CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

UM ARCABOUÇO DE RECOMENDAÇÃO MULTIMODAL DE IMAGENS BASEADO EM FATORES LATENTES DE BAIXA DIMENSIONALIDADE

ANNA CLAUDIA ALMEIDA ANÍCIO

BELO HORIZONTE
AGOSTO DE 2017

ANNA CLAUDIA ALMEIDA ANÍCIO

UM ARCABOUÇO DE RECOMENDAÇÃO MULTIMODAL DE IMAGENS BASEADO EM FATORES LATENTES DE BAIXA DIMENSIONALIDADE

Dissertação apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Sistemas Inteligentes

Orientador: Prof. Dr. Anísio Mendes Lacerda
Coorientador: Prof. Dr. Flávio Luis Cardeal Pádua

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL
BELO HORIZONTE
AGOSTO DE 2017

SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

"UM ARCABOUÇO DE RECOMENDAÇÃO MULTIMODAL DE IMAGENS BASEADO EM FATORES LATENTES DE BAIXA DIMENSIONALIDADE"

Dissertação de Mestrado apresentada por **Anna Claudia Almeida**, em 4 de agosto de 2017, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Anísio Mendes Lacerda (Orientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Flávio Luis Cardeal Pádua (Coorientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Álvaro Rodrigues Pereira Júnior
Universidade Federal de Ouro Preto

Prof. Dr. Thiago de Souza Rodrigues
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida a impressão,

Prof. Dr. José Geraldo Peixoto de Faria
Coordenador do Programa de Pós-Graduação em
Modelagem Matemática e Computacional

Anício, Anna Claudia Almeida
A597a Um arcabouço de recomendação multimodal de imagens baseado em fatores latentes de baixa dimensionalidade. / Anna Claudia Almeida Anício. – – Belo Horizonte, 2017.
xv, 72 f. : il.

Dissertação (mestrado) – Centro Federal de Educação Tecnológica de Minas Gerais, Programa de Pós-Graduação em Modelagem Matemática e Computacional, 2017.
Orientador: Prof. Dr. Anísio Mendes Lacerda
Coorientador: Prof. Dr. Flávio Luis Cardeal Pádua

Bibliografia

1. Aprendizado do Computador. 2. Mineração de Dados (Computação). 3.Processamento de Imagens. I. Lacerda, Anísio Mendes. II. Centro Federal de Educação Tecnológica de Minas Gerais. III. Título

CDD 001.535

Elaboração da ficha catalográfica pela Biblioteca-Campus II / CEFET-MG

Agradecimentos

Dedico meus sinceros agradecimentos

- À Deus, pelo amor, pelo dom da vida, pela força, sustento, inteligência e pelo cuidado em todos os momentos e à Nossa Senhora pela poderosa intercessão.
- Ao meu esposo Ádames, meu amor, pelo carinho, compreensão e apoio nos momentos em que mais precisei.
- Aos meus pais Deilson e Hermelinda e minha irmã Amanda pelo amor, dedicação e por sempre lutarem por mim, e toda minha família por todo carinho e compreensão.
- Ao meu orientador Anísio e coorientador Flávio Cardeal pelo conhecimento compartilhado e por terem acreditado em mim e neste trabalho.
- À Capes, pelo apoio financeiro.
- À Raíssa, aluna de iniciação, que tanto me ajudou no início do projeto.
- Aos meus companheiros do Piim Lab, pela convivência, companhia, ajudas e boas risadas.
- A todos os docentes e discentes do PPGMMC e a galera do Apoio, pela ajuda e amizade, e também às secretárias por sempre facilitarem a nossa vida.
- À Igreja, à Renovação Carismática Católica e aos meus irmãos do Ministério Universidades Renovadas em Belo Horizonte, o GOU-CEFET, vocês foram minha igreja em Belo Horizonte e o quanto isso foi necessário.
- À Mirele, por ter me acolhido em sua casa e pela companhia em Belo Horizonte.

