



UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

REALIMENTAÇÃO DE RELEVÂNCIA EM BUSCAS DE IMAGEM USANDO
PROGRAMAÇÃO GENÉTICA

Gregory Oliveira da Silva

Maio de 2016

Manaus - AM



UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

REALIMENTAÇÃO DE RELEVÂNCIA EM BUSCAS DE IMAGEM USANDO
PROGRAMAÇÃO GENÉTICA

Gregory Oliveira da Silva

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Computação - IComp, da Universidade Federal do Amazonas, como parte dos requisitos necessários à obtenção do título de Mestre em Informática.

Orientador: Edleno Silva de Moura

Maio de 2016

Manaus - AM

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

S586r Silva, Gregory Oliveira da
Realimentação de Relevância em Buscas de Imagens Usando
Programação Genética / Gregory Oliveira da Silva. 2016
53 f.: il.; 31 cm.

Orientador: Edleno Silva de Moura
Dissertação (Mestrado em Informática) - Universidade Federal do
Amazonas.

1. Busca de imagem. 2. Realimentação de relevância. 3.
Programação Genética. 4. Comércio eletrônico. I. Moura, Edleno
Silva de II. Universidade Federal do Amazonas III. Título

REALIMENTAÇÃO DE RELEVÂNCIA EM BUSCAS DE IMAGEM USANDO
PROGRAMAÇÃO GENÉTICA

Gregory Oliveira da Silva

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE
PÓS-GRADUAÇÃO DO INSTITUTO DE COMPUTAÇÃO DA UNIVERSIDADE
FEDERAL DO AMAZONAS COMO PARTE DOS REQUISITOS NECESSÁRIOS
PARA A OBTENÇÃO DO GRAU DE MESTRE EM INFORMÁTICA.

Aprovado por:

Prof. Edleno Silva de Moura, D.Sc.

Prof. João Marcos Bastos Cavalcanti, D.Sc.

Prof. Marcos Gonçalves, D.Sc.

MAIO DE 2016

MANAUS, AM – BRASIL

Agradecimentos

Agradeço a minha mãe Fátima e a minha namorada e amiga Emanuelle que me deram suporte emocional durante esta etapa, e aos professores João Marcos e Edleno pela orientação e ensinamentos. Em especial, dedico este trabalho ao meu pai Manoel (in memoriam), que me acompanhou durante toda minha vida e me proporcionou todas as oportunidades que me fizeram chegar onde estou hoje.

*"Existem muitas hipóteses em ciência que estão erradas. Isso é perfeitamente aceitável,
eles são a abertura para achar as que estão certas"*

Carl Sagan

Resumo

Produtos de moda são itens difíceis de ser anotados e descritos por texto, fazendo-se necessário o uso de imagens para a realização de buscas em web sites de *e-commerce*. Tais produtos detém grande apelo visual, ou seja, a apresentação de imagens referentes aos mesmos são fatores que influenciam diretamente a decisão de compra de um cliente. Estes fatos justificam o estudo do uso de CBIR (*Content Based Image Retrieval*) neste contexto, uma área já bastante estudada na comunidade científica, mas que ainda possui diversas lacunas, sendo a principal o problema do *Gap Semântico*. O uso de características extraídas da imagem por um algoritmo ainda não é eficaz o suficiente em associá-la ao seu significado, o que se reflete nos resultados de uma busca, afetando a satisfação do cliente com a loja. Este trabalho busca abordar o problema do *Gap Semântico* através do uso de Programação Genética e *Relevance Feedback*, motivado pelos bons resultados relatados na literatura referentes ao uso de tais técnicas. Foram realizados experimentos com uma base de imagens extraídas de web sites de *e-commerce*, e foram usados dois subconjuntos de imagens como consultas, sendo um formado por imagens com plano de fundo uniforme (semelhantes às presentes na base), e outro por imagens com ruído no fundo (fotografias em geral). Foram comparados o uso de *Relevance Feedback* para os dois subconjuntos de consultas, e para cada subconjunto foram comparados o uso de funções de *ranking* aprendidas com e sem o uso de *feedback*. Como resultado temos que o melhor cenário para ambos os subconjuntos é o uso da função de *ranking* aprendida sem RF. O uso de RF durante a aprendizagem torna os indivíduos dependentes do *feedback*, piorando as respostas em buscas antes da primeira interação de RF, e fazendo com que a função aprendida não seja capaz de captar a semântica da consulta original.

PALAVRAS-CHAVE: CBIR, e-commerce, Relevance Feedback, Gap Semântico, Programação Genética

Abstract

Fashion products are difficult items to be annotated and described by text, making it necessary to use images to perform searches on web sites of e-commerce. Such products hold great visual appeal, in other words the presentation of images relating to them are factors that directly influence the buying decision of a customer. These facts justify the study of the use of CBIR (Content Based Image Retrieval) in this context, an area already well studied in the scientific community, but that still has several shortcomings, the main one being the problem of Semantic Gap . The use of features extracted from the image by an algorithm is still not effective enough in associate it with its meaning, which is reflected in the results of a search, affecting the customer satisfaction with the store. This study seeks to address the problem of Semantic Gap through Genetic Programming and Relevance Feedback, motivated by the good results reported in the literature concerning the use of such techniques. Experiments were performed with an image base extracted from web sites e-commerce, and we used two subsets of images as queries, where one has images with a uniform background (as do the images of the data set), and the other has images with noisy backgrounds (photography in general). We compared the use of Relevance Feedback for both subsets, and for each subset we compared the use of ranking functions learned with and without using feedback. As the result, the best cenery for both subsets is to use the ranking function learned without usinf RF. Using RF on the learning process of GP makes the individuals dependent of the feedback, worsening the answers of searches before the first interaction with the user, and making the learned function unable to capture the semantic of the original query.

KEY-WORDS: CBIR, e-commerce, relevance Feedback, Semantic gap, Genetic Programming

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Objetivo	3
1.2 Organização do Trabalho	4
2 Fundamentação Teórica	5
2.1 Busca de imagens	5
2.1.1 Text-Based Image Retrieval (TBIR)	5
2.1.2 Content-Based Image Retrieval (CBIR)	6
2.2 Recuperação de Informação Multimodal	8
2.3 Busca de imagens em e-commerce	9
2.4 Relevance Feedback e Expansão de Consulta	10
2.4.1 Feedback Explícito	10
2.4.2 Feedback Implícito	11
2.4.3 Pseudo-Feedback	12
2.5 Programação Genética	13
2.5.1 Operações Genéticas	14
2.5.2 Função de Aptidão	15
2.5.3 Geração da População Inicial	16
3 Relevance Feedback com Programação Genética	18
3.1 Arcabouço de GP para Busca Multimodal	18
3.1.1 Terminais	19

3.1.2	Função de Aptidão	21
3.1.3	Implementação do Relevance Feedback	22
3.1.4	Consultas Compostas	23
3.1.5	Cenários e Hipóteses	24
4	Experimentos	26
4.1	Base de Imagens	26
4.1.1	Configuração do GP	28
4.2	Experimentos com o Subconjunto Easy Set	29
4.2.1	Resultados para o Subconjunto Easy Set	30
4.3	Experimentos com o Subconjunto Hard Set	37
4.3.1	Resultados para o Subconjunto Hard Set	38
4.4	Discussão sobre os resultados	45
5	Conclusão e Trabalhos Futuros	47
5.1	Conclusões	47
5.2	Trabalhos Futuros	49
	Referências Bibliográficas	50

Lista de Figuras

2.1	Abordagem da fusão precoce, fusão tardia e fusão intermediária - imagem disponível em [25]	9
2.2	Diagrama de um esquema de RF explícito simples	11
2.3	Métodos de análise local e global	12
2.4	Exemplo de um indivíduo estruturado como uma árvore binária.	14
2.5	Ilustração do processo de mutação.	15
2.6	Ilustração do processo de cruzamento.	15
2.7	Funcionamento básico da Programação Genética. Imagem disponível em [26].	16
3.1	Diagrama do processo de RF proposto.	23
4.1	Exemplos de imagens contidas nos conjuntos <i>easy set</i> e <i>hard set</i>	27
4.2	Gráficos sobre os resultados das buscas no Cenário 1 (sem usar RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas fáceis	32
4.3	Gráficos sobre os resultados das buscas no Cenário 2 (sem usar RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas fáceis	33
4.4	Gráficos sobre os resultados das buscas no Cenário 3 (usando RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas fáceis	34
4.5	Gráficos sobre os resultados das buscas no Cenário 4 (usando RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas fáceis	35

4.6	Gráficos sobre os resultados das buscas no Cenário 5 (usando RF na aprendizagem, sem remover a consulta original e combinando os métodos) usando consultas fáceis	36
4.7	Gráficos sobre os resultados das buscas no Cenário 6 (usando RF na aprendizagem, removendo a consulta original e combinando os métodos) usando consultas fáceis	36
4.8	Gráficos sobre os resultados das buscas no Cenário 1 (sem usar RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas difíceis	40
4.9	Gráficos sobre os resultados das buscas no Cenário 2 (sem usar RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas difíceis	41
4.10	Gráficos sobre os resultados das buscas no Cenário 3 (usando RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas difíceis	42
4.11	Gráficos sobre os resultados das buscas no Cenário 4 (usando RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas difíceis	43
4.12	Gráficos sobre os resultados das buscas no Cenário 5 (usando RF na aprendizagem, sem remover a consulta original e combinando os métodos) usando consultas difíceis	44
4.13	Gráficos sobre os resultados das buscas no Cenário 6 (usando RF na aprendizagem, removendo a consulta original e combinando os métodos) usando consultas difíceis	45

Lista de Tabelas

2.1	Esquemas de atribuição de peso para termos em documentos	6
3.1	Lista de descritores de imagens que são usados para gerar os terminais. . .	20
4.1	Distribuição das imagens da base entre as categorias.	28
4.2	Parâmetros do GP.	29
4.3	Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 1	31
4.4	Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 2	33
4.5	Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 3	33
4.6	Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 4	35
4.7	Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 5	36
4.8	Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 6	37
4.9	Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 1	40
4.10	Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 2	41
4.11	Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 3	42
4.12	Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 4	43

4.13	Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 5	44
4.14	Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 6	44

Capítulo 1

Introdução

Busca de imagens é uma área da computação que vem ganhando a atenção de muitos pesquisadores atualmente. A popularização de smartphones e a facilitação do acesso à internet fez com que o uso de diferentes mídias, como imagens e vídeos, se tornassem práticas comuns no dia-a-dia. Desta forma, a quantidade de conteúdo multimídia disponibilizado na internet, utilizado para diferentes finalidades, se tornou grande ao ponto de ser necessário o estudo de técnicas de recuperação de informação que pudessem lidar, de forma eficiente, com tal conteúdo.

Produtos referentes a vestuário detém um grande apelo visual, isto é, a apresentação de imagens dos produtos influencia diretamente o consumidor no momento da compra. Sabendo disso, torna-se importante explorar o uso deste tipo de mídia em web sites de *e-commerce* como forma de influenciar a decisão do cliente, podendo aumentar sua satisfação com a loja e potencializar compras futuras.

Produtos de moda, como roupas e acessórios, têm características difíceis de serem descritas pelos clientes por meio de texto, podendo ser estas relacionadas a cor, marca, estampa, entre outros fatores descritivos que variam entre os tipos de produtos. Dessa forma, o uso de sistemas de recuperação de informação tradicionais, como por exemplo buscas textuais, geram insatisfação ao tornar o processo de compra difícil, refletindo nos resultados das buscas.

Através da utilização de imagens como consulta, como figuras e fotos de produtos, é possível contornar este problema. Clientes podem utilizar imagens para encontrar outros produtos semelhantes em uma loja virtual, tornando o processo de escolha mais simples. Ao contrário das buscas textuais, onde o usuário precisa fornecer manualmente as in-

formações que descrevem, em palavras, a sua necessidade de informação, os sistemas de buscas visuais se encarregam de extrair, de forma automática, os dados que melhor representam esta necessidade.

Sistemas de buscas visuais utilizam características presentes nas imagens digitais para compará-las. Tais características se referem a dados como a frequência de determinadas cores nos pixels, padrões e texturas, formas e contornos encontrados na imagem, entre outros. Esta estratégia é conhecida como busca de imagens por conteúdo (*Content Based Image Retrieval* ou CBIR), que consiste em utilizar algoritmos (chamados também de descritores) que analisam as imagens e transformam as características em dados estruturados. A busca é feita comparando estes dados, encontrando a similaridade entre as imagens e ordenando as mesmas por este valor.

Um problema frequente em buscas de imagens é o *gap* entre a percepção do usuário sobre uma imagem e a percepção do computador sobre a mesma. Na literatura este problema é conhecido como *Gap Semântico* [8]. A subjetividade da percepção humana gera um fosso semântico entre a representação de uma imagem por um algoritmo e a interpretação da mesma pelo usuário. Por exemplo, enquanto uma pessoa é capaz de olhar uma imagem e entender o seu significado, um algoritmo entenderia a mesma como sendo um vetor numérico dentro de um espaço vetorial.

O uso de diferentes fontes de evidências combinadas, como por exemplo texto e imagem, é uma solução que provou ser eficiente para contornar o problema do *Gap Semântico* [13, 11, 21, 3]. Este tipo de abordagem é conhecida como *Busca Multimodal*, onde diferentes modalidades de evidências são combinadas para representar uma consulta. Porém, a forma como estas evidências são combinadas é algo que ainda vem sendo estudado.

Saraiva [26] mostra em seu trabalho que o uso de informações textuais presentes na base, combinadas com as características visuais das consultas, supera, em termos de precisão nas buscas, o uso isolado de descritores de imagem. Seu método, que foi chamado de *Expansão-GPI*, consiste em utilizar Programação Genética (*Genetic Programming* ou GP) para combinar os resultados de diferentes estratégias de expansão multimodal, que por fim retorna uma função de ordenação de respostas (função de *ranking*) contendo a melhor combinação encontrada. Seu método foi aplicado no contexto de buscas de imagens em comércio eletrônico, obtendo bons resultados.

Calumby et al. [3] propuseram um método de busca interativo utilizando Realimen-

tação de Relevância (*Relevance Feedback* ou RF) para melhorar buscas de imagem em bases na web. Este método consiste em utilizar diversas evidências das imagens, entre informações textuais e visuais, combinando-as através do GP. Neste método, as imagens recuperadas pela busca são selecionadas usando o *feedback* do usuário, e estas são utilizadas pelo GP para reformular a função de *ranking* tendo como base a necessidade de informação expressada pelo RF.

Santos et al. [11] aplicaram RF em conjunto com GP para abordar buscas de imagens de sensoriamento remoto. O *feedback* do usuário é usado para encontrar regiões de interesse nas imagens, as quais são aprendidas utilizando GP. Através dessa abordagem é possível atingir valores de precisão que superam os resultados de buscas usando evidências visuais isoladas.

Com base nos resultados observados na literatura, este trabalho propõe a aplicação de técnicas de RF, em conjunto com o arcabouço descrito em [26], para melhorar buscas em comércio eletrônico. Os resultados encontrados na literatura mostram que o uso de RF trás o aumento da precisão nas buscas. Isso ocorre pois, desta forma, o usuário ajuda o sistema de busca a contornar o problema do *Gap Semântico* ao indicar para o buscador onde ele acertou, possibilitando que o sistema se adapte a sua percepção.

No entanto, o uso de RF acaba sendo inconveniente para o usuário, pois é necessário que as buscas sejam refeitas a cada ciclo de interação. O uso desta técnica é indicado quando o seu ganho em precisão nas buscas compensa o esforço do usuário final.

Nosso trabalho propõe usar GP para aprender uma combinação de modalidades que, trabalhando com processos de buscas interativas com RF, seja possível melhorar a qualidade das respostas utilizando o menor número de interações possível. Este trabalho parte da hipótese de que, ao inserir RF dentro do processo de aprendizagem do GP, seja possível aprender uma função de *ranking* capaz de otimizar o buscador para este contexto.

