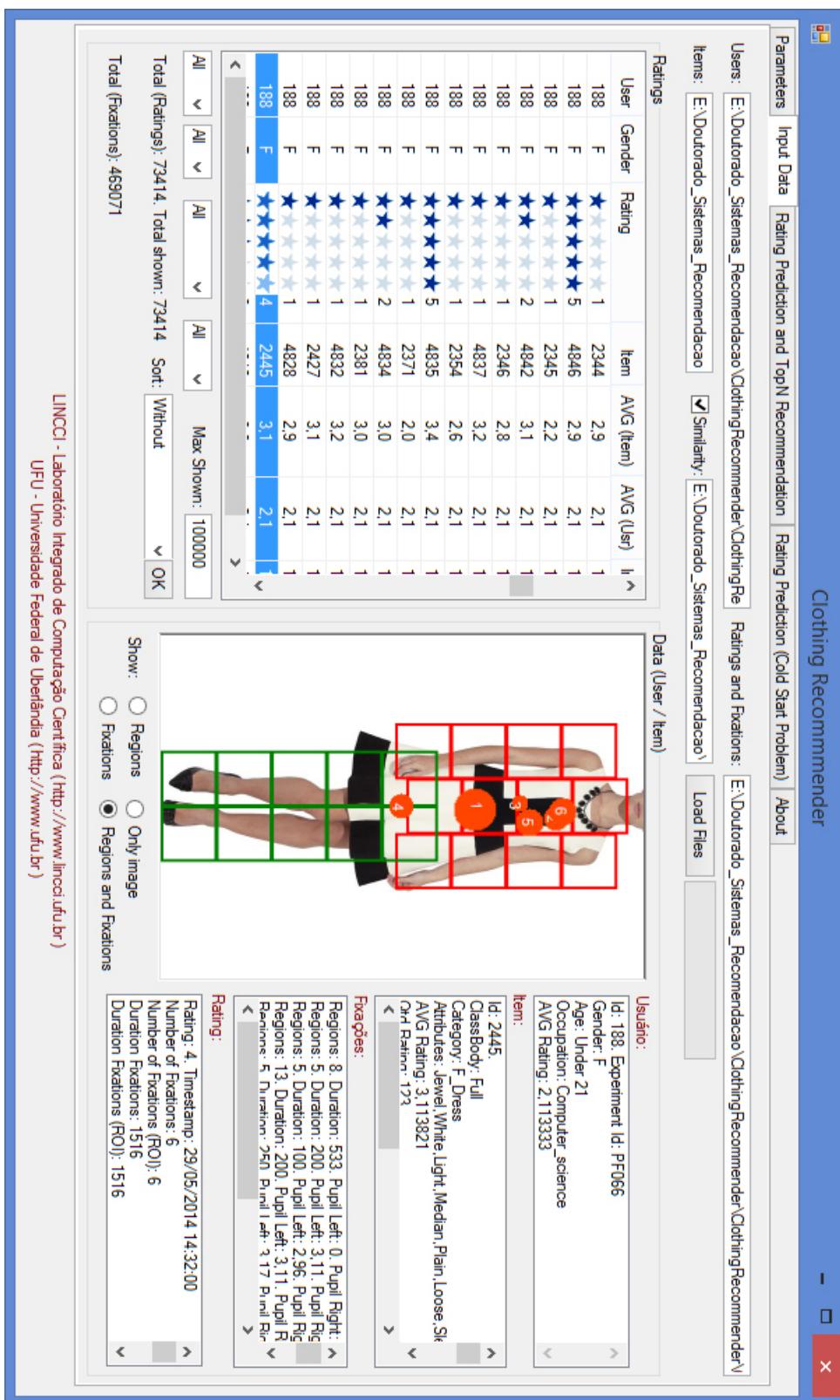


Figura 35 – Sistema de Recomendação de Roupas. Tela de entrada de dados.