Resumo

A popularização de dispositivos fotográficos, como *smartphones* e câmeras digitais, tem levado as pessoas ao hábito de registrar momentos por meio de fotos. Assim, todos os dias, uma grande quantidade de imagens são disponibilizadas em redes sociais de compartilhamento de fotos, como *Flickr*, *Pinterest* e *Instagram*. Esta dissertação propõe um arcabouço de recomendação de imagens, que visa ajudar os usuários a lidar com a enorme quantidade de informação disponível, por meio de uma experiência personalizada. Sistemas de Recomendação são ferramentas computacionais que auxiliam usuários a encontrar informação relevante (por exemplo: imagens, músicas, filmes) de acordo com seus interesses. Tradicionalmente, esses sistemas utilizam informação colaborativa das preferências dos usuários para sugerir itens. Porém, nem sempre essa informação está disponível. Por exemplo, quando novos itens são inseridos no sistema não existe informação de preferência dos usuários sobre eles, fato que leva à necessidade de explorar o conteúdo desses itens. Neste trabalho, o foco é explorar informação de natureza multimodal para recomendação de imagens, sendo essas representadas por meio de características (i) visuais e (ii) textuais. As características visuais são extraídas da própria imagem (ex., cor e textura), utilizando técnicas de Visão Computacional e Processamento de Imagens. Por outro lado, as características textuais são extraídas de metadados das imagens (ex., do título e da descrição), utilizando técnicas de Recuperação de Informação. Dessa forma, é possível descrever uma imagem sob diferentes dimensões. Os sistemas de recomendação estado-da-arte são baseados em informação colaborativa e descrição dos itens. Esses sistemas sofrem do problema da "maldição da dimensionalidade". As características de alta dimensão dos itens dificulta a identificação de itens similares, dado que a distribuição dos dados originais é extremamente esparsa. A fim de aliviar esse problema, propõe-se projetar a representação dos itens de alta dimensão em um espaço de fatores latentes de dimensão inferior. O arcabouço de recomendação proposto é baseado no algoritmo *Sparse Linear Methods with Side Information* (SSLIM) que, por sua vez, consiste na construção de matrizes de (i) relação entre usuários e itens e de (ii) relação entre itens. O trabalho proposto é avaliado em uma base de dados real coletada do *Flickr*; e comparado com técnicas do estado-da-arte. Conforme demonstrado nos experimentos, o arcabouço proposto apresenta melhora na qualidade das recomendações e fornece ganhos de 15% a 17% em termos de taxa de acertos em relação aos métodos de referência.

Palavras-chave: Sistemas de Recomendação, Representação Multimodal de Imagens, Fatores Latentes, Redução de Dimensionalidade.

Abstract

The popularization of photo-taking devices, such as smart phones and digital cameras, leads people to the common habit of recording moments by taking photos. Hence, everyday, millions of images are uploaded to photo sharing social networks, such as Flickr, Pinterest, and Instagram. This dissertation, proposes an image recommender framework, which aims to help users to deal with this huge amount of available information by means of a personalized experience. Recommender Systems are software tools that help users find relevant information (eg news, music, movies) according to their interests. Traditionally, recommender systems using collaborative information preferences users to suggest items. But not always collaborative information is available. For example, when new items are entered in the system, there is no preference information for them, in this way, this paper uses of the items contents for recommendation. Specifically, it is intended use multimodal nature information for image recommendation. The work consist to represent images by means of characteristics (i) visual and (ii) textual. The visual features are extracted from the image itself, using techniques of Computer Vision and Image Processing. On the other hand, textual features are extracted meta-data of images (e.g., title and description). Thus, it's possible describe an image in different dimensions. Traditional state-of-the-art item-based recommender systems suffer from the "curse of dimensionality" problem, such as, in the multimodal approach used in this work. The high-dimensional image features lead to a degraded recommendation performance since the distribution of the original data is extremely sparse. In order to alleviate this problem, it is proposed to project this multi-modal high-dimensional data into a common and lower dimensional latent factor space. Hence, get able to combine multi-modal information in a principled way and minimize the curse of dimensionality problem. The proposed recommendation algorithm is based on the algorithm Sparse Linear Methods with Side Information (SSLIM), which consists of the construction matrix responsible for setting the relation between the users and items and the items themselves. For the comparison, we run a series of experiments using a real-world image dataset extracted from Flickr and compared to state-of-the-art baselines. As shown in experiments, the low-dimensional image-based framework improves the quality of recommendations and provides gains from 15% to 17% in terms of hit-rate over the baselines.

