UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Rio Claro

João Gabriel Camacho Presotto

Métodos de Aprendizado de Máquina Fracamente Supervisionados Baseados em Ranqueamento

Rio Claro - SP

2021

João Gabriel Camacho Presotto

Métodos de Aprendizado de Máquina Fracamente Supervisionados Baseados em Ranqueamento

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Financiadora: FAPESP - Proc. 2019/04754-6

Orientador: Prof. Dr. Daniel Carlos Guimarães Pedronette

Rio Claro - SP

2021

P934m

Presotto, João Gabriel Camacho
Métodos de aprendizado de máquina fracamente supervisionados baseados em ranqueamento / João Gabriel Camacho Presotto. -- Rio Claro, 2021
85 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Geociências e Ciências Exatas, Rio Claro
Orientador: Daniel Carlos Guimarães Pedronette

1. Ciência da Computação. 2. Recuperação de Imagens por Conteúdo. 3. Aprendizado de Máquina. 4. Aprendizado Fracamente Supervisionado. 5. Modelo de Ranqueamento. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Geociências e Ciências Exatas, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

João Gabriel Camacho Presotto

Métodos de Aprendizado de Máquina Fracamente Supervisionados Baseados em Ranqueamento

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Financiadora: FAPESP - Proc. 2019/04754-6

Comissão Examinadora

- Prof. Dr. Daniel Carlos Guimarães Pedronette (Orientador)
  Departamento de Estatística, Matemática Aplicada e Computação
  Universidade Estadual Paulista - UNESP
- Prof. Dr. Moacir Antonelli Ponti
  Instituto de Ciências Matemáticas e de Computação
  Universidade de São Paulo - USP
- Prof. Dr. Fabricio Aparecido Breve
  Departamento de Estatística, Matemática Aplicada e Computação
  Universidade Estadual Paulista - UNESP

Resultado: Aprovado

Rio Claro - SP

27 de Agosto de 2021

Agradecimentos

Aos meus pais e irmã pelo apoio e todo o suporte.

Ao meu orientador pelos anos de parceria desde minha graduação.

Agradeço à FAPESP pela concessão da bolsa de pesquisa, sob o processo nº 2019/04754-6, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Resumo

Apesar dos impressionantes avanços recentes nas técnicas de aprendizado de máquina, principalmente na compreensão de dados multimídia, desafios significativos ainda persistem. Um dos principais desafios em cenários reais apresenta-se na escassa disponibilidade de dados rotulados. Nesse contexto, desenvolver métodos capazes de explorar as informações presentes em dados não rotulados de modo a mitigar os problemas associados à insuficiência de dados rotulados é um desafio de suma importância. Métodos de aprendizado fracamente supervisionado conseguem lidar com tais restrições ao trabalhar com rótulos estimados ou aproximados como maneira de potencializar informações úteis de treinamento. Nessa linha de pesquisa, apresentaremos dois métodos de aprendizado fracamente supervisionado capazes de analisar as relações entre os dados rotulados e não rotulados, de modo a expandir pequenos conjuntos de treinamento rotulados. Ambos recorrem a um modelo de ranqueamento e utilizam diferentes estratégias para analisar as informações de similaridade codificadas nos dados não rotulados e identificar fortes relações de similaridade com os dados rotulados. Tais relações são consideradas durante a etapa de expansão do conjunto de treinamento. Os métodos foram avaliados em conjunto com diferentes classificadores supervisionados e semi-supervisionados, incluindo uma recente rede convolucional baseada em grafos. Foram considerados cinco diferentes coleções de imagens públicas e os vetores de características de cada imagem foram obtidos através de diferentes descritores visuais. Ganhos positivos de acurácia foram obtidos por ambos os métodos nos mais diferentes cenários quando comparados aos classificadores treinados sem o auxílio de nossos métodos e a técnicas de expansão similares, evidenciando a robustez das abordagens propostas.

Palavras-chave: Aprendizado de Máquina, Aprendizado Fracamente Supervisionado, Aprendizado Semi-Supervisionado, Classificação, Hipergrafos, Modelo de Ranqueamento, Métricas de Correlação de Listas Ranqueadas, Recuperação de Imagens Baseada em Conteúdo

Abstract

Despite the impressive recent advances in machine learning techniques, especially in multimedia data understanding, significant challenges remain. One of the main challenges in real-world scenarios is the limited availability of labeled data. In this context, developing methods capable of exploiting the information encoded in the unlabeled data to mitigate the problems associated with insufficient labeled data, and to overcome this issue is something of paramount importance. Weakly supervised learning methods are capable to handle such restrictions by working with estimated or approximate labels as a way to maximize useful training information. In this line of research, we will present two weakly supervised methods that can analyze the relationships between labeled and unlabeled data to expand small labeled training sets. Both use a ranking model and different strategies to examine similarity information encoded in the unlabeled data to identify strong similarity relationships with the labeled data. Such relations will be considered during the training set expansion step. The methods were evaluated in conjunction with different supervised and semi-supervised classifiers, including a recent graph convolutional network. Five different public image datasets were considered with different visual descriptors. Positive accuracy gains were achieved by both methods in the different scenarios when compared to classifiers trained without the aid of our methods and compared to similar expansion techniques, evidencing the strength of both.

Keywords: Classification, Content-based Image Retrieval, Hipergraphs Machine Learning, Rank Correlation Measures, Ranking, Semi-Supervised Learning, Weakly Supervised Learning

Lista de ilustrações

Lista de tabelas

Lista de símbolos

α	taxa de expansão de rótulos utilizada na expansão baseada em hipergrafos.
$\mathbf{c_i}$	ranque de correlações.
$\mathbf{C}$	matriz de produtos cartesianos dos elementos do hipergrafo.
$\mathbf{H}$	matriz de incidências do hipergrafo.
$\mathbf{H}^T$	matriz de incidências transposta do hipergrafo.
$\mathbf{S}$	matriz de similaridade do hipergrafo.
$\mathbf{W}$	matriz de similaridade ou afinidade final do hipergrafo.
$\mathcal{C}_p$	conjunto de pares de candidatos utilizados na expansão baseada em hipergrafos.
$\mathcal{D}$	descritor visual.
$\mathcal{N}(q,k)$	conjunto dos primeiros k vizinhos de um determinado item de dados x_q.
$\mathcal{N}_c(x_l,k_e)$	conjunto de candidatos não rotulados utilizado na expansão baseada em hipergrafos.
$\mathcal{N}_h$	conjunto de vizinhança do hipergrafo.
$\mathcal{T}$	conjunto de listas ranqueadas obtido a partir das listas de cada item de dados em $\mathfrak{X}$.
$\mathcal{T}$	conjunto de listas ranqueadas.
$\mathfrak{L}$	conjunto de rótulos para cada item de dados.
$\mathfrak{X}$	coleção de dados.
$\mathfrak{X}_E$	conjunto rotulado estimado.
$\mathfrak{X}_L$	conjunto rotulado.
$\mathfrak{X}_U$	conjunto não rotulado.
$\mathfrak{X}_W$	conjunto rotulado expandido.

$\rho(v_{x_i}, v_{x_j})$	distância entre os vetores de características de dois itens de dados x_i e x_j.
τ_C	ranque de pares de candidatos utilizados na expansão baseada em hipergrafos.
τ_q	lista ranqueada do item de índice q.
$E(x_e)$	função de estimativa de rótulos.
G	hipergrafo com V vértices e E hiperarestas.
l	comprimento da lista ranqueada.
$r(\cdot, \cdot)$	métrica de correlação entre listas ranqueadas.
$s_L(a)$	*spline* linear de interpolação.
t_p	limiar de posição (ou posição limite) utilizado na expansão baseada em hipergrafos.
t_h	limiar de correlação.
v_{x_i}	vetor de características de um item de dados x_i.
$w(e_i)$	peso da hiperaresta e_i.
x_q	item de dados de índice q.
$Y(x_i)$	função que associa o item de dados x_i com seu rótulo correspondente.

Sumário

1 Introdução

Atualmente uma grande e crescente quantidade de conteúdo digital multimídia está disponível em diversas aplicações e domínios. Em um cenário em que usuários comuns evoluíram de meros consumidores para produtores ativos dessa categoria de conteúdo, um volume expressivo de imagens e vídeos tem sido gerados diariamente [2]. Por conta disso, tarefas para compreensão de imagens se tornaram um tópico de grande interesse, tanto no meio acadêmico quanto na indústria [3]. Diversos avanços tem sido realizados e técnicas de aprendizado de máquina modernas podem atualmente igualar o desempenho humano em tarefas como classificação de imagens [4].

Mesmo com o desenvolvimento substancial nesse domínio, que contribuiu em diversas aplicações como classificação e recuperação de imagens e vídeos, ainda existem desafios relevantes a serem explorados. De fato, grande parte dos resultados positivos tem sido obtidos a custo de grandes quantidades de dados de treinamento anotados ou rotulados [4]. Redes neurais convolucionais (*Convolutional Neural Networks* (CNNs)) são uma das principais técnicas que apresentam um expressivo sucesso, em especial em tarefas de aprendizado supervisionado aplicadas a visão computacional, mas que dependem de grandes quantidades de dados de treinamento [5]. Entretanto, obter grandes conjuntos de amostras de treinamento confiáveis é uma tarefa muito custosa. Considerando dados de vídeo, por exemplo, apesar da quantidade teoricamente ilimitada de vídeos disponíveis na rede, categorizar e processar esses vídeos de modo a criar um conjunto de dados útil é uma tarefa difícil e suscetível a erros, já que os rótulos associados a essa categoria de mídia retiradas de redes sociais são muitas vezes ruidosos [3]. Visto que rotular dados multimídia é normalmente custoso, demorado e suscetível a erros, muitos esforços têm sido feitos para contornar este problema [6].

O aprendizado fracamente supervisionado é uma área ativa de pesquisa [7, 8, 9, 10], que visa enfrentar desafios relacionados a disponibilidade de dados de treinamento. Seja trabalhar com rótulos não ideais, com rótulos que definem uma imagem na totalidade (*image-level labels*) ao invés de anotações de cada objeto pertencente a imagem (*object-level labels*) [7, 8, 9] ou com rótulos possivelmente ruidosos obtidos a partir de redes sociais [11], o aprendizado fracamente supervisionado representa uma série de métodos capazes de lidar com um nível de supervisão em que a informação de rótulos disponível é uma estimativa ou aproximação se comparado aos ideais, mas que ainda conseguem obter resultados expressivos. Essa categoria de métodos em diversas situações consideram tanto os dados rotulados quanto os dados não rotulados, podendo ser também caracterizados como métodos semi-supervisionados. Deste modo, este trabalho discute métodos capazes de estimar rótulos de elementos não rotulados a partir de suas relações com os elementos

rotulados em cenários onde o conjunto de dados possui escassez de amostras rotuladas e amostras não rotuladas em abundância.

Em tarefas de classificação, obter resultados de acurácia positivos quando apenas pequenos conjuntos rotulados estão disponíveis é uma tarefa desafiadora que vem sendo investigada por décadas [12, 13, 14, 15]. Métodos de aprendizado semi-supervisionado, os quais tentam explorar de maneira efetiva os dados rotulados e não rotulados, são um tópico amplamente estudado [6, 13, 14, 15]. Muitos desses métodos semi-supervisionados exploram os dados não rotulados para aprimorar a acurácia do modelo, conseguindo produzir resultados relevantes em diversas tarefas, incluindo classificação de imagens [16, 17]. Nos últimos anos, essa área de pesquisa tem atraído bastante atenção [6, 16, 3, 2, 18], principalmente por conta do crescimento expressivo de coleções multimídia e a consequente dificuldade de rotular esses dados, grandes conjuntos de treinamento necessários às CNNs e aos recentes avanços em técnicas não supervisionadas [19, 20].

Dado o contexto, que destaca a relevância de abordagens capazes de trabalhar com escassez de dados rotulados, o objetivo do trabalho é investigar abordagens fracamente supervisionadas, em especial considerando os avanços em estratégias não supervisionadas baseadas em ranqueamento [19, 20, 4] e como tais abordagens poderiam ser exploradas em cenários próximos. Foi realizado um levantamento da literatura e feita uma investigação sobre novos métodos dessa categoria. Dessa forma, as principais contribuições desse trabalho consistem na proposta de dois métodos de aprendizado fracamente supervisionado aplicados a cenários em que apenas pequenos conjuntos rotulados estão disponíveis. Ambos recorrem a um modelo de ranqueamento e diferentes técnicas para expandir os conjuntos rotulados. Modelos de ranqueamento (ou de listas ranqueadas) permitem representar informações de similaridade entre elementos, as quais podem ser exploradas para atribuir rótulos estimados aos elementos não rotulados. Estes que têm sido explorados com sucesso por algoritmos de aprendizado de variedades (*manifold learning*) em cenários de recuperação não supervisionada [19, 20]. Estatísticas de ranques (*rank statistics*) [4] têm sido também utilizadas para descobrir novas classes em um conjunto de dados a partir de elementos rotulados de outras classes.

O primeiro método, analisa informações de similaridade entre elementos rotulados e não rotulados através da análise de métricas de correlação de listas ranqueadas [21], capazes de atribuir um valor real a relação entre as listas ranqueadas de duas imagens. As informações de similaridade entre amostras são representadas através de um modelo de listas ranqueadas e estas são processadas por recentes métodos de aprendizado de variedades [19, 20] (métodos de re-ranqueamento). Tais algoritmos conseguem calcular medidas de similaridade globais considerando informações contextuais implícitas entre as amostras. Como resultado, é possível aprimorar a eficácia das listas ranqueadas (mais elementos da mesma classe a imagem de consulta nas primeiras posições), o que por

consequência resultará em cálculos de correlação mais eficazes. Com base nas listas ranqueadas mais eficazes, métricas de correlação de listas ranqueadas são exploradas com o intuito de identificar fortes relações de similaridade entre imagens. É possível atribuir classes aos dados não rotulados com base em tal informação, expandindo o conjunto de treinamento rotulado. Uma abordagem adaptativa capaz de definir automaticamente o limiar de correlação também é proposta. Limiar utilizado para selecionar quais amostras não rotuladas vão fazer parte do conjunto rotulado expandido. Este método é uma extensão do trabalho realizado em [22], inovando em diversos aspectos, incluindo o uso de métodos de re-ranqueamento e a seleção automática do limiar de correlação utilizado na etapa de expansão de rótulos.

O segundo método de aprendizado fracamente supervisionado proposto e que é apresentado ao longo do trabalho explora informações codificadas no modelo de ranqueamento através de uma estrutura de hipergrafo construído por um algoritmo não supervisionado de aprendizado de variedades (*manifold learning*) baseado em ranqueamento [19]. Hipergrafos são uma poderosa generalização de grafos. Enquanto as arestas em grafos representam ligações entre pares de vértices, hiperarestas permitem representar relações entre conjuntos de elementos. As estruturas do hipergrafo construído a partir das informações dos ranques são então utilizadas para identificar e selecionar relações de similaridade sólidas entre amostras rotuladas e não rotuladas e tais relações são exploradas de modo a construir um conjunto rotulado expandido.

O conjunto rotulado expandido gerado por ambos os métodos de aprendizado fracamente supervisionado pode ser utilizado tanto por classificadores supervisionados como para classificadores semi-supervisionados obtendo ganhos expressivos em termos da acurácia final, mesmo começando com poucas amostras rotuladas, se comparados com o treinamento sem utilizar dos métodos de expansão.

Considerando métodos que tentam aprimorar tarefas de classificação [15, 23], ambos os métodos apresentados ao longo do texto não dependem de um único classificador, pois apresentam uma formulação flexível que pode ser utilizada em conjunto com diferentes classificadores. De forma análoga, outros métodos exploram pseudo-rótulos [24, 25, 1, 26, 27] como uma estratégia de aumentar a quantidade de dados rotulados através de redes neurais. Novamente, os métodos propostos são uma alternativa mais versátil, pois independem de um modelo como as redes neurais para gerar mais amostrar rotuladas.

Uma ampla avaliação experimental foi conduzida com o intuito de avaliar ambos os métodos nos mais diferentes cenários. Foram consideradas cinco coleções de imagens públicas e diferentes descritores visuais, incluindo um descritor baseado em CNNs pré treinado em cenários de aprendizado de transferência (*transfer learning*). O método baseado em correlações entre listas ranqueadas foi avaliado considerando cinco métricas de correlação. A estratégia de expansão de conjuntos rotulados foi avaliada em conjunto com

diferentes classificadores supervisionados e semi-supervisionados, incluindo uma recente rede convolucional de grafos (*Graph Convolutional Network* (GCN)). Todos os experimentos foram conduzidos em um cenário de aprendizado fracamente supervisionado, considerando conjuntos de treinamento rotulados de tamanho limitado. Os resultados experimentais de ambos os métodos demonstram ganhos de acurácia significativos quando comparados ao treinamento dos classificadores sem os métodos de expansão e em comparação com métodos análogos da literatura.

O restante do trabalho está dividido da seguinte maneira. O Capítulo 2 define os aspectos teóricos fundamentais e os trabalhos relacionados ao contexto do trabalho. O Capítulo 3 apresenta o método de aprendizado fracamente supervisionado baseado em correlações. O Capítulo 4 descreve o método de aprendizado fracamente supervisionado baseado em hipergrafos. O Capítulo 5 descreve a avaliação experimental realizada para avaliar ambos os métodos propostos. Por fim, o Capítulo 6 discute as conclusões obtidas ao longo do trabalho, as publicações que ambos os métodos originaram e possíveis trabalhos futuros.

2 Aspectos Teóricos e Trabalhos Relacionados

Este capítulo discute os principais aspectos teóricos e trabalhos relacionados aos métodos de aprendizado fracamente supervisionado baseados em ranqueamento que são apresentados ao longo do texto. Na Seção 2.1 são apresentados os principais conceitos da recuperação de imagens baseada em conteúdo, e quais aspectos desta categoria de métodos estão relacionados ao trabalho. Na Seção 2.2 é apresentada uma definição formal do modelo de ranqueamento que é utilizado como principal fonte de informações de ambos os métodos de aprendizado fracamente supervisionado propostos. A Seção 2.3 apresenta o conceito de aprendizado de máquina e suas categorias que se relacionam com o trabalho. Por fim, a Seção 2.4 discute a origem do termo aprendizado fracamente supervisionado e apresenta a uma definição formal do mesmo além de discutir algumas abordagens recentes que apresentam um funcionamento similar aos métodos apresentados nos próximos capítulos.

2.1 Recuperação de Imagens Baseada em Conteúdo

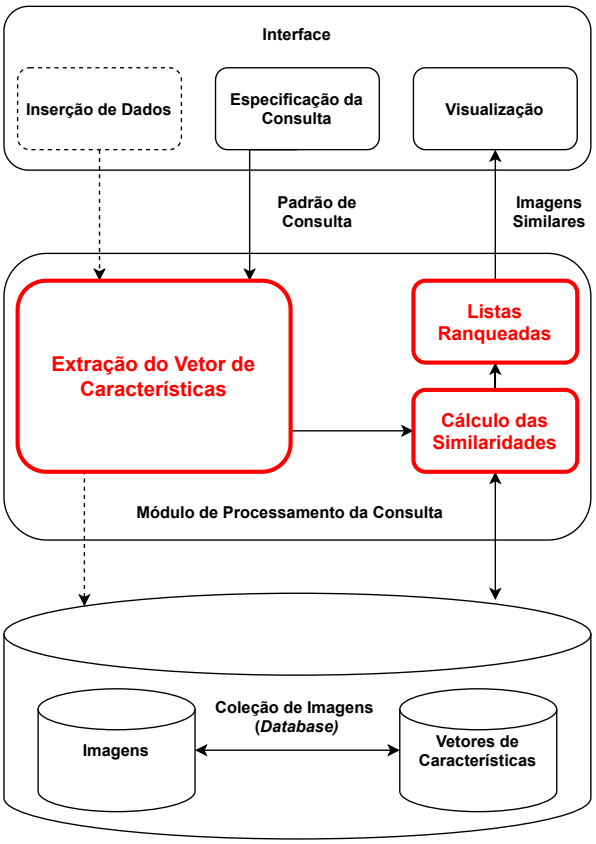
Devido ao grande crescimento das coleções de imagens, é cada vez mais necessário encontrar maneiras eficazes e eficientes para realizar a busca e indexação dessa categoria de dados [28]. Os primeiros sistemas de recuperação de imagens realizavam buscas considerando palavras-chave e metadados [29], porém, essa metodologia é bastante limitada, pois ignora o conteúdo visual disponível e pode produzir resultados de buscas ambíguos. Como solução a essa situação surge a recuperação de imagens baseada em conteúdo (CBIR - *Content-Based Image Retrieval*) [28].

A recuperação de imagens baseada em conteúdo pode ser definida como qualquer tecnologia que ajude a organizar arquivos de imagens por seu conteúdo visual, e por essa definição, qualquer aplicação desde uma função de similaridade entre imagens até um robusto mecanismo de anotação se enquadram no âmbito de CBIR [28]. O principal objetivo da recuperação baseada em conteúdo é encontrar imagens similares, e com este objetivo são utilizados os descritores visuais, capazes de representar as imagens como vetores numéricos de características. Com os vetores de características gerados é possível então comparar as imagens através de uma função de distância, como a distância Euclidiana (a mais utilizada [30]), e dessas comparações uma lista ranqueada em ordem crescente de distâncias é gerada. Para extração de características das imagens, os descritores visuais podem utilizar de diferentes estratégias como analisar a cor, a forma, a textura das imagens [31, 32] ou até se utilizar de aprendizado profundo, principalmente as redes

neurais convolucionais.

A Figura 1 apresenta a implementação clássica de recuperação de imagens baseada em conteúdo. Esta pode ser dividida em três partes: (i) a interface, (ii) o módulo de processamento da consulta e (iii) a base de dados. A interface é responsável pela interação com o usuário, com a inserção e saída dos dados. O módulo de processamento da consulta tem como tarefa extrair as características das imagens e fazer as operações necessárias de modo a compará-las. Por fim, a base de dados armazena as imagens e seus respectivos vetores de características, de modo a evitar a recalculá-los. Estão destacadas em vermelho as etapas relacionadas que são utilizadas durante a execução de nosso trabalho, onde o principal componente que dita o funcionamento dos métodos de aprendizado fracamente supervisionados propostos são as listas ranqueadas, geradas a partir do cálculo de distâncias entre os vetores de características das amostras.

Figura 1 – Arquitetura típica de recuperação de imagens baseada em conteúdo.



Fonte: Adaptado de [33].

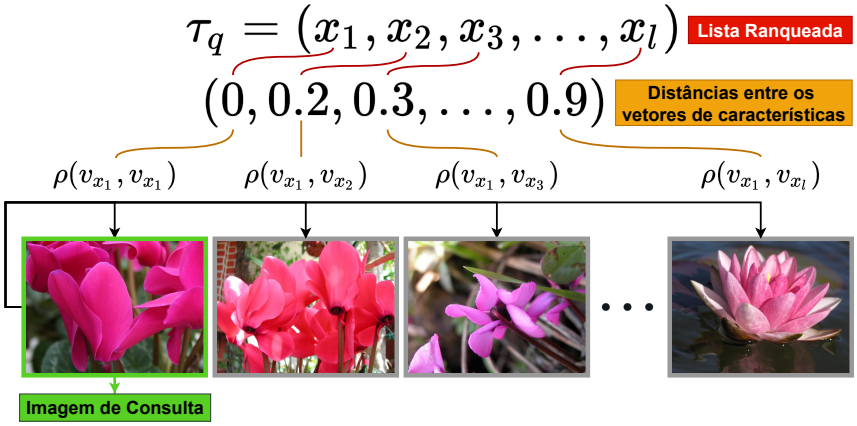
2.2 Modelo de Ranqueamento

Os métodos apresentados ao longo desse trabalho se utilizam de um modelo de ranqueamento, definido a partir de uma coleção de dados $\mathfrak{X}$. Cada item de dados $x_i \in \mathfrak{X}$

pode ser representado em um espaço de d dimensões por um vetor de características v_{x_i}, obtido a partir de um descritor $\mathcal{D}$, que pode ser definido como uma tupla (ϵ, ρ) [34], onde ϵ: $\{x_i\} \rightarrow \mathbb{R}^d$ é uma função que extrai o vetor de características v_{x_i} e ρ: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$ é uma função que calcula a distância entre dois itens de dados a partir de seus vetores de características. Sendo assim, a distância entre dois itens de dados x_i e x_j pode ser obtida por $\rho(v_{x_i}, v_{x_j})$.

Para cada item de dados x_q, uma lista ranqueada τ_q pode ser calculada a partir da função de distâncias ρ, que contêm os itens de dados da coleção $\mathfrak{X}$ mais similares a x_q. Formalmente, a lista ranqueada τ_q=$(x_1, x_2, \ldots, x_l)$ pode ser definida como uma permutação da coleção $\mathfrak{X}_l$, onde l representa o comprimento das listas ranqueadas e $\mathfrak{X}_l \subset \mathfrak{X}$ é um subconjunto que contêm os l itens de dados mais similares a x_q. A posição de x_i na lista ranqueada τ_q é indicada por $\tau_q(x_i)$. Portanto, podemos dizer se x_i está em uma posição anterior a x_j na lista ranqueada de x_q, então $\tau_q(i) < \tau_q(j)$ e $\rho(v_q, v_i) \leq \rho(v_q, v_j)$. A Figura 2 ilustra o processo de construção de uma lista ranqueada τ_q a partir dos valores de distância $\rho(\cdot, \cdot)$ entre os vetores de características dos itens de dados.

Figura 2 – Ilustração do processo de construção das listas ranqueadas a partir do cálculo de similaridade entre as amostras.



Fonte: Elaborado pelo autor.

É possível calcular uma lista ranqueada τ_i para cada item de dados $x_i \in \mathfrak{X}$ de modo a obter um conjunto de listas ranqueadas $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_N\}$. O conjunto $\mathcal{T}$ representa uma rica fonte de informações que é explorada ao longo desse trabalho.

2.3 Aprendizado de Máquina

O aprendizado de máquina é uma área que envolve uma série de algoritmos capazes de realizar tarefas sem serem diretamente programados e de se adaptar a diferentes circunstâncias. Algoritmos de aprendizado de máquina funcionam de modo a analisar um determinado conjunto de dados e extrair padrões e informações, tomando decisões

baseadas nos dados, sem haver programação explícita para a resolução de uma determinada tarefa [35]. Esta categoria de algoritmo surge para contornar situações em que algoritmos convencionais não conseguem lidar [36], como reconhecer faces em uma fotografia, converter fala em texto escrito, dirigir um carro, tarefas que são naturais aos humanos, mas que não tem uma explicação concreta de como ser feitas.

Com os avanços nas tecnologias, é cada vez mais fácil armazenar e processar grandes quantidades de dados, assim como acessá-los de locais distantes através da rede. Nesses cenários, métodos de aprendizado de máquina tem se mostrando bastante relevantes, capazes de detectar certos padrões e regularidades, facilitando o entendimento do contexto em que os dados estão inseridos ou permitindo fazer previsões [36]. Estes que também são capazes de se adaptar aos mais diferentes cenários, sem ser necessário programá-los para toda situação possível [36]. Atualmente, algoritmos de aprendizado de máquina tem sido bastante utilizados nos mais diferentes cenários, com destaque nas áreas de reconhecimento de padrões e visão computacional [37, 38, 39, 40].