1.1 Objetivo

Este trabalho tem como objetivo principal analisar os efeitos do uso de Realimentação de Relevância (*Relevance Feedback* ou RF) nos resultados de buscas em sistemas de recuperação de imagens em *e-commerce*. Propomos o uso de Programação Genética para aprender uma função de *ranking* que combine modalidades de maneira eficaz em um

sistema de buscas de imagem com RF.

Como objetivos específicos podem ser citados: Desenvolver e avaliar métodos de RF que possam ser aplicados em um processo de aprendizagem de GP; Avaliar a aplicabilidade dos métodos desenvolvidos no contexto de comércio eletrônico.

A hipótese principal deste trabalho é que, ao combinar diferentes técnicas que são comprovadamente eficazes, como expansão multimodal de consulta e RF, seja possível melhorar ainda mais a qualidade das respostas de buscas de imagem em *e-commerce*. É esperado que, ao utilizar RF no processo de GP, a evolução dos indivíduos se comporte de uma maneira diferente, possibilitando a descoberta de novas estratégias que superem as abordagens conhecidas atualmente.

1.2 Organização do Trabalho

Esta proposta está organizada da seguinte forma:

O Capítulo 2 mostra a fundamentação teórica deste trabalho, mostrando conceitos necessários para o entendimento dos demais capítulos. Conceitos sobre Recuperação de Informação e técnicas relacionadas com a área, como RF e GP, são discutidos com maior profundidade.

O Capítulo 3 mostra a metodologia proposta neste trabalho. São detalhadas as configurações a serem aplicadas no processo de aprendizagem do GP, é mostrado o contexto no qual esta pesquisa se aplica, e é mostrado o funcionamento dos métodos de RF propostos.

O Capítulo 4 retrata detalhes sobre os experimentos realizados até então, os quais buscam validar a proposta deste trabalho. São mostrados detalhes sobre a base de imagem utilizada e, posteriormente, discutidos os resultados obtidos.

O Capítulo 5 mostra as conclusões tiradas com base nos resultados obtidos pelos experimentos, e destaca algumas questões em aberto para serem tratadas em trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Neste capítulo são apresentados conceitos sobre recuperação de informação, focando em busca de imagens, *Relevance Feedback* e Programação Genética, fundamentais para o entendimento deste trabalho.

2.1 Busca de imagens

Os primeiros sistemas de recuperação de imagens utilizavam informações textuais referentes a cada imagem dentro de um modelo de recuperação de informação tradicional. Informações textuais eram obtidas de diferentes formas, podendo ser extraídas de partes textuais específicas presentes no documento de onde foram encontradas, ou até mesmo através de anotação manual. Este método é conhecido como Text-Based Image Retrieval (TBIR), ou Recuperação de imagens Baseada em texto).

2.1.1 Text-Based Image Retrieval (TBIR)

O modelo de recuperação de informação tradicional mais conhecido é o modelo do Espaço Vetorial [22, 1]. Neste modelo cada termo presente em um documento é considerado como um eixo em um espaço vetorial, e a sua frequência, combinada com sua *importância* no conjunto de documentos presentes na base de dados do sistema em questão, funcionam como coordenadas. Por fim, a consulta e os documentos são considerados como vetores em um espaço vetorial, e a similaridade entre eles é dada pelo cosseno do ângulo formado

entre os mesmos. O cálculo da similaridade é mostrado na equação a seguir:

$$sim(d_j, q) = \cos\theta = \frac{\vec{d} \cdot \vec{q}}{|\vec{d}| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (2.1)$$

onde d_j é um documento da base de dados, q é uma consulta, t é o número total de termos presentes entre os documentos e w é uma função que atribui peso ao termo i em um documento j ou na consulta.

A *importância* de um termo na base é calculada com base em sua frequência entre os documentos. A ideia é que um termo presente na grande maioria dos documentos da base tem uma importância menor que um outro presente em uma fração pequena de documentos. Dada uma consulta, uma palavra presente nesta que ocorre em todos os documentos da base não é capaz de distinguir qual é o mais relevante para a mesma.

As funções de atribuição de peso são baseadas na frequência e importância de um termo, em um esquema conhecido como *tf-idf*. Baeza-Yates & Ribeiro-Neto [1] descrevem alguns esquemas de *tf-idf*, que podem ser observados na Tabela 2.1.1.

Esquema	peso do termo no documento	peso do termo na consulta
1	$f_{i,j} * \log \frac{N}{n_i}$	$(0.5 + 0.5 \frac{f_{i,q}}{\max_i f_{i,q}}) * \log \frac{N}{n_i}$
2	$1 + \log f_{i,j}$	$\log(1 + \frac{N}{n_i})$
3	$(1 + \log f_{i,j}) * \log \frac{N}{n_i}$	$(1 + \log f_{i,q}) * \log \frac{N}{n_i}$

Tabela 2.1: Esquemas de atribuição de peso para termos em documentos

O modelo vetorial é utilizado neste trabalho no que se refere a obtenção de valores de similaridade por evidências textuais entre imagens.

2.1.2 Content-Based Image Retrieval (CBIR)

Apesar da eficiência obtida ao utilizar TBIR, evidências textuais nem sempre estão presentes em bases de imagens, e muitas vezes a anotação manual se torna um processo muito custoso. Fez-se necessário o surgimento de um método capaz de indexar imagens utilizando somente o seu conteúdo, ou informação visual. A busca de imagens baseada em conteúdo (CBIR) vem sendo aplicada em diversos trabalhos ao longo dos anos [23, 28] e continua sendo aprimorada em diversos trabalhos mais recentes [7, 21].

CBIR consiste em utilizar um descritor de imagens, ou seja, um algoritmo que processe a imagem a nível de pixels, e transformá-la em um vetor numérico que possa ser

trabalhado em um espaço vetorial. As características a serem extraídas podem ser baseadas nas cores, texturas, formas e outros aspectos da imagem. Formalmente, um descritor d é constituído por um par (ϵ_d, δ_d) onde:

- $\epsilon_d : I \rightarrow \mathbb{R}^n$ é uma função que extrai um vetor de características \vec{v}_I de uma imagem I .
- $\delta_d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ é uma função que calcula a similaridade entre duas imagens através da distância entre seus vetores de características.

A similaridade entre imagens é dada como a distância entre os vetores, podendo esta ser obtida de diferentes maneiras. A distância euclidiana, demonstrada na Equação 2.2, é a mais comum dentre as funções de similaridade. A escolha dos descritores e das funções de similaridade devem ser feitas de acordo com o contexto em que serão aplicadas, já que o tipo de características que diferem imagens pode variar entre diferentes bases.

$$d(V_Q, V_B) = \sqrt{\sum_{i=1}^{vs} (V_Q(i) - V_B(i))^2} \quad (2.2)$$

onde V_Q é um vetor de características de uma consulta, V_B é um vetor de uma imagem presente na base e vs é o tamanho do vetor, sendo este de mesmo tamanho para ambas as imagens.

O uso de características locais da imagem é capaz de encontrar imagens semelhantes em um determinado aspecto, porém é ineficiente em capturar o significado da imagem consulta. É de interesse do usuário que a busca retorne imagens que representem a consulta efetuada, como por exemplo, se o usuário usa a foto de uma camisa azul como consulta, um sistema de CBIR é capaz de retornar não só camisas, mas outros itens azuis com formato semelhante. Esse problema é conhecido na literatura como *Gap Semântico* [8]. Diversas formas de contornar esse problema vêm sendo estudadas nos últimos anos, dentre elas se destacam a busca multimodal, o uso de *Relevance Feedback* e a aplicação de métodos de aprendizado de máquina como classificação [9, 21, 8], agrupamento [12] e programação genética [3, 11, 27, 26]. Este trabalho aborda o problema do *Gap Semântico* através da busca multimodal com programação genética e RF.

2.2 Recuperação de Informação Multimodal

Busca multimodal consiste em realizar buscas utilizando diversos tipos de evidências combinadas, como texto, informações visuais e metadados. Em busca de imagens, combinar informações textuais e visuais torna possível capturar melhor o significado de cada imagem e ser resgatada. Smeulders et al. [29] demonstram que a combinação de ambas as modalidades conseguem compensar as suas limitações em um sistema de busca.

Primeiramente, é necessário definir de que forma o sistema de busca representa cada evidência individualmente. Existem diversos modelos de recuperação de informação capazes de lidar com evidências textuais, sendo um dos mais eficazes o Modelo Vetorial, mencionado na Seção 2.1.1 deste capítulo. Já para evidências visuais não há um consenso sobre qual o melhor método para representá-las, fazendo da escolha dos descritores e das funções de similaridade uma questão experimental.

Tendo definidas as formas de representação das evidências, resta saber de que forma as mesmas são combinadas. Segundo Santos [25], existem três maneiras básicas de combinar modalidades em um sistema de busca: fusão precoce (*early fusion*), fusão tardia (*late fusion*) e fusão intermediária (*intermedia fusion*).

Na fusão precoce, as informações são extraídas do documento e combinadas para formar uma representação única do mesmo, em um único vetor de características concatenado. Desta forma é possível ter uma real representação multimídia de um documento, de uma forma que todas as fontes de evidência são usadas para uma decisão. Porém esta abordagem acaba gerando vetores de características muito extensos, o que, segundo Depeursinge & Müller [10], ocasiona a dispersão de agrupamentos de instâncias pertencentes aos mesmos conceitos.

Na fusão tardia o documento é representado separadamente por cada modalidade, e então os valores de similaridade entre os documentos e a consulta, para cada modalidade, são combinados através de alguma expressão matemática. A forma como os valores de similaridade são combinados pode ser aprendida, experimentada ou simplesmente baseada em outros trabalhos presentes na literatura.

Fusão intermediária funciona como um esquema de *pseudo-relevance feedback*, onde uma modalidade é responsável por gerar um *rank* com possíveis resultados para uma consulta, e as demais modalidades servem para expandir ou reordenar o *rank* principal, tendo como base os resultados do próprio *rank*. As três abordagens são apresentadas no

diagrama da Figura 2.2.

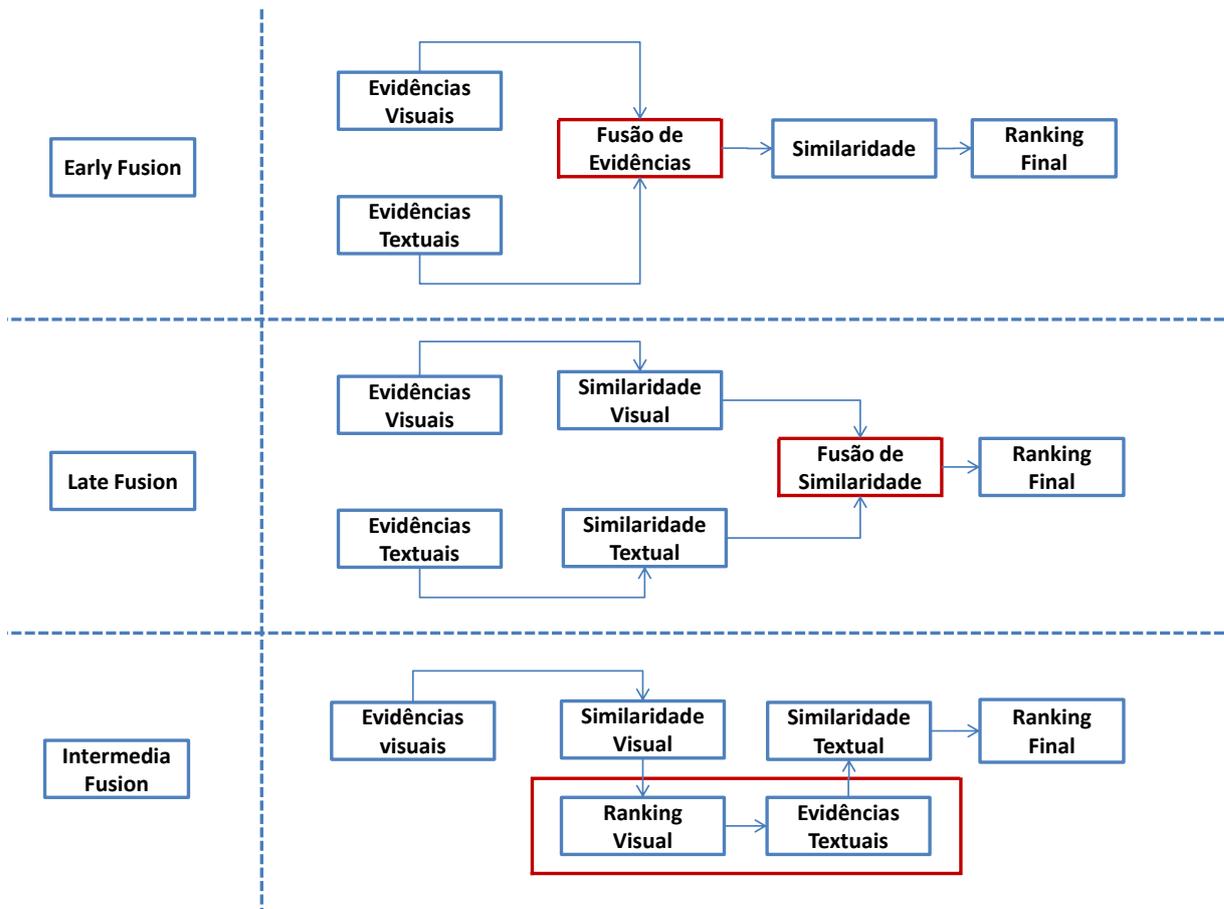


Figura 2.1: Abordagem da fusão precoce, fusão tardia e fusão intermediária - imagem disponível em [25]

Neste trabalho usa-se o esquema de fusão tardia, onde são usados *ranks* gerados por valores de similaridades visual, junto com *ranks* obtidos com base em consultas textuais geradas a partir das informações presentes nos documentos do topo dos *ranks* visuais. Os *ranks* são combinados a partir de expressões matemáticas obtidas através de Programação Genética.

2.3 Busca de imagens em e-commerce

E-commerce significa comércio eletrônico, e abrange vendas de todos os tipos de produtos na internet. De um modo geral as buscas feitas em *e-commerce* utilizam modelos de recuperação de informação tradicionais, já que usuários costumam procurar marcas, especificações, preços e outros aspectos. Porém algumas consultas são difíceis de serem

representadas por texto, principalmente no setor de vestuário. O uso de imagens como consulta torna-se necessário em buscas de roupas e outros itens de vestuário, pois desta forma a representação da necessidade de informação do usuário é demonstrada com maior facilidade.

Diferentemente de sistemas de recuperação de imagens da web, onde utiliza-se dados extraídos das páginas onde foram encontradas as imagens para indexação, as bases de imagens em sites de comércio eletrônico geralmente são estruturadas, tendo informações textuais organizadas, sendo estas referentes à descrição do produto, categoria, preço e outros aspectos. Desta forma é possível otimizar o uso de tais evidências para melhorar a qualidade das respostas das buscas. Existem muitos estudos referentes a recuperação de informação em *e-commerce* [15, 27, 26, 25], além de ser o foco deste trabalho.

2.4 Relevance Feedback e Expansão de Consulta

Como uma alternativa para a solução do problema do *Gap Semântico*, existem técnicas que utilizam informação de *feedback* de usuários para capturar o significado das consultas submetidas pelos mesmos. *Relevance feedback* (RF) se refere ao uso de dados fornecidos pelo usuário para o melhoramento da consulta. A ideia por detrás do RF é usar dados das respostas que são retornadas inicialmente por uma consulta, sabendo quais são relevantes, para serem incorporados à consulta original em um processo conhecido como Expansão de Consulta. Existem três tipos de *feedback*: *feedback explícito*, *feedback implícito* e *pseudo-feedback*.

2.4.1 Feedback Explícito

Relevance feedback explícito é o método mais eficaz para expansão de consulta, trazendo melhores resultados para as buscas. Esta técnica funciona através de um processo iterativo onde, com base na consulta original, são mostradas algumas respostas, onde as relevantes são selecionados manualmente pelo usuário e incorporadas na consulta para uma nova busca. Este método tem a desvantagem de requerer a interação explícita do usuário, porém ainda é utilizado devido ao ganho na qualidade das respostas [16, 18]. A interação do usuário também pode ser simulada, o que torna possível o uso desta técnica em processos de aprendizagem de máquina [3, 21].