Keywords: Recommender Systems, Multimodal Image Representation, Latent Factors, Dimensionality Reduction.

Lista de Figuras

Lista de Tabelas

Lista de Algoritmos

Lista de Abreviaturas e Siglas

API	Application Programming Interface
ARHR	Average-Reciprocal Hit Rate
BoW	Bag of Words
CAN	Content-Aware Network
DoG	Diferença das Gaussianas
HOG	Histogram of Oriented Gradients
HR	Hit Rate
HSV	Hue, Saturation and Value
ISOMAP	Isometric Mapping
LDSSLIM	Low-Dimensional Sparse Linear Methods with Side Information
LSA	Latent Semantic Analysis
LSI	Latent Semantic Indexing
MAP	Mean Average Precision
MAPE	Mean Absolute Percentage Error
MDS	Multidimensional Scaling
NDCG	Normalized Discounted Cumulative Gain
PCA	Principal Component Analysis
RMSE	Root Mean Squared Error
SIFT	Scale-Invariant Feature Transform
SLIM	Sparse Linear Methods
SSLIM	Sparse Linear Methods with Side Information
SURF	Speeded Up Robust Features
SVD	Singular Value Decomposition
SVM	Support Vector Machine

TF	Term Frequency
TF-IDF	Term Frequency–Inverse Document Frequency

Sumário

1 Introdução

Com a crescente quantidade de informação disponível na Web, encontrar informação relevante é um desafio. Diariamente, usuários disponibilizam milhões de fotos em redes sociais de compartilhamento de imagens. Por exemplo, segundo um levantamento realizado em 2016 (LOWE, 2016), por dia, mais de 350 milhões de imagens são postadas no *Facebook*[1], 14 milhões de imagens são postados no *Pinterest*[2], quase 85 milhões de imagens são publicadas no *Instagram*[3], e, em média, 1.68 milhões de imagens são postadas no *Flickr*[4] (MICHEL, 2016).

Claramente, este grande volume de informação representa uma sobrecarga para os usuários e torna-se inútil mediante a impossibilidade de encontrar efetivamente conteúdo relevante de acordo com as preferências dos usuários. Nesse contexto, os Sistemas de Recomendação são uma alternativa, uma vez que ajudam os usuários a encontrar informação relevante de acordo com seu perfil de interesse (SCHAFFER J.B., 2001).

A Figura 1 mostra a interface de duas redes sociais de compartilhamento de imagens - *Flickr* e *Pinteret*, respectivamente - dado um determinado usuário e as imagens pertencentes ao seu perfil.