Considerando os métodos de aprendizado de máquina relevantes ao nosso trabalho, é possível destacar três categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semi-supervisionado. Tais categorias são discutidas em mais detalhes a seguir.

2.3.1 Aprendizado Supervisionado e Não Supervisionado

Métodos de aprendizado supervisionados são treinados a partir da utilização de dados rotulados e tomam decisões considerando dados não rotulados [35]. Amostras rotuladas possuem informação de classe ou categoria associada, enquanto os dados não rotulados não tem essa informação disponível. A partir de um conjunto de dados rotulados, o objetivo dessa categoria de método é construir um modelo capaz de estimar os rótulos para novas amostras desconhecidas.

Um tipo comum de método de aprendizado supervisionado são os classificadores supervisionados, como redes neurais [41] e máquinas de vetores de suporte (SVM - *Support Vector Machines*) [42], ambos que tentam encontrar um hiperplano que melhor separa os dados considerados relevantes dos não relevantes. Podemos também destacar abordagens baseadas em programação genética [43, 44], grafos [45] e redes neurais profundas [46].

Apesar de populares, os métodos de aprendizado supervisionados não são adequados em todos os cenários. Situação que ocorre por conta da sensibilidade dessa categoria de método aos dados de treinamento, pois se a quantidade de dados rotulados disponível for muito pequena, o modelo pode não obter um nível de generalização adequado. Outro problema que pode ocorrer é o sobreajuste (*overfitting*), em que o modelo se especializa em somente uma tarefa, não conseguindo classificar adequadamente novos dados [35].

Outra limitação apresentada por modelos supervisionados está relacionada ao tempo de treinamento, pois se a quantidade de dados de treinamento for muito elevada, esse tempo pode ser muito alto, tornando certas tarefas inviáveis.

Os métodos de aprendizado não supervisionado não dependem de intervenção do usuário ou de informação de classe das amostras. Por conta disso, essa categoria de método tenta identificar estruturas relevantes a partir do conjunto de dados (normalmente representados como vetores de valores numéricos). Nesse contexto podemos destacar os métodos de agrupamento (*clustering*), em que as instâncias são separadas em grupos a partir da análise de similaridade entre as amostras [47]. No contexto de recuperação de imagens, algoritmos de aprendizado não supervisionado tem entre os principais objetivos substituir medidas de distâncias par a par por medidas globais e mais eficazes, realizando uma análise da informação do contexto em que as amostras estão inseridas [48]. Dentre as diversas abordagens de aprendizado não supervisionado podemos destacar: processos de difusão [49, 50], grafos [51, 48], agrupamento [52], frequência de padrões [53] e análises de listas ranqueadas [54].

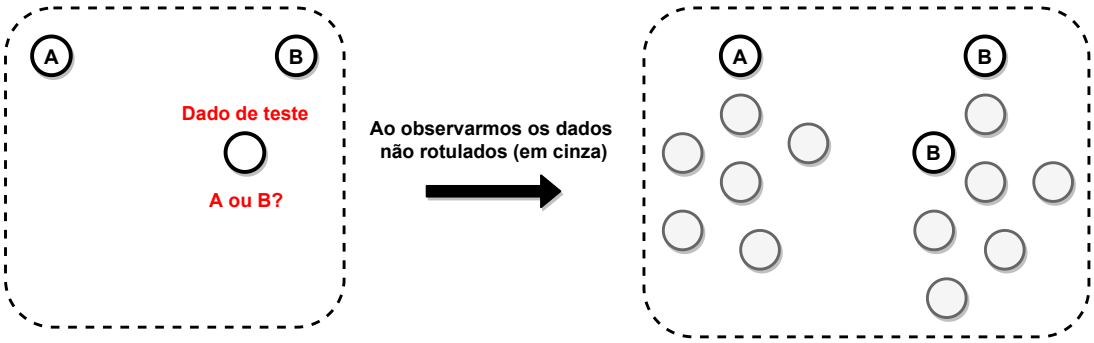
Os métodos de aprendizado não supervisionado baseados em listas ranqueadas (ou modelo de ranqueamento) têm atraído interesse expressivo na literatura [55, 56, 19, 20]. Situação que pode ser justificada pelo baixo custo computacional requerido, pois estes métodos não necessitam das listas ranqueadas calculadas por completo (somente até um tamanho $l \ll |\mathfrak{X}|$), já que os resultados relevantes se concentram nas primeiras posições dos ranques. Dentre métodos baseados em ranqueamento que exploram diferentes estratégias, pode-se destacar: *Correlation Graph* [57], CPRR [58], *Ranked List Graph Distance* [59], ReckNNGraph [60], RL-Recom [61] e RL-Sim [54, 62].

2.3.2 Aprendizado Semi-Supervisionado

O aprendizado semi-supervisionado é uma vertente de aprendizado de máquina que tenta combinar as ideias utilizadas tanto no aprendizado supervisionado quanto no aprendizado não supervisionado [13, 63]. A título de exemplo, em um problema de classificação, pontos de dados não rotulados podem ser utilizados para auxiliar no processo de classificação. Tal exemplo é ilustrado na Figura 3, em que é possível prever o rótulo **B** da amostra com um certo grau de confiança ao observar as amostras não rotuladas em cinza, que formam dois conjuntos distintos. Os métodos de classificação semi-supervisionados se mostram bastante relevantes por conseguirem trabalhar em cenários em que a quantidade de dados rotulados é muito pequena. Em cenários análogos, a tarefa de construir um classificador supervisionado confiável baseado em um número reduzido de amostras rotuladas é uma tarefa muito difícil.

Podemos dividir os métodos de aprendizado semi-supervisionados em duas categorias principais: métodos indutivos e transdutivos. Métodos indutivos, assim como os

Figura 3 – Exemplificação de como dados não rotulados podem ajudar no processo de classificação.



Fonte: Adaptado de [10], FIG. 3.

métodos de aprendizado supervisionados produzem um modelo de classificação capaz de prever rótulos de amostras previamente desconhecidas. Situação similar ao objetivo dos métodos supervisionados [64]: construir um modelo durante a etapa de treinamento e utilizá-lo para prever os rótulos de novas amostras. Os métodos transdutivos não produzem um modelo de classificação, mas conseguem fazer previsões dos rótulos diretamente, limitando-se às amostras que encontram durante a processo de treinamento [64]. Por conta disso, se novas amostras forem adicionadas a um conjunto de dados, um modelo transdutivo teria de ser novamente executado para considerar essas novas amostras. Nessa categoria, a informação é diretamente propagada via conexões diretas entre os pontos de dados. Situação que pode ser naturalmente representada por grafos, onde pontos similares são conectados e os rótulos podem ser propagados ao longo das arestas. Na prática, grande parte dos métodos transdutivos são diretamente modelados por grafos ou podem ser implicitamente entendidos como [64].

De maneira geral, métodos semi-supervisionados consideram uma série de hipóteses a respeito da distribuição dos dados em seu funcionamento [13]. Estas que podem são consideradas individualmente ou em conjunto, a depender do método [64]. A primeira das hipóteses, a hipótese da suavidade (*smoothness assumption*) considera que amostras próximas no espaço de características tem uma alta probabilidade de pertencerem a mesma classe. A hipótese da baixa densidade (*low-density assumption*) considera que a fronteira de decisão não deve ser colocada em regiões de alta densidade de amostras. Outra hipótese utilizada é de que amostras próximas entre si em uma representação de menor dimensionalidade devem possuir o mesmo rótulo (*manifold assumption*). Considerando o cenário de agrupamento (*clustering*), podemos também destacar a hipótese de agrupamentos (*cluster assumption*), em que pontos de dados pertencentes ao mesmo grupo pertencem a mesma classe.

2.4 Aprendizado Fracamente Supervisionado

O termo aprendizado fracamente supervisionado (*weakly supervised learning*) é utilizado na literatura para representar diferentes situações. Em [7, 8, 9] o termo é utilizado para representar o uso de rótulos não ideais para suas aplicações, onde são utilizados rótulos que definem a imagem como um todo (*image-level labels*) ao invés de anotações de cada objeto que compõe as mesmas. O trabalho realizado em [11] apresenta uma diferente abordagem de aprendizado de transferência (*transfer learning*), em que os modelos são previamente treinados utilizando de bilhões de imagens "rotuladas" por etiquetas (*hashtags*) em redes sociais. Em termos práticos, essas etiquetas podem conter uma quantidade excessiva de ruído ou até gerar uma distribuição com um certo viés, o que pode prejudicar a tarefa de aprendizado de transferência [11]. Contudo, tal expectativa não se concretizou, e experimentos indicaram que o uso desses rótulos "fracos" se mostrou muito promissor. Além disso, em [10] o termo é definido como uma série de métodos de aprendizado de máquina capazes de trabalhar em cenários de supervisão fraca, em que rótulos completos e verdadeiros (*fully ground-truth labels*) não estão sempre disponíveis ou obtê-los manualmente tem um custo muito alto.

Considerando as situações apresentadas, pode-se destacar como característica comum o desafio de lidar com um nível de supervisão em que os rótulos utilizados não são ideais ou são "fracos". Cada uma das estratégias utilizam de rótulos que se comparado aos esperados são uma estimativa ou aproximação, mas que mesmo assim conseguem atingir resultados relevantes.

No contexto deste trabalho, são apresentados métodos capazes de estimar rótulos de elementos não rotulados a partir das suas relações com os elementos rotulados, criando assim um conjunto rotulado expandido. O uso do termo aprendizado fracamente supervisionado advém do fato de que o conjunto rotulado expandido tem rótulos estimados, que podem ser considerados "fracos". Sendo assim, o conjunto rotulado expandido ou fracamente rotulado, é composto da união dos rótulos reais, já disponíveis no conjunto de dados inicial, e dos rótulos "fracos".

Ao longo do texto são apresentados dois métodos de aprendizado fracamente supervisionado baseados em ranqueamento. Ambos funcionam a partir da análise da estrutura do conjunto de dados, explorando as informações de similaridade codificadas nas listas ranqueadas. Com diferentes estratégias, mas com a premissa de utilização do modelo de ranqueamento, os métodos tentam estabelecer relações entre os elementos rotulados e os não rotulados, de modo transmitir rótulos ao conjunto não rotulado quando uma forte relação de similaridade for identificada. Dessa forma, os métodos propostos também podem ser vistos como semi-supervisionados, visto que utilizam dados rotulados e não rotulados e são adequados para cenários em que há escassez de dados para treinamento.

Ao transmitir (ou expandir) os rótulos dos elementos rotulados aos não rotulados, um conjunto rotulado expandido é gerado. O conjunto expandido é utilizado para o treinamento de classificadores supervisionados ou semi-supervisionados, com o intuito de melhorar a acurácia obtida quando comparados com o treinamento sem a utilização do conjunto expandido. Os métodos apresentados analisam todo o conjunto de dados para criação das estruturas relacionadas e expansão de rótulos. Por conta disso, caso uma nova amostra seja adicionada, é necessário recalcular as listas ranqueadas e o restante das estruturas relacionadas a cada um dos métodos. Situação que representa um método semi-supervisionado transdutivo, em que se realizam previsões dos rótulos diretamente e limitando-se às amostras iniciais. Porém, caso os métodos de expansão sejam utilizados com seu objetivo principal, aprimorar o treinamento e resultados dos classificadores, esses classificadores podem ser utilizados em amostras desconhecidas, caracterizando um cenário indutivo.

2.4.1 Definição Formal

Esta seção apresenta uma definição formal do problema considerado. Seja $\mathfrak{X} = \{x_1, x_2, \ldots, x_L, x_{L+1}, \ldots, x_N\}$ uma coleção de dados onde cada elemento x_i representa um item de dados. A coleção $\mathfrak{X}$ pode ser definida como uma coleção parcialmente rotulada, i.e., $\mathfrak{X} = \mathfrak{X}_L \cup \mathfrak{X}_U$, onde $\mathfrak{X}_L = \{x_i\}_{i=1}^{L}$ representa um subconjunto de itens rotulados e $\mathfrak{X}_U = \{x_i\}_{i=L+1}^{N}$ o subconjunto não rotulado. Normalmente, a quantidade disponível de dados rotulados é muito menor do que a quantidade de dados não rotulados, ou seja, $|\mathfrak{X}_L| \ll |\mathfrak{X}_U|$.

Podemos também definir formalmente os classificadores utilizados em conjunto com os métodos de expansão. Seja, $\mathfrak{L} = \{1, \ldots, C\}$ um conjunto que contêm os rótulos do conjunto de dados. Seja $Y : \mathfrak{X} \to \mathfrak{L}$ uma função que associa cada item de dados $x_i \in \mathfrak{X}$ a seu rótulo correspondente $Y(x_i)$. Portanto, a tarefa de classificação realizada pode ser definida como a estimativa da função $Y(x_i)$ para cada item não rotulado $x_i \in \mathfrak{X}_U$ a partir do modelo aprendido.

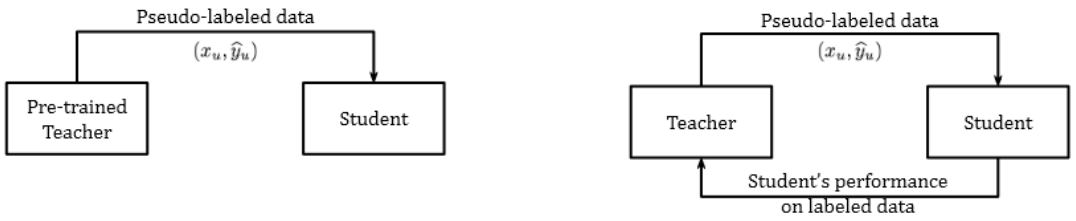
O principal objetivo desse trabalho é aprender um subconjunto dos dados não rotulados que pode ser agregado a um conjunto rotulado expandido a partir de uma estimativa da função Y. Formalmente, seja $\mathfrak{X}_E \subseteq \mathfrak{X}_U$ um conjunto rotulado estimado. O conjunto fracamente rotulado ou expandido $\mathfrak{X}_W$ é definido como $\mathfrak{X}_W = \mathfrak{X}_L \cup \mathfrak{X}_E$. Para cada $x_e \in \mathfrak{X}_E$, uma função de estimativa $E(x_e)$ é utilizada como estimativa da função $Y(x_e)$. Posteriormente, o conjunto rotulado expandido $\mathfrak{X}_W$ é utilizado para treinamento dos classificadores. Sendo assim, a questão central que irá ser discutida ao longo do trabalho é como definir o subconjunto $\mathfrak{X}_E$ e a função de estimativa de rótulos $E(x_e)$, de modo a permitir bons resultados ao combiná-los com os classificadores.

2.4.2 Abordagens Recentes

A tentativa de atribuir rótulos aos dados não rotulados de modo a construir um conjunto rotulado expandido é algo amplamente estudado na literatura [65, 66, 1, 26, 27]. As técnicas de auto-rotulagem (*self-labeled techniques*) [65], uma categoria de classificadores semi-supervisionados, consideraram que suas previsões de rótulos são em maior parte verdadeiras e assim conseguem produzir um conjunto rotulado expandido. Mais recentemente, uma nova categoria de método de auto-rotulagem foi proposto, em [66] o conceito de pseudo-rótulos foi introduzido. O trabalho utiliza de uma rede neural profunda que é treinada utilizando de amostras rotuladas e de amostras pseudo-rotuladas, geradas em cada atualização de pesos da rede, ao analisar a classe que teve a maior probabilidade prevista pelo modelo. O conceito de pseudo-rótulos foi originalmente utilizado através do uso de redes neurais profundas, mas este pode ser utilizado com qualquer rede ou classificador que permita obter para cada amostra informações de probabilidade dessa pertencer a cada amostra.

O conceito de pseudo-rótulos tem sido amplamente estudado [1, 26, 27], tentando minimizar cenários em que o método que gera os pseudo-rótulos está mal otimizado, o que pode gerar a treinamentos com amostras ruidosas, prejudicando os resultados. Em [1] o conceito de "meta" pseudo-rótulos é proposto, onde o modelo que gera os rótulos é adaptado conforme o desempenho obtido no conjunto rotulado ao se treinar com os "meta" pseudo-rótulos. A Figura 4 apresenta a diferença entre o conceito de pseudo-rótulos e os "meta" pseudo-rótulos. No primeiro um método previamente treinado gera os dados pseudo-rótulados utilizados para treinar o estudante (*student*). Já os "meta" pseudo-rótulos o professor (*teacher*) é treinado em conjunto do estudante, a partir da performance obtida por este no conjunto rotulado.

Figura 4 – Diferença entre a implementação dos pseudo-rótulos e dos "meta" pseudo-rótulos apresentada em [1].



Fonte: Retirado de [1], Fig. 1

O trabalho realizado em [26] aprimora o aprendizado por pseudo-rótulos no contexto de segmentação de imagens ao trabalhar também com o grau de incerteza que o modelo tem a respeito das previsões que faz. O método gera esse grau de incerteza a partir da análise da variância da previsão. De maneira similar, em [27] se utiliza do grau de incerteza

das previsões para selecionar uma maior quantidade de amostras corretas ao conjunto expandido que é utilizado na etapa de treinamento.

Considerando os métodos propostos ao longo do trabalho, esta é a primeira vez na literatura em que modelos de ranqueamento são explorados com o intuito de expandir conjuntos rotulados de treinamento e aprimorar tarefas de classificação. Além disso, ambos os métodos possuem uma formulação flexível, em que a etapa de expansão independe do classificador, permitindo sua combinação com diferentes classificadores nos mais diferentes cenários.

3 Aprendizado Fracamente Supervisionado Baseado em Correlações

Explorar efetivamente as relações entre os dados rotulados e os dados não-rotulados é um desafio crucial presente nos métodos fracamente supervisionados. Em diversos métodos de aprendizado de máquina, visão computacional e tarefas de recuperação, as amostras de dados são comumente representadas como pontos em um espaço multidimensional. Em tal cenário, estimar a similaridade entre pontos de dados se mantêm como um ponto central de pesquisa, mesmo para representações obtidas por métodos recentes de aprendizado profundo. Métricas de distância tradicionais como a distância Euclidiana são frequentemente utilizadas, mas por considerar apenas a análise de pares acabam limitando avaliações de similaridade em dados mais complexos.

Em contrapartida, listas ranqueadas fornecem uma representação contextual inerente, estabelecendo relações entre todos os elementos dos dados. Sendo assim, este capítulo apresentará um método capaz de tirar proveito dessa representação e aproveitar informações de similaridade entre dados rotulados e não rotulados. Sendo assim, a ideia central que vai ser explorada pelo método proposto ao longo desse capítulo é:

(i) A informação de contexto em que os dados estão inseridos presente nas listas ranqueadas pode ser analisada através de métricas de correlação de modo a identificar fortes relações de similaridade entre os dados;

(ii) Identificadas as amostras com uma alta similaridade, estas relações podem ser utilizadas para expandir pequenos conjuntos rotulados de treinamento, permitindo um treinamento e classificação mais efetivos por classificadores supervisionados ou semi-supervisionados;

(iii) Se tratando de listas ranqueadas, é considerada a utilização de métodos de re-ranqueamento, capazes de gerar listas mais efetivas (com mais amostras de mesma classe nas posições iniciais), permitindo gerar informações de similaridade com mais confiança a partir da análise das métricas de correlação.

Como discutido ao longo das Seções 2.1 e 2.2, as listas ranqueadas são geradas a partir dos vetores de características extraídos por um descritor visual. Por conta disso, a qualidade das listas ranqueadas está fortemente associada a qualidade dos vetores de características. Considerando a qualidade das listas ranqueadas, o método de aprendizado fracamento supervisionado baseado em correlações também depende da qualidade dos

vetores de características. Assim, a análise entre amostras que será realizada a partir das métricas de correlação depende de um espaço de características capaz de representar as classes de maneira minimamente adequada. Além disso, a análise das listas ranqueadas tem de ser limitadas a uma constante l, de modo a manter o método escalável e adequado para conjuntos de dados com muitas amostras. Apesar das limitações, por conta da informação contextual que pode ser explorada por meio das listas ranqueadas, o método consegue rotular instâncias incertas, onde há uma alta sobreposição com outras classes, situação em que normalmente esse tipo de amostra é desconsiderado.

A Figura 5 apresenta o método fracamente supervisionado baseado em correlações proposto, dividido em suas principais etapas desde a preparação dos dados até a classificação. Primeiro são extraídas as características de cada imagem de modo a obter uma representação para cada imagem. Em seguida, o método proposto passa por cinco etapas principais:

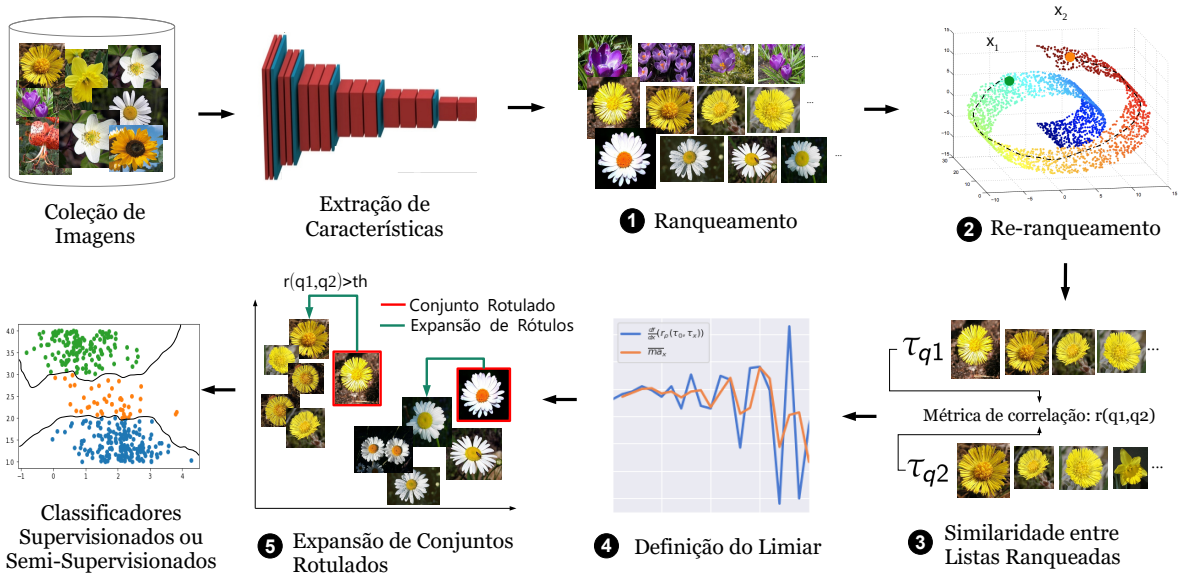
1. **Ranqueamento:** listas ranqueadas são calculadas para cada uma das imagens com base nas distâncias entre os vetores de características extraídos [1];
2. **Re-Ranqueamento:** um algoritmo de re-ranqueamento explora informações de similaridade entre as imagens de modo a calcular distâncias mais globais e por consequência gerar listas ranqueadas mais efetivas. Essa etapa pode ser opcional, uma vez que as listas ranqueadas originais podem ser utilizadas na próxima etapa;
3. **Similaridade entre Listas Ranqueadas:** a similaridade entre as imagens é avaliada por meio de métricas de correlação de ranqueamento;
4. **Definição de Limiar:** a variação das métricas de correlação nas primeiras posições das listas ranqueadas é explorada de modo a definir um limiar que é utilizado na etapa de expansão;
5. **Expansão de Conjuntos Rotulados:** relações de similaridade acima do limiar estimado na etapa anterior são selecionadas de modo a expandir os conjuntos rotulados de treinamento[2].

Por fim, o conjunto de treino ampliado é utilizado para treinamento de classificadores supervisionados ou semi-supervisionados. Cada etapa do método proposto é formalmente definida e discutida nas seções a seguir. O método proposto é flexível e independe tanto do extrator de características quanto do classificador. Os métodos selecionados para

[1] É possível também obter as listas ranqueadas através da utilização de métodos de indexação como, por exemplo, os métodos *BallTree* [67] e *K-DTree* [68].

[2] O item **5** na Figura 5 ilustra a expansão de conjuntos rotulados em um cenário de classificação binária. Em verde as setas indicam uma alta relação de similaridade encontrada entre as imagens através da análise das correlações.

Figura 5 – Diagrama de funcionamento do método de aprendizado fracamente supervisionado baseado em correlações.



Fonte: Elaborado pelo autor.

essas tarefas nesse trabalho são discutidos durante a avaliação experimental, onde são considerados métodos tradicionais e mais recentes.

3.1 Aprendizado de Variedades (*Manifold Learning*)

Medir a similaridade entre objetos de forma precisa e eficaz é um ponto fundamental em muitas tarefas de ranqueamento e aprendizado de máquina. Entretanto, a mais comum comparação entre pares de vetores de características, muitas vezes baseada em funções de distância como a distância Euclidiana, ignora os complexos arranjos de similaridade e as informações estruturais do conjunto de dados. De modo a contornar tais limitações, métodos de aprendizado de variedades (*manifold learning*) tem sido propostos. Estes trabalham com medidas mais globais, capazes de considerar a estrutura dos conjuntos de dados e fornecer medidas de similaridade mais eficazes. Mais recentemente, métodos de aprendizado de variedades baseados em ranques tem apresentado avanços consideráveis [69, 19, 20].

Formalmente, o objetivo principal dos métodos de aprendizado de variedades baseados em ranques é explorar informações de similaridade codificadas no conjunto de listas ranqueadas $\mathcal{T}$, capaz de capturar de uma melhor maneira a estrutura dos dados. Através de tal análise, um novo e mais eficaz conjunto de listas ranqueadas $\mathcal{T}_m$ é calculado de maneira não supervisionada de modo a melhorar a eficácia de métodos que trabalham com ranques. Portanto, podemos descrever os métodos de aprendizado de variedades como

uma função f_m:

$$\mathcal{T}_m = f_m(\mathcal{T}). \tag{3.1}$$

Por gerar um novo conjunto de listas ranqueadas, ao longo do texto, os referir aos métodos de aprendizado de variedades baseados em ranques são referidos como métodos de re-ranqueamento. A abordagem proposta é flexível e permite o uso de diversos métodos de re-ranqueamento que atendam à definição apresentada [19, 20, 57, 58, 59, 54, 61]. Esse trabalho considera dois métodos de aprendizado de variedades baseados em ranques para gerar a função f_m: LHRR (*Log-based Hypergraph of Ranking References*) [19] e BFS-Tree *Manifold Learning* [20]. A escolha desses métodos deve-se ao fato de que ambos foram propostos recentemente e apresentam resultados de alta eficácia em diferentes conjuntos de dados. Tais métodos são discutidos brevemente a seguir.