A Figura 2.2 mostra o funcionamento de um esquema de RF explícito simples, retratando a interação entre o usuário e um sistema de buscas. A consulta fornecida pelo usuário é processada pelo sistema, o qual retorna um *rank* inicial com documentos possivelmente relevantes. O usuário então seleciona os relevantes para que o sistema possa expandir a consulta e retornar novos resultados, repetindo o processo até alcançar um critério de parada previamente estabelecido.

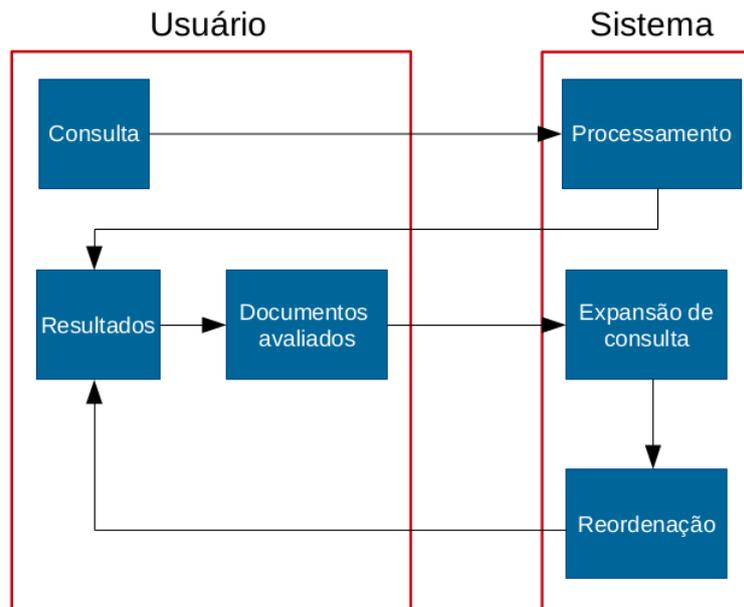


Figura 2.2: Diagrama de um esquema de RF explícito simples

2.4.2 Feedback Implícito

Uma alternativa para expandir a consulta de uma forma mais eficiente em termos de usabilidade, sem requerer a interação explícita do usuário, é através do uso de *feedback* implícito [31]. Este método consiste na geração de uma base de dados de ações de usuários, que são associadas às consultas previamente realizadas pelos mesmos, a qual é utilizada para aprender padrões, ou associações que possam ajudar na expansão de consultas futuras. Os dados mais comumente coletados para esta tarefa são dados de navegação e cliques de usuários, além do tempo em que permanecem em uma página e ações de rolagem da página.

2.4.3 Pseudo-Feedback

Pseudo-feedback providencia uma análise local automática dos resultados, automatizando a interação manual que seria fornecida pelo usuário. Este método consiste em assumir que uma parte da resposta contém informações relevantes, como por exemplo o topo do *rank*, e utilizá-las para expandir a consulta inicial. Outra forma de expandir a consulta de maneira implícita é através de uma análise global, utilizando fontes de dados externas, como por exemplo o *thesaurus*¹, para associar termos relacionados à mesma. Este processo busca melhorar os resultados sem que hajam iterações prolongadas. A Figura 2.3 mostra o funcionamento dos esquemas de análise local (a) e global (b).

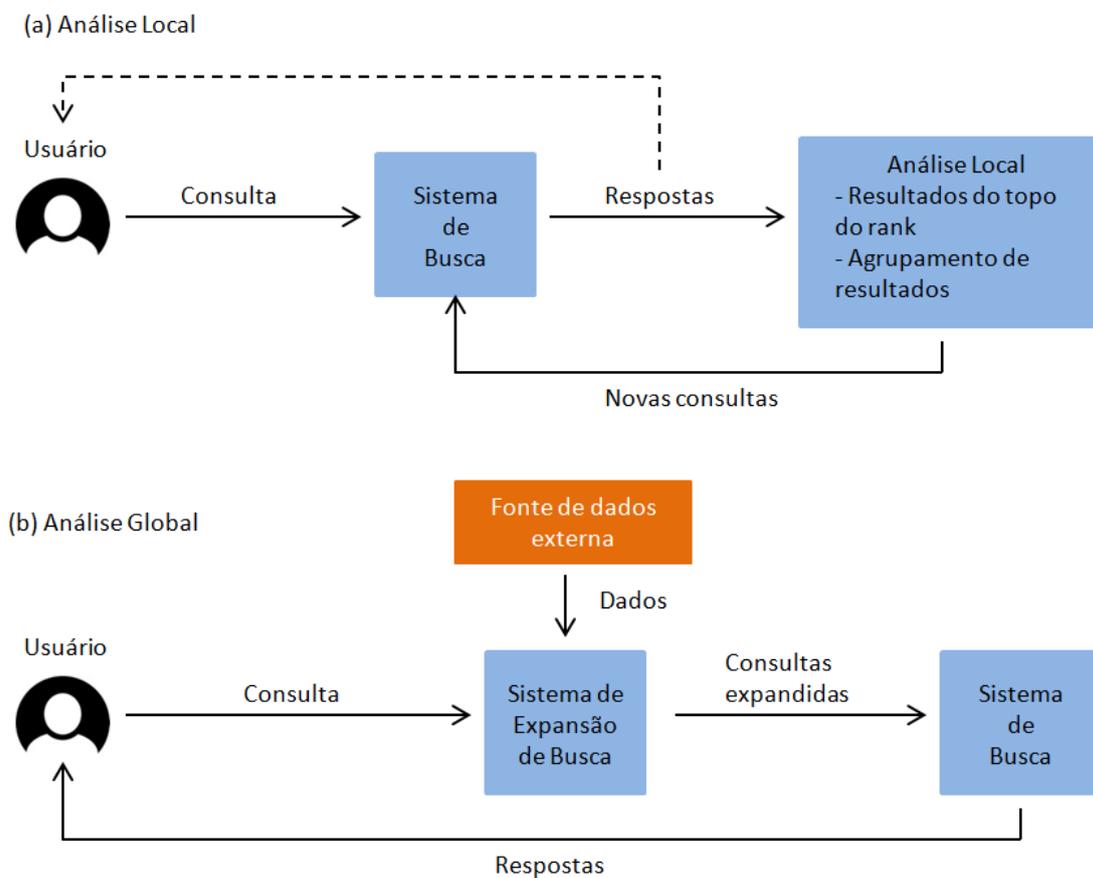


Figura 2.3: Métodos de análise local e global

Em (a) observamos que o processo de expansão de consulta pode se repetir para aprimorar ainda mais os resultados, enquanto a qualquer momento, obedecendo um critério de parada estabelecido, os resultados são mostrados para o usuário. Em (b) A consulta

¹Base de dados contendo sinônimos, podendo ser usada para generalizar uma consulta textual. Mais informações estão disponíveis em <http://www.thesaurus.com>

fornecida pelo usuário passa por um processo de expansão antes da realização das buscas, o que ocorre apenas uma vez, retornando os resultados direto para o usuário.

2.5 Programação Genética

Programação Genética (GP - *Genetic Programming*) é uma técnica de aprendizagem de máquina baseada na teoria evolutiva de Charles Darwin, na qual ocorre o fenômeno de adaptação das espécies na luta pela sobrevivência [19]. Assim como em sua teoria, programas podem ser considerados indivíduos que evoluem com o passar de gerações, onde são passivos de sofrer operações genéticas. Esta técnica vem sendo bastante estudada ao longo dos anos [24, 3, 11, 27, 26] e apresenta bons resultados.

No paradigma de Programação Genética, populações de programas vão evoluindo ao longo de gerações, a partir dos princípios de Darwin de sobrevivência dos mais aptos e do cruzamento de espécies. Neste processo são usadas estruturas hierárquicas com forma e tamanho variáveis, isto é, os programas a serem avaliados detêm diferentes características, podendo alguns serem representados por menos características e outros por mais. Os mesmos não são necessariamente ótimos, mas esta diversidade garante a variabilidade genética, que ao ser combinada de acordo com valores de aptidão obtidos por cada programa, geram soluções eficientes para problemas onde não é possível, ou é muito difícil, obter uma boa resposta.

Mais precisamente, programas são expressões matemáticas estruturadas como árvores binárias. Os nós internos representam operadores aritméticos ou lógicos, e as folhas, chamadas de terminais, são variáveis ou constantes a serem combinadas. Na Figura 2.4 é mostrada a estrutura de um programa representado a expressão $(CEDD + GCH)/(log(10))$, onde $CEDD$ e GCH são variáveis, 10 é um valor constante e $+$, $/$ e log são operadores.

Um programa estruturado é chamado, dentro do processo de GP, de *indivíduo*. Um indivíduo representa uma possível solução dentro de um espaço de possibilidades das soluções para um determinado problema. Seus nós internos são chamados de *funções*, as quais atuam sobre os valores nos terminais ou sobre outras funções de forma hierárquica. Terminais são as características atribuídas a cada programa, guardando valores de diferentes fontes referentes ao problema abordado. Como exemplo, é possível usar valores de

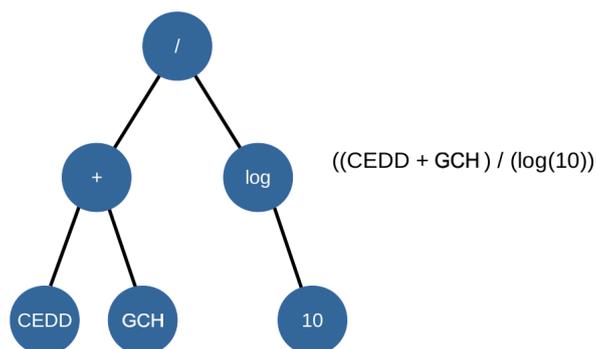


Figura 2.4: Exemplo de um indivíduo estruturado como uma árvore binária.

similaridades obtidos entre diferentes descritores de imagens como terminais, no intuito de descobrir a melhor representação para imagens ao comparar uma com a outra.

2.5.1 Operações Genéticas

As operações genéticas atuam na estrutura dos programas, alterando fragmentos dos indivíduos aleatoriamente ou combinando-os entre si. Em GP existem três operadores genéticos: *Reprodução*, *Mutação* e *Cruzamento*.

Reprodução é uma operação genética que clona indivíduos de uma geração para a seguinte, sem modificar sua estrutura. A probabilidade de um indivíduo ser clonado é diretamente proporcional ao valor obtido para sua aptidão. Isso garante que as características presentes nos indivíduos mais aptos estejam presentes nas próximas gerações.

Enquanto a Reprodução mantém indivíduos sem modificações, as operações de Mutação e Cruzamento se encarregam de alterar ou combinar as características presentes nos indivíduos de uma geração.

Mutação atua em um único indivíduo da população. Esta operação consiste em escolher uma sub-árvore aleatoriamente e substituí-la por uma nova ramificação. Em geral esta operação deve ocorrer com uma probabilidade menor, pois sua ação pode desconstruir indivíduos com altos valores de aptidão. Ao mesmo tempo, Mutação é capaz de reconstruir indivíduos menos aptos, adicionando novas características para a próxima geração. A Figura 2.5 mostra o funcionamento da Mutação.

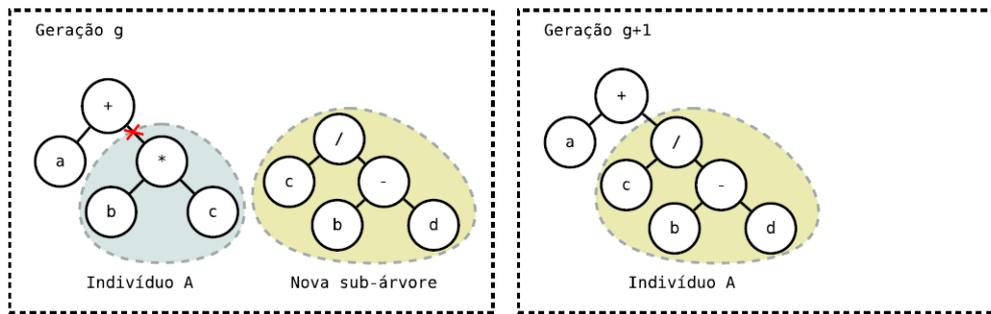


Figura 2.5: Ilustração do processo de mutação.

Cruzamento, também chamado de *Crossover*, consiste em selecionar dois programas que obtiveram bons resultados na geração atual, selecionar aleatoriamente uma sub-árvore em cada e troca-las entre si, gerando dois filhos. São formados dois grupos de indivíduos escolhidos aleatoriamente, onde os mais aptos são selecionados para passarem pelo processo de cruzamento. A ideia é que, ao combinar as características responsáveis pelos altos valores de aptidão, seja possível gerar filhos que superem os seus pais. Esse processo é ilustrado pela Figura 2.6.

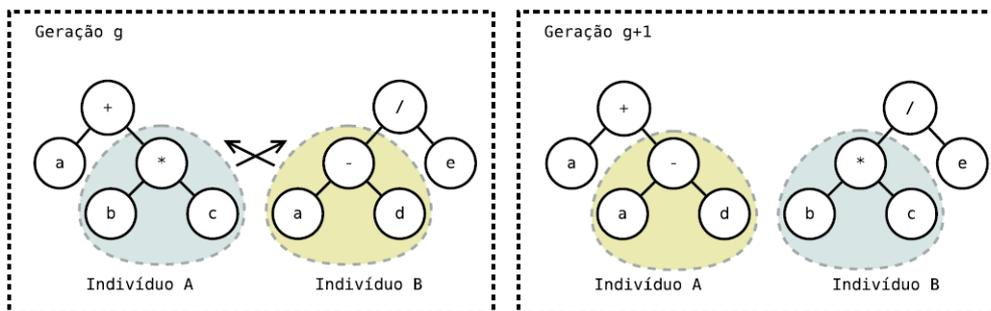


Figura 2.6: Ilustração do processo de cruzamento.

A probabilidade com a qual cada operação ocorre é definida pelo usuário, através de experimentos ou com base em trabalhos relacionados.

2.5.2 Função de Aptidão

Os programas são avaliados a partir de uma Função de Aptidão a cada geração, sendo esta implementada pelo usuário de acordo com o problema em questão. Os programas são ordenados de acordo com o valor de aptidão atribuído a cada um, e as operações genéticas garantem que os mais bem adaptados sobrevivam para as novas gerações. O experimento termina quando ocorrem um determinado número de gerações, ou quando

um programa alcança determinado valor de aptidão.

A Figura 2.7 mostra o funcionamento básico do GP. O processo se inicia com a criação da população inicial (Passo 1), a qual já é avaliada logo em seguida usando-se a Função de Aptidão (Passo 2). Finaliza-se o processo caso haja um indivíduo que cumpra o critério de parada, caso contrário há uma seleção de indivíduos (Passo 3) para sofrerem a ação das operações genéticas (Passo 4). Por fim, cria-se uma nova geração (Passo 5) a ser avaliada, repetindo o processo.

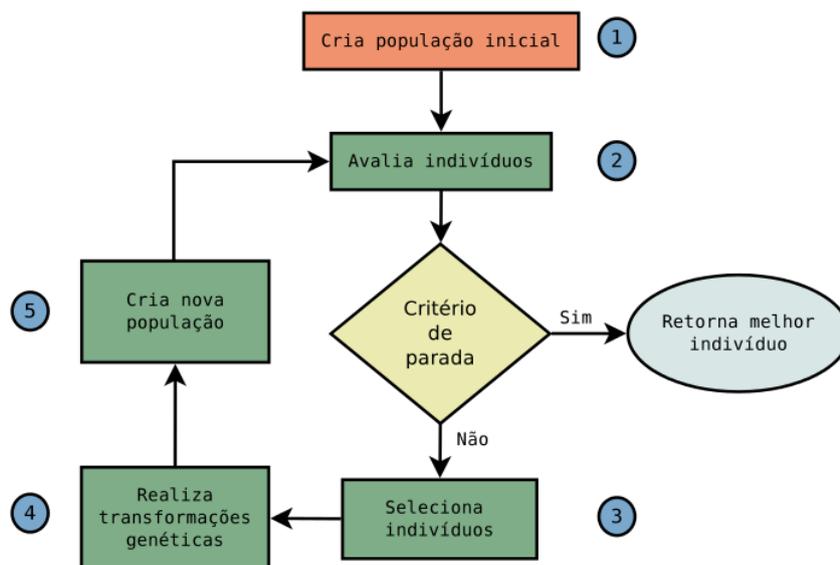


Figura 2.7: Funcionamento básico da Programação Genética. Imagem disponível em [26].