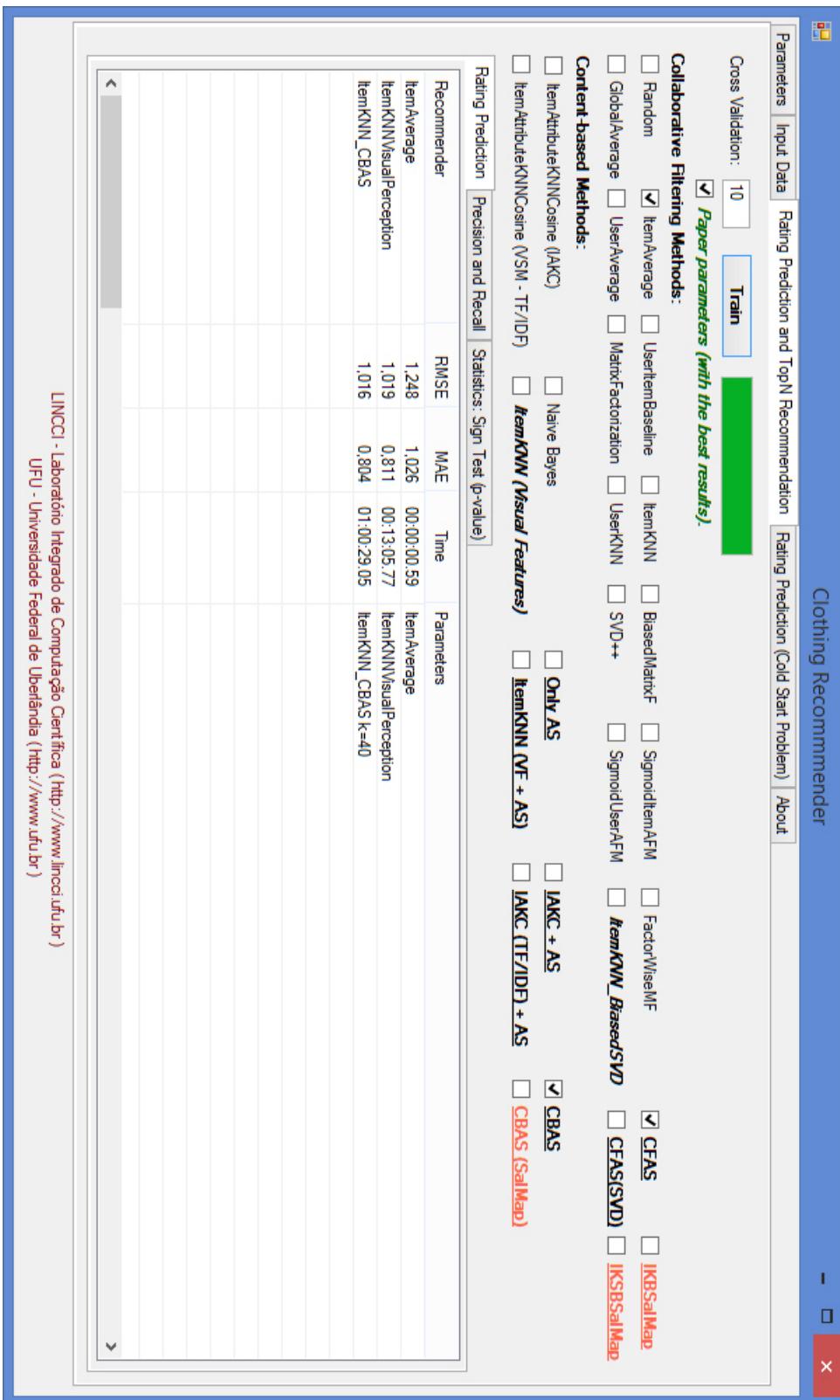


Figura 36 – Sistema de Recomendação de Roupas. Execução dos processos de predição e recomendação.