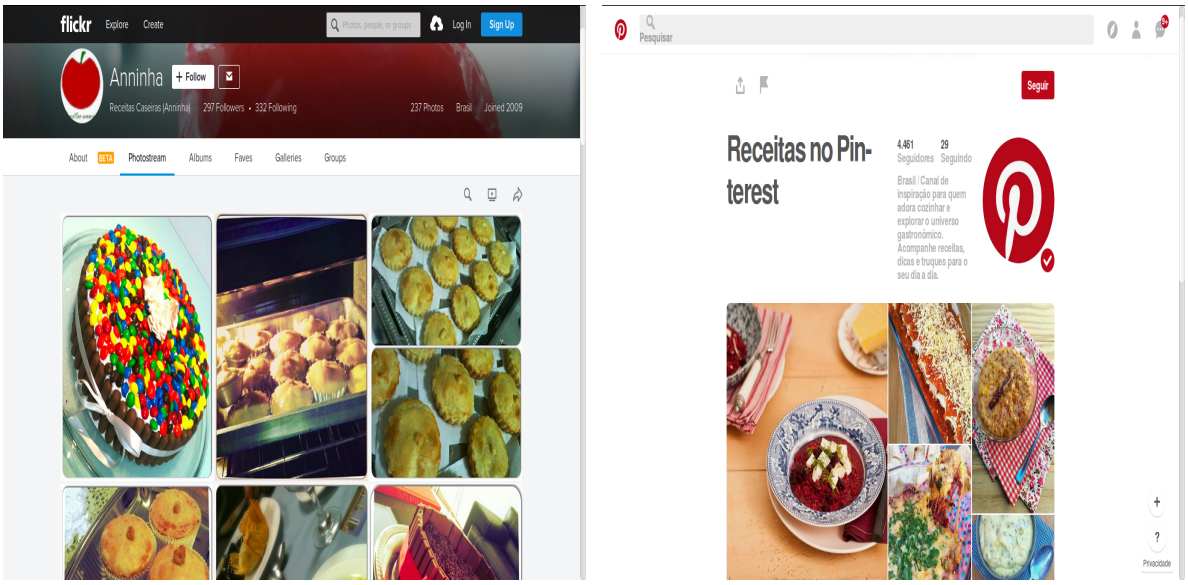

Figura 1 – Redes sociais de compartilhamento de imagens.

Em (BURKE, 2002), o autor define Sistemas de Recomendação como sistemas computacionais capazes de produzir recomendações individualizadas, ou que tenha o efeito

[1] https://www.facebook.com/
[2] https://www.pinterest.com/
[3] https://www.instagram.com/
[4] https://www.flickr.com/

de guiar o usuário de forma personalizada a itens interessantes, diante de uma grande variedade de opções. A personalização da informação está relacionada ao modo pelo qual a informação e os serviços podem ser ajustados às necessidades específicas de um usuário ou comunidade (CALLAN, 2002).

Os sistemas de recomendação tem se tornado muito importantes em aplicações Web, uma vez que auxiliam os usuários a lidar com o problema da sobrecarga de informações (BOBADILLA et al., 2013). Vários algoritmos de recomendação foram propostos na literatura (NING; KARYPIS, 2011; NING; KARYPIS, 2012; DESHPANDE; KARYPIS, 2004; CREMONESI; KOREN; TURRIN, 2010) e podem ser categorizados como (i) métodos baseados a espaços de fatores latentes e (ii) métodos baseados em vizinhança (seja ela referente a usuários ou itens). Tais métodos podem ser usados para resolver tarefas de *predição da avaliação*, na qual tenta se inferir a avaliação que um usuário faria sobre um dado item, ou tarefas de *recomendação top-N*, na qual se recomenda uma lista de N itens para o usuário, ordenados conforme o grau de relevância previsto pelo método de recomendação. Os métodos baseados em espaços de fatores latentes demonstram ser superiores para resolver a tarefa de avaliação da predição, enquanto os métodos de vizinhança são melhores na tarefa de recomendação *top-N* (BOBADILLA et al., 2013; KABBUR; ; KARYPIS, 2013). Desses, isto é, as abordagens baseadas em vizinhança, os métodos baseados em itens - onde a semelhança entre os itens é calculada por uma medida predefinida - se mostram superiores às abordagens baseadas em usuários (BOBADILLA et al., 2013).