2.5.3 Geração da População Inicial

O conjunto de indivíduos que forma a população inicial é gerado aleatoriamente com base nas funções e terminais predeterminados. As funções e terminais são combinados em estruturas de árvores, preocupando-se em permanecer em um limite de profundidade definido pelo arcabouço GP. A geração aleatória dos indivíduos da população inicial pode ser feita utilizando os métodos *Full*, *Grow* e *Ramped-falf-and-half*.

O método *Full* gera árvores completas, onde todos os nós-folhas possuem a mesma distância até o nó-raiz. A altura das árvores geradas é igual à profundidade máxima definida pelo arcabouço.

O método *Grow* cria árvores de profundidade variável. A profundidade das árvores ainda é restrita aos limites mínimo e máximo, porém o método não busca gerar árvores completas.

O método *Ramped-falf-and-half* é a fusão entre os métodos *Full* e *Grow*. Neste método são geradas árvores com base em diferentes profundidades, sendo que, para cada valor entre os limites mínimo e máximo, existam quantidades iguais de indivíduos com a respectiva profundidade. Para cada fração de indivíduos, usa-se em 50% deles o método *Full*, e nos outros 50% o método *Grow*.

Capítulo 3

Relevance Feedback com Programação Genética

A proposta deste trabalho consiste em explorar o ganho semântico obtido ao usar *Relevance Feedback* (RF), junto com o paradigma de aprendizagem de máquina, para aprender uma função de *ranking* mais eficiente para ser usada no contexto de busca de imagens em *e-commerce*. O método proposto utiliza programação genética, explorando diferentes modalidades com base no arcabouço utilizado pelo método *Expansão-GPI* [26], em conjunto com o RF. O *feedback* do usuário é simulado durante o processo de aprendizado, fazendo com que o GP aprenda uma função de *ranking* que seja mais adequada para o uso de múltiplas imagens de consulta. Este capítulo mostra detalhes sobre o arcabouço de GP a ser utilizado, sobre como os métodos de RF foram implementados, sobre os tipos de evidências a serem avaliados pelo GP, e como é feita a simulação da interação do usuário com o sistema.

3.1 Arcabouço de GP para Busca Multimodal

Diferentemente da abordagem apresentada em [26], o método proposto neste trabalho admitirá consultas formadas por múltiplas imagens. O valor atribuído a um terminal passa a ser a média entre os valores de similaridade entre as imagens presentes na consulta e a imagem da base. Este cálculo é feito para cada método de expansão multimodal ou descritor visual isolado utilizado como terminal pelo método, gerando vários valores de similaridades que são combinados pela função de *ranking*.

O arcabouço de GP usado neste trabalho é dividido em duas fases: treino e validação. São utilizadas consultas previamente avaliadas como fonte de conhecimento para o processo de aprendizagem. Uma fração do conjunto de imagens de consulta é utilizado pela função de aptidão durante a fase de treinamento, enquanto outra fração é usada para validar a função de *ranking* obtida ao final do processo de aprendizagem. A base de imagens e os conjuntos de consultas são detalhados na Seção 4.1 deste capítulo.

Utiliza-se um processo de validação cruzada de *5 folds* para calcular o resultado final do método. Neste processo o conjunto de imagens de consulta é dividido em 5 subconjuntos mutuamente exclusivos, e para cada um dos subconjuntos são usadas as imagens restantes como treino, sendo estas utilizadas pela função de aptidão do GP. Desta forma, o processo retorna 5 funções de *ranking* que são avaliadas por seus respectivos subconjuntos de imagens de consulta. O uso de validação cruzada nesse contexto é necessário devido ao fato de que a população inicial, gerada aleatoriamente no começo do processo de aprendizagem do GP, pode influenciar diretamente o resultado final.

Ao final do processo de aprendizagem, consideramos como resultado as 5 funções de *ranking* aprendidas, as quais são avaliadas em conjunto.

3.1.1 Terminais

O uso de uma grande variedade de terminais em GP tende a gerar bons resultados devido a maior variabilidade das combinações dos mesmos. Através dos dados disponíveis na base de imagens usada neste trabalho é possível extrair evidências de diferentes modalidades, as quais podem ser combinadas através de técnicas já descritas no Capítulo 2. As evidências visuais isoladas, juntas com as combinações das mesmas com informações textuais associadas geram uma quantidade de terminais grandes o suficiente para garantir a variabilidade dos indivíduos.

Tendo como contexto busca de imagens em comércio eletrônico de itens de vestuário, as características dos produtos usadas para as buscas são referentes a forma (por exemplo, ao se comparar uma camisa com um sapato), cor e estampa (textura do produto). para que sejam realizadas buscas com base nessas características é necessário descrever as imagens com base nas mesmas.

O terminais responsáveis por representar estas características visuais dos produtos são obtidos a partir do valor de similaridade entre as imagens utilizando vetores de caracterís-

ticas extraídos por descritores focados em cada aspecto. Os descritores selecionados para os experimentos neste trabalho são capazes de extrair as propriedades visuais descritas anteriormente. A Tabela 3.1 mostra a relação entre os descritores usados neste trabalho e suas respectivas propriedades visuais.

Descritor de Imagem	Propriedade Visual
GCH(<i>Global Color Histogram</i>)[20]	Cor
ACC(<i>Auto Color Correlogram</i>)[14]	Cor
CLD(<i>Color Layout Descriptor</i>)[17]	Cor
PHOG(<i>Pyramid Histogram of Oriented Gradients</i>)[2]	Forma
CEDD(<i>Color and Edge Directivity Descriptor</i>)[4]	Cor e Textura
FCTH(<i>Fuzzy Color and Texture Histogram</i>)[5]	Cor e Textura
JCD(<i>Joint Composite Descriptor</i>)[6]	Cor e Textura
BIC(<i>Border/Interior pixel Classification</i>)[30]	Cor

Tabela 3.1: Lista de descritores de imagens que são usados para gerar os terminais.

Os descritores de imagem e o modelo de busca multimodal são baseados nos estudos realizados em [26]. A modalidade textual precisa ser explorada de maneira implícita. O método proposto admite consultas através de imagens, portanto para que a modalidade textual seja usada durante as buscas é preciso que uma consulta textual seja inferida. A maneira adotada neste trabalho de inferir tal informação utiliza o próprio *ranking* obtido pela busca visual, para cada descritor de imagem, como fonte para a extração da consulta textual.

Dado um *ranking* obtido por busca visual, são selecionadas as K primeiras imagens do topo. Cada imagem da base apresenta uma descrição textual, um pequeno texto que descreve o produto. É extraída a primeira palavra da descrição para cada imagem selecionada, e esta é concatenada em uma string, que servirá como consulta textual. O valor de K é alternado entre 1, 5, 10 e 20, como proposto em [26], e este processo se repete para cada um dos 8 *ranks* obtidos por CBIR, totalizando 32 consultas textuais.

O modelo do espaço vetorial é utilizado para realizar as buscas textuais. As descrições de cada produto da base de dados são indexadas junto com a informação de categoria. Dada uma consulta textual, a similaridade é calculada e seu valor é utilizado como terminal no GP. A pesagem dos termos, tanto para as consultas, como para os termos da base, é dada por $tf \times idf$, onde o *idf* definido pela seguinte equação:

$$idf = \log \frac{N}{df_i} \quad (3.1)$$

sendo N o número total de imagens na base, e df_i o número de imagem que contem o termo i na descrição ou categoria.

Pode-se inferir a probabilidade de uma dada consulta pertencer a determinada categoria analisando as frequências de cada categoria presente em um *rank*, e esta probabilidade pode ser usada como peso para ser feito o *rerank* de um resultado obtido por busca visual. Neste trabalho são calculadas as frequências com base no topo do *rank* de tamanho K , com K variando entre 1, 5, 10 e 20. Portanto, este peso é calculado para os resultados obtidos por cada descritor de imagem, e utilizado como terminal no GP, totalizando mais 32 terminais.

Por fim, dada uma imagem de consulta e o seu resultado na busca visual, são usados como terminais o menor e o maior valor de similaridade encontrado nas 100 primeiras respostas do *rank*. Estes terminais também são usados em [26] para uma possível normalização dentro do GP. No total são utilizados 80 terminais mais uma variável que alterna entre números de 0 a 100.

3.1.2 Função de Aptidão

Um indivíduo recebe um valor de aptidão com base na avaliação das buscas, que utilizam RF e um processo iterativo. Existe um conjunto de consultas previamente avaliadas, onde são conhecidas as suas respostas relevantes, que são utilizadas dentro da função de aptidão para que sejam avaliadas as precisões de buscas que utilizam a função de *ranking* representada pelo indivíduo. Desta forma, cada consulta é processada pelo sistema de busca, passando por um processo de RF onde a interação do usuário é simulada, e o *ranking* obtido é avaliado com base nas imagens relevantes conhecidas.

A simulação da interação do usuário é feita utilizando o conjunto de imagens relevantes previamente coletado. Após o processo de busca utilizando uma das imagens de consulta, é recuperado um *ranking* que seria mostrado ao usuário. As respostas da busca são analisadas do topo do *ranking* para baixo, e a primeira imagem relevante, que ainda não está contida no conjunto de consultas, é inserida à consulta (inicialmente formada por uma única imagem fornecida pelo usuário). O *ranking* é percorrido até que seja encontrada uma imagem relevante entre as 100 primeiras imagens. A busca é realizada novamente considerando as duas imagens.

A imagem selecionada pelo usuário é então desconsiderada durante a próxima avaliação, isto é, ela é removida da resposta. A forma como o valor de similaridade é calculado entre a consulta e uma imagem da base, sabendo que a consulta agora é composta por duas imagens, é feita usando a média aritmética entre os dois valores de similaridade.

Os *rankings* obtidos para cada consulta são avaliados usando a métrica $p@5$. Ao final da avaliação de todas as consultas separadas para o treinamento do GP, é tirada a Média Aritmética entre os valores de $p@5$ para os *rankings* obtidos após a adição de uma imagem da base à consulta, selecionada pelo *feedback* simulado.

Considera-se a métrica $p@5$ para o valor da aptidão de um indivíduo devido a capacidade de avaliar a qualidade das primeiras imagens presentes na respostas. Isso faz com que o GP aprenda funções capazes de concentrar mais imagens relevantes no topo do *ranking*.

3.1.3 Implementação do Relevance Feedback

Na Figura 3.1 é possível observar como é realizado o processo de RF. A base de imagens passa por um pré-processamento onde são calculados os vetores de características para cada descritor, além da indexação das informações textuais de cada produto (passos 1 e 4). A imagem usada como consulta é processada por cada descritor de imagem, onde os vetores gerados por cada descritor são usados para realizar as buscas que recuperam os primeiros *ranks* (passo 2). Todos os métodos de expansão multimodal apresentados no arcabouço GP são aplicados, gerando *ranks* expandidos, para isso os *ranks* gerados pelo sistema de busca são passados para o sistema de expansão de consulta (passo 3), onde são usadas as respostas do topo do *rank* para realizar a expansão junto com os dados fornecidos pela base. Por fim os *ranks* são combinados usando a função de *ranking* a ser avaliada, formando uma resposta que é mostrada ao usuário (passo 6), onde esse fornece o *feedback* necessário para a obtenção de uma nova consulta (passo 7).

O *indivíduo* mostrado no diagrama é responsável por combinar todos os dados obtidos e gerar um único *rank*, e para isso recebe os *ranks* vindos do sistema de busca (passo 3) e do sistema de expansão de busca (passo 5).

O processo de RF ocorre apenas uma vez durante o processo de GP. O método proposto busca a função de *ranking* capaz de obter a maior melhoria das respostas usando RF com o menor número de interações possíveis. Desta forma, limitamos o número de

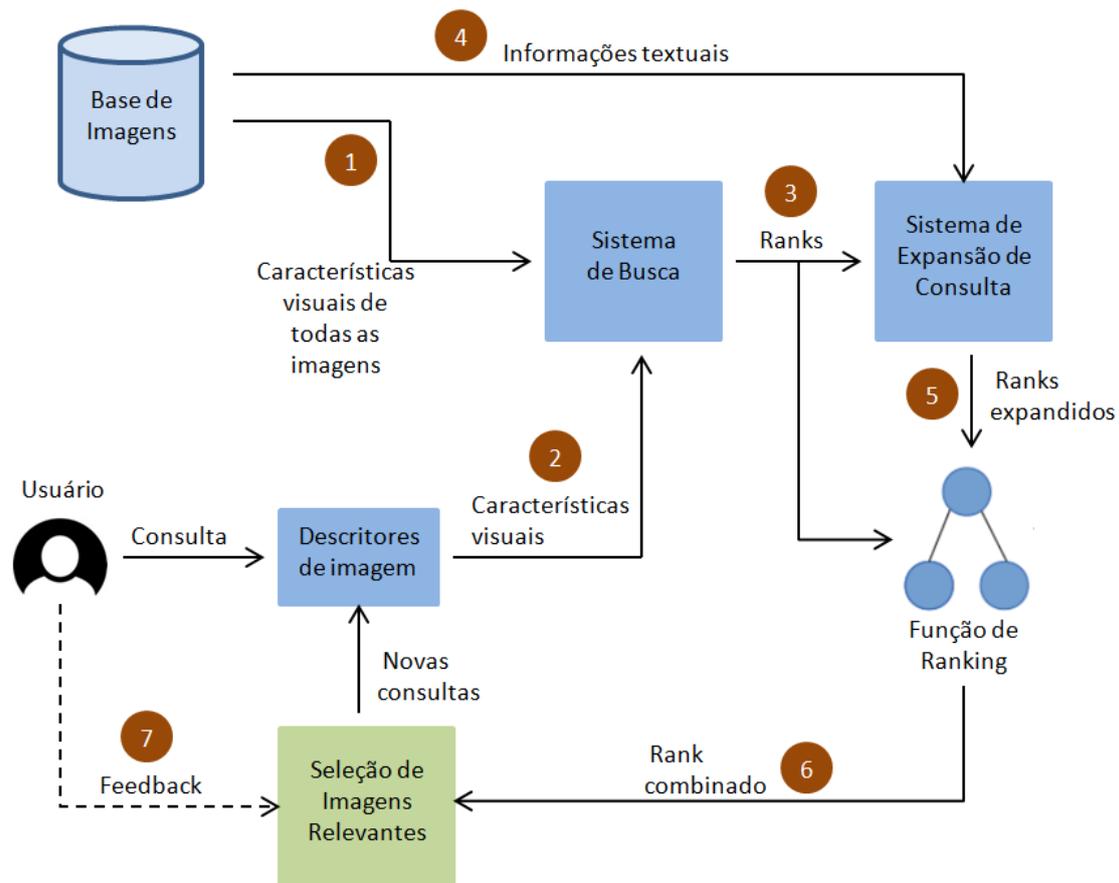


Figura 3.1: Diagrama do processo de RF proposto.

interações para que o GP dê preferência aos indivíduos com o maior valor de precisão já no primeiro ciclo de RF.

Este trabalho propõe experimentar o uso de *feedback* dentro do processo de aprendizado do GP. Será avaliado o *feedback* explícito, onde a interação do usuário é simulada dentro da função de aptidão. Desta forma é possível observar o desempenho do método proposto aplicado a um sistema de buscas iterativo.