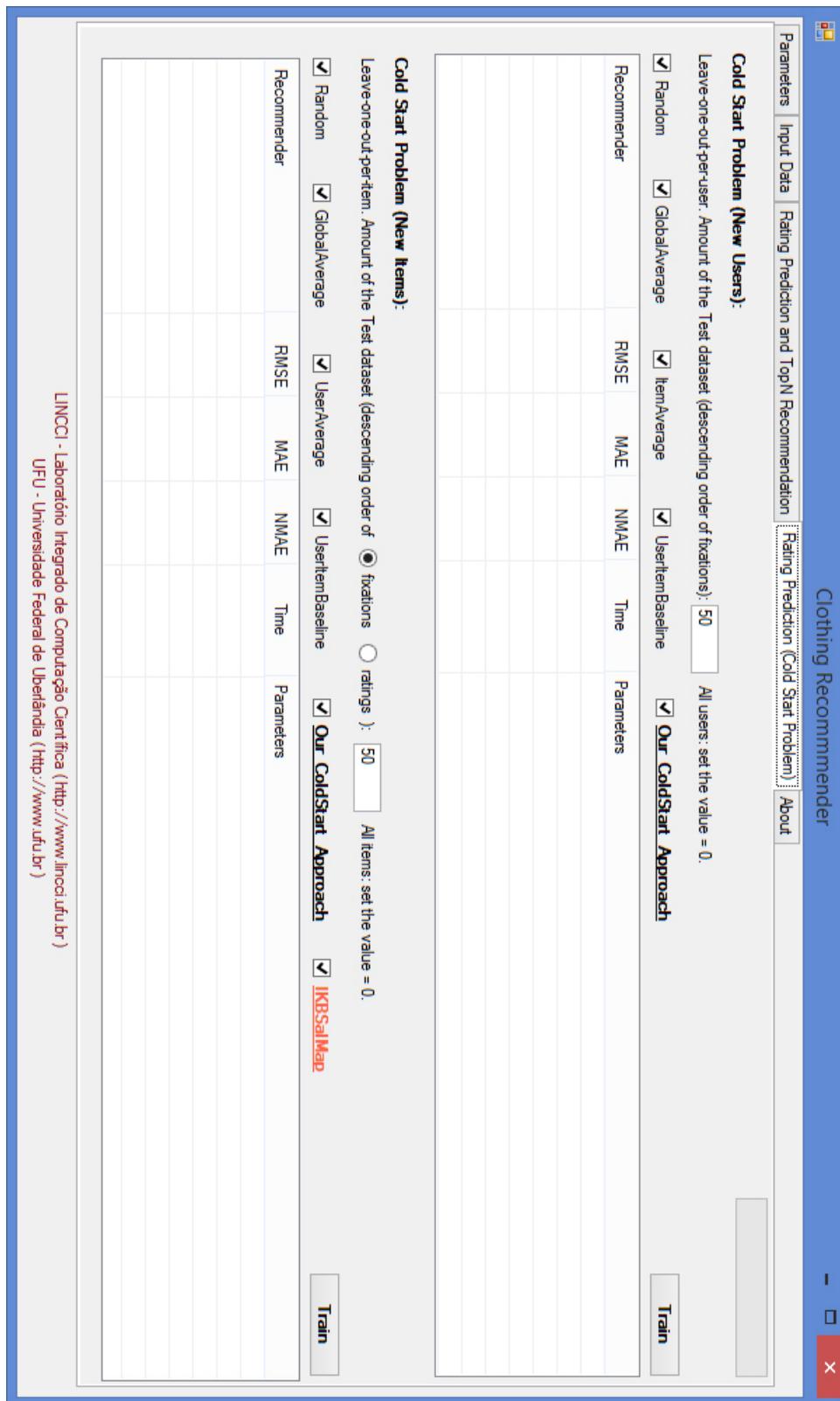


Figura 37 – Sistema de Recomendação de Roupas. Cenário de partida a frio.