No entanto, os métodos baseados em itens sofrem de problemas de desempenho quando os itens são descritos em espaços de alta dimensionalidade que, geralmente, também são muito esparsos. A fim de aliviar o problema da alta dimensionalidade, os autores em (NING; KARYPIS, 2011) apresentaram o método *Sparse LInear Methods* (SLIM). O SLIM trata dados de alta dimensionalidade computando uma pontuação da recomendação do item com base em uma matriz de agregação esparsa que relaciona os usuários e os seus itens (Seção 3.3). O SSLIM (Side information SLIM) (NING; KARYPIS, 2012) acrescenta ao SLIM informações a cerca dos itens, como o diretor de um filme ou o gênero de uma música, não se concentrando apenas nas preferências passadas dos usuários por seus itens.

Outro grande desafio dos sistemas de recomendação é conhecido como o problema *cold-start* de itens, que consiste na recomendação de itens que nunca foram avaliados ou com pouca avaliação de preferência. Como exemplo de itens que se encaixam nesse contexto, podemos citar os que são disponibilizados a cada minuto na web.

Com base nesse cenário, este trabalho propõe a exploração do conteúdo das imagens por meio de características multimodais, que também são extremamente esparsas e possuem alta dimensão, para tratar a recomendação de itens, sendo eles, imagens. A recomendação baseia-se nos princípios gerais da abordagem clássica de recomendação

baseada em conteúdo, explorando-se características textuais (por exemplo, descrição) e visuais (por exemplo, cor) das imagens, como detalhados no Capítulo 3 e no Capítulo 4. Embora as representações multimodais sejam esparsas e de alta dimensão, uma aplicação direta do SLIM ou SSLIM implica na fusão de dois espaços de características diferentes (por exemplo, texto e cor) em um único espaço.

Nesse sentido, este trabalho propõe o LDSSLIM *(Low-Dimensional Side information SLIM)*. O LDSSLIM é um arcabouço desenvolvido para problemas de recomendação *top-N*. Especificamente, dado o histórico de preferências do usuário e descrições multimodais das imagens, são retornadas as N imagens mais relevantes para o usuário. Uma vez que as semelhanças entre os itens são representadas diretamente em uma matriz de informação auxiliar, a abordagem tem a vantagem de capturar as relações com base nas características entre os itens. O arcabouço propõe o uso de métodos de redução de dimensionalidade para combinar a descrição multimodal das imagens a partir de um espaço latente comum aos itens. Sendo essa a proposta nesse trabalho, descrever as imagens com o propósito de aliviar o problema da alta dimensionalidade. Ainda é válido ressaltar que o arcabouço é flexível para considerar diferentes métodos de redução de dimensionalidade, medidas de similaridade entre os itens, bem como a forma de representá-los.

1.1 Caracterização do Problema

Os sistemas de recomendação tem por objetivo sugerir itens a usuários, com base em seu histórico de preferências. Tais itens podem ser de naturezas diversas como fotos, filmes ou viagens. É possível fazer recomendações comparando as preferências de um usuário com um grupo de outros usuários com preferências semelhantes, bem como procurando itens com características similares aos que o usuário já demonstrou interesse no passado.

Conforme descrito anteriormente, este trabalho consiste em explorar características multimodais para a recomendação de imagens, combiná-las a fim de que a recomendação se torne mais precisa e redimensioná-las, através do uso de um redutor de dimensionalidade, a fim de diminuir a esparsidade e melhorar o desempenho do recomendador. As características exploradas são de natureza visual e textual. As características visuais consistem em descritores de imagens obtidos por meio de técnicas de Visão Computacional, ao passo que, as características textuais consistem em metadados das imagens, extraídos do título e da descrição, obtidos através de técnicas de mineração de textos e recuperação de informação.

Assim, dado um grupo de usuários, a solução proposta deverá identificar as top-N melhores recomendações, baseando-se (i) nas preferências dos usuários em relação à um conjunto de itens e (ii) em informação adicional dos itens. Essa tarefa está descrita na

Figura 2.