3.1.4 Consultas Compostas

Outra questão abordada na metodologia deste trabalho é a forma como é calculada a similaridade entre a consulta e uma imagem da base. Após a seleção de itens relevantes na primeira iteração em um processo de RF, são agregadas à consulta novas imagens. Durante a aprendizagem do GP, usamos Média Aritmética para calcular esta similaridade. Dada uma consulta com N imagens, onde para cada uma existe um valor de similaridade obtido pela função δ do indivíduo em questão, a similaridade para a consulta é dada pela

equação a seguir:

$$\text{Média Aritmética} = \frac{1}{N} \sum_{i=1}^N \delta(I_i, d_j) \quad (3.2)$$

onde d_j é uma imagem da base e I é a consulta composta.

3.1.5 Cenários e Hipóteses

Sabemos que as imagens presentes em bases de imagens de web sites de *e-commerce* muitas vezes seguem um padrão, como por exemplo, o uso de um plano de fundo igual para todas as fotos, ou todas as fotos serem tiradas com a mesma iluminação. No entanto o usuário pode fornecer ao sistema uma imagem que não segue esse mesmo padrão, como uma foto que ele mesmo tenha tirado. Também é possível que o usuário use uma imagem encontrada e outro web site, que segue um padrão parecido.

Neste caso temos dois cenários, onde o usuário submete consultas difíceis para o sistema (fora do padrão) e onde ele submete consultas fáceis (padrão semelhante). É necessário usar GP para aprender as funções de *ranking* que mais se adequem a estes dois cenários, e para isso usamos o método Expansão-GPI sem o uso de RF no processo de aprendizagem, e também o método Expansão-GPI com o uso de RF no processo de aprendizagem.

Como não é possível garantir que, para ambos os cenários, a imagem consulta siga o mesmo padrão que as imagens presentes na base, existe a hipótese de que esta cause ruído na busca. Para sabermos se isso é verdade, precisamos remover a consulta original (fornecida pelo usuário) após a seleção da primeira imagem relevante usando RF, e usar somente as imagens selecionadas pelo *feedback* como consulta. Isso deve ser verificado para cada cenário mencionado anteriormente, e para cada método sendo avaliado.

Outra hipótese a ser avaliada é o uso da função de *ranking* aprendida sem o uso de RF para a realização da primeira busca (apenas com a consulta original), e no momento em que a primeira imagem selecionada por *feedback* entrar para a consulta, alternamos a função para uma outra aprendida com o uso de RF. Essa hipótese se baseia na possibilidade de que o método que usa RF no processo de aprendizagem funcione bem para as buscas com *feedback*, mas que seja ineficaz durante a primeira busca, onde ainda não há o *feedback* do usuário. Também é necessário avaliar esta última hipótese removendo

a consulta original, baseando-se na mesma possibilidade de que a consulta original gere ruído na busca.

Quando removemos a consulta original enquanto avaliamos os métodos usando consultas difíceis, é necessário alternar as funções para aquelas aprendidas usando consultas fáceis, já que, em teoria, estas consultas são mais semelhantes às imagens da base, as quais passam a ser usadas como consulta neste contexto.

Capítulo 4

Experimentos

Neste capítulo são mostrados os experimentos realizados para avaliar os possíveis cenários onde podemos aplicar os métodos Expansão-GPI sem RF na aprendizagem e Expansão-GPI com RF na aprendizagem.

Os experimentos consistiram em avaliar processos de aprendizagem de GP, aplicando técnicas de RF e busca Multimodal já mencionadas no Capítulo 3. Para a realização dos experimentos foi utilizada uma base de imagens de produtos de moda, avaliando a aplicação dos métodos no contexto de buscas interativas em *e-commerce*.

4.1 Base de Imagens

As imagens utilizadas neste trabalho foram extraídas de web sites de comércio eletrônico de itens de vestuário, ou seja, roupas, calçados e acessórios. A base de imagens contém 23155 imagens, com *background* uniforme (por exemplo, com fundo branco), além de dados associados a cada uma referentes às especificações do respectivo produto, como o preço, categoria e descrição. Os dados disponíveis nesta base foram estudados em [25, 26] e mostram que o uso de diferentes modalidades, explorando a informação visual e os dados textuais das imagens, geram resultados melhores em sistemas de busca quando comparados a buscas meramente visuais ou textuais.

As imagens de consulta são compostas por dois subconjuntos de imagens, cada um contendo 50 imagens distintas, as quais serão chamadas de *easy set* e *hard set*. O primeiro conjunto consiste em imagens semelhantes às presentes na base de imagem, ou seja, fotos de itens de vestuário com *background* uniforme, extraídas de outros web sites

de *e-commerce*. O segundo conjunto é formado por imagens consideradas difíceis para o buscador, consistindo em fotos sem pré-processamento algum, apresentando *background* variado. Alguns exemplos de imagens contidas nesses conjuntos de consultas podem ser observados na Figura 4.1.



Figura 4.1: Exemplos de imagens contidas nos conjuntos *easy set* e *hard set*

A coleção de imagens é composta por 6 categorias, as quais são listadas na Tabela 4.1. A Tabela 4.1 mostra a distribuição das imagens entre as categorias. É importante ressaltar que a distribuição entre as mesmas é desigual, e por causa disso é feita uma divisão mais inteligente para o conjunto de consultas durante o processo de validação cruzada do GP, onde ocorre uma divisão justa das categorias.

O conjunto de consultas fáceis tem uma quantidade média de respostas relevantes por consulta de 60 imagens, enquanto para o conjunto de imagens difíceis temos uma média de 20 imagens relevantes por consulta. A diferença na quantidade de itens relevantes

por consulta pode influenciar os valores de precisão das buscas, principalmente quando desconsideramos da resposta a consulta relevante selecionada como *feedback*.

Categorias	Quantidade (%)
Roupas masculinas	18,31
Roupas femininas	39,37
Bolsas	5,47
Calçados masculinos	7,34
Calçados femininos	27,34
Acessórios	2,17

Tabela 4.1: Distribuição das imagens da base entre as categorias.

O processo de aprendizagem com GP foi realizado utilizando o método Expansão-GPI para os conjuntos de consultas fáceis e difíceis, para que fossem obtidos os dados a serem comparados com o método proposto. Em seguida o processo se repetiu com a utilização de RF com ciclos de uma iteração, ou seja, o *feedback* acontecia uma única vez para cada consulta.

Utilizando o GP, foram obtidas 16 funções de *ranking*, 5 especializadas em buscar consultas fáceis sem utilizar RF, outras 5 em consultas difíceis, e outras 10 especializadas em buscas interativas utilizando RF explícito, para consultas fáceis e difíceis. Os experimentos a seguir foram realizados para explorar diferentes cenários, onde é avaliado o desempenho de cada método individualmente, utilizando diversas métricas de avaliação de *ranking*, e também cenários onde é possível combinar as funções aprendidas.

4.1.1 Configuração do GP

A configuração do GP teve como base a configuração usada em [26], e foi adequada à complexidade do novo problema. A Tabela 4.2 mostra os valores determinados para cada parâmetro da configuração.

É importante ressaltar que os experimentos não levam em consideração o seu tempo de execução. Os resultados obtidos são funções de *ranking* que podem ser aplicadas em sistemas de busca online, porém a execução do GP ocorre offline. Isso significa que parâmetros como o tamanho da população e número máximo de gerações podem ser altos, e que os experimentos podem levar longos períodos de tempo para terminar.

Parâmetro	Valor
Tamanho da população	500
Número máximo de gerações	40
Profundidade inicial mínima	2
Profundidade inicial máxima	6
Profundidade máxima	7
Probabilidade de reprodução	0,9
Probabilidade de mutação	0,05
Probabilidade de cruzamento	0,05
Operadores	$\log, \times, +, -, /, \sqrt{}, \min, \max$

Tabela 4.2: Parâmetros do GP.

4.2 Experimentos com o Subconjunto Easy Set

Esta seção aborda os métodos mencionados anteriormente no contexto de buscas usando consultas semelhantes as presentes na base. O objetivo dos experimentos é descobrir se o uso de RF durante a aprendizagem do GP influencia os resultados das buscas quando aplicamos RF fora do processo de aprendizagem.

Seguindo a metodologia explicada no Capítulo 3, o processo de aprendizagem com GP foi aplicado, resultando em 5 funções de *ranking* para cada método a ser avaliado. Para cada método, são usadas as 5 funções para calcular a precisão do uso do método nas buscas. Isso é feito utilizando a média entre as precisões das buscas para cada partição de teste.

As funções de *ranking* selecionadas para os experimentos seguintes são mostradas a seguir, sendo (a) referente ao método Expansão-GPI sem o uso de RF no aprendizado, e (b) referente ao método Expansão-GPI usando RF durante o aprendizado:

$$(a.1): (((jcd + bic\ cat5) - (cedd\ cat20 / acc\ max)) * (cedd\ cat10 + cedd\ sim)) * ((cedd\ cat10 + cedd) - (cedd\ cat20 / acc\ max))$$

$$(a.2): ((sqrt(jcd\ cat20)) + ((min(((max((min((max(fcth\ min, cld\ min)), (sqrt(gch\ cat1)))), (max((sqrt(fcth\ txt20)), gch\ cat5)))) + (cedd + (cedd * fcth\ cat20))), (jcd\ cat20 + (cedd * (jcd\ cat20 + (cedd * cedd)))))) + (((sqrt(cedd * fcth\ cat20)) + (max((sqrt(fcth\ txt20)), (sqrt(gch\ cat1)))) + (cedd * fcth\ cat20))))$$

$$(a.3): cedd$$

$$(a.4): (((acc\ min + (log(sqrt(cedd)))) + (((acc\ min + (jcd\ cat20 + bic\ cat10)) + ((log(sqrt(cedd))) + ((sqrt(cedd)) + cedd\ cat1))) + (gch\ cat20 + cedd\ cat1))) + bic\ cat10) + ((log(jcd\ cat20 + bic\ cat10)) + (log(sqrt(cedd))))$$

(a.5): $jcd\ cat20$

(b.1): $((phog\ cat5 + ((\min(bic\ max, ((\min(bic\ max, cedd)) + (\max(((\max(((cld + jcd) * gch\ cat20), (fcth + fcth))) + jcd), phog\ cat5)))))) + jcd\ txt10)) * cedd) * (fcth + acc)$

(b.2): $\max(fcth\ cat20, bic\ txt10) * (\min((\min(((gch\ cat20 * bic\ max) * ((fcth - fcth\ txt10) * (((\min((gch\ cat20 * cedd), (fcth - fcth\ txt10))) * cedd) - fcth\ txt10) * (\min(bic\ max, jcd\ cat5))))), (\max(fcth\ cat20, bic\ txt10))), (\max(fcth\ cat20, bic\ txt10)))) * (fcth - fcth\ txt10)$

(b.3): $\max((\sqrt{cedd}), (\min((bic\ txt5 + jcd\ cat5), (\sqrt{(\sqrt{jcd}) + (jcd\ txt5 + cedd)}))))$

(b.4): $\min((\min(jcd, jcd\ cat20)), (\min(jcd, (\min((\min(jcd\ cat20, (\min((\max(fcth\ cat10, cedd\ txt10)), jcd\ cat20))))), jcd\ cat20))))$

(b.5): $(((((\min((\log(fcth\ cat10)), acc)) + (((cedd\ cat5 * jcd) + (jcd + (cld\ txt10 + bic\ cat5))) + (jcd + (\min((\log(fcth\ cat10)), acc)))))) + ((\min((\max(jcd\ txt5, phog\ txt20), cedd)) + (\min((\log(fcth\ cat10)), acc)))) + (cedd\ cat5 * jcd)) + (\min((cld\ txt10 + bic\ cat5), acc)))$

Podemos constatar que, assim como foi documentado em [26], ocorre uma presença maior do descritor CEDD dentro das funções de *ranking* aprendidas pelo GP em (a). Este descritor aparece menos em (b), enquanto o descritor BIC passa a aparecer mais vezes, o que ocorre com menos frequência em (a).

4.2.1 Resultados para o Subconjunto Easy Set

Nos experimentos a seguir as funções de *ranking* são avaliadas em diferentes cenários. É realizado teste estatístico para avaliar os resultados encontrados, para isso usamos o teste de Wilcoxon.

Os cenários avaliados são:

Cenário 1 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem usar RF na aprendizagem, mantendo a consulta original durante todo o processo de RF.

Cenário 2 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem usar RF na aprendizagem, removendo a consulta original após o primeiro *feedback* do usuário.

Cenário 3 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI usando RF durante a aprendizagem, mantendo a consulta original durante todo o processo de RF.

Cenário 4 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI + RF usando RF durante a aprendizagem, removendo a consulta original após o primeiro *feedback* do usuário.

Cenário 5 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem RF na aprendizagem antes da primeira interação do usuário, trocando a função para a aprendida pela Expansão-GPI que usou RF na aprendizagem após o primeiro *feedback*, mantendo a consulta original durante todo o processo de RF.

Cenário 6 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem RF na aprendizagem antes da primeira interação do usuário, trocando a função para a aprendida pela Expansão-GPI que usou RF na aprendizagem após o primeiro *feedback*, removendo da consulta original após o primeiro *feedback* do usuário.

A figura 4.2 mostra o desenvolvimento das respostas das buscas no *Cenário 1*, ao receberem *feedback* do usuário, adicionando uma resposta relevante à consulta a cada interação. Foram utilizadas as métricas $P@5$, $P@10$ e $P@20$ para avaliar o impacto do RF para diferentes métricas de avaliação de *rank*. A Tabela 4.3 mostra os resultados dos testes estatísticos de Wilcoxon sobre os resultados obtidos pelas buscas considerando os valores de $p@5$.

	Sem RF	1° RF	2° RF	3° RF	4° RF
Sem RF	0	61	50	62	82
1° RF	61	0	17	38	63
2° RF	50	17	0	16	69
3° RF	62	38	16	0	62
4° RF	82	63	69	62	0

Tabela 4.3: Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 1

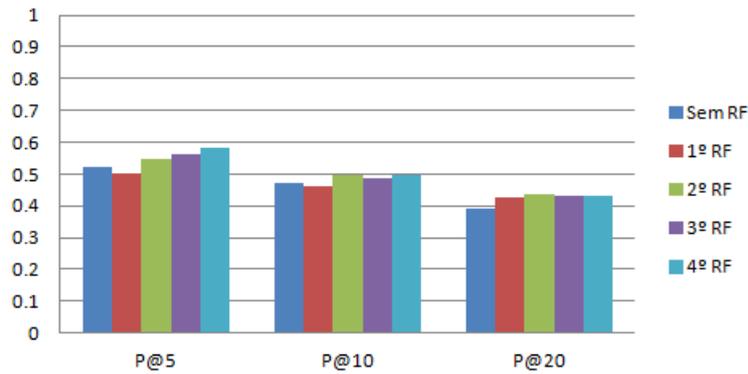


Figura 4.2: Gráficos sobre os resultados das buscas no Cenário 1 (sem usar RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas fáceis

Podemos ver que, independente da métrica de avaliação de *rank*, o uso de RF não melhorou significativamente a precisão das buscas. São necessárias 2 interações de RF para que haja alguma melhora na qualidade das respostas, sendo esta de 2.8% para $p@5$, 2.6% para $p@10$ e 4.4% para $p@20$. Após o segundo ciclo de RF, a qualidade das respostas não se altera em $p@10$ e $p@20$, mas podemos ver que as respostas relevantes tendem a ir para o topo do *rank*, o que é evidenciado pelo aumento da precisão em $p@5$, subindo de 54.8% no segundo ciclo para 58% no quarto.

Os resultados dos testes estatísticos mostram que não há um alto grau de confiança na diferença entre os resultados encontrados durante o processo de RF. Isto é, não há garantia de que haja uma real diferença entre os resultados das buscas. Para que possamos ter confiança de que os resultados realmente melhoraram ou pioraram, é necessário um valor igual ou acima de 95% de diferença estatística, enquanto para este cenário a maior diferença foi entre os resultados das buscas sem RF e os resultados no quarto ciclo de RF, com 82%. Os valores encontrados nos testes de Wilcoxon, junto com a pequena diferença nos valores de precisão das buscas, deixa evidente que nesse cenário o uso de RF é ineficaz.