- **LHRR (*Log-based Hypergraph of Ranking References*) [19]:** explora a informação similaridade presente nas listas ranqueadas por meio de um hipergrafo. Hipergrafos são uma poderosa generalização dos grafos, a qual permite a criação de hiperarestas capazes de conectar um número qualquer de vértices e por consequência representar similaridades de ordem superiores entre conjuntos de objetos.

  O hipergrafo e suas hiperarestas são construídos com base nas referências entre os ranques, e os pesos das hiperarestas são atribuídos com base em uma função baseada em logaritmos. A abordagem usa das hiperarestas para construir uma representação dos dados que considera o contexto e explora das informações codificadas nesse contexto para gerar uma maneira mais eficaz de similaridade. A similaridade entre objetos é então calculada como um produto das similaridades entre suas respectivas hiperarestas. A função de similaridade é então utilizada para construir um novo e mais eficaz conjunto de listas ranqueadas.

- **BFS-Tree *Manifold Learning* [20]:** usa das informações de similaridade codificadas nas referências entre listas ranqueadas através de uma estrutura de árvore de busca em largura. A estrutura de árvore fornece uma representação hierárquica das primeiras k posições das listas ranqueadas, codificando relações de vizinhança de primeira e segunda ordem obtidas com as referências entre as listas ranqueadas. Os pesos das arestas são atribuídos aos elementos da árvore representam a similaridade sendo calculados com base em métricas de correlação entre listas ranqueadas.

  Assim que construída, a estrutura de árvore é explorada de modo a descobrir relações de similaridade ocultas entre os dados. Cada elemento é representado com base no caminho do mesmo até raiz da árvore junto dos pesos das arestas. Novas conexões são então estabelecidas entre os nós folha, estas que não são inicialmente feitas considerando somente as referências entre os ranques. Tais conexões permitem a descoberta de novas relações de similaridade.

Além dessas novas conexões de similaridade, a estrutura de árvore permite analisar a frequência com que cada elemento aparece. De maneira geral, um indicativo sólido de similaridade pode ser obtido através da coocorrência de elementos em diferentes níveis da estrutura de árvore, enquanto poucas aparições são indicativo de ruído. O algoritmo constrói uma árvore de busca em largura para cada item de dados do conjunto. Por fim, uma métrica de similaridade mais global e eficaz é calculada usando a informação de similaridade extraída de cada uma das árvores.

3.2 Métricas de Correlação de Listas Ranqueadas

Com base em um conjunto de listas ranqueadas, uma similaridade contextual entre as amostras de dados pode ser calculada. Nesse cenário, uma questão fundamental consiste em como calcular as métricas de correlação entre listas ranqueadas. Métricas de correlação podem ser definidas como uma função r: $\mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}^+$ que compara duas listas ranqueadas e atribui um valor real a essa comparação.

Essa comparação é feita considerando as primeiras k posições das listas ranqueadas. A notação $\mathcal{N}(q,k)$ é utilizada ao longo da seção para se referir aos primeiros k vizinhos de um item de dados x_q. O conjunto $\mathcal{N}(q,k)$ pode ser definido formalmente baseado na informação presente na lista ranqueada τ_q:

$$\mathcal{N}(q,k) = \{\mathfrak{X}_k \subset \mathfrak{X}, \mid \exists x_i \in \mathfrak{X}_k \wedge \exists x_j \in \mathfrak{X} \backslash \mathfrak{X}_k \wedge \tau_q(i) < \tau_q(j) \wedge |\mathfrak{X}_k| = k\}. \tag{3.2}$$

São discutidas a seguir diferentes maneiras de se calcular uma métrica de correlação $r(\cdot,\cdot)$. Foram consideradas métricas de abordagens distintas como a interseção nas primeiras k posições (RBO [70] e *Intersection* [71]), a distância entre as posições (*Kendall*τ [71]) e operações entre conjuntos (*Jaccard* [72] e *Jaccard*$_k$ [73]).

- ***Intersection* [71]:** calcula a quantidade de sobreposição entre duas listas ranqueadas τ_i e τ_j de tamanho k, considerando as sobreposições nas profundidades de 1 a k. Para cada profundidade $d \in \{1 \ldots k\}$, a quantidade de sobreposição é calculada e a média de todas as sobreposições é computada, gerando uma métrica de similaridade. A métrica *Intersection* atribui um maior peso as primeiras posições das listas ranqueadas, já que estas são consideradas mais vezes durante o cálculo da métrica. A Equação 3.3 apresenta uma definição formal da métrica:

$$r(\tau_i, \tau_j, k) = \frac{\sum_{d=1}^{k} \mid \mathcal{N}(i,d) \cap \mathcal{N}(j,d) \mid}{k}. \tag{3.3}$$

- ***Jaccard* [72]:** o coeficiente de *Jaccard* [72] é calculado através da análise de dois conjuntos não vazios de tamanho k. O coeficiente calcula a probabilidade de que um

elemento de pelo menos um dos dois conjuntos é elemento de ambos. Este é definido a seguir:

$$r(\tau_i, \tau_j, k) = \frac{|\mathcal{N}(i,k) \cap \mathcal{N}(j,k)|}{|\mathcal{N}(i,k) \cup \mathcal{N}(j,k)|}. \tag{3.4}$$

- ***Jaccard***$_k$ **[73]:** o coeficiente de *Jaccard* [72] considera somente uma profundidade k de uma lista ranqueada, portanto ignora a informação fornecida pelas primeiras posições das listas ranqueadas com profundidade menor do que k. Surge então a métrica *Jaccard*$_k$ [73], sendo um índice de *Jaccard* [72] acumulado que considera profundidades de 1 a k, atribuindo pesos maiores às primeiras posições:

$$r(\tau_i, \tau_j, k) = \frac{1}{k} \sum_{d=1}^{k} \frac{|\mathcal{N}(i,d) \cap \mathcal{N}(j,d)|}{|\mathcal{N}(i,d) \cup \mathcal{N}(j,d)|}. \tag{3.5}$$

- ***Kendall*τ [71]:** tradicional métrica de distância entre permutações, é calculada com base no número de trocas necessárias em uma ordenação por flutuação (*bubble sort*) para converter uma permutação na outra [71]. A métrica é *Kendall*τ é definida a seguir:

$$r(\tau_i, \tau_j, k) = \frac{(k \cdot (k-1))}{2 \cdot \sum_{x,y \in \mathcal{N}(i,k) \cup \mathcal{N}(j,k)} \bar{K}_{x,y}(\tau_i, \tau_j)}, \tag{3.6}$$

onde $\bar{K}_{x,y}(\tau_i, \tau_j)$ é uma função que determina se duas imagens x e y estão na mesma ordem nas primeiras k posições das listas ranqueadas τ_i e τ_j, e pode ser definida como:

$$\bar{K}_{x,y}(\tau_i, \tau_j) = \begin{cases} 0 & \text{se } (\tau_i(x) \leqslant \tau_i(y) \wedge \tau_j(x) \leqslant \tau_j(y)), \\ 0 & \text{se } (\tau_i(x) \geqslant \tau_i(y) \wedge \tau_j(x) \geqslant \tau_j(y)), \\ 1 & \text{caso contrário.} \end{cases}$$

- **RBO (*Rank-Biased Overlap*) [70]:** assim como a métrica *Intersection* [71], também considera as sobreposições de duas listas ranqueadas em diferentes profundidades. Entretanto, a métrica RBO [70] se utiliza de um parâmetro que determina a probabilidade de considerar a sobreposição na próxima profundidade, e é definida a seguir:

$$r(\tau_i, \tau_j, k) = (1 - p_r) \sum_{d=1}^{k} p_r^{(d-1)} \frac{|\mathcal{N}(i,d) \cap \mathcal{N}(j,d)|}{d}, \tag{3.7}$$

onde $p_r \in [0, 1]$ é utilizada para calcular a probabilidade em uma profundidade d.

3.3 Expansão de Conjuntos Rotulados por Correlações

Espera-se que valores altos de correlação sejam atribuídos a listas ranqueadas similares, valores superiores ao limiar definido para tarefa de expansão. A hipótese principal explorada pelo nosso trabalho é de que valores altos de métricas de correlação indicam fortes

relações de similaridade, informação que pode ser utilizada para expandir os conjuntos rotulados.

Formalmente, seja $x_l \in \mathfrak{X}_L$ um item rotulado e seja $x_u \in \mathfrak{X}_U$ um item não rotulado. Portanto, se a métrica de correlação entre τ_l e τ_u ultrapassa um determinado limiar t_h apenas a elementos rotulados de mesma classe, o conjunto rotulado é expandido de modo a incorporar x_u. Podemos definir o conjunto rotulado estimado $\mathfrak{X}_E$ como:

$$\begin{aligned}\mathfrak{X}_E = \{x_e : x_e \in \mathfrak{X}_U, x_l \in \mathfrak{X}_L \wedge r(\tau_l, \tau_e) \geq t_h \wedge \\ \nexists x_z \in \mathfrak{X}_L | r(\tau_z, \tau_e) \geq t_h \wedge Y(x_l) \neq Y(x_z)\}.\end{aligned} \tag{3.8}$$

Sendo assim, classe estimada do elemento x_e é a mesma que o item rotulado x_l, logo:

$$E_C(x_e) = Y(x_l). \tag{3.9}$$

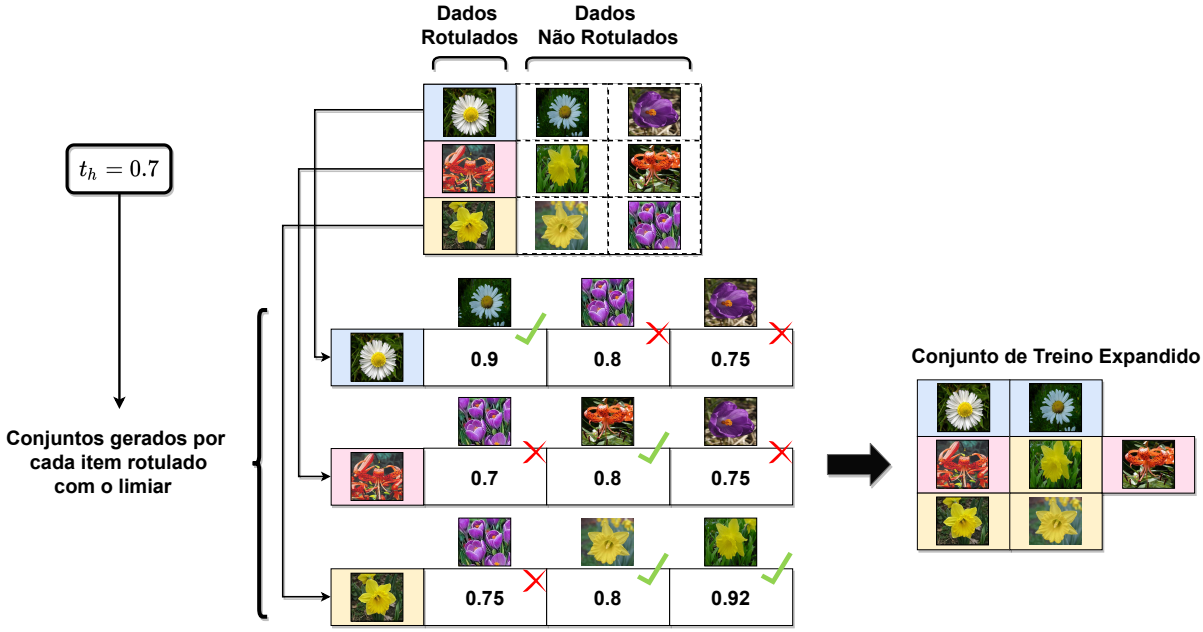
O conjunto rotulado expandido $\mathfrak{X}_W$ pode ser então definido através da união do conjunto rotulado original e do conjunto rotulado estimado:

$$\mathfrak{X}_W = \mathfrak{X}_L \cup \mathfrak{X}_E. \tag{3.10}$$

A Figura 6 ilustra o processo de expansão de conjuntos rotulados. Começando com um pequeno conjunto rotulado, calculamos as métricas de correlação $r(\tau_l, \tau_u)$ entre os itens rotulados e os não rotulados. Então, considerando o limiar t_h, para cada imagem rotulada, são selecionadas quais imagens não rotuladas obtêm uma correlação que seja maior ou igual ao limiar t_h, gerando uma série de conjuntos. Após isso, para remover inconsistências, no caso de um mesmo item ter um rótulo atribuído duas vezes ou mais, as interseções entre esses conjuntos são removidas. Por fim, o conjunto rotulado expandido é utilizado para treinamento de um classificador supervisionado ou semi-supervisionado.

Considerando a complexidade computacional do método, a expansão de rótulos está intimamente associada a quantidade de amostras rotuladas. Seja L igual à quantidade de elementos do conjunto rotulado $\mathfrak{X}_L$ ($L = |\mathfrak{X}_L|$) e l o comprimentos das listas ranqueadas considerado. Para cada item rotulado (L), uma métrica de correlação é calculada para cada item não rotulado ($N - L$). A complexidade das métricas de correlação varia, mas todas as métricas são limitadas a analisar somente as primeiras k posições das listas ranqueadas, e como k é uma constante, sua complexidade é portanto $O(1)$. A complexidade total então é $O(L(N - L))$ se considerarmos as listas ranqueadas inteiras (de tamanho N), porém, as listas ranqueadas são normalmente definidas apenas até um tamanho constante l, de modo que $k < l \ll N$. Deste modo, as métricas de correlações são calculadas considerando somente as primeiras l posições das listas ranqueadas dos elementos rotulados, logo, a complexidade total pode ser reduzida para $O(L)$.

Figura 6 – Ilustração do método de expansão de conjuntos rotulados baseado em correlações de listas ranqueadas.



Fonte: Elaborado pelo autor.

3.4 Definição do Limiar de Correlação

As métricas de correlação de listas ranqueadas fornecem uma maneira efetiva de analisar a informação de similaridade contextual codificada nos conjuntos de dados. Entretanto, de modo a executar uma expansão de conjuntos rotulados adequada, é essencial selecionar um limiar para os valores de correlação capaz de separar amostras relevantes das não relevantes. Caso esse limiar seja muito alto, expansões muito pequenas ou insignificantes podem ocorrer. Ao contrário, caso o valor de limiar seja muito baixo, amostras incorretas podem ser incorporadas ao conjunto de treinamento, o que pode diminuir os resultados de acurácia obtidos. A Figura 7 apresenta uma análise visual desse cenário. Considerando uma imagem de consulta (em azul), a partir de diferentes valores do limiar são selecionadas imagens em que a correlação entre estas e a imagem de consulta ultrapasse o limiar. Podemos observar que para valores de limiar muito altos, poucos elementos são selecionados, e conforme diminuímos esse valor mais imagens são selecionadas até que imagens incorretas começam a ser selecionadas (em vermelho).

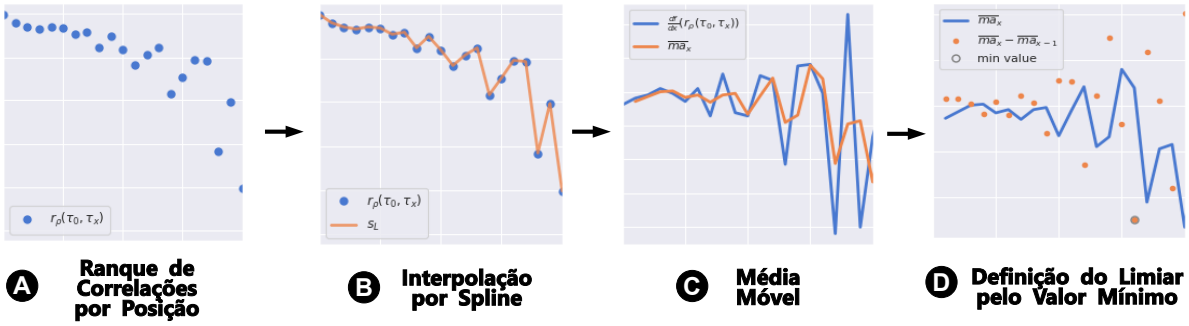
Com isso em mente, um método não supervisionado capaz de estimar um limiar adequado em cada cenário através da análise das correlações em diferentes posições nas listas ranqueadas. As primeiras posições das listas ranqueadas normalmente concentram amostras relevantes de mesma classe. A transição das amostras relevantes para não relevantes tende a causar mudanças bruscas nos valores de correlação, situação que pode ser identificada e explorada para definir um limiar adequado.

Figura 7 – Diferentes conjuntos gerados pela análise de uma métrica de correlação considerando diferentes limiares.

Fonte: Elaborado pelo autor.

A sequência de métricas de correlação é representada por splines lineares, derivadas e diferenças de médias móveis, cujos valores são utilizados para identificar tendências e estimar o limiar. A análise é conduzida para cada lista ranqueada e a média de cada limiar estimado é calculada para definir o parâmetro t_h utilizado na etapa de expansão do método de correlações. O método de obtenção do limiar pode ser dividido em quatro etapas principais, como está ilustrado na Figura 8. Cada uma das etapas é discutida com detalhes a seguir.

Figura 8 – Método proposto para definição automática do limiar de expansão.



Fonte: Elaborado pelo autor.

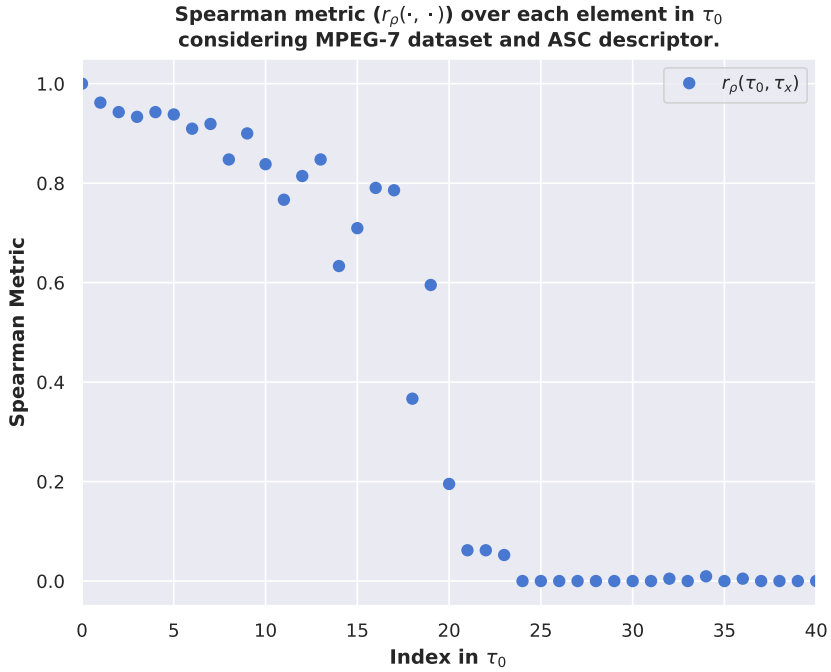
A) Ranque de Correlações por Posição:

O primeiro passo do método de definição do limar é criar um vetor de métricas de correlação calculadas a partir das posições das listas ranqueadas. Formalmente,

seja $x_i \in \mathfrak{X}$ um item de dados e τ_i a sua lista ranqueada de tamanho l. O vetor $\mathbf{c_i} = [c_{i1}, c_{i2}, \ldots, c_{il}]$ denota um ranque de métricas de correlação definido a partir de τ_i. O valor de cada elemento c_{ip_o} é definido como $c_{ip_o} = r(\tau_i, \tau_j)$, onde p_o é a posição de x_j na lista ranqueada τ_i, de modo que $p_o = \tau_i(j)$ e $r(\cdot, \cdot)$ denota uma métrica de correlação entre listas ranqueadas, definida na seção anterior (Seção 3.2).

A Figura 9 ilustra os valores de um vetor de métricas de correlação conforme as posições em uma lista ranqueada de uma imagem da coleção de imagens MPEG-7 [74]. Uma queda brusca nos valores de correlação pode ser observada perto do número de elementos por classe da coleção de imagens (perto da posição 20) pode ser observada.

Figura 9 – Valores de correlação segundo a posição na lista ranqueada τ_0.



Fonte: Elaborado pelo autor.

B) Interpolação por *Spline*:

Seja f uma função definida no intervalo $[c_{i1}, c_{il}]$, a qual representa a relação entre as posições do ranque e os valores de correlação dados pelo vetor $\mathbf{c_i}$. Para estimar a função f, utilizamos de uma interpolação por *spline* linear $s_L(a)$. O intervalo $[c_{i1}, c_{il}]$ pode ser dividido em $[a_j, a_{j+1}]$ subintervalos com $j = 1, \ldots, l-1$, $a_1 = c_{i1}$ e $a_l = c_{il}$. É possível definir um polinômio $p(a)$ no intervalo $[c_{i1}, c_{il}]$:

$$p(a) = p_i(a), a_{j-1} < a < a_j, \tag{3.11}$$

onde para $j = 1, 2, \ldots, l$ cada $p_j(a)$ é uma função polinomial definida no intervalo $[a_{j-1}, a_j]$.

Considerando os pontos $a_1, a_2, \ldots, a_l$, a *spline* linear $s_L(a)$ que interpola a função f nesses pontos pode ser definida como:

$$s_L(a) = f(a_{j-1})\frac{a - a_j}{a_{j-1} - a_j} + f(a_j)\frac{a - a_{j-1}}{a - a_{j-1}}, \tag{3.12}$$

onde $a \in [a_{j-1}, a_j]$ e $j = 2, \ldots, l$.

Um exemplo de interpolação pode ser observado na Figura 10, que apresenta uma *spline* linear estimada com base nos valores de correlação obtidos no passo anterior.

Figura 10 – Interpolação por *spline* linear $s_L(a)$ feita sobre os valores de correlação do vetor $\mathbf{c_i}$.

Fonte: Elaborado pelo autor.

C) Média Móvel:

Através da *spline* linear gerada no passo anterior, podemos calcular as derivadas para cada valor em $\mathbf{c_i}$. Para cada valor de correlação, a derivada correspondente é calculada:

$$d_{ij} = \frac{df}{da}(c_{ij}), \tag{3.13}$$

gerando um vetor de derivativas $\mathbf{d_i} = [d_{i1}, d_{i2}, \ldots, d_{il}]$.

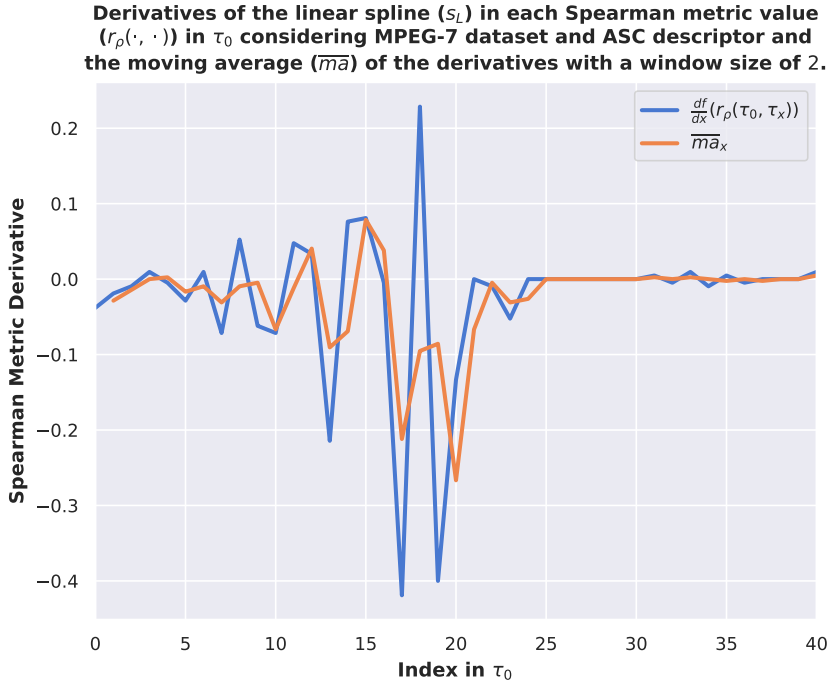
Com base em cada valor do vetor de derivadas, as médias móveis são calculadas considerando um tamanho de janela m[3]. Seja $\overline{\mathbf{ma}}_i = [\overline{ma}_{i1}, \overline{ma}_{i2}, \ldots, \overline{ma}_{iN}]$ o vetor de médias móveis, definido por:

$$\overline{ma}_{ij} = \frac{1}{m}\sum_{l=0}^{m-1} d_{ij+l}. \tag{3.14}$$

[3] Em nossos experimentos, utilizamos $m = 2$.

A Figura 11 apresenta as derivadas calculadas a partir dos valores da *spline* linear e suas respectivas médias móveis.

Figura 11 – Derivadas da *spline* linear $s_L(a)$ e as médias móveis desses valores considerando um tamanho de janela $m = 2$.



Fonte: Elaborado pelo autor.

D) Definição do Limiar pelo Valor Mínimo:

Ao considerar as médias móveis das derivadas, avaliamos as tendências nos valores de derivadas das métricas de correlação. Nosso método tenta verificar onde os valores de derivadas começam a diminuir a uma taxa que indica baixos valores de confiança nas informações de similaridade. Para analisar essas tendências nos valores de médias móveis, são analisadas as diferenças entre valores consecutivos das médias. Ou seja, dado o vetor de médias móveis $\overline{\mathbf{ma}}_\mathbf{i}$, a diferença entre valores consecutivos é definida por:

$$\text{dif}_{\overline{ma}_{ij}} = \overline{ma}_{ij+1} - \overline{ma}_{ij}, \tag{3.15}$$

onde $j = 1, 2, \ldots, l-1$.

Podemos definir um vetor de diferenças como: $\mathbf{dif}_{\overline{\mathbf{ma}}_\mathbf{i}} = [\text{dif}_{\overline{ma}_{i1}}, \text{dif}_{\overline{ma}_{i2}}, \ldots, \text{dif}_{\overline{ma}_{il}}]$ e com as diferenças calculadas, precisamos identificar a posição/índice que contém o menor valor. Essa posição é utilizada como o índice do vetor de correlações $\mathbf{c_i}$ e define o limiar t_{hi}, e pode ser definida formalmente como:

$$t_{hi} = \mathbf{c_i}[\arg\min \text{dif}_{\overline{ma}_i}]. \tag{3.16}$$

A Figura 12 apresenta essa análise. Além das diferenças entre posições consecutivas das médias móveis, o menor valor dessas diferenças é destacado. O índice do menor valor na Figura 12 é igual a 16, logo, ao analisar os valores de correlação do ranque de correlações presente na Figura 9 podemos definir um limiar por volta de 0.8 nesse cenário.

Figura 12 – Médias móveis $\overline{\mathbf{ma}}_i$ dos valores de derivada da *spline* linear s_L, as diferenças entre valores consecutivos e o valor mínimo dessas diferenças.

Fonte: Elaborado pelo autor.