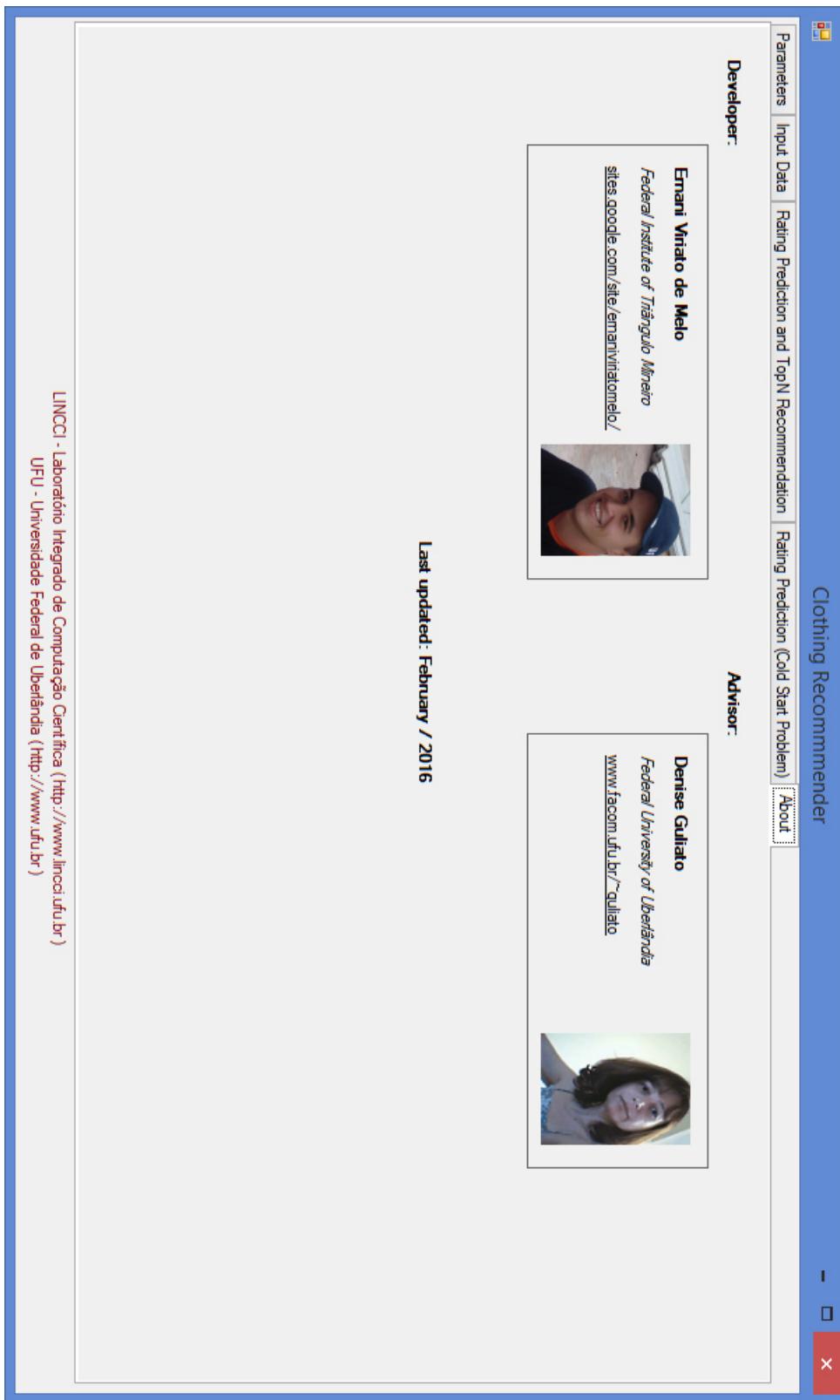


Figura 38 – Sistema de Recomendação de Roupas. Informações sobre os desenvolvedores.

## APÊNDICE C

---

### **Sistema de Recomendação de Pinturas com os Métodos CFAS e CBAS**

O Sistema de Recomendação de Pinturas (SRP) possui a mesma estrutura do SRR descrito no Apêndice B. Apenas a aba *Input Data* é diferente. Nesta tela, exibe-se a pintura segmentada com as fixações, conforme ilustra a Figura 39.