A Figura 4.3 mostra os valores de precisão das buscas no *Cenário 2*, ainda utilizando o método Expansão-GPI sem o uso de RF na aprendizagem, removendo a consulta original após a primeira interação do usuário. Podemos ver que a remoção da consulta original prejudica as buscas, o que indica que perdemos parte da semântica da consulta ao trabalharmos apenas com o *feedback*. A Tabela 4.4 mostra os resultados dos testes estatísticos de Wilcoxon.

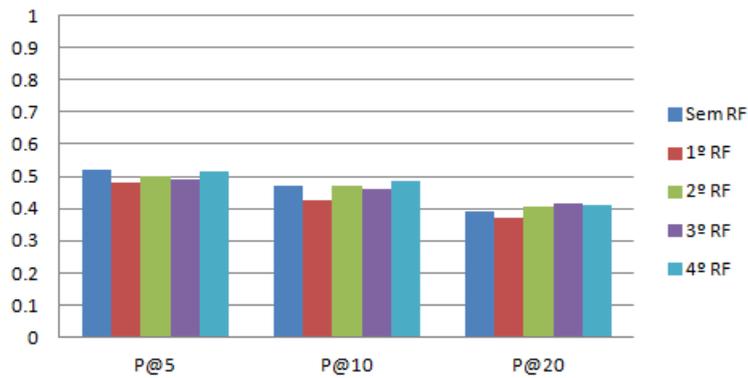


Figura 4.3: Gráficos sobre os resultados das buscas no Cenário 2 (sem usar RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas fáceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	33	3	22	26
1º RF	33	0	55	14	0
2º RF	3	55	0	37	50
3º RF	22	14	37	0	19
4º RF	26	0	50	19	0

Tabela 4.4: Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 2

Os resultados das buscas pioram ao remover a consulta original, e são necessários 4 ciclos de RF para que a precisão volte ao valor obtido antes da adição de *feedback* à consulta. Neste cenário as diferenças estatísticas obtiveram valores ainda menor que no *Cenário 1*, onde todos os ciclos de RF obtiveram respostas com valores de diferença estatística menores que 33% quando comparadas com as respostas das buscas sem RF.

Em seguida verificamos o desenvolvimento das respostas das buscas no *Cenário 3*, quando utilizamos o método Expansão-GPI com o uso de RF na aprendizagem. A Figura 4.4 mostra os valores de precisão das buscas quando utilizamos o método Expansão-GPI + RF, mantendo a consulta original, e a Tabela 4.5 mostra os resultados dos testes estatísticos de Wilcoxon.

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	85	84	75	56
1º RF	85	0	11	3	62
2º RF	84	11	0	56	95
3º RF	75	3	56	0	69
4º RF	56	62	95	69	0

Tabela 4.5: Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 3

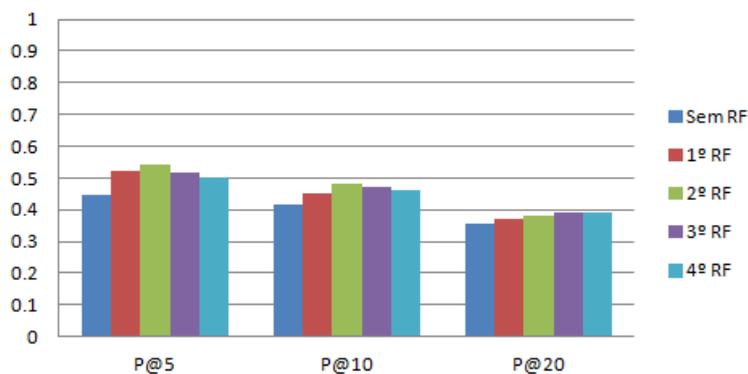


Figura 4.4: Gráficos sobre os resultados das buscas no Cenário 3 (usando RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas fáceis

Podemos ver que o método Expansão-GPI usando RF durante a aprendizagem consegue desenvolver a resposta com a adição do *feedback* do usuário à consulta, porém a precisão das buscas antes do RF é inferior a obtida pelo método Expansão-GPI sem o uso de RF na aprendizagem, o que se reflete nos ciclos seguintes. Ao compararmos os valores de precisão para os primeiros ciclos de ambos os métodos, sem remover a consulta original, temos a diminuição de 8.8% em $P@5$, 5.2% em $p@10$ e 3.4% em $p@20$.

Percebe-se também que o método desenvolve a resposta até o segundo ciclo, chegando ao maior valor de precisão entre os 4 ciclos. Porém, quando consideramos a precisão no segundo ciclo do *Cenário 1* (o qual obteve os maiores valores de precisão), ainda temos as diferenças de 0.4%, 1.4% e 5.6%, para $p@5$, $p@10$ e $p@20$ respectivamente.

Os testes estatísticos mostram que existe a diferença estatística de 85% entre os resultados sem RF e o primeiro ciclo de RF no *Cenário 3*. Assim como os resultados mostrados no gráfico pioram após o segundo ciclo de RF, os testes estatísticos também passam a obter valores menores. A diferença estatística entre as buscas sem RF e o segundo ciclo de RF é de 84%, semelhante à diferença entre o primeiro ciclo. No entanto, quando comparamos o primeiro com o segundo ciclo, os resultados obtiveram 11% de diferença estatística. Isso significa que podemos garantir a melhoria das respostas, ainda que com um valor baixo de confiança, após até o primeiro ciclo de RF.

A Figura 4.5 mostra os resultados do *Cenário 4*, ao removermos a consulta original após realizar as buscas usando Expansão-GPI com RF na aprendizagem, para então refazer a busca usando apenas o *feedback* do usuário. Assim como ocorreu no *Cenário 2*, o RF não desenvolveu as respostas tão bem trabalhando apenas com o *feedback*. Houve

uma queda significativa na precisão das buscas no segundo ciclo (logo após a remoção) para as métricas $p@10$ e $p@20$, enquanto para $p@5$ não houve queda, mas o aumento da precisão foi pouco. A Tabela 4.6 mostra os resultados dos testes estatísticos de Wilcoxon.

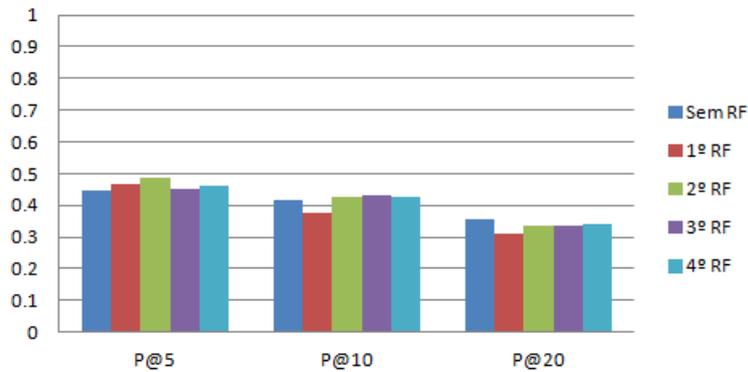


Figura 4.5: Gráficos sobre os resultados das buscas no Cenário 4 (usando RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas fáceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	9	16	15	22
1º RF	9	0	36	24	21
2º RF	16	36	0	77	36
3º RF	15	24	77	0	42
4º RF	22	21	36	42	0

Tabela 4.6: Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 4

Os testes estatísticos obtiveram valores muito baixos para este cenário. Fica evidente que a remoção da consulta original não faz diferença para a qualidade das respostas quando usamos consultas fáceis, pois tanto no *Cenário 2* como no *Cenário 4* foram obtidos valores menores de precisão após o primeiro ciclo de RF, e em ambos os casos observa-se uma queda nos valores das diferenças estatísticas.

A Figura 4.6 mostra os resultados das buscas no *Cenário 5*. Quando combinamos as funções de *ranking* durante o processo de RF, percebemos que os valores de precisão pouco se alteram com a adição de imagens da base à consulta, independente de quantas imagens são adicionadas. A Tabela 4.7 mostra os resultados dos testes estatísticos de Wilcoxon para este cenário.

Vemos que, ao alternar a função de *ranking*, as buscas não se desenvolvem mesmo com a adição de mais *feedback* do usuário. Os testes estatísticos também mostram que não há grande diferença estatística entre as respostas.

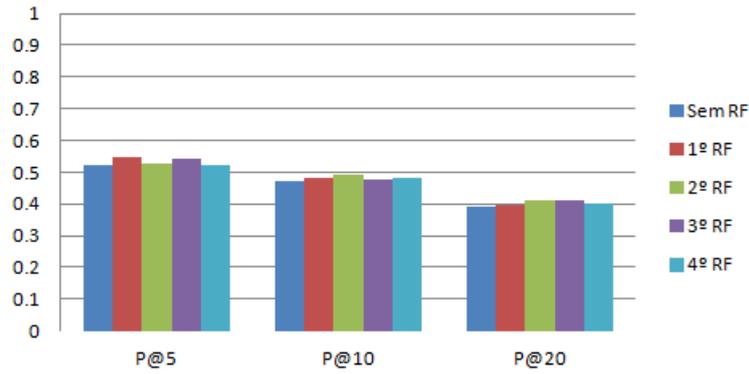


Figura 4.6: Gráficos sobre os resultados das buscas no Cenário 5 (usando RF na aprendizagem, sem remover a consulta original e combinando os métodos) usando consultas fáceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	53	18	19	3
1º RF	53	0	80	54	66
2º RF	18	80	0	5	19
3º RF	19	54	5	0	0
4º RF	3	66	19	0	0

Tabela 4.7: Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 5

Quando fazemos a troca de funções de *ranking* e também removemos a consulta original, podemos observar um comportamento que já havia sido percebido nos cenários 2 e 4. No *Cenário 6* observamos um comportamento semelhante aos observados em outros cenários onde a consulta original foi removida e usa-se apenas o *feedback* para as buscas. A Figura 4.7 mostra os resultados encontrados no *Cenário 6*, e a Tabela 4.8 mostra os resultados dos testes estatísticos de Wilcoxon.

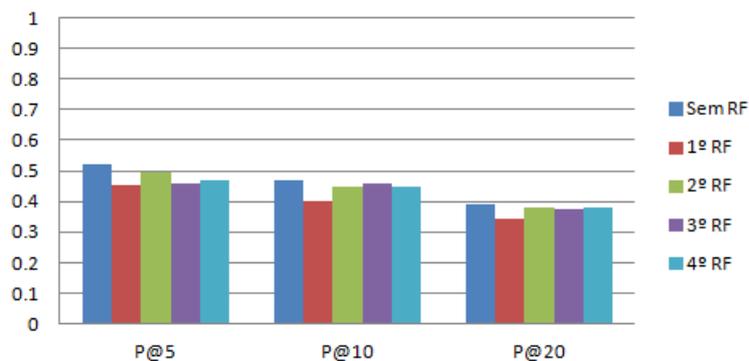


Figura 4.7: Gráficos sobre os resultados das buscas no Cenário 6 (usando RF na aprendizagem, removendo a consulta original e combinando os métodos) usando consultas fáceis

Os valores de precisão das buscas foram piores após a remoção da consulta original,

	Sem RF	1° RF	2° RF	3° RF	4° RF
Sem RF	0	40	18	62	16
1° RF	40	0	46	19	14
2° RF	18	46	0	72	27
3° RF	62	19	72	0	78
4° RF	16	14	27	78	0

Tabela 4.8: Teste de Wilcoxon para os resultados dos experimentos usando consultas fáceis no Cenário 6

e os valores encontrados nos testes estatísticos foram baixos. Isso reforça a ideia de que a consulta original detém um grande valor semântico sobre a necessidade de informação do usuário.

Em geral, o uso de RF não foi eficaz em nenhum dos cenários avaliados. Porém, devemos analisar também os resultados dos experimentos realizados usando as consultas difíceis para termos uma visão mais geral sobre os métodos. Os resultados destes experimentos são discutidos na seção 4.4.

4.3 Experimentos com o Subconjunto Hard Set

Os experimentos foram replicados utilizando o conjunto de consultas difíceis. O roteiro permanece o mesmo, porém, nos cenários onde removemos a consulta original, passamos a utilizar as funções de *ranking* aprendidas com o uso das consultas fáceis, devido à semelhança entre estas e as imagens da base que são usadas como consulta. As funções aprendidas pelo GP são listadas a seguir, sendo (a) referente ao método Expansão-GPI e (b) referente ao método Expansão-GPI + RF:

$$(a.1): \min(\left(\log(\min(\text{cld cat10}, (\sqrt{((\log(\text{fcth max})) + (\log(\text{bic cat5}))))}\right)) * \text{fcth}), (\text{cld cat10} + ((\min(\left(\log(\max(\left(\log(\max(\left(\max(\text{cld cat10}, \text{phog txt5})), \text{phog txt5})), \text{phog txt5}))) * \text{fcth}), (\log(\text{fcth})))) - \text{fcth}))\right)$$

$$(a.2): (((\text{cld max} / \text{phog min}) + (\min(\text{jcd cat20}, \text{phog txt10}))) + (\text{fcth cat10} + \text{jcd})) * (\text{cld cat10} / \text{bic txt10}) - (\text{cedd cat1} / \text{gch txt20}))$$

$$(a.3): (\text{cedd} * (\text{cedd} * ((\sqrt{(\max(\text{phog cat5}, \text{gch cat1})) * (\text{bic max} - ((\text{cedd} * (\max(\text{phog cat5}, \text{gch cat1})) * (\text{bic max} - (\sqrt{(\sqrt{(\text{jcd}))))}))} * \text{bic max}))))$$

$$(a.4): ((\text{cld cat5} * (\text{acc} / \text{acc txt1})) * (\text{cld cat5} / (\text{cedd txt20} + (\text{cld cat5} / \text{acc txt1})))) * (((\sqrt{(\text{fcth txt20}))} + (\text{acc cat5} * (\max(\text{cld cat10}, (\text{cedd txt20} + ((\text{cld cat5} / \text{acc txt1}) +$$

$((\text{phog cat20} - (\max((\text{cld cat5} / \text{acc txt1}), \text{gch txt20}))) + \text{cld cat1})))))) + (\min(\text{cld txt1}, \text{cld txt20}))))$

(a.5): $\text{sqrt}(\log(\text{cedd} + ((\max(\text{bic max}, \text{cld txt10})) + (\text{sqrt}((\min((\text{jcd txt10} + ((\text{cedd} + (\text{sqrt}((\text{cld txt1} * \text{jcd txt10}) * \text{cld txt10}))) + ((\min(\text{phog cat5}, \text{cedd min})) + \text{phog}))) + \text{phog}), (\text{cedd min} * \text{jcd})) + \text{phog}))))))$

(b.1): $((\max(\text{cld cat10}, (\text{jcd txt5} - \text{phog max}))) * (\text{acc cat20} + (\min(\text{jcd}, \text{gch cat10})))) + ((\text{sqrt}(\text{fcth cat20} + \text{cld cat10})) + (\max(\text{jcd txt1}, (\max(\text{jcd txt1}, (\text{sqrt}((\text{bic cat1} + \text{acc cat20}) * (\text{cld cat20} + \text{phog cat5}))))))))))$

(b.2): $((\max(\text{cld cat10}, (\text{jcd txt5} - \text{phog max}))) * (\text{acc cat20} + (\min(\text{jcd}, \text{gch cat10})))) + ((\text{sqrt}(\text{fcth cat20} + \text{cld cat10})) + (\max(\text{jcd txt1}, (\max(\text{jcd txt1}, (\text{sqrt}((\text{bic cat1} + \text{acc cat20}) * (\text{cld cat20} + \text{phog cat5}))))))))))$

(b.3): cld cat20

(b.4): $\text{sqrt}(\text{sqrt}(\text{sqrt}(\text{cedd})))$

(b.5): $\min(((\min((\min(\text{jcd} * \text{cld cat20}), \text{cedd cat10})), ((\text{cld cat20} * \text{cld cat5}) - \text{gch txt1}))) + (\text{sqrt}(\text{acc cat5})), ((\max(\text{cld cat20}, \text{cedd cat1})) * \text{acc cat10}))$

Quando usamos consultas difíceis na aprendizagem, há uma presença maior do descritor CLD em (a) e (b). Ainda encontramos com frequência o descritor CEDD, principalmente em (b), onde são usadas imagens da base (com plano de fundo uniforme) como consulta. Isso reforça a eficiência desse descritor no contexto de buscas de imagens em *e-commerce*, como foi constatado em [26].