Por fim, de modo a estimar um limiar capaz de representar o conjunto de dados na totalidade, as etapas apresentadas anteriormente são repetidas para todos os itens de dados, gerando um limiar t_{hi} para cada $x_i \in \mathfrak{X}$. O limiar global t_h que é considerado na tarefa de expansão é calculado a partir da média de todos os limiares t_{hi} estimados individualmente, sendo definido como:

$$t_h = \frac{1}{N} \sum_{x_i \in \mathfrak{X}} t_{hi}. \tag{3.17}$$

4 Aprendizado Fracamente Supervisionado Baseado em Hipergrafos

A falta de dados de treinamento em cenário de supervisão fracamente supervisionada traz diversos desafios para aprimorar a acurácia em tarefas de classificação. Entre elas, a questão principal consiste em explorar ao máximo a informação presente nos dados não rotulados, os quais são abundantes, e como estabelecer relações confiáveis entre os escassos dados de treinamento disponíveis e as amostras não rotuladas. Uma solução para contornar esse problema é proposta ao longo desse capítulo, em que utilizaremos de um método de aprendizado de variedades baseado em ranqueamento e de uma estratégia de expansão de rótulos.

A estratégia utilizada pelo método apresentado ao longo do capítulo se da através da análise de uma estrutura de hipergrafo construído a partir da análise do modelo de ranqueamento. Hipergrafos conseguem relacionar diversas amostras através de suas hiperarestas (diferente de pares em grafos convencionais), e essa representação pode ser explorada de modo a atribuir rótulos aos elementos não rotulados. De modo similar ao método de aprendizado fracamento supervisionado baseado em correlações (Capítulo 3), o desempenho do método está fortemente associada a qualidade das listas ranqueadas, e por consequência a qualidade dos vetores de características. Além disso, a análise das listas ranqueadas também deve ser limitada a uma constante l para tornar o método adequado a todos os cenários. Por trabalhar com um modelo de ranqueamento, o método é capaz de considerar instâncias com uma alta sobreposição com outras classes no espaço de características e ao utilizar uma estrutura de hipergrafo, o método apresenta um bom comportamento em conjuntos de dados grandes, por conseguir relacionar diversos conjuntos de elementos.

A Figura 13 apresenta o método de aprendizado fracamente supervisionado baseado em hipergrafos proposto, dividido desde a preparação dos dados até a etapa final de classificação. A primeira etapa para funcionamento do método é extrair os vetores de características de cada imagem, após isso o método proposto passa por quatro etapas principais:

1. **Ranqueamento:** um conjunto de listas ranqueadas é criado para cada uma das imagens do conjunto de dados a partir das distâncias entre os vetores de características de cada uma das amostras;

2. **Construção do Hipergrafo:** é construído um hipergrafo e estruturas associadas a partir do método LHRR [19] de aprendizado de variedades. As informações fornecidas

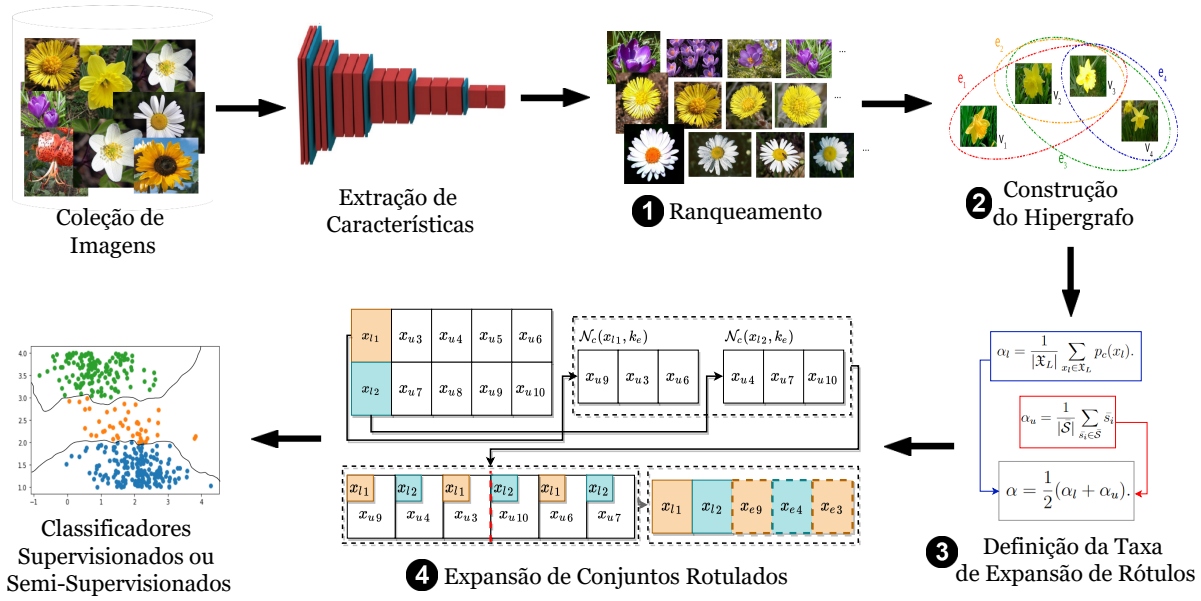
pelas estruturas do hipergrafos são utilizadas pelo método de expansão de rótulos;

3. **Definição da Taxa de Expansão de Rótulos:** o método de expansão proposto depende de um parâmetro que seleciona qual porcentagem das novas amostras rotuladas geradas é realmente incorporada ao conjunto expandido. Uma estratégia para definir essa taxa de expansão de rótulos é explorada;

4. **Expansão de Conjuntos Rotulados:** A informação fornecida pelo hipergrafo é explorada de modo a identificar hiperarestas e relações de similaridade de alta confiança entre os dados. Essas informações são utilizadas para construir um conjunto de pares (rotulados, não rotulados) de candidatos à expansão de rótulos. Os pares são ordenados de acordo com seu grau de confiança estimado e os melhores pares são selecionados para a expansão.

Por fim, o conjunto rotulado de treino expandido é utilizado para treinamento de classificadores supervisionados ou semi-supervisionados.

Figura 13 – Diagrama de funcionamento do método de aprendizado fracamente supervisionado baseado em hipergrafos.



Fonte: Elaborado pelo autor.

A Seção 4.1 discute o método de aprendizado de variedades que constrói o modelo de hipergrafo que é utilizado para realizar a tarefa de expansão. Na Seção 4.2 o método de expansão de conjuntos rotulados por hipergrafos é apresentado. Por fim, a Seção 4.3 apresenta uma estratégia para definir automaticamente um dos parâmetros do método de expansão.

4.1 Construção do Hipergrafo

Hipergrafos fornecem uma técnica robusta para representar relações de similaridade entre os conjuntos de objetos. Além disso, listas ranqueadas concentram relevantes informações de similaridade nas primeiras posições. O algoritmo LHRR [19] se utiliza de um hipergrafo que é construído a partir das informações presentes nas listas ranqueadas para representar relações de similaridade codificadas na estrutura do conjunto de dados. Uma vez que as estruturas do hipergrafo define a base do método de expansão de rótulos, sua definição é discutida em detalhes a seguir. O algoritmo pode ser dividido em cinco etapas principais:

1. **Normalização das Listas Ranqueadas:**

   O primeiro passo para funcionamento do algoritmo é normalizar cada lista ranqueada considerando uma nova métrica de similaridade p_n calculada sobre posições reciprocas dos ranques. Seja x_i e x_j dois itens de dados, a nova métrica de similaridade é:

   $$\rho_n(x_i, x_j) = 2l - (\tau_i(x_j) + \tau_j(x_i)), \tag{4.1}$$

   onde l é o tamanho da lista ranqueada que considerada. Através dessa nova métrica, os itens de dados nas primeiras l posições das listas ranqueadas são ordenados se utilizando um algoritmo de ordenação estável.

2. **Construção do Hipergrafo:**

   Seja $G = (V, E, w)$ um hipergrafo onde V é um conjunto finito de vértices e E o conjunto de hiperarestas. O conjunto de hiperarestas E é uma família de subconjuntos de V tal que $\bigcup_{e \in E} = V$. Para cada vértice $v_i \in V$ um item $x_i \in \mathfrak{X}$ está associado. Cada hiperaresta tem um peso $w(e_i)$ atribuído, que representa o grau de confiança das relações estabelecidas pela hiperaresta e_i. Dizemos que uma hiperaresta e_i é incide no vértice v_j quando $v_j \in e_i$. Com isso em mente, podemos representar o hipergrafo através de uma matriz de incidências $\mathbf{H}_b$ de tamanho $|E| \times |V|$:

   $$h_b(e_i, v_i) = \begin{cases} 1 & \text{se } v_j \in e_i, \\ 0 & \text{caso contrário.} \end{cases}$$

   Podemos também definir uma hiperaresta e_i como um conjunto de vértices $e_i = \{v_1, v_2, \ldots, v_m\}$. A matriz de incidências $\mathbf{H}_b$ permite apenas atribuições binárias de um vértice à hiperaresta, não considerando cenários onde é desejável considerar um certo grau de incerteza. Para contornar esse problema, podemos definir um hipergrafo probabilístico, onde a probabilidade de que um vértice pertence ao hipergrafo é considerada. Seja $r : E \times V \to \mathbb{R}^+$ uma função, podemos definir uma matriz de incidências contínua $\mathbf{H}$ onde:

   $$h(e_i, v_i) = \begin{cases} r(e_i, v_j) & \text{se } v_j \in e_i, \\ 0 & \text{caso contrário.} \end{cases}$$

Para cada item de dados $x_i \in \mathfrak{X}$ uma hiperaresta é definida com base no conjunto dos k vizinhos de x_i. Seja $x_z \in \mathcal{N}(x_i, k)$ um vizinho de x_i e seja $x_j \in \mathcal{N}(x_z, k)$ um vizinho de x_z. Podemos definir uma medida de associação $r(e_i, v_j)$, que indica o grau de que um vértice v_j pertence a hiperaresta e_i:

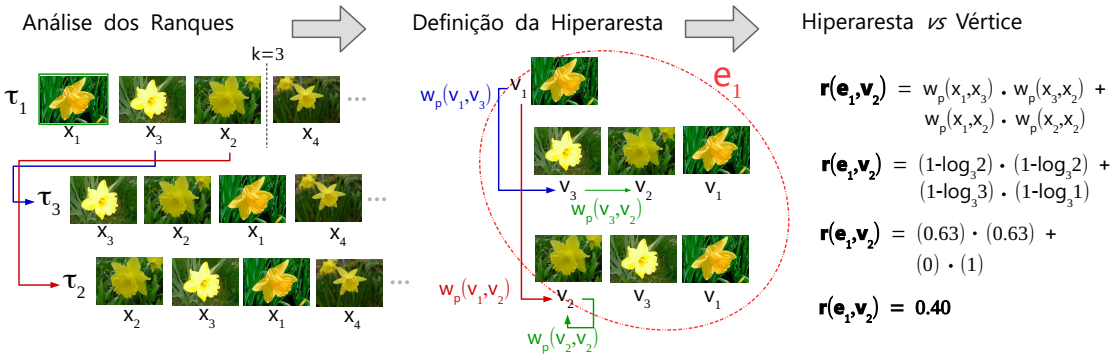
$$r(e_i, v_j) = \sum_{x_z \in \mathcal{N}(x_i,k) \wedge x_j \in \mathcal{N}(x_z,k)} w_p(x_i, x_z) \cdot w_p(x_z, x_j), \tag{4.2}$$

onde $w_p(x_i, x_z)$ atribui um peso de relevância a x_z com base em sua posição em τ_i. O peso $w_p(i, z)$ atribuído a x_z com base em sua posição em τ_i pode ser definido como:

$$w_p(x_i, x_z) = 1 - \log_k \tau_i(x_z). \tag{4.3}$$

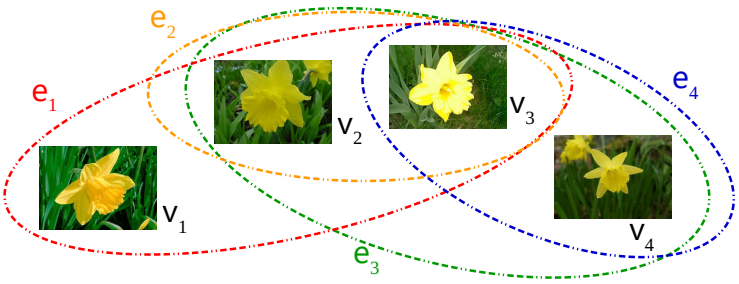
A Figura 14 ilustra como as listas ranqueadas são exploradas para construção da hiperaresta e_1 e como a relação entre a hiperaresta e os vértices é calculada. A hiperaresta e_1 é criada a partir da lista ranqueada τ_1 considerando as refêrencias aos ranques τ_2 e τ_3. Cada função w_p atribui um peso a cada vértice de acordo com sua posição na lista ranqueada. Entre a hiperaresta e_1 e o vértice v_2 uma associação é definida por $r(e_1, v_2)$, calculada usando dos pesos gerados por w_p (Equação 4.3). A Figura 15 apresenta um exemplo de um hipergrafo construído pelo método LHRR [19] em sua totalidade com quatro vértices e quatro hiperarestas.

Figura 14 – Definição da hiperaresta e_1 através das referências entre os ranques considerando um tamanho de vizinhança igual 3. A Função w_p atribui pesos de acordo com as posições das imagens nas hiperarestas e são utilizadas para definir uma relação entre a hiperaresta e seus vértices respectivos ($r(e_1, v_2)$).



Fonte: Elaborado pelo autor.

Figura 15 – Hipergrafo construído pelo método LHRR [19] com os vértices $\{v_1, v_2, v_3, v_4\}$ e as hiperarestas $\{e_1, e_2, e_3, e_4\}$.



Fonte: Elaborado pelo autor.

Cada hiperaresta possui um peso $w(e_i)$ respectivo, o qual representa o grau de confiança das relações estabelecidas entre os vértices. Antes de calcularmos os pesos $w(e_i)$ é preciso definir o conjunto de vizinhança do hipergrafo $\mathcal{N}_h$. Dado uma hiperaresta e_i, o conjunto $\mathcal{N}_h$ com os vértices com os valores mais altos de $h(e_i, \cdot)$ pode ser definido como:

$$\mathcal{N}_h(x_q, k) = \{\mathcal{S} \subseteq e_q, |\mathcal{S}| = k \wedge \forall x_i \in \mathcal{S}, x_j \in e_q - \mathcal{S} : h(e_q, v_i) > h(e_q, v_j)\}. \quad (4.4)$$

É esperado que uma hiperaresta altamente efetiva contenha poucos vértices, que estão relacionados com altos valores de $h(e_i, \cdot)$ [19]. Com isso em mente, o peso $w(e_i)$ pode ser definido como:

$$w(e_i) = \sum_{j \in \mathcal{N}_h(x_i, k)} h(e_i, v_j). \quad (4.5)$$

Como cada hiperaresta é criada a partir da análise das listas ranqueadas τ_i, o peso $w(e_i)$ de cada hiperaresta pode ser considerado como uma medida de eficácia não supervisionada da lista ranqueada τ_i. Logo, altos valores de $w(e_i)$ podem ser interpretados como uma lista ranqueada de alta eficácia [19].

3. **Similaridade entre as Hiperarestas:**

O próximo passo do método LHRR é criar uma matriz de similaridade $\mathbf{S}$. Sua construção se baseia em duas hipóteses [19]. A primeira é que itens de dados similares tem listas ranqueadas similares e, portanto, hiperarestas similares. Considerando que a matriz de incidências codifica todas as informações de similaridade, uma métrica de similaridade entre duas hiperarestas e_i e e_j pode ser calculada através da soma de todos os valores de h multiplicados pelos seus vértices correspondentes. Para considerar todos os elementos, essa operação pode ser realizada multiplicando a matriz de incidências pela sua transposta.

$$\mathbf{S}_h = \mathbf{H}\mathbf{H}^T. \quad (4.6)$$

A segunda hipótese é a de que itens similares tem uma probabilidade alta de participarem da mesma hiperaresta. Portanto, para calcular a similaridade entre dois vértices v_i e v_j, os valores h da hiperaresta correspondente devem ser multiplicados. De modo similar a primeira hipótese, é possível realizar essa operação através da matriz de incidências, multiplicando a matriz transposta pela matriz não transposta:

$$\mathbf{S}_v = \mathbf{H}^T\mathbf{H}. \tag{4.7}$$

Por fim, as duas similaridades são combinadas em uma matriz de similaridade $\mathbf{S}$ através do produto de Hadamard das matrizes $\mathbf{S}_h$ e $\mathbf{S}_v$:

$$\mathbf{S} = \mathbf{S}_h \circ \mathbf{S}_v. \tag{4.8}$$

4. **Produto Cartesiano dos Elementos das Hiperarestas:**

   Construída a matriz de similaridade $\mathbf{S}$, a próxima etapa é executar operações de produto cartesiano para extrair relações entre pares de elementos diretamente das hiperarestas. Operação feita com o objetivo de maximizar as informações de similaridade, e essa nova informação pode ser combinada com as similaridades das hiperarestas. O produto cartesiano entre duas hiperarestas $e_q \in E$ e $e_i \in E$ é:

   $$e_q \times e_i = \{(v_y, v_z) : v_y \in e_q \wedge v_z \in e_i\} \tag{4.9}$$

   O produto cartesiano entre os elementos da mesma hiperaresta e_q pode ser descrito como e_q^2. Cada par de vértices $(v_i, v_j) \in {e_q}^2$ tem uma relação $p : E \times V \times V \rightarrow \mathbb{R}^+$ estabelecida. Essa relação é calculada através do peso $w(e_q)$, o grau de confiança da hiperaresta. É possível também definir um grau de associação entre os vértices v_i and v_j:

   $$p(e_q, v_i, v_j) = w(e_q) \times h(e_q, v_i) \times h(e_q, v_j). \tag{4.10}$$

   Com o grau de associação p calculado, a matriz $\mathbf{C}$ é construída considerando as relações entre todas as hiperarestas. Cada posição da matriz $\mathbf{C}$ é definida a seguir:

   $$\mathbf{C}_{i,j} = \sum_{e_q \in E \wedge (v_i, v_j) \in {e_q}^2} p(v_i, v_j). \tag{4.11}$$

5. **Similaridade Baseada em Hipergrafos:**

   Calculadas as matrizes $\mathbf{S}$ e $\mathbf{C}$ nos passos 3 e 4, uma nova matriz de similaridade ou afinidade $\mathbf{W}$ é formada ao combiná-las:

   $$\mathbf{W} = \mathbf{C} \circ \mathbf{S}. \tag{4.12}$$

   A matriz $\mathbf{W}$ concentra as informações de similaridade e é utilizada para calcular um novo conjunto de listas ranqueadas, o qual apresenta uma melhor eficácia em tarefas

de ranqueamento. Deste modo, ambas a entrada e saída do modelo de hipergrafo são criadas com base na informação dos ranques, e os cálculos podem ser repetidos iterativamente para obter ranques e representações por hipergrafo cada vez mais eficazes.

A notação $^{(t)}$ vai ser utilizada para representar a iteração atual do algoritmo. O conjunto de listas ranqueadas $\mathcal{T}^{(t+1)}$ é calculado através da matriz de similaridade $\mathbf{W}^{(t)}$ e o conjunto $\mathcal{T}^{(0)}$ é construído a partir dos vetores de características iniciais. O processo é repetido ao longo de T iterações e o hipergrafo e estruturas relacionadas (a matriz $\mathbf{W}^{(T)}$) finais são utilizados no processo de expansão de rótulos, detalhado a seguir.

4.2 Expansão de Conjuntos Rotulados por Hipergrafos

Como o método LHRR [19] cria um hipergrafo e estruturas capazes de fornecer informações mais eficazes com relação à similaridade entre itens de dados, o método de expansão de conjuntos rotulados que será discutido ao longo dessa seção explora das informações codificadas em tais estruturas. A informação presente nas hiperarestas vai ser explorada de modo a construir um conjunto de pares, que representa os candidatos a expansão de rótulos. Os pares de candidatos são ordenados com base no peso das hiperarestas e na similaridade entre os elementos. Os pares nas primeiras posições são então selecionados com base em um limiar de posição, de modo a compor o conjunto expandido. Cada passo do método de expansão será detalhado a seguir.

A) Conjunto de Candidatos para a Expansão de Rótulos:

Para cada item rotulado $x_l \in \mathfrak{X}_L$ um conjunto de candidatos é definido pelos itens não rotulados com os maiores valores na hiperaresta de x_l. Formalmente, seja $\mathcal{N}_h(x_l, k_e)$ um conjunto que contêm os k_e maiores valores de $h(\cdot,\cdot)$ em sua respectiva hiperaresta e_{x_l} (Equação 4.4). Portanto, podemos definir um conjunto de candidados não rotulados para cada item rotulado, dado por $\mathcal{N}_c(x_l, k_e)$, como:

$$\begin{aligned}\mathcal{N}_c(x_q, k_e) = \{\mathcal{S} \subseteq e_q^{(T)}, |\mathcal{S}| \leq k_e \wedge \forall x_i \in \mathcal{S},\\ x_j \in e_q^{(T)} - \mathcal{S} : h^{(T)}(x_q, x_i) > h^{(T)}(x_q, x_j)\}.\end{aligned} \tag{4.13}$$

Um conjunto de pares de candidatos é então gerado considerando todo o conjunto de dados, a partir da união dos conjuntos de cada item rotulado. Sendo assim, o conjunto de pares de candidatos $\mathcal{C}_p$ pode ser definido como:

$$\mathcal{C}_p = \bigcup_{x_l \in \mathfrak{X}_L} \bigcup_{x_u \in \mathcal{N}_c(x_l, k_e)} \{(x_l, x_u)\}. \tag{4.14}$$

B) Ranque de Pares de Candidatos:

Dado o conjunto de pares de candidatos $\mathcal{C}_p$, queremos selecionar um subconjunto de pares que será utilizado nos procedimentos de expansão. Portanto, os pares são ordenados conforme o grau de confiança da hiperaresta que gerou o par (a hiperaresta do elemento rotulado) e com a similaridade entre os elementos (rotulado e não rotulado). O grau de confiança é obtido através do peso da hiperaresta do item rotulado $w^{(T)}(e_{x_l})$, já a similaridade entre os itens rotulados e não rotulados é obtida pelo valor na matriz de similaridade $\mathbf{W}^{(T)}_{x_l,x_u}$.

O ranque de pares de candidatos τ_C pode ser definido como uma permutação do conjunto $\mathcal{C}_p$. A permutação τ_C é uma bijeção do conjunto $\mathcal{C}_p$ no conjunto $\{1, 2, \ldots, |\mathcal{C}_p|\}$. A posição do par (x_l, x_u) na lista ranqueada de pares de candidatos é indicada por $\tau_C((x_l, x_u))$. O ranque τ_C é calculado de acordo com uma função $s_c(.,.)$ a qual considera informações de confiança e similaridade. Se no ranque de pares de candidatos (x_l, x_u) está em uma posição anterior a (x_i, x_j), isto é, $\tau_C((x_l, x_u)) < \tau_C((x_i, x_j))$, então $s_c(x_i, x_j) \geq s_c(x_i, x_j)$. Formalmente, podemos definir a função $s_c(.,.)$ como:

$$s_c(x_l, x_u) = w^{(T)}(e_{x_l}) \times \mathbf{W}^{(T)}_{x_l,x_u} \tag{4.15}$$

C) Seleção dos Pares de Candidatos:

Assim que a lista ranqueada de pares de candidatos estiver construída, é necessário definir qual conjunto de pares participará do processo de expansão. Essa seleção será feita com base nas primeiras posições de τ_C até uma posição limite t_p. O número de amostras selecionadas é proporcional ao tamanho do conjunto rotulado e ao tamanho de vizinhança k_e:

$$t_p = |\mathfrak{X}_L| \times k_e \times \alpha, \tag{4.16}$$

onde α é um parâmetro chamado taxa de expansão de rótulos, definido no intervalo $[0, 1]$, que define a porcentagem de amostras selecionadas para a expansão. Uma estratégia para estimar o valor de α será discutida na Seção 4.3.

D) Expansão de Conjuntos Rotulados:

A partir do ranque τ_C e da posição limite t_p, o conjunto rotulado estimado $\mathfrak{X}_E$ pode ser definido como:

$$\mathfrak{X}_E = \{x_e : x_e \in \mathfrak{X}_U | (x_e, x_l) \in \mathcal{C}_p \wedge x_l \in \mathfrak{X}_L \wedge \tau_C((x_e, x_l)) \leq t_p\} \tag{4.17}$$

Assim que o conjunto rotulado estimado $\mathfrak{X}_E$ é gerado, a classe estimada do item x_e será a mesma do item rotulado x_l do seu par. Elementos do conjunto estimado que aparecem no conjunto de pares de candidatos em pares com imagens de diferentes classes são removidos, de modo a remover inconsistências (uma mesma imagem ter

dois rótulos diferentes atribuídos). Sendo assim, podemos definir uma função de estimativa de rótulos baseada na informação de hipergrafos H_E como:

$$E_H(x_e) = Y(x_l), \tag{4.18}$$

onde $x_e \in \mathfrak{X}_E$ and $(x_e, x_l) \in \mathcal{C}_p$. O conjunto rotulado expandido pode ser então definido pela união do conjunto rotulado original e do conjunto rotulado estimado:

$$\mathfrak{X}_W = \mathfrak{X}_L \cup \mathfrak{X}_E. \tag{4.19}$$

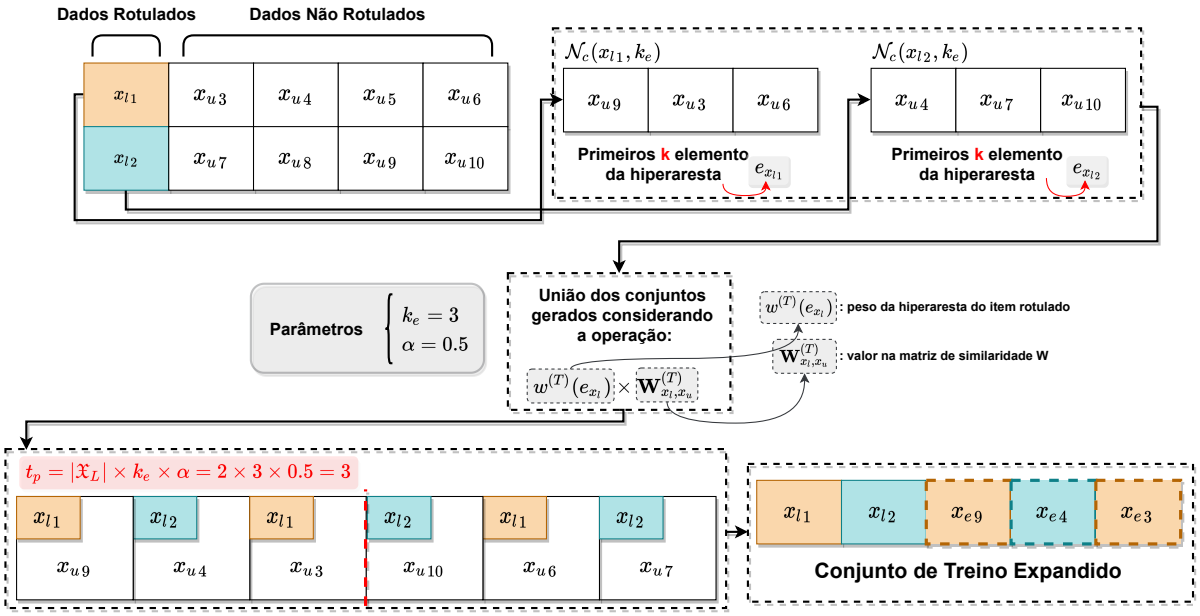
Por fim, o conjunto $\mathfrak{X}_W$ será utilizado no treinamento de classificadores supervisionados ou semi-supervisionados.