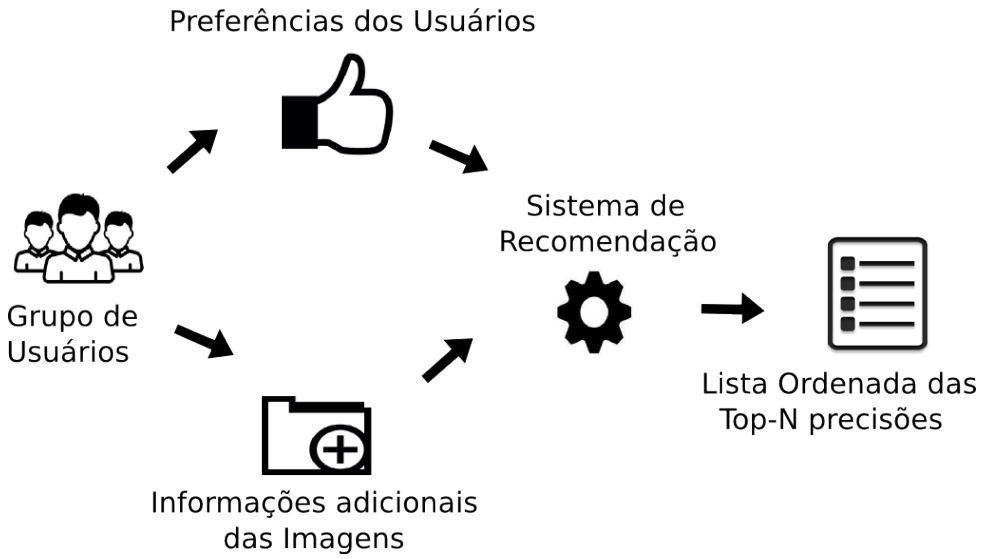

Figura 2 – Tarefa de recomendação das top-N precisões. A partir das preferências do usuário e de informações adicionais dos itens, gera-se uma lista ordenada dos itens recomendados.

Para obter os valores da recomendação, é necessário a execução de uma série de etapas, que são detalhadas no Capítulo 4. A Figura 3 traz uma visão geral dessas etapas que consistem em: (i) selecionar as imagens e demais informações significativas das bases de dados, (ii) com base nas informações coletadas, estabelecer a relação entre os usuários e seus itens, (iii) representar cada imagem por meio da extração das características visuais e textuais, (iv) reduzir a dimensionalidade dos vetores de representação das imagens, (v) calcular as similaridades entre os itens e, por fim, (vi) a execução do algoritmo de recomendação, recebendo como parâmetros de entrada as similaridades e associações entre os usuários e seus itens.

1.2 Motivação

Uma série de fatos podem ser listados para o desenvolvimento de um sistema como o proposto. Em primeiro lugar, a vasta quantidade de informação disponível na Web leva à necessidade por sistemas que auxiliem os usuários a filtrar conteúdo relevante às suas preferências. Nessa dissertação se escolheu os estudos e investigação dos sistemas de recomendação para ajudar os usuários na tarefa de filtragem de conteúdo de interesse.

Em segundo lugar, o problema de recomendação de imagens tem despertado interesse na comunidade (XIE et al., 2015) (PUEYO, 2010), isso devido a grande afeição das pessoas em registrar momentos através de fotografias e compartilhá-las em redes sociais. Tais ações foram facilitadas graças à popularização da *internet*, bem como de dispositivos fotográficos, gerando uma imensidão de informações disponíveis, o que dificulta