4.3.1 Resultados para o Subconjunto Hard Set

Os cenários avaliados são:

Cenário 1 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem usar RF na aprendizagem, mantendo a consulta original durante todo o processo de RF.

Cenário 2 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem usar RF na aprendizagem, removendo a consulta original após o primeiro *feedback* do usuário, também trocando a função de *ranking* por uma função aprendida usando as consultas fáceis sem o uso de RF na aprendizagem.

Cenário 3 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI usando RF durante a aprendizagem, mantendo a consulta original durante todo o processo de RF.

Cenário 4 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI + RF usando RF durante a aprendizagem, removendo a consulta original após o primeiro *feedback* do usuário, também trocando a função de *ranking* por uma função aprendida usando as consultas fáceis e usando RF durante a aprendizagem.

Cenário 5 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem RF na aprendizagem antes da primeira interação do usuário, trocando a função para a aprendida pela Expansão-GPI com consultas fáceis e que usou RF na aprendizagem após o primeiro *feedback*, mantendo a consulta original durante todo o processo de RF.

Cenário 6 Buscas interativas usando as funções de *ranking* aprendidas com Expansão-GPI sem RF na aprendizagem antes da primeira interação do usuário, trocando a função para a aprendida pela Expansão-GPI com consultas fáceis e que usou RF na aprendizagem após o primeiro *feedback*, removendo da consulta original após o primeiro *feedback* do usuário.

No *Cenário 1*, as funções aprendidas apresentaram bons resultados quando aplicadas em um processo de busca com RF. A Figura 4.8 mostra os resultados das buscas avaliadas com diferentes métricas. Vemos que, logo na primeira interação do usuário, temos um ganho de 27.6% em $p@5$, 16.2% em $p@10$ e 9.7% em $p@20$. A Tabela 4.9 mostra os resultados dos testes estatísticos de Wilcoxon para a métrica $p@5$.

Para as 3 métricas, os valores de precisão alcançam o maior valor ao terceiro ciclo de RF, sendo estes 40.8% em $p@5$, 24.3% em $p@10$ e 14.9% em $p@20$. Podemos ver que quando analisamos um número maior de respostas, como é o caso das métricas $p@10$ e $p@20$, conseguimos valores menores, o que pode indicar que hajam poucos itens relevantes para as consultas dentro da base de imagens analisada neste experimento. Podemos

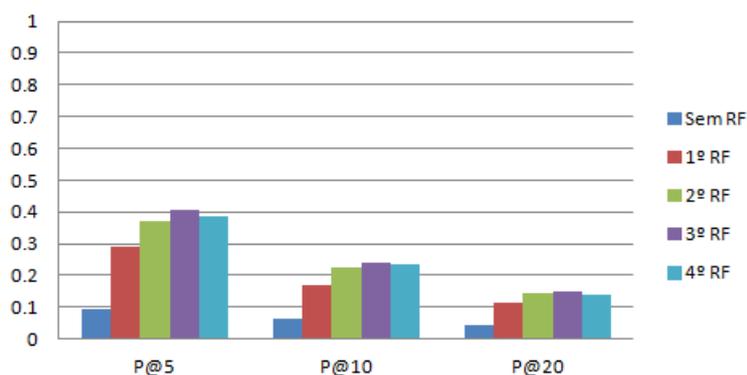


Figura 4.8: Gráficos sobre os resultados das buscas no Cenário 1 (sem usar RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas difíceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	100	100	100	100
1º RF	100	0	89	95	70
2º RF	100	89	0	67	22
3º RF	100	95	67	0	90
4º RF	100	70	22	90	0

Tabela 4.9: Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 1

analisar melhor o desempenho do método observando a concentração de respostas relevantes no topo do *rank*, como é o caso da métrica $p@5$.

Os testes estatísticos mostram que o resultados mostrado nos gráficos para o primeiro ciclo de RF é 100% confiáveis. As respostas para todos os ciclos de RF são 100% diferentes, estatisticamente, quando comparadas às respostas sem RF. A diferença estatística entre o primeiro e o segundo ciclo de RF cai para 89%, e do segundo para o terceiro ciclo cai para 67%. Isso nos mostra que o use de RF para consultas difíceis, no *Cenário 1*, é eficaz quando a interação é feita uma vez, e assim como os valores de precisão passam a aumentar menos com a adição de mais *feedback*, a diferença estatística entre as respostas passa a diminuir também.

No *Cenário 2* testamos os efeitos da remoção da consulta original durante as buscas. Vemos que o uso de RF foi eficaz no *Cenário 1*, porém os resultados passam a piorar quando removemos a consulta original. Assim como foi observado para as consultas fáceis, a remoção da consulta original para as buscas com consultas difíceis, usando Expansão-GPI sem o uso de RF durante a aprendizagem, também prejudicou o desenvolvimento das buscas. A Figura 4.9 mostra os resultados das buscas, e a Tabela 4.10 mostra

os resultados dos testes estatísticos de Wilcoxon.

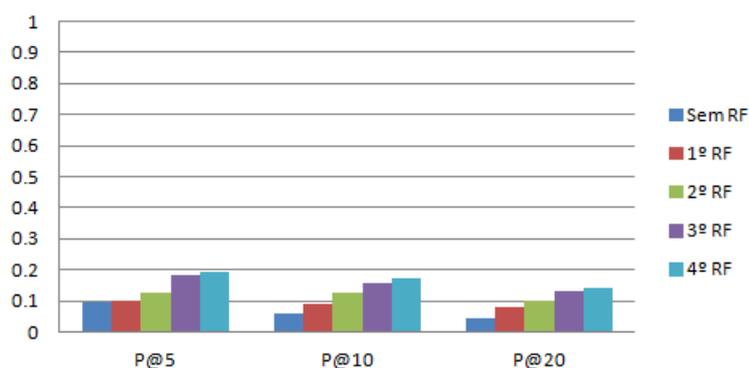


Figura 4.9: Gráficos sobre os resultados das buscas no Cenário 2 (sem usar RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas difíceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	75	36	75	76
1º RF	75	0	38	99	98
2º RF	36	38	0	97	98
3º RF	75	99	97	0	0
4º RF	76	98	98	0	0

Tabela 4.10: Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 2

Assim como ocorreu com as buscas usando as consultas fáceis, a remoção da consulta original piorou as respostas. Os valores de precisão em $p@5$, que antes da remoção eram de 28.8%, 37.2%, 40,8% e 38.8% para os 4 ciclos de RF, passaram para 10%, 12.8%, 18.4% e 19.6%. Para as métricas $p@10$ e $p@20$ os valores de precisão mantiveram-se abaixo de 20% até o quarto ciclo de RF, e nota-se que levam mais ciclos para que haja uma diferença significativa no valor de precisão comparando com os resultados nos primeiros ciclos, enquanto a diferença é notável já no segundo ciclo quando não removemos a consulta original.

Em seguida, avaliamos o *Cenário 3*. A Figura 4.10 mostra os resultados obtidos para as buscas nesse cenário, e a Tabela 4.11 mostra os resultados dos testes estatísticos de Wilcoxon.

Ao contrário do esperado, o método Expansão-GPI usando RF durante a aprendizagem não desenvolveu as buscas quando usamos RF. Os resultados ruins no primeiro ciclo mostram que o método, ao avaliar os indivíduos pela precisão das buscas somente após

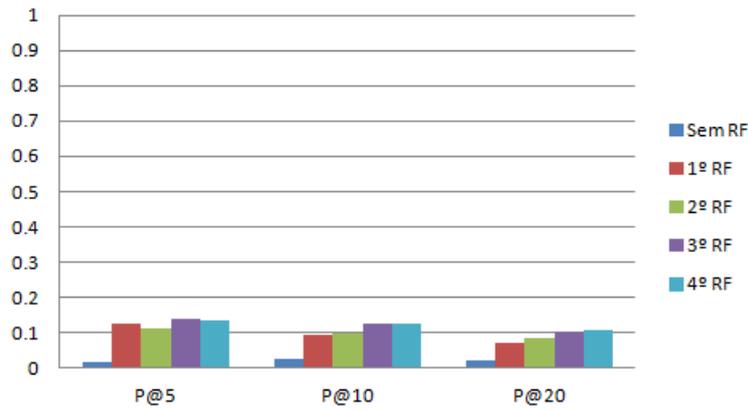


Figura 4.10: Gráficos sobre os resultados das buscas no Cenário 3 (usando RF na aprendizagem, sem remover a consulta original e sem combinar os métodos) usando consultas difíceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	98	99	100	100
1º RF	98	0	0	16	42
2º RF	99	0	0	26	54
3º RF	100	16	26	0	19
4º RF	100	42	54	19	0

Tabela 4.11: Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 3

a adição de *feedback* à consulta, fez com que o método se tornasse muito dependente do *feedback*. Isso é evidenciado pela valor muito baixo na precisão das buscas sem RF.

Os testes estatísticos confirmam a diferença estatística entre os resultados antes e após o uso de RF, com valores acima de 98% para todos os casos. No entanto, quando comparamos os resultados do primeiro com o segundo ciclo, a diferença estatística cai para 0%, e os ciclos seguintes continuam com diferenças muito pequenas até o quarto ciclo. Isso indica que o RF, nesse cenário, é capaz de melhorar as buscas, mas a adição de mais de uma imagem da base não melhora a resposta.

No *Cenário 4* avaliamos a remoção da consulta original após a inclusão de *feedback* à consulta. Para este cenário, a remoção da consulta original após o primeiro ciclo de RF foi capaz de melhorar a qualidade das respostas, embora os valores de precisão não ultrapassem os valores obtidos no *Cenário 1*. A Figura 4.11 mostra os resultados das buscas, onde é possível ver a melhoria das respostas, e a Tabela 4.12 mostra os resultados dos testes estatísticos de Wilcoxon.

Apesar da remoção da consulta original ter melhorado as respostas, o ganho na precisão não foi o suficiente para ultrapassar os valores de precisão obtidos ao usar o método

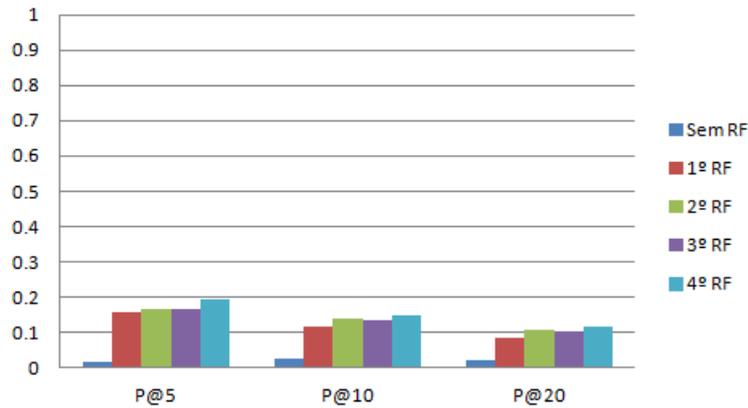


Figura 4.11: Gráficos sobre os resultados das buscas no Cenário 4 (usando RF na aprendizagem, removendo a consulta original e sem combinar os métodos) usando consultas difíceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	99	100	99	100
1º RF	99	0	31	16	64
2º RF	100	31	0	16	69
3º RF	99	16	16	0	75
4º RF	100	64	69	75	0

Tabela 4.12: Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 4

Expansão-GPI. Os maiores valores de precisão alcançados por esta abordagem foram 15.2% em $p@5$, 13.4% em $p@10$ e 11.4% em $p@20$, no quarto ciclo, enquanto a abordagem anterior obteve 13.6% em $p@5$, 12.6% em $p@10$ e 10.9% em $p@20$. No *Cenário 1* obtivemos 40.8% em $p@5$, 24.2% em $p@10$ e 14.9% em $p@20$, ainda na terceira interação, superando as demais abordagens.

A melhoria nos resultados após a remoção da consulta deixa mais evidente a possibilidade do método avaliado ter se tornado dependente do *feedback*. Isso mostra que a função de *ranking* funciona melhor ranqueando as imagens com base no *feedback* do que usando a consulta original. Os testes estatísticos comprovam a melhoria nos resultados das buscas após a remoção da consulta, com 99% de diferença estatística entre os resultados antes e após a adição de uma imagem da base e remoção da consulta original. A partir do segundo ciclo de RF não podemos concluir que houve diferença nas respostas.

No *Cenário 5* fazemos a combinação dos métodos Expansão-GPI sem RF na aprendizagem, e Expansão-GPi com RF na aprendizagem, usando o primeiro método para realizar a primeira busca, e em seguida trocando o método para trabalhar com as buscas que usar o *feedback* do usuário. Os resultados após a primeira busca obtiveram valores

de precisão semelhantes aos obtidos no *Cenário 3*, não melhorando os resultados e obtendo diferenças estatísticas pequenas. A Figura 4.12 mostra os resultados das buscas, e a Tabela 4.13 mostra os resultados dos testes estatísticos de Wilcoxon.

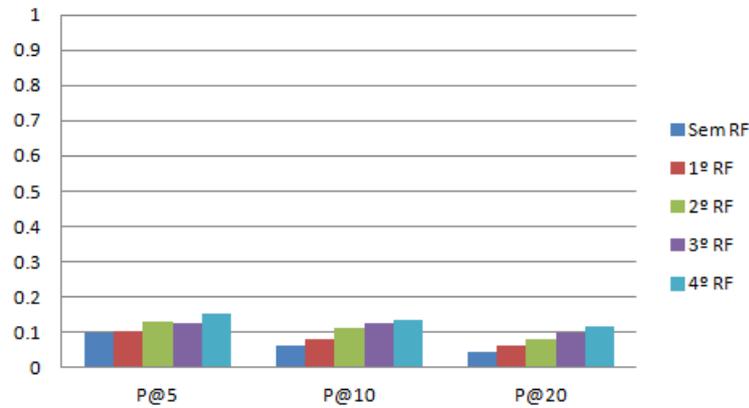


Figura 4.12: Gráficos sobre os resultados das buscas no Cenário 5 (usando RF na aprendizagem, sem remover a consulta original e combinando os métodos) usando consultas difíceis

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	36	0	5	56
1º RF	36	0	45	12	76
2º RF	0	45	0	19	62
3º RF	5	12	19	0	89
4º RF	56	76	62	89	0

Tabela 4.13: Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 5

No *Cenário 6*, removemos a consulta original enquanto usamos o mesmo contexto do *Cenário 5*. Os resultados das buscas mostram uma melhoria muito pequena na precisão das buscas, enquanto os testes estatísticos não mostram uma diferença significativa. A Figura 4.13 nos mostra os resultados das buscas, e a Tabela 4.14 mostra os resultados dos testes estatísticos de Wilcoxon.