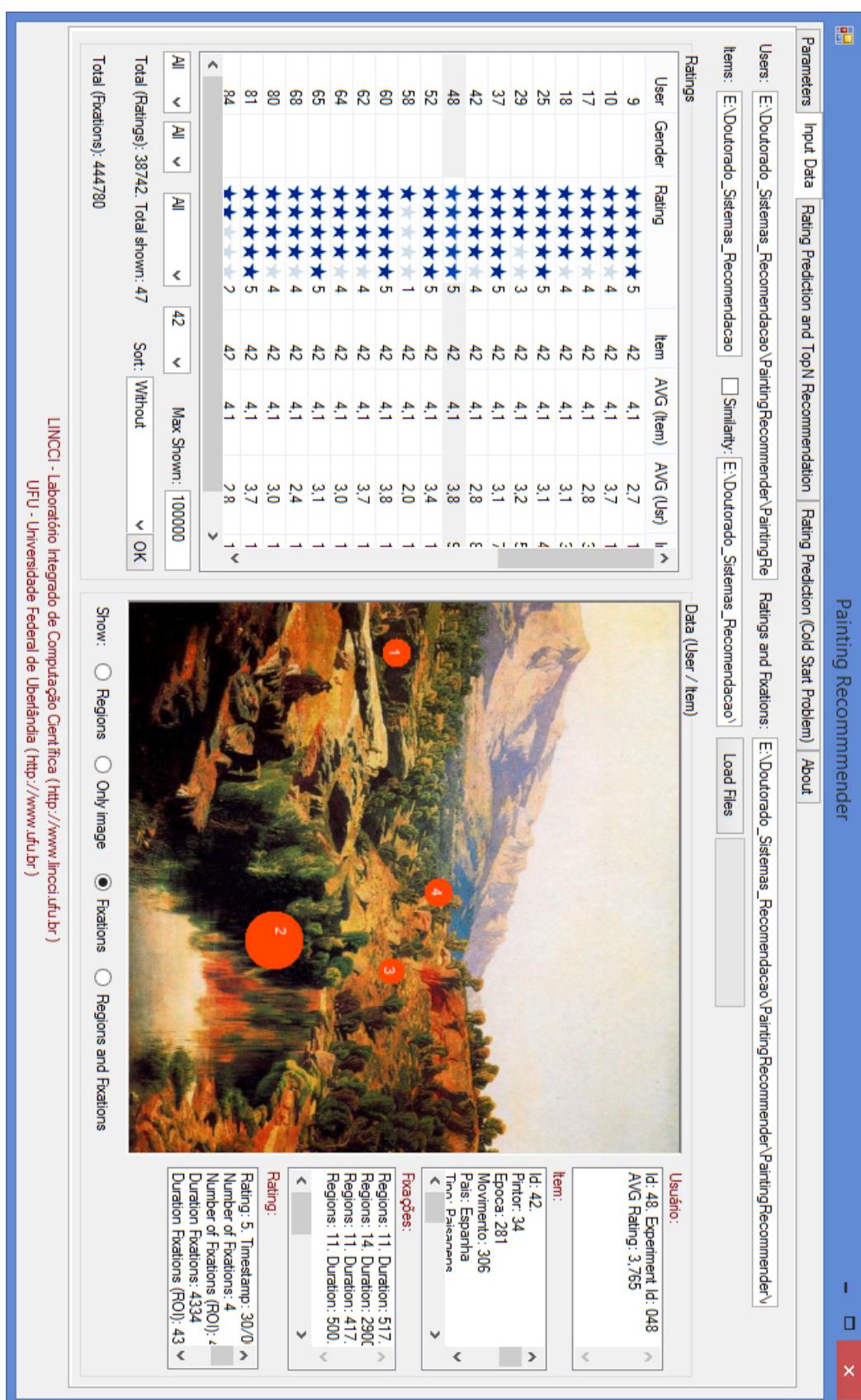


Figura 39 – Sistema de Recomendação de Pinturas. Tela de entrada de dados.