A Figura 16 ilustra como a expansão de conjuntos rotulados baseada em hipergrafos é feita. Começando com um pequeno conjunto rotulado são gerados conjuntos de candidatos não rotulados $\mathcal{N}_c(x_q, k_e)$ que contêm os maiores valores de $h(\cdot, \cdot)$ nas hiperarestas dos itens rotulados e_{x_l}. No nosso exemplo, criamos conjuntos de candidatos a partir dos itens rotulados x_{l_1} e x_{l_2} com os $k_e = 3$ maiores valores em suas respectivas hiperarestas. Isto irá criar um conjunto de pares de candidatos $\mathcal{C}_p$ que será utilizado para gerar um ranque $\tau_\mathcal{C}$ considerando informações em ambas as hiperarestas e na matriz de similaridade $\mathbf{W}$ (Equação 4.15). A seguir, são selecionados os pares de candidatos do ranque $\tau_\mathcal{C}$ na posição limite t_p. Considerando os parâmetros especificados no exemplo: $k_e = 3$, $\alpha = 0.5$ e $|\mathfrak{X}_L| = 2$, temos que $t_p = 3$. A última etapa do método é estimar a classe de cada entrada selecionada em $\tau_\mathcal{C}$, a qual será a classe do item rotulado que gerou aquela entrada, por exemplo, o item não rotulado x_{u_9} terá atribuída a classe do item rotulado x_{l_1}. Além disso, caso alguma entrada apresente itens rotulados associados a diferentes classes, indicam cenários de baixa confiança e são removidos para que não sejam atribuídos rótulos diferentes a um mesmo item. Por fim, o conjunto de treino expandido $\mathfrak{X}_W$ é criado, o qual será utilizado para treinamento de classificadores supervisionados os semi-supervisionados.

4.3 Definição da Taxa de Expansão de Rótulos

A fim de que o método de expansão baseado em hipergrafos opere de forma eficaz é necessário utilizar um valor adequado para a uma taxa de expansão de rótulo (α). Caso utilizemos de uma taxa de expansão de rótulos muito alta, uma quantidade alta de falsos positivos de τ_C podem ser selecionados para a expansão, o que diminuirá a acurácia obtida pelos classificadores. Por outro lado, caso o valor de α seja muito baixo, expansões com um número reduzido de amostras podem ocorrer, ocasionando impacto reduzido nos resultados de classificação se compararmos com o cenário sem expansão. Sendo assim, esta seção apresenta uma maneira de estimar o valor de α.

Figura 16 – Ilustração do método de expansão de conjuntos rotulados baseado em hipergrafos.



Fonte: Elaborado pelo autor.

A ideia central consiste em definir a magnitude de α de acordo com a eficácia das listas ranqueadas do conjunto de dados. Para os dados rotulados, o método irá analisar a eficácia a partir da quantidade de itens de dados da mesma classe nas primeiras posições das listas ranqueadas. O método também considera a análise dos dados não rotulados a partir das estruturas geradas durante a construção do hipergrafo para auxiliar na definição do parâmetro.

A primeira etapa para estimar o parâmetro é analisar os dados rotulados, gerando um parâmetro α_l intermediário. Para cada $x_l \in \mathfrak{X}_L$, será calculada a porcentagem de vizinhos da mesma classe de x_l no conjunto de vizinhança correspondente. Formalmente, essa porcentagem é obtida pela função $p_c(x_l)$:

$$p_c(x_l) = \frac{|\{x_c : x_c \in \mathcal{N}(x_l, k_e) \cap \mathfrak{X}_L \wedge Y(x_c) = Y(x_l)\}|}{|\{x_m : x_m \in \mathcal{N}(x_l, k_e) \cap \mathfrak{X}_L\}|}. \tag{4.20}$$

A taxa α_l é definida pela porcentagem média ao analisar o conjunto rotulado $\mathfrak{X}_L$:

$$\alpha_l = \frac{1}{|\mathfrak{X}_L|} \sum_{x_l \in \mathfrak{X}_L} p_c(x_l). \tag{4.21}$$

A segunda parte do método irá estimar outro parâmetro α_u intermediário, calculado com base no grau de confiança médio obtido por $s_c(.,.)$ para os pares de candidatos $\mathcal{C}_p$. Seja $\mathcal{S}$ um conjunto de valores de confiança definido por: $\mathcal{S} = \{s_c(x_l, x_u) \in \mathbb{R} | (x_l, x_u) \in \mathcal{C}_p\}$. Feito isso, será realizada uma normalização min-max sobre o conjunto $\mathcal{S}$, gerando um

conjunto $\bar{\mathcal{S}}$, que será utilizado para gerar o parâmetro α_u, a média dos valores do novo conjunto:

$$\alpha_u = \frac{1}{|\bar{\mathcal{S}}|} \sum_{\bar{s}_i \in \bar{\mathcal{S}}} \bar{s}_i. \tag{4.22}$$

Por fim, a taxa de expansão de rótulos α estimada é definida como a média entre os parâmetros intermediários α_l e α_u:

$$\alpha = \frac{1}{2}(\alpha_l + \alpha_u). \tag{4.23}$$

5 Avaliação Experimental

Este capítulo apresenta a avaliação experimental realizada para aferir a eficácia de ambos os métodos de expansão propostos em diversos aspectos. Os dois métodos foram avaliados considerando as mesmas coleções de imagens e descritores e os mesmos classificadores com os mesmos parâmetros. A Seção 5.1 apresenta o protocolo experimental, descrevendo a metodologia de avaliação utilizada nos experimentos. A Seção 5.2 descreve as coleções de imagens e descritores considerados. A Seção 5.3 apresenta os classificadores considerados durante a avaliação experimental dos métodos. Na Seção 5.4 são discutidos os experimentos conduzidos para avaliar o impacto do método de aprendizado fracamente supervisionado por correlações em tarefas de classificação. De forma análoga, a Seção 5.5 avalia o impacto do método de aprendizado fracamente supervisionado por hipergrafos em tarefas de classificação. A Seção 5.6 apresenta experimentos de comparação entre ambos os métodos propostos e métodos recentes. Por fim, a Seção 5.7 apresenta uma análise visual do comportamento da expansão de rótulos realizada.

5.1 Protocolo Experimental

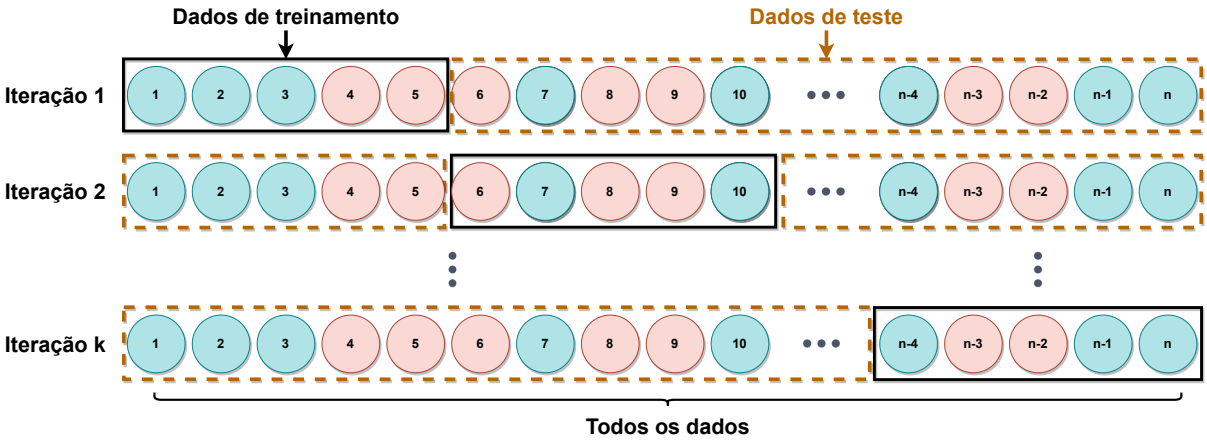
Todos os experimentos foram conduzidos considerando um protocolo de validação cruzada com 10 subconjuntos de dados (*folds*) simulando um cenário fracamente supervisionado, com poucos dados rotulados disponíveis para treinamento. Para cada etapa da validação cruzada um subconjunto é utilizado para treinamento e os nove restantes são utilizados para teste da acurácia do classificador. Ou seja, consideraremos uma proporção de 10% dos dados para treinamento (os dados "rotulados") e 90% para teste. Uma ilustração do protocolo utilizado pode ser observada na Figura 17.

Os resultados de acurácia apresentados são a média de acurácia obtida ao treinar os classificadores com todos os 10 subconjuntos. Considerando os métodos de expansão, a expansão é feita a partir dos conjuntos de 10% dos dados e os dados que vão ser rotulados são parte do conjunto de dados restante. Como os métodos podem expandir rótulos incorretamente, consideraremos os mesmos 90% restantes para teste da acurácia do classificador, sem excluir nenhuma amostra que foi rotulada.

5.2 Coleções de Imagens e Descritores

Esta seção exibe os conjuntos de dados e descritores utilizados durante a avaliação experimental dos métodos de expansão de conjuntos rotulados propostos. Os conjuntos de dados são distintos entre si, de modo que ambos os métodos são avaliados em diferentes

Figura 17 – Exemplo do protocolo de validação cruzada considerado nos experimentos, com pequenos conjuntos de treinamento.



Fonte: Elaborado pelo autor.

cenários. Consideramos 5 coleções de imagens públicas diferentes com tamanho variando de 1.360 a 70.000 imagens, considerando diferentes descritores visuais para cada coleção. As Figuras 18 a 22 ilustram exemplos de imagens de cada uma das coleções de imagem. Cada um dos conjuntos de dados são descritos a seguir:

- **MPEG-7 [74]:** coleção de imagens bastante popular, composta por 1400 imagens de contorno divididas em 70 classes. Para extração das características das imagens da coleção, foram utilizados os descritores visuais ASC - *Aspect Shape Context* [75] e CFD - *Contour Features Descriptor* [76]. Ambos estes descritores fornecem apenas matrizes de distâncias (distância entre as imagens) ao contrário dos vetores de características de cada imagem esperados, por conta disso, durante os experimentos realizados, consideraremos as matrizes de distâncias obtidas por ambos os descritores como as características que utilizaremos para treinamento dos classificadores.

Figura 18 – Exemplos de imagens do conjunto de dados MPEG-7 [74].

Fonte: Elaborado pelo autor.

- ***Flowers* [77]:** coleção de imagens distribuída pela Universidade de Oxford, contém imagens de 17 espécies de flores com 80 amostras para cada espécie, variando ângulo e a iluminação. Consideramos características extraídas das imagens obtidas pelos seguintes descritores visuais: ACC - *Auto Color Correlogram* [32] e CNN-Resnet [78].

Figura 19 – Exemplos de imagens do conjunto de dados *Flowers* [77].

Fonte: Elaborado pelo autor.

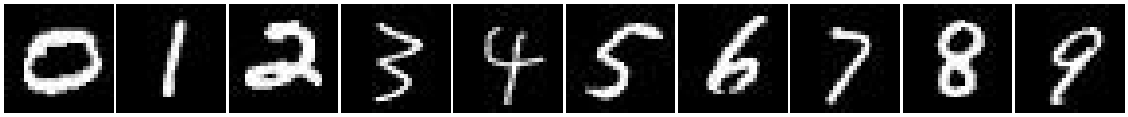
- **Corel5k [79]:** o conjunto de dados possue imagens de diferentes cenários, como fogos de artifício, animais, azulejos, árvores, entre outros. É composto de 5000 imagens divididas em 50 categorias, com 100 imagens cada. Consideramos para esta coleção de imagens os mesmos descritores visuais utilizados na coleção Flowers: ACC - *Auto Color Correlogram* [32] e CNN-Resnet [78].

Figura 20 – Exemplos de imagens do conjunto de dados Corel5k [79].

Fonte: Elaborado pelo autor.

- **MNIST [80]:** conjuntos de dados de dígitos manuscritos, possui um total de 70000 imagens (60000 no conjunto de treinamento e 10000 no conjunto de teste) divididas em 10 categorias de números de 0 a 9. Para esta coleção de imagens, os vetores de características considerados são construídos a partir dos valores brutos dos píxeis de cada uma das imagens.

Figura 21 – Exemplos de imagens do conjunto de dados MNIST [80].



Fonte: Elaborado pelo autor.

- **CUB-200-2011 [81]:** esta coleção de imagens contêm fotos de 200 espécies de pássaros em um total de 11788 imagens. Utilizamos do descritor visual CNN-Resnet [78] para extrair as características das imagens desse conjunto de dados.

Figura 22 – Exemplos de imagens do conjunto de dados CUB-200-2011 [81].

Fonte: Elaborado pelo autor.

A respeito da configuração das listas ranqueadas, apesar de que as listas ranqueadas podem assumir diferentes tamanhos, é indispensável limiar este tamanho a uma constante l de modo a manter ambos os métodos propostos escaláveis e adequados para aplicação em coleções de imagens maiores. Sendo assim, consideramos dois valores diferentes da constante l: $l = 400$ para os conjuntos de dados MPEG-7 [74] e Flowers [77] e $l = 1000$ para os conjuntos Corel5k [79], MNIST [80] e CUB-200-2011 [81].

5.3 Classificadores

Com o intuito de avaliar os métodos de aprendizado fracamente supervisionado propostos nos mais diferentes cenários, selecionamos uma série de classificadores supervisionados e semi-supervisionados. Estes são avaliados quando submetidos ao treinamento com e sem o método de expansão no protocolo de validação cruzada.

5.3.1 Métodos Supervisionados

Nossos experimentos consideraram três classificadores supervisionados: OPF (*Optimum Path Forest*) [45], SVM (*Support Vector Machines*) [42] e kNN (k-*Nearest Neighbors*). Estes são descritos brevemente a seguir.

- **OPF[1] (*Optimum Path Forest*) [45]:** constrói um grafo onde cada elemento é representado por um nó do grafo e o peso das arestas corresponde a distância entre os nós. O algoritmo visa encontrar o caminho ótimo entre os nós para classificá-los em uma determinada classe. Uma vantagem desse classificador em comparação a outros classificadores supervisionados se deve ao fato de que o OPF [45] independe de parâmetros, ou seja, não é necessário para cada cenário encontrar os melhores parâmetros.

- **SVM (*Support Vector Machines*) [42]:** as máquinas de vetores de suporte tentam encontrar a melhor superfície possível que divide os dados a partir dos exemplos que possuem rótulo. Ou seja, a superfície de decisão criada pelo método é colocada em uma posição onde a maior parte das amostras de cada classe ficam separadas entre si. Para implementação do método nos nossos experimentos, utilizamos a biblioteca *scikit-learn* [82], com os seguintes parâmetros: *kernel* polinomial de grau 2, $\gamma = 0.001$ e custo de classificação incorreta (C) igual a 10. O restante dos parâmetros do método são os parâmetros padrões presentes na implementação na biblioteca. Esse conjunto de parâmetros foi estimado a partir da análise da acurácia do classificador quando aplicado as coleções de imagem MPEG-7 [74], Flower [77] e Corel5k [79]. Tentou-se encontrar uma configuração da máquina de vetores de suporte que se

[1] Implementação utilizada foi retirada de: https://github.com/marcoscleison/PyOPF

comportasse relativamente bem nessas três coleções de dados. Há a possibilidade de se testar diversas combinações de parâmetros em cada um dos classificadores considerados para cada coleção de imagens e descritor visual, porém, nesse trabalho será priorizado otimizar os parâmetros de ambos os métodos de expansão propostos, o foco primário, ao invés dos classificadores.

- **kNN (k-*Nearest Neighbors*):** classificador supervisionado amplamente conhecido, que atribui rótulos aos elementos de acordo com seus k vizinhos mais próximos. De maneira similar a SVM [42] utilizada nos experimentos, a implementação utilizada foi feita a partir da biblioteca *scikit-learn* [82], e se utilizou de $k = 20$ nos nossos experimentos, além dos parâmetros padrões da implementação na biblioteca.

5.3.2 Métodos Semi-Supervisionados

Além dos métodos supervisionados, três métodos semi-supervisionados foram considerados na nossa avaliação: SGC (*Simple Graph Convolution*) [83], *Label Spreading* [84] e *Pseudo-Label* [66]. O primeiro dos três métodos foi utilizado de maneira similar aos métodos, tendo sua acurácia avaliada com e sem a utilização dos nossos métodos de expansão de rótulos. Os outros dois métodos são utilizados como comparação aos métodos de expansão combinados com os primeiros quatro classificadores apresentados (OPF [45], SVM [42], kNN e SGC [83]).

- **SGC [2] (*Simple Graph Convolution*) [83]:** implementação simplificada de uma rede convolucional de grafos proposta recentemente, a qual apresenta resultados comparáveis ao estado da arte em muitos cenários. É um classificador que aprende uma estrutura de dependência discreta e esparsa entre os dados (os nós do grafo) enquanto em simultâneo atualiza os pesos das camadas da rede convolucional de grafos. Diferente da maior parte das redes convolucionais de grafo, a SGC [83] remove não-linearidades e agrupa as matrizes entre camadas consecutivas da rede com o intuito de acelerar o modelo. Para todos os experimentos realizados com o classificador, utilizamos dos parâmetros padrões do método disponíveis em sua implementação e para treinamento do classificador consideramos a seguinte configuração: 50 épocas, otimizador *Adam* com taxa de aprendizado (*learning rate*) de 10^{-3}, decaimento de peso (*weight decay*) no valor de $5 \cdot 10^{-4}$ e $k = 40$ para construir os grafos kNN do método.

- ***Label Spreading* [84]:** algoritmo semi-supervisionado similar ao algoritmo de propagação de rótulos (*Label Progagation*) [12], mas que se utiliza de uma matriz de afinidade construída a partir de uma matriz laplaciana normalizada. O método

[2] Implementação utilizada foi retirada de: `https://github.com/rusty1s/pytorch_geometric` [85]

propaga rótulos dos nós rotulados aos seus vizinhos de acordo com sua proximidade. Para implementação do método, utilizamos também da biblioteca *scikit-learn* [82], considerando um *kernel* RBF com $\alpha = 0.4125$, $\gamma = 0.1$ e um número máximo de iterações igual a 100. Este método é utilizado como comparação aos métodos propostos por ter um funcionamento similar de expansão de conjuntos de treinamento.

- ***Pseudo-Label* [66]:** método de aprendizado semi-supervisionado originalmente utilizado com redes neurais profundas. A principal ideia dessa técnica é se utilizar das probabilidades obtidas pelo modelo para atribuir rótulos aos dados não-rotulados. O modelo então é re-treinado utilizando os dados rotulados originais e os dados pseudo rotulados. Apesar desse método ter sido originalmente proposto para ser utilizado em conjunto de redes neurais profundas, ele pode ser aplicado com diferentes classificadores. Sendo assim, nos nossos experimentos utilizamos de uma implementação do método de pseudo-rótulos disponível publicamente[3] em conjunto de regressão logística com gradiente descendente estocástico como classificador. Este classificador foi retirado da bliblioteca *scikit-learn* [82] [4], e os parâmetros $loss = log$ e $\alpha = 0.00001$ foram considerados. Assim como no algoritmo de *Label Spreading* [84], o método de *Pseudo-Label* [66] é utilizado como termo de comparação a ambos os métodos propostos nesse trabalho.

5.4 Método de Expansão por Correlações

Esta seção discute os experimentos conduzidos para avaliar o impacto do método de expansão por correlações proposto em tarefas de classificação. Primeiro, a subseção 5.4.1 faz uma análise da performance do método para diferentes tamanhos de vizinhança nas listas ranqueadas, de forma a encontrar o melhor parâmetro em cada cenário para os experimentos de classificação. Por fim a subseção 5.4.2 apresenta e discute os resultados obtidos por nosso método de aprendizado fracamente supervisionado por correlações nos diferentes classificadores e coleções de imagens apresentados anteriormente (Seções 5.1 e 5.3).

5.4.1 Definição do Tamanho de Vizinhança

Além do limiar ótimo estimado pelo método apresentado na Seção 3.4, cada métrica de correlação $r(\cdot,\cdot)$ é calculada considerando um tamanho de vizinhança k. Com isso em mente, foi conduzida uma análise experimental com o intuito de avaliar o impacto da

[3] Implementação utilizada foi retirada de: https://github.com/anirudhshenoy/pseudo_labeling_small_datasets

[4] O classificador utilizado pode ser encontrado como: **sklearn.linear_model.SGDClassifier** ("*Logistic Regression with Stochastic Gradient Descent* (SGD)")

variação do tamanho de vizinhança nos resultados de acurácia e por consequência definir um valor de k adequado para diferentes coleções de imagens e características.

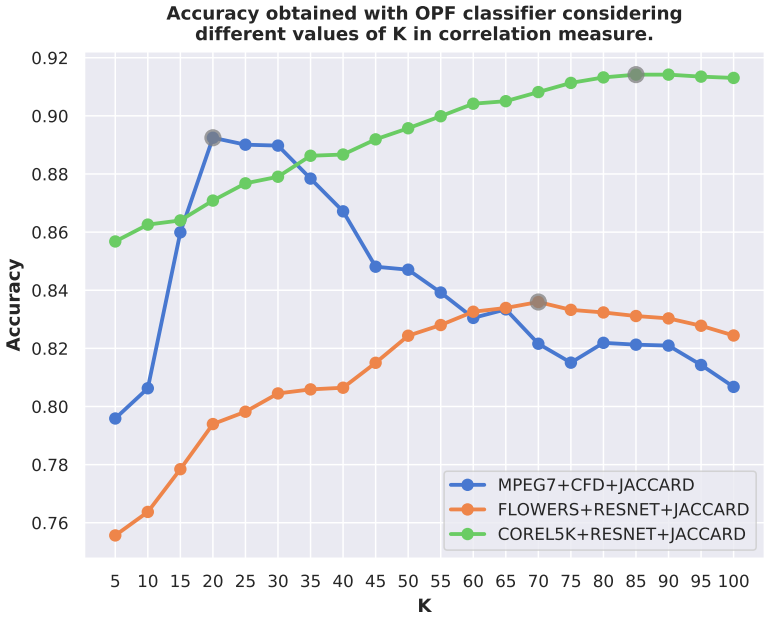
Nesses experimentos consideramos três das cinco coleções de imagens apresentadas anteriormente: MPEG-7 [74] com o descritor CFD [76], *Flowers* [77] com o descritor CNN-Resnet [78] e por último a coleção Corel5k [79] considerando as características obtidas com o descritor CNN-Resnet [78]. As listas ranqueadas consideradas nesses experimentos foram obtidas utilizando o método de re-ranqueamento LHRR [19] implementado no *framework* UDLF (*Unsupervised Distance Learning Framework*) [86], onde nós consideramos $l = |\mathfrak{X}|$ (l igual à quantidade de elementos da coleção) além dos parâmetros padrões da implementação do método no *framework*. Para esses experimentos, foi escolhida a métrica para *Jaccard* [72] como base para o método de expansão de rótulos. Quanto aos classificadores, avaliamos a acurácia de três classificadores supervisionados apresentados anteriormente: OPF [45], SVM [42] e kNN, para diferentes valores de vizinhança. Em todos os experimentos, utilizamos de validação cruzada com 10 *folds*, onde consideramos uma proporção de 10%/90% dos dados para os conjuntos de treino e teste, respectivamente. Foram avaliados tamanhos de vizinhança $k \in [5, 100]$, com incrementos de 5.

As Figuras de 23 a 25 apresentam os resultados obtidos com cada um dos três classificadores. A maior acurácia obtida para cada combinação de coleção de imagens, descritor e métrica de correlação em cada uma das figuras está destacada. De modo a facilitar a leitura, a Tabela 1 apresenta os valores de tamanho de vizinhança que apresentaram os melhores resultados em cada combinação de coleção de imagens e descritor visual. Os valores de k destacados nas Figuras de 23 a 25 são bastante similares ao tamanho de cada classe em cada um dos três conjuntos de dados considerados: 20 no MPEG-7 [74], 80 no *Flowers* [77] e 100 no Corel5k [79]. Portanto, no restante dos experimentos com o método de expansão baseado em correlações, será considerado o valor k como sendo o tamanho da classe da coleção de imagens que é considerada. Caso o conjunto de dados não tenha um tamanho de classe fixo, a parte inteira do tamanho médio das classes é considerada como o valor de k. Situação a qual acontece na coleção CUB-200-2011 [81], onde usamos $k = 58$. A única exceção a essa regra é o conjunto de dados MNIST [80], onde o tamanho de cada uma das classes varia de 5421 a 6742 imagens, e por conta do tamanho elevado das classes dessa coleção, é considerado o maior valor utilizado entre os outros conjuntos de dados, $k = 100$.

5.4.2 Resultados de Classificação

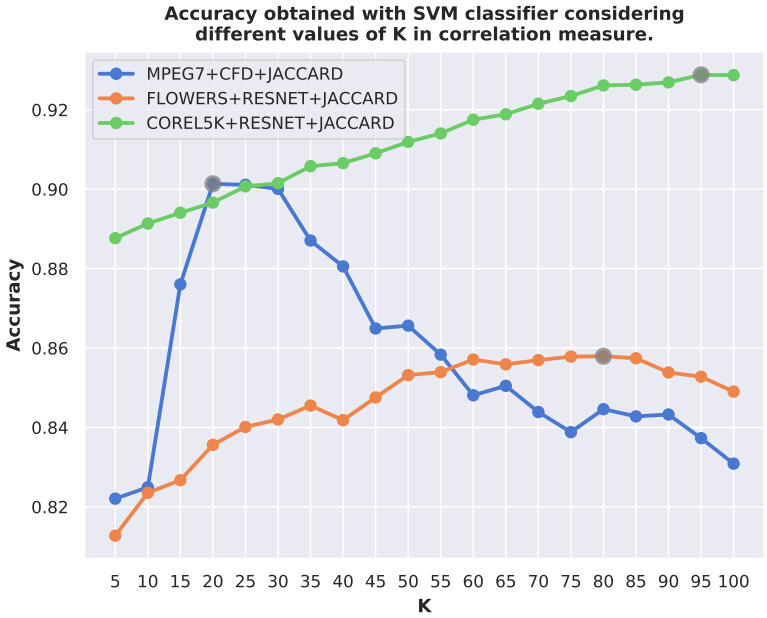
Esta subseção apresenta os resultados de classificação obtidos pelos classificadores supervisionados e semi-supervisionados apresentados na Seção 5.3 quando combinados ao método proposto de expansão de conjuntos de treinamento baseado em correlações. Para cada classificador são apresentados dois resultados diferentes: o primeiro considera o

Figura 23 – Avaliação do tamanho de vizinhança considerando o classificador OPF [45].