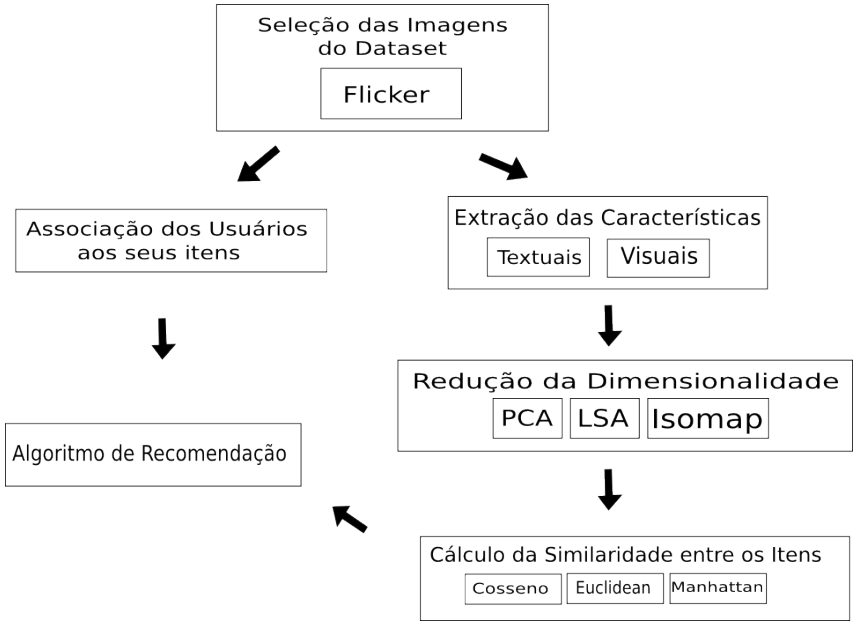

Figura 3 – Visão geral das principais etapas do arcabouço proposto.

uma busca com exatidão por parte dos usuários. Porém, na maior parte dos trabalhos propostos, apenas um tipo de característica é considerada. Em (BARRILERO M. ; ALDUAN, 2011) e (SUNDARESAN, 2014), por exemplo, os autores utilizaram apenas características visuais para descrever as imagens.

Por fim, sistemas de recomendação baseados em comparação de itens sofrem do problema de alta-dimensionalidade da representação dos itens. Assim, uma motivação para esse trabalho é investigar técnicas que permitam aliviar o problema da representação dos itens em espaços de alta-dimensionalidade e, assim, melhorar a eficácia dos itens recomendados.

Nesse trabalho, são propostos o uso de métodos distintos para a extração das características que descrevem as imagens, uma vez que a utilização de informação multi-modal é uma forma de aprimorar a descrição do item, além do estudo de como diferentes técnicas para a redução da dimensionalidade dos vetores de descrição dos itens impactam a sugestão de itens.

1.3 Objetivos: Geral e Específicos

O objetivo geral desta dissertação é investigar técnicas para representar e combinar características de imagens e, assim, aumentar a relevância da recomendação. Além disso, se propõe testar técnicas distintas de redução de dimensionalidade a fim de melhorar a acurácia do sistema por meio da redução da dimensionalidade na descrição das imagens. Como objetivos específicos, pretende-se:

- Propor novas estratégias de recomendação de imagens que explorem informações visuais, textuais e combinações dessas;
- Propor o uso de métodos diversos de redução de dimensionalidade em relação ao vetor de características das imagens;
- Avaliar experimentalmente o uso de diferentes características de representação de imagens em coleções de dados reais;
- Avaliar experimentalmente o uso de diferentes técnicas de redução de dimensionalidade em coleções de dados reais;
- Comparar o desempenho das estratégias propostas com abordagens do estado-da-arte em bases de dados reais.

Assim, este trabalho visa responder a seguinte questão:

- Características descritivas de natureza multimodal de imagens aliadas à métodos de redução de dimensionalidade podem ser utilizadas para melhorar a relevância da recomendação?