	Sem RF	1º RF	2º RF	3º RF	4º RF
Sem RF	0	15	59	56	72
1º RF	15	0	69	53	80
2º RF	59	69	0	12	50
3º RF	56	53	12	0	56
4º RF	72	80	50	56	0

Tabela 4.14: Teste de Wilcoxon para os resultados dos experimentos usando consultas difíceis no Cenário 6

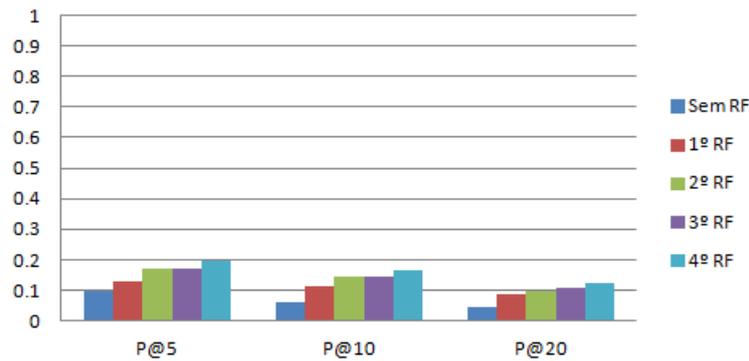


Figura 4.13: Gráficos sobre os resultados das buscas no Cenário 6 (usando RF na aprendizagem, removendo a consulta original e combinando os métodos) usando consultas difíceis

4.4 Discussão sobre os resultados

A partir dos resultados dos experimentos podemos tirar algumas conclusões sobre o uso de RF nas buscas. A diferença entre os resultados para as consultas fáceis e difíceis deixa evidente que, mesmo combinando diferentes estratégias de busca multimodal, o fundo presente nas imagens consideradas difíceis gera ruído na busca, levando a uma descrição errada da consulta (do ponto de vista do usuário) e, conseqüentemente, a aparição de imagens não relevantes na resposta. Isso é evidenciado ao compararmos os valores de precisão obtidos nos melhores cenários para os dois casos, onde para as consultas fáceis conseguimos a precisão de 58% em $p@5$, enquanto conseguimos apenas 40.8% em $p@5$ para as consultas difíceis, ambos obtidos no *Cenário 1*.

No entanto, o uso de RF (fora da aprendizagem) para as consultas difíceis se mostrou mais vantajoso que o seu uso com consultas fáceis. Em todos os cenários, para as consultas fáceis, os resultados foram estatisticamente inconclusivos, enquanto as buscas mostravam diferenças pequenas nos valores de precisão no decorrer dos processos de RF. Porém, para as consultas difíceis, os experimentos obtiveram valores altos nos testes estatísticos para os cenários onde a consulta original era mantida durante todo o processo.

Os melhores cenários foram encontrados utilizando apenas o método Expansão-GPI sem o uso de RF na aprendizagem. O método Expansão-GPI que usa RF na aprendizagem, proposto neste trabalho, não foi capaz de desenvolver as buscas durante o processo de RF. Este método torna as buscas ineficazes antes da obtenção do primeiro *feedback* do usuário, e não consegue melhorar a qualidade das respostas das buscas com o acréscimo de mais *feedback* além do primeiro.

Quando avaliamos indivíduos no GP pela qualidade das respostas apenas após a utilização do *feedback*, não temos a garantia de que o método é capaz de gerar uma primeira resposta com qualidade usando somente a consulta fornecida inicialmente. Isso é evidenciado pelos resultados ruins encontrados durante as primeiras buscas no processo de RF para as consultas difíceis, mostrados nos experimentos deste capítulo, os quais obtiveram graus de confiança estatística altos com o teste de Wilcoxon. Os resultados mostram que nas primeiras buscas, para grande parte das consultas, não existem imagens relevantes no topo do *rank*, gerando valores de precisão inferiores a 2%, enquanto este valor aumenta a partir do primeiro ciclo de RF.

Para as consultas fáceis também é percebida uma queda nos valores de precisão para as primeiras buscas, quando comparamos os método Expansão-GPI sem RF na aprendizagem e com RF na aprendizagem, embora a diferença tenha sido pequena e os testes estatísticos tenham mostrado um grau de confiança baixo nos resultados. Em geral, avaliar funções de *ranking* pela qualidade das respostas após o uso de RF em GP torna o método muito dependente do *feedback* do usuário. As funções aprendidas passam a se especializar em captar a semântica das imagens da própria base, ao invés de aprender a lidar com a consulta fornecida pelo usuário, a qual detém a verdadeira representação da necessidade de informação do mesmo. Isso faz com que, mesmo com a adição de mais imagens á consulta, não haja melhoria significativa na qualidade das respostas.

Isso pode ser evidenciado também pelo aumento dos valores de precisão nas buscas por consultas difíceis ao remover a consulta original após o primeiro ciclo de RF. Como, neste caso, a consulta original é muito diferente das imagens presentes na base, ela acaba gerando ruído no conjunto de consultas (conjunto formado com a inserção de imagens da base à consulta inicial). Já para as consultas fáceis, a remoção da consulta original diminui a precisão do método, pois neste contexto as consultas são semelhantes às imagens da base, fazendo com que o método tire proveito de ambas as imagens durante a aprendizagem do GP.

Capítulo 5

Conclusão e Trabalhos Futuros

Neste Capítulo são mostradas as conclusões tiradas sobre o resultado dos experimentos, destacando a contribuição deste trabalho e mostrando trabalhos a serem realizados no futuro.

5.1 Conclusões

Este trabalho abordou o uso de GP para aprender funções de *ranking* que, dado o contexto de comércio eletrônico, fossem capazes de combinar as similaridades obtidas por diferentes descritores de imagem e diferentes abordagens de expansão multimodal, dada uma consulta e uma imagem da base, de uma forma que fosse possível obter a melhoria das respostas com o acréscimo de imagens ao conjunto de consultas, selecionadas pelo usuário a partir do *rank* mostrado pelo sistema de busca, em um processo iterativo de *Relevance Feedback*. Este método consiste em adicionar RF as buscas dentro do GP, usando o arcabouço da Expansão-GPI.

O intuito deste trabalho foi continuar os estudos realizados em [26], referentes a buscas de imagens em *e-commerce* usando GP e expansão multimodal para aprender uma função de *ranking* eficaz, de maneira offline, para as buscas. Estudamos o efeito que a inclusão de RF durante a aprendizagem com o mesmo arcabouço teria, ao aplicar as funções aprendidas em processos de busca iterativos. Para tal, consideramos como valor de aptidão de um indivíduo em GP como sendo a precisão das buscas após uma interação de *feedback* do usuário.

Foram realizados experimentos onde avaliamos diferentes cenários aos quais o método

desenvolvido poderia ser inserido, comparando o uso do método mostrado em [26] nos mesmos cenários. Verificamos os efeitos que a aplicação destes métodos causariam em buscas interativas com RF, as diferenças nas respostas ao remover a consulta original e usar somente o *feedback* como consulta, e o que acontece ao combinar os dois métodos.

Os resultados dos experimentos mostram que, ao avaliar a aptidão de uma função de *ranking* pelo valor da precisão após a interação do usuário, o método não aprende a extrair a semântica da consulta original, ficando dependente da adição de novas imagens à consulta. As respostas das buscas antes da adição do *feedback* foram ruins, para as consultas fáceis e difíceis.

No entanto, percebemos que o método Expansão-GPI sem o uso de RF durante o aprendizado foi capaz de desenvolver as buscas com o uso de poucos ciclos de interação do usuário quando usamos consultas difíceis. Este método não foi capaz de aprender uma função de *ranking* que, sem o uso de RF, obtivesse respostas com qualidade satisfatória, porém há um aumento significativo na precisão das buscas após o primeiro ciclo de RF, aumentando ainda mais até chegar ao quarto ciclo. Neste caso, o ganho na qualidade das respostas com a aplicação de RF durante as buscas (fora do processo de aprendizagem) justifica o uso do *feedback*.

Podemos concluir que o uso de RF em buscas é uma estratégia que, independente do método de aprendizagem em GP aplicado, consegue melhorar a qualidade de buscas de imagens em *e-commerce* quando são usadas consultas difíceis. Embora essa estratégia seja eficaz, a aplicação de RF em processos de GP, avaliando os indivíduos apenas pela precisão das buscas após o uso do *feedback* do usuário, torna a função de *ranking* muito dependente do *feedback*, deixando-a deficiente quando a obtenção da semântica da consulta fornecida pelo usuário. Percebemos também que a melhora nos resultados das buscas é mais evidente quando a consulta contém informações não relacionadas à necessidade de informação do usuário, como acontece com as consultas difíceis (onde o fundo da imagem contém informações irrelevantes). Neste caso o *feedback* é capaz de confirmar as características da imagens que são relevantes em meio ao ruído causado pelo plano de fundo. Para as consultas fáceis não há uma melhoria tão grande nas respostas, pois devido a semelhança das consultas com as imagens da base, o *feedback* acaba se tornando redundante.

5.2 Trabalhos Futuros

Este trabalho abordou buscas de imagens em *e-commerce*, usando GP com RF para aprender funções de *ranking* eficazes em buscas interativas. No entanto, existem diversas formas de explorar este paradigma. Além disso, é possível utilizar outras técnicas de aprendizagem de máquina, além do GP, para melhorar buscas de imagem, as quais não foram exploradas neste trabalho. Sabendo disso, são listadas algumas lacunas a serem exploradas em trabalhos futuros:

1. Considerar o resultado das buscas de cada descritor isolado utilizando RF como sendo um terminal no processo do GP. Dado um indivíduo como uma função de *ranking*, cada terminal seria a similaridade de uma consulta composta pela imagem original e uma outra selecionada por RF, usando uma abordagem de expansão multimodal específica, ou buscas sem expansão com cada descritor de imagem. Juntamente a estes terminais estariam os terminais usados neste trabalho. As imagens relevantes seriam selecionadas após uma primeira busca usando alguma função de *ranking* que não usasse os terminais que dependem de RF.
2. Desenvolver um classificador que identifique quando uma consulta é "fácil" ou "difícil", para que seja usada a função de *ranking* correta nas buscas.
3. Estudar uma forma de remover o fundo das imagens submetidas como consulta, submetendo-as a uma nova aprendizagem do GP com RF.
4. Testar outras abordagens para a função de aptidão do GP, que levem em consideração não só a precisão das buscas no último ciclo de RF, mas que pontue também a precisão do primeiro *rank*, e as diferenças de precisão entre um ciclo e o seguinte.
5. Aplicar algoritmos de agrupamento nas imagens da base, submetendo as consultas ao mesmo algoritmo para reduzir o escopo da busca.
6. Usar classificação para descobrir a categoria da consulta, e dividir a base de imagens por categoria, aprendendo uma função de *ranking* diferente para cada categoria, que seja especializada na mesma, e em seguida comparar com as buscas sem esta divisão.

Referências Bibliográficas

- [1] BAEZA-YATES, R., AND RIBEIRO-NETO, B. Modern information retrieval: the concepts and technology behind search, harlow, 2011.
- [2] BOSCH, A., ZISSERMAN, A., AND MUNOZ, X. Representing shape with a spatial pyramid kernel. *ACM CIVR* (2007), 401–408.
- [3] CALUMBY, R. T., DA SILVA TORRES, R., AND GONÇALVES, M. A. Multimodal retrieval with relevance feedback based on genetic programming. *Multimedia Tools and Applications* 69, 3 (June 2012), 991–1019.
- [4] CHATZICHRISTOFIS, S., AND BOUTALIS, Y. Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. *ICVS* (2008), 312–322.
- [5] CHATZICHRISTOFIS, S., AND BOUTALIS, Y. Fcth: Fuzzy color and texture histogram—a low level feature for accurate image retrieval. *WIAMIS* (2008), 191–196.
- [6] CHATZICHRISTOFIS, S., BOUTALIS, Y., AND LUX, M. Selection of the proper compact composite descriptor for improving content based image retrieval. *6th IASTED International Conference* (2009).
- [7] CHOWDHURY, M., DAS, S., AND KUNDU, M. K. Novel cbir system based on ripplelet transform using interactive neuro-fuzzy technique. *Electronic Letters on Computer Vision and Image Analysis* 11, 1 (2012), 1–13.
- [8] CIOCCA, G., CUSANO, C., SANTINI, S., AND SCHETTINI, R. Halfway through the semantic gap: Prosemanic features for image retrieval. *Information Sciences* 181, 22 (Nov. 2011), 4943–4958.

- [9] DA SILVA, A. T., FALCÃO, A. X., AND MAGALHÃES, L. P. Active learning paradigms for CBIR systems based on optimum-path forest classification. *Pattern Recognition* 44, 12 (Dec. 2011), 2971–2978.
- [10] DEPEURSINGE, A., AND MÜLLER, H. Fusion techniques for combining textual and visual information retrieval. *ImageCLEF (2010)* (2010), 95–114.
- [11] DOS SANTOS, J., FERREIRA, C., TORRES, R. D. S., GONÇALVES, M., AND LAMPARELLI, R. A relevance feedback method based on genetic programming for classification of remote sensing images. *Information Sciences* 181, 13 (July 2011), 2671–2684.
- [12] DUAN, L., LI, W., TSANG, I. W.-H., AND XU, D. Improving Web image search by bag-based reranking. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 20, 11 (Nov. 2011), 3280–90.
- [13] HE, R., XIONG, N., YANG, L. T., AND PARK, J. H. Using Multi-Modal Semantic Association Rules to fuse keywords and visual features automatically for Web image retrieval. *Information Fusion* 12, 3 (July 2011), 223–230.
- [14] HUANG, J., KUMAR, S., MITRA, M., ZHU, W., AND ZABIH, R. Image indexing using color correlograms. *IEEE Computer Vision and Pattern Recognition* (2001), 762–768.
- [15] JI, X. Research on the Relevance Feedback-based Image Retrieval in E-commerce. *Advances in information Sciences and Service Sciences(AISS)* 3, November (2011), 452–460.
- [16] JONES, S., SHAO, L., ZHANG, J., AND LIU, Y. Relevance feedback for real-world human action retrieval. *Pattern Recognition Letters* 33, 4 (Mar. 2012), 446–452.
- [17] KASUTANI, E., AND YAMADA, A. The mpeg-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. *ICIP* (2001), 674–677.

- [18] KO, B. C., LEE, J., AND NAM, J.-Y. Automatic medical image annotation and keyword-based image retrieval using relevance feedback. *Journal of digital imaging* 25, 4 (Aug. 2012), 454–65.
- [19] KOZA, J. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.
- [20] LUX, M. Content based image retrieval with lire. *ACM MM* (2011), 735–738.
- [21] MARAKAKIS, A., SIOLAS, G., GALATSANOS, N., LIKAS, A., AND STAFYLOPATIS, A. Relevance feedback approach for image retrieval combining support vector machines and adapted Gaussian mixture models. *IET Image Processing* 5, 6 (2011), 531.
- [22] MCGILL, M., AND SALTON, G. *Introduction to modern information retrieval*, 1983.
- [23] NIBLACK, W., BARBER, R., EQUITZ, W., FLICKNER, M., GLASMAN, E., PETKOVIC, D., YANKER, P., FALOUTSOS, C., AND TAUBIN, G. The qbic project: Querying images by content, using color, texture, and shape. *Storage and Retrieval for Image and Video Databases (SPIE)* (1993).
- [24] QI-MING FU*, QUAN LIU, X.-Y. W., AND ZHANG, L. Relevance feedback techniques and genetic algorithm for image retrieval based on multiple features. *Int. J. Modelling, Identification and Control, Vol. 14, No. 4* 14, 4 (2011), 2011.
- [25] SANTOS, J. Re-ranking de busca visual de produtos usando informaÃ§Ã£o multimodal. Mestrado, Icomp, Universidade Federal do Amazonas, UFAM, 2013.
- [26] SARAIVA, P. *ProgramaÃ§Ã£o Genética Aplicada à Busca de Imagens*. Doutorado, Icomp, Universidade Federal do Amazonas, UFAM, 2014.
- [27] SARAIVA, P. C., CAVALCANTI, J. A. M. B., GONÇALVES, M. A., DOS SANTOS, K. C. L., DE MOURA, E. S., AND TORRES, R. D. S. Evaluation of parameters for combining multiple textual sources of evidence for Web image retrieval using genetic programming. *Journal of the Brazilian Computer Society* 19, 2 (Sept. 2012), 147–160.

- [28] SCLAROFF, S., TAYCHER, L., AND CASCIA, M. Imagerover: A content-based image browser for the world wide web. *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries* (1997).
- [29] SMEULDERS, A., WORRING, M., SANTINI, S., GUPTA, A., AND JAIN, R. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2000).
- [30] STEHLING, R., NASCIMENTO, M., AND FALCÃO, A. A compact and efficient image retrieval approach based on border/interior pixel classification. *ACM CIKM* (2002), 102–109.
- [31] SU, J.-H., HUANG, W.-J., YU, P. S., AND TSENG, V. S. Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns. *IEEE Transactions on Knowledge and Data Engineering* 23, 3 (Mar. 2011), 360–372.