Fonte: Elaborado pelo autor.

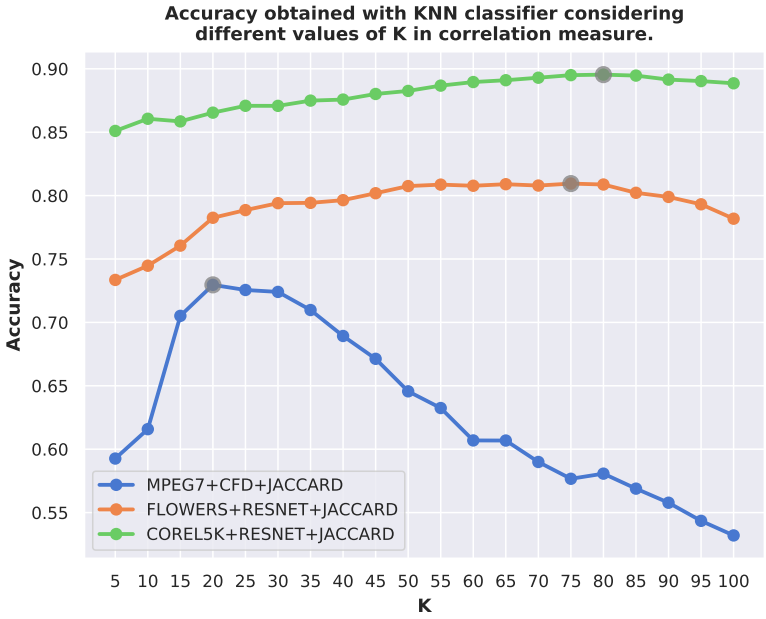
Figura 24 – Avaliação do tamanho de vizinhança considerando o classificador SVM [42].



Fonte: Elaborado pelo autor.

cenário de treinamento dos classificadores com poucas amostras rotuladas (um dos 10 *folds*, equivalente a 10% dos dados), e o segundo resultado irá considerar a acurácia obtida pelos classificadores quando submetidos a treinamento com o conjunto expandido pelo método baseado em correlações. De modo a manter o mesmo protocolo de avaliação para todas as coleções de imagens, todas as amostras são consideradas para gerar todos os 10 *folds*

Figura 25 – Avaliação do tamanho de vizinhança considerando o classificador kNN.



Fonte: Elaborado pelo autor.

Tabela 1 – Melhores tamanhos de vizinhança k obtidos para cada combinação de coleção de imagens e descritor.

		MPEG-7	Flowers	Corel5k
		CFD	CNN-Resnet	CNN-Resnet
WS-OPF		20	70	85
WS-SVM	***Jaccard***	20	80	95
WS-kNN		20	75	80
Tamanho de classe		**20**	**80**	**100**

Fonte: Elaborado pelo autor.

que são utilizados durante os experimentos, mesmo que o conjunto de dados tenha um conjunto de treino e teste especificado previamente. Também é avaliada a acurácia obtida considerando cada uma das métricas de correlação de listas ranqueadas apresentadas na Seção 3.2.

A Tabela 2 apresenta os resultados obtidos para cada coleção de imagens, características, método de re-ranqueamento e métrica de correlação considerando o classificador OPF [45]. Em cada cenário, considerando as listas ranqueadas geradas a partir dos vetores de características e as listas re-ranqueadas pelos métodos LHRR [19] e BFS-Tree [20], a maior acurácia obtida em cada combinação de conjunto de dados e descritor está destacada em negrito. A maior acurácia obtida considerando todos os cenários para cada combinação de coleção de imagens e descritor está destacada em vermelho. Para ambos os métodos de re-ranqueamento, utilizamos a implementação presente no *framework* UDLF (*Unsupervised*

Distance Learning Framework) [86] para gerar as novas listas ranqueadas. Consideramos os parâmetros padrões presentes nas implementações dos métodos com exceção do parâmetro l, em que consideramos $l = |\mathfrak{X}|$ para as coleções MPEG-7 [74], *Flowers* [77] e Corel5k [79], $l = 4000$ para a coleção CUB200-2011 [81] e $l = 3000$ para a coleção MNIST [80]. Além disso, na tabela é apresentado o ganho relativo entre a acurácia obtida pelo classificador sem expansão e combinado ao método de expansão. A média entre os resultados de acurácia e o maior ganho relativo obtido em todos os cenários também são apresentados. Note que o método apresentado quando combinado ao classificador OPF [45] obtêm resultados melhores do que o classificador sem a sua utilização.

Quanto ao cenário sem a utilização de listas ranqueadas re-ranqueadas (**WS-OPF**), as métricas de correlação *Intersection*, $Jaccard_k$ e $Kendall\tau$ apresentaram os melhores resultados em cada uma das coleções de imagens e a métrica *Intersection* apresentou os melhores resultados em média, apresentando um ganho relativo de +9,50% comparado a média do classificador sem a utilização do nosso método (de 63,61% para 69,65%). Considerando o cenário com as listas ranqueadas processadas pelo método LHRR [19] (**WS-OPF+LHRR**), as métricas *Intersection*, *Jaccard* e $Jaccard_k$ foram responsáveis pelos melhores resultados. Nesse contexto, o melhor resultado em média foi obtido através da utilização da métrica *Jaccard*, apresentando um ganho relativo de +12,37% em comparação com o caso padrão (acurácia média de 63,61% para 71,48%). Por fim, para os resultados obtidos com o método de re-ranqueamento BFS-Tree [20] (**WS-OPF+BFS-Tree**) todas as cinco correlações atingiram os melhores resultados em pelo menos uma combinação de conjunto de dados e descritor visual. Similar ao cenário com o método LHRR [19], a métrica *Jaccard* atingiu os melhores resultados em média, com um ganho relativo de +11,82% comparado ao cenário padrão (63,61% de acurácia para 71,13%).

Considerando os resultados apresentados podemos observar que praticamente todos os resultados obtidos pelo classificador quando utilizamos das listas ranqueadas re-ranqueadas são superiores a utilizar as listas originais. Tal observação confirma que fazer o uso de listas ranqueadas de melhor qualidade, geradas por métodos de re-ranqueamento implicam em uma melhor qualidade das expansões, o que por consequência gera melhores resultados de classificação. Essa melhor qualidade nas expansões será melhor explorada posteriormente na subseção 5.7, onde realizamos experimentos de visualização do comportamento das expansões.

De maneira similar, a Tabela 3 apresenta os resultados obtidos para cada coleção, descritor, correlação e método de re-ranqueamento para o classificador SVM [42]. Os resultados obtidos pelo classificador em combinação com nosso método de expansão baseado em correlações atingiu resultados de acurácia maiores do que o cenário sem expansão na maior parte dos casos, com exceção para alguns cenários com as características extraídas pelo descritor ACC [32]. Situação que pode ser explicada por conta da pior

Tabela 2 – Acurácia (%) obtida pelo classificador OPF [45] para cada coleção de imagens, descritor com e sem a utilização do método de expansão baseado em correlações. São também reportados os resultados obtidos pelo método de expansão em conjunto dos métodos de re-ranqueamento LHRR [19] e BFS-Tree [20].

		MPEG-7		Flowers		Corel5k		MNIST	CUB200	Média
		ASC	CFD	ACC	Resnet	ACC	Resnet	-	Resnet	
OPF		82,95	67,75	30,54	71,77	40,21	83,56	93,49	38,59	63,61
WS-OPF	**Intersection**	**88,91**	**86,21**	31,70	78,93	42,25	89,21	**95,45**	44,50	**69,65**
	Jaccard	84,71	77,56	31,31	76,63	41,49	88,08	95,11	43,13	67,25
	Jaccard$_k$	86,84	82,75	31,99	**79,25**	**42,58**	**89,22**	95,10	**45,13**	69,11
	Kendallτ	87,99	84,87	**32,38**	78,70	41,99	89,16	95,22	45,05	69,42
	RBO	87,07	82,20	31,45	78,74	41,53	87,52	94,94	44,42	68,48
Ganho Relativo Médio		+5,01	+22,09	+4,01	+9,31	+4,37	+6,08	+1,79	+15,17	+8,48
WS-OPF + LHRR	**Intersection**	**90,68**	88,13	30,60	81,99	43,64	91,16	**96,46**	45,09	70,97
	Jaccard	89,15	**89,25**	**31,27**	**83,24**	**45,06**	**91,30**	96,16	46,44	**71,48**
	Jaccard$_k$	88,15	82,67	30,85	81,36	44,30	90,32	95,72	**46,52**	69,99
	Kendallτ	88,40	84,04	30,88	82,60	44,54	90,65	96,14	44,63	70,24
	RBO	87,79	82,40	30,39	77,91	42,72	86,72	95,19	46,30	68,68
Ganho Relativo Médio		+7,09	+25,90	+0,84	+13,45	+9,55	+7,74	+2,61	+18,67	+10,73
WS-OPF + BFS-Tree	**Intersection**	89,79	86,94	30,23	82,32	43,55	**91,58**	**96,34**	45,36	70,76
	Jaccard	**90,62**	**87,34**	31,00	82,22	**44,66**	91,38	95,93	45,86	**71,13**
	Jaccard$_k$	87,91	81,75	**31,62**	81,87	44,52	91,10	95,58	46,15	70,06
	Kendallτ	88,22	82,00	31,00	**82,67**	44,72	91,36	95,87	43,68	69,94
	RBO	87,28	80,13	30,31	77,46	42,51	86,71	95,11	**46,39**	68,24
Ganho Relativo Médio		+7,01	+23,44	+0,96	+13,29	+9,41	+8,22	+2,43	+17,88	+10,33
Ganho Relativo Máximo		+9,32	+31,73	+6,02	+15,98	+12,06	+9,60	+3,18	+20,55	+13,56

Fonte: Elaborado pelo autor.

qualidade do descritor se comparado ao CNN-Resnet [78], o qual apresentou resultados positivos em ambos os conjuntos de dados MPEG-7 [74] e Corel5k [79]. A baixa acurácia do descritor ACC [32] pode resultar em mais falsos positivos durante a etapa de expansão, o que gera resultados abaixo do esperado pelo classificador. Para o cenário com as listas ranqueadas padrões (**WS-SVM**), nosso método obteve as melhores acurácias com as métricas *Intersection* e RBO, com a métrica *Intersection* obtendo o melhor resultado em média, apresentando um ganho relativo de +5,31% comparado ao cenário sem expansão (68,79% para 72,44%). Considerando o re-ranqueamento pelo método LHRR [19] (**WS-SVM+LHRR**), as métricas *Intersection* e *Jaccard* atingiram os melhores resultados nas coleções de imagens, e em média, a métrica *Jaccard* se saiu melhor, com um ganho relativo de +7,37% em comparação com a acurácia média do classificador no cenário sem expansão (68,79% de acurácia para 73,86%). Resultados que confirmam novamente que listas ranqueadas de melhor qualidade geradas pelos métodos de re-ranqueamento resultam em melhores acurácias. Para o re-ranqueamento com o método BFS-Tree [20] (**WS-SVM+BFS-Tree**) as métricas de correlação *Intersection*, *Jaccard* e *Kendallτ* os melhores resultados em cada coleção de imagens. A métrica *Jaccard* apresentou os melhores

resultados em média, com um ganho relativo de +7,20% com relação ao cenário padrão (68,79% para 73,74%).

Tabela 3 – Acurácia (%) obtida pelo classificador SVM [42] para cada coleção de imagens, descritor com e sem a utilização do método de expansão baseado em correlações. São também reportados os resultados obtidos pelo método de expansão em conjunto dos métodos de re-ranqueamento LHRR [19] e BFS-Tree [20].

		MPEG-7		**Flowers**		**Corel5k**		**MNIST**	**CUB200**	**Média**
		ASC	**CFD**	**ACC**	**Resnet**	**ACC**	**Resnet**	**-**	**Resnet**	
SVM		83,12	68,56	37,50	80,65	45,27	88,33	95,34	51,51	68,79
WS-SVM	**Intersection**	**89,30**	**87,60**	35,65	83,10	44,97	**91,51**	95,62	51,74	**72,44**
	Jaccard	85,72	79,61	35,13	81,40	44,23	90,77	95,37	50,78	70,38
	Jaccard$_k$	87,15	84,61	35,90	83,24	45,22	91,38	95,64	**52,29**	71,93
	Kendallτ	88,18	86,25	36,00	82,69	44,88	91,36	95,58	51,69	72,08
	RBO	87,32	83,95	35,48	**83,50**	44,69	90,26	**95,80**	51,67	71,58
Ganho Relativo Médio		+5,31	+23,11	-4,98	+2,65	-1,04	+3,09	+0,27	+0,24	+3,58
WS-SVM + LHRR	**Intersection**	**91,05**	89,94	34,53	84,72	46,40	92,39	**96,91**	50,78	73,34
	Jaccard	90,21	**90,13**	34,54	**85,79**	**47,85**	**92,87**	96,88	**52,58**	**73,86**
	Jaccard$_k$	88,56	84,77	34,68	84,28	47,00	91,50	96,62	52,18	72,45
	Kendallτ	88,83	85,96	34,66	85,10	46,95	91,91	96,81	50,84	72,63
	RBO	88,04	84,54	33,51	83,08	45,55	89,32	96,21	51,99	71,53
Ganho Relativo Médio		+7,48	+27,00	-8,31	+4,89	+3,27	+3,70	+1,41	+0,32	+4,97
WS-SVM + BFS-Tree	**Intersection**	90,12	**88,63**	34,16	84,82	46,69	93,08	**96,87**	51,46	73,23
	Jaccard	**91,34**	88,57	34,46	85,19	47,60	**93,21**	96,73	**52,81**	**73,74**
	Jaccard$_k$	88,31	83,70	35,75	84,73	47,45	92,36	96,54	52,58	72,68
	Kendallτ	88,22	83,71	35,20	**85,45**	**47,62**	92,76	96,69	50,28	72,49
	RBO	87,48	82,38	34,49	82,87	45,76	89,35	96,11	52,68	71,39
Ganho Relativo Médio		+7,19	+24,56	-7,17	+4,91	+3,87	+4,33	+1,31	+0,88	+4,98
Ganho Relativo Máximo		+9,89	+31,46	-4,00	+6,37	+5,70	+5,52	+1,65	+2,52	+7,39

Fonte: Elaborado pelo autor.

A Tabela 4 apresenta a mesma análise feita nas duas tabelas anteriores, porém considerando o classificador kNN. Por ser um classificador mais simples, os resultados obtidos pelo classificador kNN, principalmente no cenário padrão quando realizamos o treinamento com somente 10% dos dados, são inferiores aos obtidos com os classificadores OPF [45] e SVM [42], que tiveram seus resultados apresentados anteriormente. Porém, quando o conjunto de treinamento é expandido por nosso método, ganhos muito positivos são obtidos. Quanto ao cenário sem re-ranqueamento (**WS-kNN**), os melhores resultados foram obtidos pela expansão com as métricas de correlação *Intersection*, *Jaccard$_k$* e *Kendallτ*, e a métrica *Intersection* obteve o melhor resultado em média, apresentando um ganho relativo de +46,73% em comparação com a média de acurácia no cenário padrão (44,75% para 65,66%). Para o re-ranqueamento com o método LHRR [19] (**WS-kNN+LHRR**), o método atingiu os melhores resultados com as métricas *Intersection*, *Jaccard* e *Jaccard$_k$*. Em média, a métrica que melhor se saiu nesse cenário foi a *Jaccard*, a qual apresentou um ganho relativo de +49,56% ao cenário sem expansão (44,75% para

66,93%). Finalmente, para o método BFS-Tree [20] (**WS-kNN+BFS-Tree**), as métricas *Intersection*, *Jaccard*, $Jaccard_k$ e $Kendall\tau$ apresentaram os melhores resultados, com a métrica *Jaccard* apresentando a maior média de acurácia com ganho relativo de +49,38% em relação ao cenário sem expansão (44,75% de acurácia para 66,85%). Novamente podemos observar que a utilização de métodos de re-ranqueamento implica em melhores resultados de acurácia em praticamente todos os cenários.

Tabela 4 – Acurácia (%) obtida pelo classificador kNN para cada coleção de imagens, descritor com e sem a utilização do método de expansão baseado em correlações. São também reportados os resultados obtidos pelo método de expansão em conjunto dos métodos de re-ranqueamento LHRR [19] e BFS-Tree [20].

		MPEG-7		**Flowers**		**Corel5k**		**MNIST**	**CUB200**	**Média**
		ASC	**CFD**	**ACC**	**Resnet**	**ACC**	**Resnet**	**-**	**Resnet**	
kNN		13,92	12,39	28,47	63,67	34,05	76,80	92,04	36,67	44,75
WS-kNN	**Intersection**	**78,94**	**71,03**	32,66	74,37	41,41	**87,52**	**95,18**	44,17	**65,66**
	Jaccard	66,63	58,60	33,10	71,58	40,75	84,96	94,55	42,70	61,61
	$Jaccard_k$	75,00	66,97	33,58	77,30	**41,94**	87,10	94,47	45,58	65,24
	$Kendall\tau$	76,57	68,77	**33,67**	76,36	41,66	87,09	94,62	**47,32**	65,76
	RBO	74,46	64,93	33,46	**78,06**	41,02	86,49	94,81	44,07	64,66
Ganho Relativo Médio		+433,91	+433,17	+16,94	+18,63	+21,46	+12,80	+2,92	+22,08	+120,24
WS-kNN + LHRR	**Intersection**	**81,45**	72,65	29,80	78,83	39,68	88,84	**95,98**	43,67	66,36
	Jaccard	78,52	**72,96**	30,96	**80,87**	**42,18**	88,86	95,65	45,44	**66,93**
	$Jaccard_k$	74,45	66,31	31,05	79,99	41,91	89,27	95,35	**47,12**	65,68
	$Kendall\tau$	75,38	67,69	**31,36**	80,85	41,62	**89,32**	95,71	46,41	66,04
	RBO	72,42	65,14	30,22	76,81	41,20	86,65	95,03	46,18	64,21
Ganho Relativo Médio		+449,17	+456,50	+7,76	+24,82	+21,35	+15,35	+3,81	+24,80	+125,44
WS-kNN + BFS-Tree	**Intersection**	80,21	**72,14**	28,73	78,19	39,73	88,85	**95,84**	44,05	65,97
	Jaccard	**82,13**	71,99	31,14	79,49	41,64	89,21	95,37	43,79	**66,85**
	$Jaccard_k$	73,56	64,97	32,02	80,11	**41,94**	89,66	95,14	**46,58**	65,50
	$Kendall\tau$	73,96	65,60	**31,71**	**80,35**	41,80	**89,68**	95,39	46,21	65,59
	RBO	68,16	61,56	31,14	76,36	41,20	86,42	94,93	46,26	63,25
Ganho Relativo Médio		+443,13	+442,79	+8,70	+23,92	+21,18	+15,58	+3,58	+23,75	+122,83
Ganho Relativo Máximo		+490,01	+488,86	+18,26	+27,01	+23,88	+16,77	+4,28	+29,04	+137,27

Fonte: Elaborado pelo autor.

Por fim, a Tabela 5 apresenta os resultados obtidos pelo classificador SGC [83]. Diferente dos outros três classificadores que tiveram seus resultados apresentados anteriormente, a acurácia obtida pelo classificador SGC [83] varia de exucução para execução. Por conta disso, os resultados reportados na Tabela 5 são a média de acurácia obtida em 5 execuções do protocolo de validação cruzada de 10 *folds*. Apesar dos resultados ruins obtidos na coleção MPEG-7 [74] e os baixos valores de acurácia na coleção MNIST [80] se comparados com os outros classificadores, o classificador SGC [83] apresentou resultados extremamente positivos nas coleções de imagens *Flowers* [77], Corel5k [79] e CUB200-2011 [81]. Os valores ruins de acurácia obtidos pelo classificador na coleção MPEG-7 [74] podem ser explicados pelo uso das distâncias entre as imagens como as características em ambos os descritores ASC [75] e CFD [76].

Para o cenário com a utilização das listas ranqueadas originais (**WS-SGC**), todas as cinco métricas de correlação apresentaram os melhores resultados em pelo menos uma combinação de coleção de imagens e descritor visual. A métrica RBO se saiu melhor em média, apresentando um ganho relativo de +8,13% em comparação com não utilizar do método de expansão (48,31% de acurácia para 52,24%). Para as listas re-ranqueadas pelo método LHRR [19] (**WS-SGC+LHRR**) as métricas *Intersection*, *Jaccard*, *Kendall*τ e RBO obtiveram os melhores resultados, com a métrica *Kendall*τ apresentando a melhor média, com ganho relativo de +9,04% em comparação ao cenário sem expansão (48,31% para 52,68%). Finalmente, considerando os resultados obtidos com o método BFS-Tree [20] (**WS-SGC+BFS-Tree**), todas as correlações com exceção da métrica *Jaccard*$_k$ apresentaram os melhores resultados em pelo menos uma combinação de coleção de imagens e descritor. A métrica *Kendall*τ foi a que melhor se comportou em média com ganho relativo de +8,80% para o cenário padrão (48,31% de acurácia obtido para 52,56%). Novamente confirmamos que utilizar dos métodos de re-ranqueamento é benéfico aos resultados de acurácia na maioria dos cenários.

Tabela 5 – Acurácia (%) obtida pelo classificador SGC [83] para cada coleção de imagens, descritor com e sem a utilização do método de expansão baseado em correlações. São também reportados os resultados obtidos pelo método de expansão em conjunto dos métodos de re-ranqueamento LHRR [19] e BFS-Tree [20].

		MPEG-7		**Flowers**		**Corel5k**		**MNIST**	**CUB200**	**Média**
		ASC	**CFD**	**ACC**	**Resnet**	**ACC**	**Resnet**	**-**	**Resnet**	
SGC		1,88	3,25	31,33	80,02	42,37	89,74	90,02	47,90	48,31
WS-SGC	**Intersection**	2,10	3,12	38,34	**85,29**	50,50	92,95	90,25	54,53	52,14
	Jaccard	2,05	**3,71**	37,60	84,01	50,40	92,88	90,02	53,83	51,81
	Jaccard$_k$	**2,20**	3,61	38,69	85,01	50,49	93,12	90,15	54,52	52,22
	Kendallτ	1,95	3,26	38,92	84,76	**50,54**	**93,26**	90,08	54,46	52,15
	RBO	2,11	3,10	**39,17**	85,13	49,79	92,26	**90,34**	**56,05**	**52,24**
Ganho Relativo Médio		+10,74	+3,38	+23,03	+6,02	+18,82	+3,51	+0,16	+14,15	+9,98
WS-SGC + LHRR	**Intersection**	**2,01**	3,27	38,09	85,76	50,39	**93,62**	90,60	55,10	52,36
	Jaccard	1,94	3,36	38,64	85,55	50,92	92,85	**90,66**	54,82	52,34
	Jaccard$_k$	1,82	3,34	39,14	85,74	50,75	93,29	90,52	55,61	52,53
	Kendallτ	1,95	3,46	**38,92**	**86,33**	**51,03**	93,42	90,54	55,75	**52,68**
	RBO	1,85	**3,50**	38,37	84,57	49,88	92,01	90,63	**55,81**	52,08
Ganho Relativo Médio		+1,81	+4,18	+23,31	+6,96	+19,41	+3,68	+0,63	+15,70	+9,46
WS-SGC + BFS-Tree	**Intersection**	1,88	3,38	38,08	85,66	50,78	**93,52**	90,59	55,48	52,42
	Jaccard	1,87	3,59	**39,31**	84,93	50,55	92,93	**90,60**	54,28	52,26
	Jaccard$_k$	1,80	3,64	38,59	85,66	50,33	93,19	90,48	55,47	52,40
	Kendallτ	1,89	3,24	39,26	**85,95**	**50,86**	93,22	90,55	55,52	**52,56**
	RBO	1,86	**3,68**	37,89	84,75	49,60	92,14	90,32	**55,96**	52,03
Ganho Relativo Médio		-1,06	+7,88	+23,29	+6,71	+19,01	+3,63	+0,54	+15,54	+9,44
Ganho Relativo Máximo		+17,02	+14,15	+25,47	+7,89	+20,44	+4,32	+0,71	+17,01	+13,38

Fonte: Elaborado pelo autor.

5.5 Método de Expansão por Hipergrafos

Esta seção apresenta os resultados obtidos com o método de aprendizado fracamente supervisionado por hipergrafos combinado com classificadores supervisionados e semi-supervisionados nas coleções de imagens apresentadas anteriormente. A subseção 5.5 descreve e discute os resultados obtidos.

5.5.1 Resultados de Classificação

Esta subseção exibe os resultados obtidos pelo método de expansão baseado em hipergrafos quando combinado com os classificadores supervisionados e semi-supervisionados apresentados na Seção 5.3. O protocolo experimental é o mesmo utilizado na avaliação do método de expansão baseado em correlações, considerando validação cruzada com 10 *folds* para avaliação da acurácia dos classificadores. Para cada classificador são reportados dois resultados, um considerando a acurácia obtida pelos classificadores com 10% dos dados (1 *fold*), e outra considerando o treinamento com os conjuntos rotulados expandidos pelo método criados a partir de um *fold*.

A Tabela 6 apresenta os resultados de acurácia obtidos em nossos experimentos com o método de expansão baseado em hipergrafos. Primeiro, são apresentados os resultados dos classificadores kNN, OPF [45], SVM [42] e SGC [83] em todas as coleções de imagens apresentadas na subseção 5.2 sem a utilização do método de expansão de conjuntos rotulados, ou seja, treinamento somente com 10% dos dados. Em seguida são apresentados os resultados obtidos pelas versões fracamente supervisionadas de cada um dos classificadores citados anteriormente, se utilizando de diferentes valores de k_e, que é a profundidade das hiperarestas analisadas durante o método de expansão. Para implementação do método de expansão baseado em hipergrafos, as estruturas do método LHRR [19] que utilizamos para realizar essa expansão foram construídas considerando o seguinte conjunto de parâmetros: $l = |\mathfrak{X}|$ para os conjuntos de dados MPEG-7 [74], *Flowers* [77] e Corel5k [79], $l = 3000$ para a coleção MNIST [80] e $l = 4000$ para o conjunto CUB200-2011 [81], $k = 18$ e $T = 2$. Considerando os parâmetros do método de expansão, foram avaliados dois valores de k_e diferentes, $k_e = 18$, o mesmo que utilizamos para construir as estruturas do método LHRR [19] e k_e igual à quantidade de elementos de cada classe do conjunto de dados ou a parde inteira do tamanho médio das classes caso a coleção de imagens caso tenha classes de tamanho variável. A única exceção será para a coleção de imagens MNIST [80], em que a quantidade de elementos de cada classe varia de 5421 a 6742 imagens. Sendo assim, são considerados $k_e = 20$ para a coleção MPEG-7 [74], $k_e = 80$ para a coleção *Flowers* [77], $k_e = 100$ para o conjunto Corel5k [79], $k_e = 58$ para o conjunto CUB200-2011 [81] e $k_e = 100$ para a coleção MNIST [80], o maior valor de k_e utilizado nos outros conjuntos de dados, $k_e = 100$. Para cada um dos resultados com expansão, também são apresentados o ganho relativo médio obtido pelo treinamento com expansão comparado ao treinamento

sem utilizar do método de expansão. A média entre os resultados obtidos e o maior ganho relativo obtido pelas expansões também são destacados. Os melhores de resultados de acurácia obtidos por cada um dos classificadores combinados ao método de expansão em cada combinação de coleção de imagens e descritor estão destacados em negrito, e a maior acurácia obtida entre todos os cenários para cada combinação de conjunto de dados e descritor está destacada em vermelho. Note que os resultados de classificação obtidos pelos classificadores quando combinados ao método de expansão baseado em hipergrafos são em sua maioria melhores do que os resultados obtidos pelos classificadores sem a utilização do método.

Considerando o classificador kNN (**LHRR-kNN**), um classificador mais simples, os resultados apresentados na maior parte dos conjuntos de dados são bastante positivos, situação evidenciada pelos valores de ganho relativo médio observados, onde na coleção de imagens MPEG-7 [74] os ganhos apresentados são muito positivos. Nesse cenário utilizar de valores maiores de k_e, ou seja, o k_e igual ao tamanho de classe se mostrou superior, apresentando um ganho relativo de +47,75% se comparado a não utilização do método de expansão (acurácia de 44,75% para 66,12%).

Para a versão fracamente supervisionada do classificador OPF [45] (**LHRR-OPF**) novamente os melhores resultados ocorrem quando utilizamos k_e igual ao tamanho de classe de cada coleção de imagens, apresentando ganhos em todas as combinações de coleções de imagens e descritores. Os resultados obtidos são superiores ao classificador kNN, apresentando um ganho relativo +11,40% com relação ao OPF [45] padrão (63,61% para 70,86%).

Considerando o classificador SVM [42] (**LHRR-SVM**), os resultados são positivos em praticamente todos os cenários com exceção quando utilizamos do descritor ACC [32] na coleção de imagens *Flowers* [77], situação que também aconteceu na avaliação do método de expansão baseado em correlações, que pode ser observada na Tabela 3. Esta piora na acurácia pode ser explicada pela baixa acurácia do descritor, o que pode adicionar falsos positivos ao conjunto expandido. No mesma coleção de imagens, quando utilizamos do descritor CNN-Resnet [78], de maior qualidade, o classificador apresentou resultados positivos. As duas variações do parâmetro k_e apresentaram resultados positivos e utilizar de k_e igual ao tamanho de classe se mostrou superior na maior parte dos cenários, apresentando em média o melhor resultado entre todos os classificadores com um ganho relativo de +6,08% com relação ao cenário padrão (68,79% de acurácia para 72,97%). Podemos também observar que a combinação do classificador SVM [42] com o método de expansão baseado em hipergrafos apresentou os melhores resultados de acurácia entre todos os classificadores avaliados nas coleções MPEG-7 [74], *Flowers* [77] com o descritor CNN-Resnet [78] e MNIST [80].

Para o classificador SGC [83] (**LHRR-SGC**). A avaliação do classificador foi

realizada do mesmo modo quando avaliamos o método de expansão baseado em correlações, ou seja, os resultados reportados são a média de 5 execuções da validação cruzada com 10 *folds*, considerando a proporção de 10% dos dados para treinamento e os outros 90% para testes. O classificador apresentou resultados ruins na coleção de dados MPEG-7 [74], estes que podem ser explicados pela utilização de matrizes de distâncias como as características para treinamento dos dois descritores ASC [75] e CFD [76]. Apesar disso, o classificador apresentou bons resultados, principalmente ao utilizarmos do parâmetro k_e igual ao tamanho de classe, onde apresentou os melhores resultados entre todos os outros classificadores nas coleções de imagens *Flowers* [77] com o descritor ACC [32], Corel5k [79] e CUB200-2011 [81]. Com relação a não utilização do método de expansão, a versão fracamente supervisionada do classificador SGC [83] quando utilizamos de k_e igual ao tamanho da classe obteve um ganho relativo de +7,28%, aumentando o valor de acurácia de 48,31% para 51,83%.

Tabela 6 – Acurácia (%) obtida pelos classificadores kNN, OPF [45], SVM [42] e SGC [83] para cada coleção de imagens, descritor com e sem a utilização do método de expansão baseado em hipergrafos. Foram considerados valores de $k_e = 18$ e k_e igual ao tamanho de cada classe.

		MPEG-7		Flowers		Corel5k		MNIST	CUB200	Média
		ASC	CFD	ACC	Resnet	ACC	Resnet	-	Resnet	
kNN		13,92	12,39	28,47	63,67	34,05	76,80	92,04	36,67	44,75
OPF		82,95	67,75	30,54	71,77	40,21	83,56	93,49	38,59	63,61
SVM		83,12	68,56	37,50	80,65	45,27	88,33	95,34	51,51	68,79
SGC		1,88	3,25	31,33	80,02	42,37	89,74	90,02	47,90	48,31
LHRR-kNN	$k_e = 18$	73,80	64,11	32,25	77,90	41,80	87,20	95,44	46,53	64,88
	$k_e =$ **tam(classe)**	**75,18**	**65,50**	**32,86**	**80,79**	**42,06**	**89,08**	**95,90**	**47,60**	**66,12**
Ganho Relativo Médio		+435,13	+423,04	+14,35	+24,62	+23,14	+14,77	+3,94	+28,35	+120,92
LHRR-OPF	$k_e = 18$	89,37	83,76	31,61	79,06	43,96	88,21	95,82	45,78	69,70
	$k_e =$ **tam(classe)**	**89,87**	**84,58**	**31,70**	**82,65**	**44,48**	**90,04**	**96,55**	**47,03**	**70,86**
Ganho Relativo Médio		+8,04	+24,24	+3,65	+12,66	+9,97	+6,66	+2,88	+20,25	+11,04
LHRR-SVM	$k_e = 18$	90,01	85,60	34,67	83,97	**46,94**	90,49	96,61	**52,55**	72,61
	$k_e =$ **tam(classe)**	**90,28**	**86,25**	35,02	**84,78**	46,75	**91,39**	**97,05**	52,27	**72,97**
Ganho Relativo Médio		+8,45	+25,33	-7,08	+4,62	+3,48	+2,95	+1,56	+1,75	+5,13
LHRR-SGC	$k_e = 18$	1,82	2,73	37,36	82,30	48,26	91,67	**91,46**	51,15	50,84
	$k_e =$ **tam(classe)**	1,71	2,88	**38,98**	**84,24**	**50,12**	**92,64**	91,37	**52,67**	**51,83**
Ganho Relativo Médio		-6,12	-13,69	+21,83	+4,06	+16,10	+2,69	+1,55	+8,37	+4,35
Ganho Relativo Máximo		+440,09	+428,65	+24,42	+26,89	+23,52	+15,99	+4,19	+29,81	+115,40

Fonte: Elaborado pelo autor.

5.6 Comparação entre os Métodos Propostos e Outras Abordagens

Esta seção apresenta uma comparação entre os resultados obtidos por ambos os métodos de aprendizado fracamente supervisionados propostos e outros métodos recentes.

Todos os métodos seguiram o mesmo protocolo experimental que utilizamos para avaliar ambos os métodos de expansão baseado em correlações e hipergrafos. Os resultados reportados do método baseado em correlações são os melhores resultados obtidos em cada combinação de coleção de imagens e descritor, por exemplo, para o classificador OPF [45] (Tabela 2 sem re-ranqueamento considerando a coleção MPEG-7 [74] e descritor ASC [75] o resultado de acurácia reportado sera o valor correspondente a métrica de correlação *Intersection*, 88,91%. Isso é repetido para cada um dos cenários com e sem re-ranqueamento.

O mesmo é feito com os resultados obtidos pelo método baseado em hipergrafos, em que o melhor resultado de acurácia de cada classificador obtido para cada combinação de coleção de imagens e descritor visual é apresentado. Por exemplo, para o classificador kNN (Tabela 6) na coleção MPEG-7 [74] e descritor ASC [75] o resultado reportado é o equivalente ao parâmetro k_e igual ao tamanho de classe, ou seja, 75,18%.

A Tabela 7 apresenta os resultados obtidos. Primeiro são apresentados os resultados em cada uma das combinações de coleção de imagem e descritor para os três classificadores supervisionados considerados durante todos os nossos experimentos (OPF [45], SVM [42], and kNN) obtidos utilizando da validação cruzada de 10 *folds* sem a utilização dos nossos métodos de expansão de conjuntos de treinamento. Em seguida são exibidos os resultados obtidos pelo classificador semi-supervisionado SGC [83] e pelos métodos semi-supervisionados *Label-Spreading* [84] combinado com os classificadores OPF [45], SVM [42], kNN e SGC [83] e Pseudo-Label [66]. Os maiores valores de acurácia obtidos em cada combinação de coleção de imagens e descritor (maior valor em cada coluna) está destacado em negrito e vermelho.

É possível observar que ambos os métodos de expansão quando combinados com cada um dos quatro classificadores considerados apresentam resultados de acurácia melhores do que ambos os classificadores supervisionados e semi-supervisionados. Os melhores resultados obtidos se concentraram no método baseado em correlações quando combinado com os classificadores SVM [42] e SGC [83] considerando as listas ranqueadas re-ranqueadas pelos métodos LHRR [19] e BFS-Tree [20]. O que confirma que utilizar listas ranqueadas de melhor qualidade, geradas pelos métodos de re-ranqueamento resultam em melhores resultados de acurácia. O método baseado em hipergrafos obteve o melhor resultado na coleção de imagens MNIST [80] quando combinado com o classificador SVM [42], apresentando um ganho relativo de +1,79% com relação ao classificador SVM [42] isolado e uma acurácia bem superior aos métodos semi-supervisionados considerados. Considerando todos os resultados, o método baseado em correlações teve sua melhor desempenho no classificador SVM [42] utilizando das listas ranqueadas geradas pelo método LHRR [19], este que apresentou um ganho relativo de +7,54% comparado ao classificador SVM [42] sem a utilização do da expansão de conjuntos de treinamento.

Se desconsiderarmos os resultados ruins obtidos pelo classificador SGC [83] na coleção de dados MPEG-7 [74] por conta da utilização das matrizes de distâncias, os resultados obtidos pelo classificador quando combinado aos métodos de expansão apresentam em muitos casos ganhos expressivos com relação aos outros métodos com exceção na coleção MNIST [80], em que as acurácias se mostraram inferiores.

Tabela 7 – Acurácia (%) obtida com classificadores supervisionados, semi-supervisionados e os melhores resultados obtidos por ambos os métodos de expansão propostos com cada classificador (OPF, SVM, kNN e SGC). Os resultados fracamente supervisionados exibidos são os melhores resultados de cada cenário (os valores em negrito em cada uma das tabelas de 2 a 6).

			MPEG-7		Flowers		Corel5k		MNIST	CUB200	Média
			ASC	CFD	ACC	Resnet	ACC	Resnet	-	Resnet	
Supervisionado	kNN		13,92	12,39	28,47	63,67	34,05	76,80	92,04	36,67	44,75
	OPF		82,95	67,75	30,54	71,77	40,21	83,56	93,49	38,59	63,61
	SVM		83,12	68,56	37,50	80,65	45,27	88,33	95,34	51,51	68,79
Semi-Supervisionado	SGC		1,88	3,25	31,33	80,02	42,37	89,74	90,02	47,90	48,31
	Label Spreading	kNN	81,94	67,28	33,77	73,49	43,44	83,98	9,28	36,99	53,84
		OPF	84,94	71,90	33,37	72,66	46,51	82,32	9,89	39,29	55,11
		SVM	84,94	71,90	33,38	72,75	46,52	82,37	9,89	39,29	55,13
		SGC	1,67	2,39	31,45	73,34	42,03	86,53	16,75	34,71	36,11
	Pseudo-Label+SGD		23,23	18,26	31,83	82,22	35,79	89,62	86,95	22,35	48,78
Expansão Baseada em Correlações	WS-kNN	-	78,94	71,03	33,67	78,06	41,94	87,52	95,18	47,32	66,71
		LHRR	81,45	72,96	31,36	80,87	42,18	89,32	95,98	47,12	67,66
		BFS-Tree	82,13	72,14	31,71	80,35	41,94	89,68	95,84	46,58	67,55
	WS-OPF	-	88,91	86,21	32,38	79,25	42,58	89,22	95,45	45,13	69,89
		LHRR	90,68	89,25	31,27	83,24	45,06	91,30	96,46	46,52	71,72
		BFS-Tree	90,62	87,34	31,62	82,67	44,66	91,58	96,34	46,39	71,40
	WS-SVM	-	89,30	87,60	36,00	83,50	45,22	91,51	95,80	52,29	72,65
		LHRR	91,05	**90,13**	34,68	85,79	47,85	92,87	96,91	52,58	**73,98**
		BFS-Tree	**91,34**	88,63	35,75	85,45	47,62	93,21	96,87	52,81	73,94
	WS-SGC	-	2,20	3,71	39,17	85,29	50,54	93,26	90,34	56,05	52,56
		LHRR	2,01	3,50	38,92	**86,33**	**51,03**	**93,62**	90,66	55,81	52,74
		BFS-Tree	1,89	3,68	**39,31**	85,95	50,86	93,52	90,60	**55,96**	52,72
Expansão Baseada em Hipergrafos	LHRR-kNN		75,18	65,50	32,86	80,79	42,06	89,08	95,90	47,60	66,12
	LHRR-OPF		89,87	84,58	31,70	82,65	44,48	90,04	96,55	47,03	70,86
	LHRR-SVM		90,28	86,25	35,02	84,78	46,94	91,39	**97,05**	52,55	73,03
	LHRR-SGC		1,82	2,88	38,98	84,24	50,12	92,64	91,46	52,67	51,85

Fonte: Elaborado pelo autor.

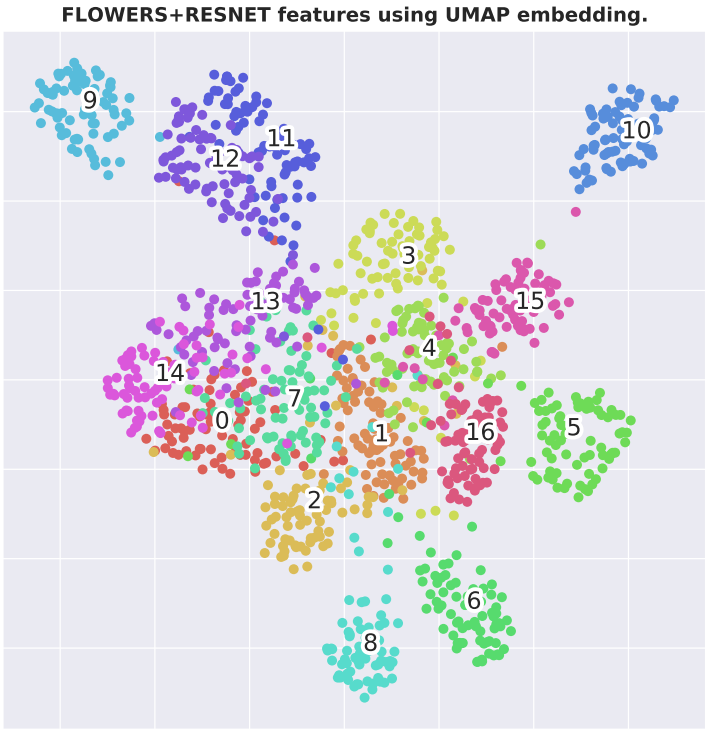
5.7 Análise Visual

Além dos experimentos quantitativos de classificação, também foi realizado um experimento para permitir uma análise visual e mais qualitativa do comportamento de ambos os métodos de expansão de conjuntos rotulados. Considerando as coleções de imagens utilizadas anteriormente e apresentadas na subseção 5.2, a coleção *Flowers* [77] apresenta o maior potencial para experimentos de visualização por conta da baixa quantidade de classes (17) e amostras (1360) na coleção. Para visualizar as características de cada imagem da coleção utilizamos do método UMAP (*Uniform Manifold Approximation and*

Projection) [87] de redução de dimensionalidade para reduzir os vetores de características de cada imagem para duas dimensões. Nesses experimentos consideramos os vetores de características extraídos pelo descritor CNN-Resnet [78].

A Figura 26 apresenta a visualização das características em duas dimensões geradas pelo algoritmo UMAP [87] considerando o descritor CNN-Resnet [78] e o conjunto *Flowers* [77]. Cada um dos elementos possui uma classe associada representada por cada uma das cores. Essa projeção é utilizada para avaliar as expansões realizadas por ambos os métodos baseados em correlações e hipergrafos.

Figura 26 – Visualização em duas dimensões dos vetores de características obtidos pelo descritor CNN-Resnet [78] na coleção de imagens *Flowers* [77] gerada pelo algoritmo UMAP [87].



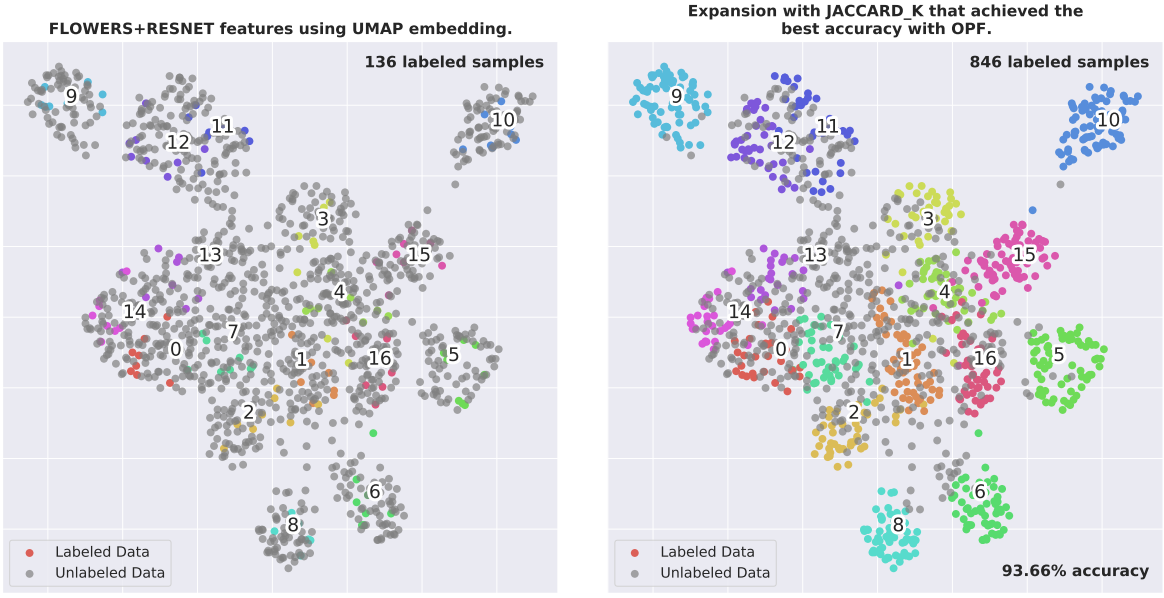
Fonte: Elaborado pelo autor.

Cada gráfico de visualização das expansões vai apresentar a visualização de um dos 10 *folds* considerados durante a avaliação experimental, e foi selecionado o *fold* que obteve a melhor acurácia entre os 10. Essa análise foi feita para o classificador OPF [45] em ambos os métodos de expansão, ou seja, é feita a visualização da expansão da melhor métrica de correlação com e sem a utilização dos métodos de re-ranqueamento (Tabela 2) e a visualização da expansão com ambos os parâmetros k_e para o método de hipergrafos (Tabela 6). Para cada visualização os pontos coloridos representam os itens rotulados e os pontos em cinza representam os itens não rotulados.

Primeiro considerando o método de expansão baseado em correlações sem a utiliza-

ção das listas ranqueadas re-ranqueadas. A Figura 27 apresenta o conjunto de treino inicial composto por 10% dos dados que é considerado como rotulado e o conjunto expandido gerado pela métrica de correlação $Jaccard_k$ [73]. Para cada uma das visualizações das expansões é também reportado o número de elementos rotulados no conjunto de treino inicial e no conjunto de treino expandido, além de reportar a acurácia obtida nos novos rótulos, comparado-os com os rótulos reais presentes na Figura 26. Podemos observar que mesmo com poucas amostras rotuladas, nosso método consegue rotular os dados não rotulados de maneira satisfatória, gerando 710 novas amostras rotuladas com uma acurácia de 93.66%. Informações que corroboram com os resultados apresentados na Tabela 2, em que a expansão realizada com a métrica $Jaccard_k$ [73] apresentou um ganho relativo de +10,42% em todos os *folds* em comparação com o cenário sem expansão (71,77% para 79,25%).

Figura 27 – Visualização em duas dimensões do conjunto de treino inicial (10% dos dados) e do conjunto de treino expandido gerado pela correlação $Jaccard_k$ [73] que obteve a melhor acurácia com o classificador OPF [45].



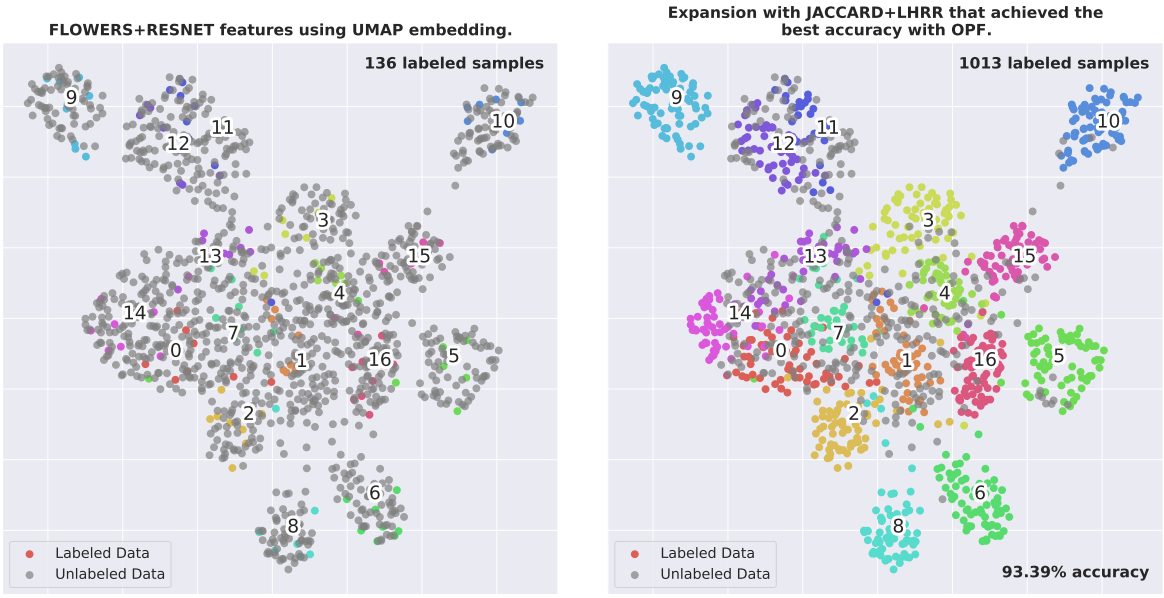
(a) **Conjunto de Treino Inicial** (b) **Conjunto de Treino Expandido**

Fonte: Elaborado pelo autor.

A Figura 28 apresenta uma análise similar, mas considerando a utilização das listas re-ranqueadas pelo método LHRR [19]. São apresentados o conjunto de treino inicial e o conjunto expandido pela métrica *Jaccard* [72]. Essa expansão gerou 877 novas amostras rotuladas com uma acurácia de 93,39%, acurácia similar a primeira expansão apresentada, mas com 167 novas amostras. Em comparação com utilizar as listas ranqueadas padrões (Figura 27), as regiões mais centrais da figura tem uma maior densidade de amostras rotuladas. A maior qualidade e quantidade dessa expansão é evidenciada pelos resultados apresentados na Tabela 2, em que a expansão com a métrica *Jaccard* [72] e as listas

ranqueadas geradas pelo método LHRR [19] obteve um ganho relativo de +15,98% em todos os *folds* em comparação a não se fazer a expansão (71,77% para 83,24%), valor +5,56% maior do que o cenário sem a utilização das listas re-ranqueadas.

Figura 28 – Visualização em duas dimensões do conjunto de treino inicial (10% dos dados) e do conjunto de treino expandido gerado pela correlação *Jaccard* [72] se utilizando das listas re-ranqueadas pelo método LHRR [19] que obteve a melhor acurácia com o classificador OPF [45].



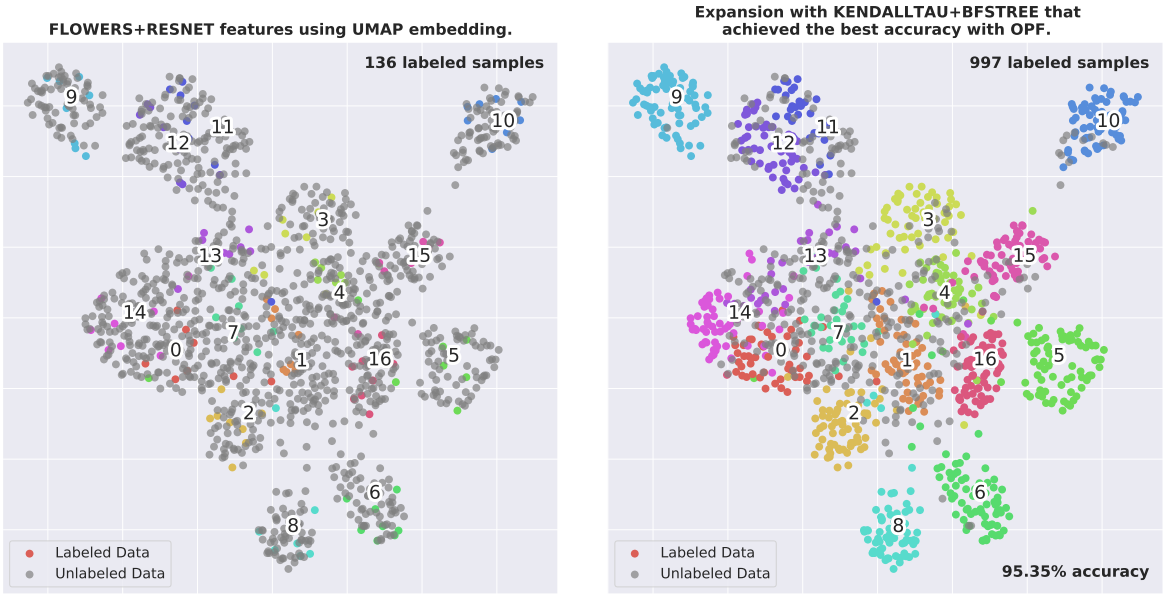
(a) **Conjunto de Treino Inicial** (b) **Conjunto de Treino Expandido**

Fonte: Elaborado pelo autor.

Na Figura 29 são apresentados o conjunto de treino inicial e o conjunto de treino expandido pela métrica *Kendall*τ mas se utilizando das listas re-ranqueadas pelo método BFS-Tree [20]. Assim como observamos no exemplo anterior considerando o método LHRR [19], podemos verificar que ao utilizar de listas ranqueadas de melhor qualidade a quantidade de novos elementos rotulados é maior e nesse caso a acurácia obtida é maior. Essa expansão gerou 861 novas amostras rotuladas com uma acurácia de 95,35%, maior do que os últimos dois casos. Ao considerar os resultados de acurácia presentes na Tabela 2, a qualidade da expansão é confirmada, onde se utilizar da métrica *Kendall*τ e das listas re-ranqueadas pelo método BFS-Tree [20] atingiu um ganho relativo de +15,19% em comparação com o treinamento do classificador sem expansão (71,77% para 82,67%), valor +4,77% maior do que a expansão com as listas ranqueadas originais.

Por fim, considerando a expansão baseada em hipergrafos com o parâmetro $k_e = 18$, A Figura 30 apresenta o conjunto de treino inicial e o conjunto expandido pelo método. Nesse caso, quando utilizamos de um valor de k_e, que é a profundidade das hiperarestas que é considerada durante o processo de expansão, igual ao valor de k que utilizamos para construir o hipergrafo do método LHRR [19] que foi utilizado para realizar essa expansão.

Figura 29 – Visualização em duas dimensões do conjunto de treino inicial (10% dos dados) e do conjunto de treino expandido gerado pela correlação *Kendall*τ [71] se utilizando das listas re-ranqueadas pelo método BFS-Tree [20] que obteve a melhor acurácia com o classificador OPF [45].



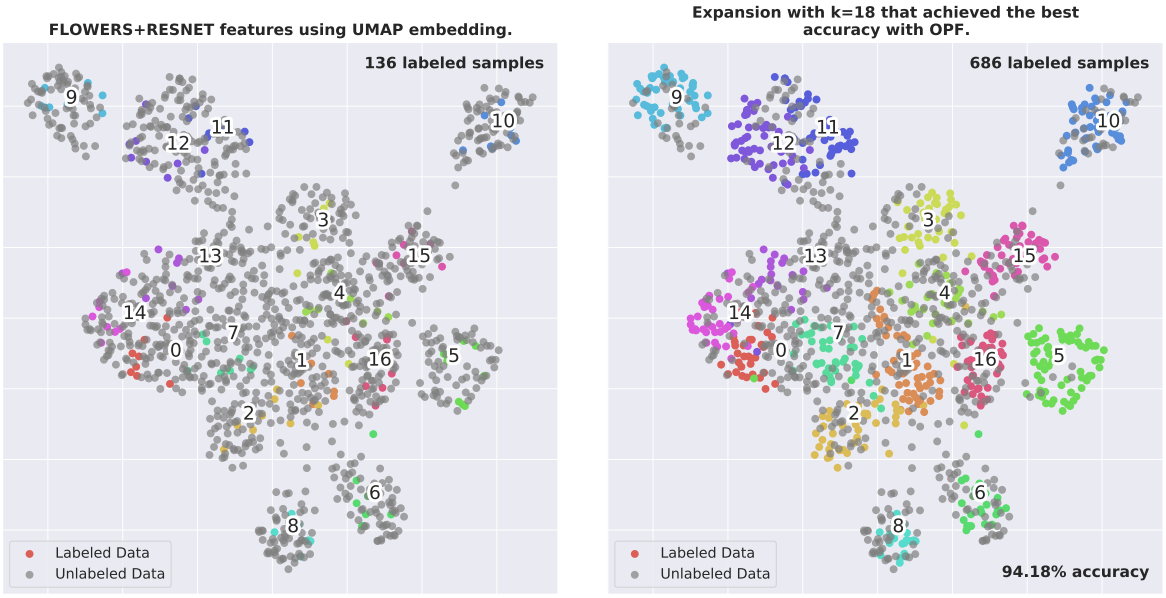
(a) **Conjunto de Treino Inicial** (b) **Conjunto de Treino Expandido**

Fonte: Elaborado pelo autor.

Essa expansão gerou 550 novas amostras rotuladas com uma acurácia de 94,18%. As novas amostras rotuladas resultaram em uma melhor acurácia obtida pelo classificador, como podemos observar na Tabela 6, onde a versão fracamente supervisionada do classificador atingiu ganho relativo de +10,16% em comparação com o cenário padrão (71,77% para 79,06%).

Em seguida, a Figura 31 apresenta a expansão realizada pelo método baseado em hipergrafos se utilizando do parâmetro k_e igual ao tamanho de classe da coleção de imagens *Flowers* [77], igual a 80. São apresentados o conjunto de treino inicial, contendo 10% de dados rotulados e o conjunto de treino expandido obtido com este novo parâmetro. Em comparação para quando utilizamos $k_e = 18$, essa nova expansão está mais próxima a original (Figura 26), com 350 novas amostras rotuladas em comparação com a primeira expansão. Esta expansão resultou em 900 novas amostras que foram rotuladas com uma acurácia de 92,44%. Apesar de menos acurácia, os resultados obtidos pelo classificador com este conjunto comparado com o obtido com $k_e = 18$ são superiores como podemos observar na Tabela 6, onde esta apresentou um ganho de +15,16% em comparação com não utilizar do método de expansão (71,77% de acurácia para 82,65%).

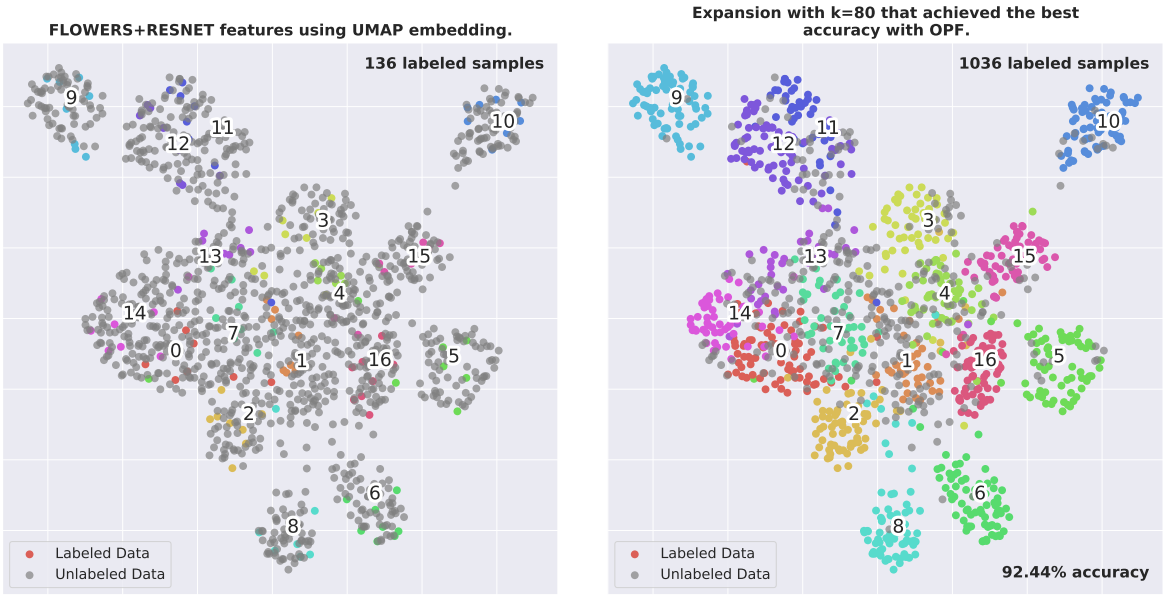
Figura 30 – Visualização em duas dimensões do conjunto de treino inicial (10% dos dados) e do conjunto de treino expandido gerado pelo método de expansão baseado em hipergrafos se utilizando de $k_e = 18$ que obteve a melhor acurácia com o classificador OPF [45].



(a) **Conjunto de Treino Inicial** (b) **Conjunto de Treino Expandido**

Fonte: Elaborado pelo autor.

Figura 31 – Visualização em duas dimensões do conjunto de treino inicial (10% dos dados) e do conjunto de treino expandido gerado pelo método de expansão baseado em hipergrafos se utilizando de $k_e = 80$ que obteve a melhor acurácia com o classificador OPF [45].



(a) **Conjunto de Treino Inicial** (b) **Conjunto de Treino Expandido**

Fonte: Elaborado pelo autor.

6 Conclusões

Considerando os desafios e restrições impostos pela escassez de dados rotulados, torna-se cada vez mais necessário o desenvolvimento de métodos capazes de operar em tais cenários. Nesse contexto, ao longo do trabalho foram apresentados dois métodos de aprendizado fracamente supervisionado. Ambos se baseiam da análise de um modelo de ranqueamento e diferentes análises para considerar quais amostras podem ser incluídas em um conjunto rotulado expandido que pode ser utilizado para treinamento de classificadores. Ambos foram avaliados no mesmo cenário, considerando as mesmas combinações de classificadores, coleções de imagens e descritores. Além da análise de acurácia obtida por ambos os métodos quando combinados a diferentes classificadores também avaliamos o desempenho da expansão realizada por ambos ao analisar resultados visuais.

O primeiro método, baseado na análise de métricas de correlação entre listas ranqueadas, explora a informação contextual codificada nas listas ranqueadas para analisar a estrutura do conjunto de dados e decidir quais amostras podem ser incluídas em um conjunto rotulado expandido. O método foi avaliado em dois cenários: considerando as listas ranqueadas geradas a partir dos vetores de características originais, publicado em [22], e considerando listas ranqueadas mais efetivas geradas por métodos de re-ranqueamento. Para cada um desses cenários analisamos a capacidade de expansão do método ao se utilizar de cinco diferentes métricas de correlação. Os resultados obtidos por meio de extensa avaliação experimental confirmam que se utilizar de listas ranqueadas mais efetivas re-ranqueadas resulta em melhores expansões de conjuntos rotulados e melhores resultados de classificação. Também foi apresentado um método não supervisionado capaz de estimar o limiar de correlação responsável por selecionar quais amostras terão rótulo atribuído e farão parte do conjunto rotulado expandido. O método apresentado preenche uma lacuna relevante de trabalho anterior [22], em que a estimativa desse limiar requeria uma extensa avaliação experimental para cada combinação de métrica de correlação, coleção de imagens e descritor visual. Os resultados de acurácia obtidos por este método quando combinado com os diferentes classificadores com e sem a utilização das listas re-ranqueadas foram bastante positivos, destacando a robustez do método proposto. Os resultados visuais obtidos da expansão realizada pelo método baseado em correlações também foram bastante positivos, em especial quando consideradas as listas ranqueadas re-ranqueadas, em que a expansão das amostras rotuladas foi bastante acertada, se aproximando muito dos rótulos reais. O trabalho realizado acerca do método de aprendizado fracamente supervisionado baseado em correlações originou um artigo [88] submetido, em avaliação.

O segundo método proposto ao longo do trabalho explora de informações de similaridade contidas em uma estrutura de hipergrafo construída a partir da análise das

listas ranqueadas. Essas informações de similaridade contidas no hipergrafo são analisadas de modo a selecionar quais amostras não rotuladas podem fazer parte do conjunto rotulado expandido. Para selecionar uma quantidade adequada de amostras ao conjunto rotulado expandido, o método depende de um parâmetro α que representa a taxa de expansão de rótulos. Deste modo, propomos uma maneira de estimar esse valor analisando somente o pequeno conjunto rotulado inicial os dados não rotulados. O método foi avaliado considerando diferentes valores do parâmetro k_e, que é a profundidade das hiperarestas analisadas durante o método de expansão, e em todos os cenários atingiu resultados positivos em comparação com o uso dos classificadores sem combiná-los com o método de expansão ou com abordagens similares. Os resultados visuais obtidos pelo método também confirmam a sua robustez, principalmente quando utilizamos de um valor maior do parâmetro k_e, em que o conjunto rotulado expandido se mostrou muito similar ao conjunto rotulado original, resultando em bons resultados de classificação. De maneira similar, os experimentos realizados com o método de aprendizado fracamente supervisionado baseado em hipergrafos originaram um artigo [89] submetido, em fase de avaliação, que também considera a aplicação do método em coleção de vídeos (UCF-101 [90]).

Dentre os trabalhos futuros, há a possibilidade de explorar os métodos propostos em outros cenários de classificação e recuperação multimídia, especialmente em cenários de re-identificação de pessoas. Além disso, pretende-se combiná-los com estratégias de autotreinamento (*self-training*). Com relação ao método de aprendizado fracamente supervisionado baseado em correlações também planejamos investigar maneiras de selecionar automaticamente qual métrica de correlação deve ser utilizada em cada cenário.

Referências

[1] H. Pham, Z. Dai, Q. Xie, and Q. V. Le, "Meta pseudo labels," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 557–11 568.

[2] D. C. G. Pedronette, Y. Weng, A. Baldassin, and C. Hou, "Semi-supervised and active learning through manifold reciprocal knn graph for image retrieval," *Neurocomputing*, vol. 340, pp. 19–31, 2019.

[3] L. Jing, T. Parag, Z. Wu, Y. Tian, and H. Wang, "Videossl: Semi-supervised learning for video classification," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 1110–1119.

[4] K. Han, S. Rebuffi, S. Ehrhardt, A. Vedaldi, and A. Zisserman, "Automatically discovering and learning new visual categories with ranking statistics," in *8th International Conference on Learning Representations, ICLR 2020*, 2020.

[5] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[6] A. Abuduweili, X. Li, H. Shi, C.-Z. Xu, and D. Dou, "Adaptive consistency regularization for semi-supervised transfer learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 6923–6932.

[7] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Is object localization for free?-weakly-supervised learning with convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 685–694.

[8] H. Bilen and A. Vedaldi, "Weakly supervised deep detection networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2846–2854.

[9] J. Peyre, J. Sivic, I. Laptev, and C. Schmid, "Weakly-supervised learning of visual relations," in *Proceedings of the ieee international conference on computer vision*, 2017, pp. 5179–5188.

[10] Z. H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018.

[11] D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. Van Der Maaten, "Exploring the limits of weakly supervised pretraining," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 181–196.

[12] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," *School Comput Sci Carnegie Mellon Univ Tech Rep CMUCALD02107*, 2002.

[13] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*. Cambridge, MA: MIT Press, 2006. [Online]. Available: http://www.kyb.tuebingen.mpg.de/ssl-book

[14] F. Breve, L. Zhao, M. Quiles, W. Pedrycz, and J. Liu, "Particle competition and cooperation in networks for semi-supervised learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 9, pp. 1686–1698, 2012.

[15] F. A. Breve and D. C. G. Pedronette, "Combined unsupervised and semi-supervised learning for data classification," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. Ieee, 2016, pp. 1–6.

[16] C. Wei, K. Sohn, C. Mellina, A. Yuille, and F. Yang, "Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 10 857–10 866.

[17] F. P. dos Santos, C. Zor, J. Kittler, and M. A. Ponti, "Learning image features with fewer labels using a semi-supervised deep convolutional network," *Neural Networks*, vol. 132, pp. 131–143, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608020303051

[18] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *arXiv preprint arXiv:1905.02249*, 2019.

[19] D. C. G. Pedronette, L. P. Valem, J. Almeida, and R. d. S. Torres, "Multimedia retrieval through unsupervised hypergraph-based manifold ranking," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5824–5838, 2019.

[20] D. Carlos Guimarães Pedronette, L. P. Valem, and R. da S. Torres, "A bfs-tree of ranking references for unsupervised manifold learning," *Pattern Recognition*, vol. 111, p. 107666, 2021.

[21] L. P. Valem, C. R. D. Oliveira, D. C. G. Pedronette, and J. Almeida, "Unsupervised similarity learning through rank correlation and knn sets," *TOMCCAP*, vol. 14, no. 4, pp. 80:1–80:23, 2018.

[22] J. G. C. Presotto, L. P. Valem, N. G. de Sá, D. C. G. Pedronette, and J. P. Papa, "Weakly supervised learning through rank-based contextual measures," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 5752–5759.

[23] L. C. Afonso, D. C. Pedronette, A. N. De Souza, and J. P. Papa, "Improving optimum-path forest classification using unsupervised manifold learning," in *2018 24th International Conference on Pattern Recognition (ICPR)*. Ieee, 2018, pp. 560–565.

[24] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Pseudo-labeling and confirmation bias in deep semi-supervised learning," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[25] J. Choi, M. Jeong, T. Kim, and C. Kim, "Pseudo-labeling curriculum for unsupervised domain adaptation," *arXiv preprint arXiv:1908.00262*, 2019.

[26] Z. Zheng and Y. Yang, "Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1106–1120, 2021.

[27] M. N. Rizve, K. Duarte, Y. S. Rawat, and M. Shah, "In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning," *arXiv preprint arXiv:2101.06329*, 2021.

[28] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 5:1–5:60, 2008. [Online]. Available: http://doi.acm.org/10.1145/1348246.1348248

[29] L. Piras and G. Giacinto, "Information fusion in content based image retrieval: A comprehensive overview," *Information Fusion*, vol. 37, no. Supplement C, pp. 50 – 60, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1566253517300076

[30] S. Patil and S. Talbar, *Content Based Image Retrieval Using Various Distance Metrics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 154–161. [Online]. Available: https://doi.org/10.1007/978-3-642-27872-3_23

[31] N. Arica and F. T. Y. Vural, "BAS: a perceptual shape descriptor based on the beam angle statistics," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1627–1639, 2003.

[32] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *CVPR*, 1997, pp. 762–768.

[33] R. D. S. Torres and A. X. Falcão, "Content-based image retrieval: Theory and applications," *Revista de Informática Teórica e Aplicada*, vol. 13, pp. 161–185, 2006.

[34] R. da S. Torres and A. X. Falcão, "Content-Based Image Retrieval: Theory and Applications," *Revista de Informática Teórica e Aplicada*, vol. 13, no. 2, pp. 161–185, 2006.

[35] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms.* Wiley, 2004.

[36] E. Alpaydin, *Introduction to machine learning.* MIT press, 2020.

[37] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "Panns: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2880–2894, 2020.

[38] D. Melati, Y. Grinberg, M. K. Dezfouli, S. Janz, P. Cheben, J. H. Schmid, A. Sánchez-Postigo, and D.-X. Xu, "Mapping the global design space of nanophotonic components using machine learning pattern recognition," *Nature communications*, vol. 10, no. 1, pp. 1–9, 2019.

[39] J. Janai, F. Güney, A. Behl, A. Geiger *et al.*, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *Foundations and Trends® in Computer Graphics and Vision*, vol. 12, no. 1–3, pp. 1–308, 2020.

[40] J. Lemley, S. Bazrafkan, and P. Corcoran, "Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision." *IEEE Consumer Electronics Magazine*, vol. 6, no. 2, pp. 48–56, 2017.

[41] G. Hepner, T. Logan, N. Ritter, and N. Bryant, "Artificial neural network classification using a minimal training set- comparison to conventional supervised classification," *Photogrammetric Engineering and Remote Sensing*, vol. 56, no. 4, pp. 469–473, 1990.

[42] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995. [Online]. Available: https://doi.org/10.1023/A:1022627411411

[43] C. D. Ferreira, J. A. dos Santos, R. da S. Torres, M. A. Gonçalves, R. C. Rezende, and W. Fan, "Relevance feedback based on genetic programming for image retrieval," *Pattern Recogninion Letters*, vol. 32, no. 1, pp. 27–37, 2011.

[44] R. da S. Torres, A. X. Falcão, M. A. Gonçalves, J. P. Papa, B. Zhang, W. Fan, and E. A. Fox, "A genetic programming framework for content-based image retrieval," *Pattern Recognition*, vol. 42, no. 2, pp. 283 – 292, 2009, learning Semantics from Multimedia Content. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320308001623

[45] J. Papa, A. Falcão, and C. Suzuki, "Supervised pattern classification based on optimum-path forest," *Int. J. Imaging Syst. Technol.*, vol. 19, no. 2, pp. 120–131, Jun. 2009. [Online]. Available: http://dx.doi.org/10.1002/ima.v19:2

[46] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proceedings of the 22Nd ACM International Conference on Multimedia*, ser. MM '14. New York, NY, USA: ACM, 2014, pp. 157–166. [Online]. Available: http://doi.acm.org/10.1145/2647868.2654948

[47] S. B. Kotsiantis and P. E. Pintelas, "Recent Advances in Clustering: A Brief Survey," *WSEAS Transactions on Information Science and Applications*, 2004.

[48] X. Yang and L. J. Latecki, "Affinity learning on a tensor product graph with applications to shape and image retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2011)*, 2011, pp. 2369–2376.

[49] X. Yang, S. Koknar-Tezel, and L. J. Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval." in *CVPR*, 2009, pp. 357–364.

[50] X. Yang, X. Bai, L. J. Latecki, and Z. Tu, "Improving shape retrieval by learning graph transduction," in *ECCV*, vol. 4, 2008, pp. 788–801.

[51] P. Kontschieder, M. Donoser, and H. Bischof, "Beyond pairwise shape similarity analysis," in *ACCV*, 2009, pp. 655–666.

[52] G. Park, Y. Baek, and H.-K. Lee, "Re-ranking algorithm using post-retrieval clustering for content-based image retrieval," *Information Processing and Management*, vol. 41, no. 2, pp. 177–194, 2005.

[53] W. Voravuthikunchai, B. Crémilleux, and F. Jurie, "Image re-ranking based on statistics of frequent patterns," in *Proceedings of International Conference on Multimedia Retrieval*, ser. ICMR '14. New York, NY, USA: ACM, 2014, pp. 129:129–129:136. [Online]. Available: http://doi.acm.org/10.1145/2578726.2578743

[54] D. C. G. Pedronette and R. da S. Torres, "Image re-ranking and rank aggregation based on similarity of ranked lists," *Pattern Recognition*, vol. 46, no. 8, pp. 2350–2360, 2013.

[55] Y. Chen, X. Li, A. Dick, and R. Hill, "Ranking consistency for image matching and object retrieval," *Pattern Recognition*, vol. 47, no. 3, pp. 1349 – 1360, 2014.

[56] X. Bai, S. Bai, and X. Wang, "Beyond diffusion process: Neighbor set similarity for fast re-ranking," *Information Sciences*, vol. 325, pp. 342 – 354, 2015.

[57] D. C. G. Pedronette and R. da S. Torres, "A correlation graph approach for unsupervised manifold learning in image retrieval tasks," *Neurocomputing*, vol. 208, no. Supplement C, pp. 66 – 79, 2016, sI: BridgingSemantic. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231216304726

[58] L. P. Valem and D. C. G. Pedronette, "Unsupervised similarity learning through cartesian product of ranking references for image retrieval tasks," *SIBGRAPI*, 2016.

[59] D. C. G. Pedronette, J. Almeida, and R. da S. Torres, "A graph-based ranked-list model for unsupervised distance learning on shape retrieval," *Pattern Recognition Letters*, vol. 83, Part 3, pp. 357 – 367, 2016, efficient Shape Representation, Matching, Ranking, and its Applications. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865516301052

[60] D. C. G. Pedronette, O. A. Penatti, and R. da S. Torres, "Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks," *Image and Vision Computing*, vol. 32, no. 2, pp. 120 – 130, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885613001819

[61] L. P. Valem, D. C. G. Pedronette, R. d. S. Torres, E. Borin, and J. Almeida, "Effective, efficient, and scalable unsupervised distance learning in image retrieval tasks," *ICMR*, 2015.

[62] C. Y. Okada, D. C. G. a. Pedronette, and R. da S. Torres, "Unsupervised distance learning by rank correlation measures for image retrieval," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ser. ICMR '15. New York, NY, USA: ACM, 2015, pp. 331–338. [Online]. Available: http://doi.acm.org/10.1145/2671188.2749335

[63] X. Zhu, "Semi-Supervised Learning Literature Survey - TR 1530," *Sciences-New York*, 2005.

[64] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[65] I. Triguero, S. García, and F. Herrera, "Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study," *Knowledge and Information systems*, vol. 42, no. 2, pp. 245–284, 2015.

[66] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, no. 2, 2013.

[67] S. M. Omohundro, *Five balltree construction algorithms*. International Computer Science Institute Berkeley, 1989.

[68] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, p. 509–517, Sep. 1975. [Online]. Available: https://doi.org/10.1145/361002.361007

[69] D. C. G. Pedronette, F. M. F. Gonçalves, and I. R. Guilherme, "Unsupervised manifold learning through reciprocal knn graph and connected components for image retrieval tasks," *Pattern Recognition*, vol. 75, pp. 161 – 174, 2018.

[70] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Transactions on Information Systems*, vol. 28, no. 4, pp. 20:1–20:38, 2010.

[71] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *ACM-SIAM Symposium on Discrete algorithms (SODA'03)*, 2003, pp. 28–36.

[72] M. Levandowsky and D. Winter, "Distance between sets," *Nature*, vol. 243, pp. 34 – 35, 1971.

[73] C. Y. Okada, D. C. G. a. Pedronette, and R. da S. Torres, "Unsupervised distance learning by rank correlation measures for image retrieval," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ser. ICMR '15. New York, NY, USA: ACM, 2015, pp. 331–338. [Online]. Available: http://doi.acm.org/10.1145/2671188.2749335

[74] L. J. Latecki, R. Lakamper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *CVPR*, 2000, pp. 424–429.

[75] H. Ling, X. Yang, and L. J. Latecki, "Balancing deformability and discriminability for shape matching," in *ECCV*, vol. 3, 2010, pp. 411–424.

[76] D. C. G. Pedronette and R. da S. Torres, "Shape retrieval using contour features and distance optmization," in *VISAPP*, vol. 1, 2010, pp. 197 – 202.

[77] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," *Computer Vision, Graphics and Image Processing*, pp. 722–729, 2008.

[78] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016.

[79] G.-H. Liu, J.-Y. Yang, and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recogn.*, vol. 48, no. 8, pp. 2554–2566, Aug. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2015.02.005

[80] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[81] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The caltech-ucsd birds-200-2011 dataset," California Institute of Technology, Tech. Rep. CNS-TR-2011-001, 2011.

[82] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[83] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 6861–6871. [Online]. Available: http://proceedings.mlr.press/v97/wu19e.html

[84] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems 16*. MIT Press, 2004, pp. 321–328.

[85] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch Geometric," in *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[86] L. P. Valem and D. C. G. a. Pedronette, "An unsupervised distance learning framework for multimedia retrieval," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '17. New York, NY, USA: ACM, 2017, pp. 107–111. [Online]. Available: http://doi.acm.org/10.1145/3078971.3079017

[87] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2020.

[88] J. G. C. Presotto, L. P. Valem, N. G. de Sá, D. C. G. Pedronette, and J. P. Papa, "Weakly supervised classification through manifold learning and rank-based contextual measures," 2021, submitted to: Pattern Recognition.

[89] J. G. C. Presotto, S. F. dos Santos, L. P. Valem, F. A. Faria, J. Almeida, J. P. Papa, and D. C. G. Pedronette, "Weakly supervised learning based on hypergraph manifold ranking," 2021, submitted to: IEEE Transactions on Multimedia.

[90] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," *CoRR*, vol. abs/1212.0402, 2012.