1.4 Contribuições

As principais contribuições desse trabalho são:

- Apresentação de um arcabouço geral para o problema de recomendação de imagens que é flexível em combinar representações multimodais dessas imagens;
- Propor e discutir um conjunto de alternativas para instanciar o arcabouço de recomendação de imagens proposto, a fim de destacar o comportamento e a eficácia de cada escolha;
- A investigação de novas estratégias de recomendação de imagens, por meio da utilização de características de diversas modalidades, como, por exemplo, modalidade visual, textual e combinações dessas, aliada ao uso de técnicas distintas de métodos de redução de dimensionalidade;
- A avaliação experimental dos métodos propostos que é realizada em termos de diferentes técnicas de avaliação de recomendações *top-N*, em dados provenientes de coleção de dados real, composta por imagens do *Flickr*[5];
- Avançar o estado-da-arte dos métodos baseados em itens usando técnicas de redução de dimensionalidade no problema de recomendação de imagem.

1.5 Organização do trabalho

Este trabalho está organizado em sete capítulos, incluindo a presente introdução. No Capítulo 2, são apresentados os trabalhos relacionados a essa pesquisa. A fundamentação

[5]https://www.flickr.com/

teórica utilizada para o desenvolvimento do trabalho é apresentada no Capítulo 3. No Capítulo 4, é descrita a metodologia proposta. No Capítulo 5, são apresentados e discutidos os experimentos e resultados obtidos. Por fim, no Capítulo 6, são apresentadas as conclusões da pesquisa e possíveis linhas de trabalhos futuros.

2 Trabalhos Relacionados

No Capítulo 1 foram apresentadas as motivações desse trabalho, bem como a contextualização do mesmo e as principais contribuições. Neste capítulo são apresentados os trabalhos relacionados ao tema dessa pesquisa. Os trabalhos citados foram agrupados de acordo com a categoria a que pertencem: (i) sistemas de recomendação no domínio específico de imagens, (ii) sistemas de recomendação que utilizam de aspectos multimodais, independentemente do domínio no qual é aplicado, (iii) sistemas de recomendação multimodal para imagens, (iv) métodos lineares esparsos, e, por fim, (v) sistemas de recomendação com uso de redutores de dimensionalidade.

2.1 Recomendação de Imagens

Sistemas de recomendação de imagens tem sido utilizados em larga escala e, por isso, são alvos de diversos estudos. Tradicionalmente, estes trabalhos utilizam conteúdo visual das imagens como forma de descrevê-las. A seguir são listados trabalhos que possuem por finalidade a recomendação de imagens, bem como a comparação desses trabalhos com o trabalho proposto nessa dissertação.

Em (SUNDARESAN, 2014), os autores apresentam um sistema de recomendação em larga escala para moda, cujas características das imagens foram extraídas utilizando-se informação de cor representadas de duas formas: (i) histograma HSV e (ii) BoW (*Bag of Words*) de cores. Também foram realizados experimentos explorando-se características de textura (transformada de *Wavelets* e Dense-SIFT), mas foram instáveis para a classificação de textura em roupas, devido à natureza deformável da roupa que leva, muitas vezes, a uma estimativa de textura não confiável, sendo mesmo rugas e dobras em roupas sólidas mal interpretadas como textura. Além disso, também foram examinados recursos baseados em estilo com base em atributos de roupas como cortes, manga, etc., mas não foi possível encontrar um descritor de estilo robusto na literatura. Desse modo, apenas representações de imagens baseadas em cores foram utilizadas. O objetivo da proposta é prever o que poderia ser interessante para o usuário com base em históricos passados, cuja suposição é que as imagens de moda da rua (ou seja, imagens de moda usual, e não de passarelas) contêm padrões de co-ocorrência que refletem os gostos atuais da moda dos compradores. Os autores propuseram um conjunto de algoritmos em duas classes de modelos: (i) DFR (*Deterministic Fashion Recommender*) e (ii) SFR (*Stochastic Fashion Recommender*), e avaliaram seus prós e contras. Para isso, os autores montaram uma coleção de imagens que compreendem roupas superiores e inferiores que co-ocorrem na mesma imagem. Este conjunto é utilizado para o treinamento, onde são observados relacionamentos de co-

ocorrência entre as roupas superiores e inferiores, como partes de fundos cinza escuro/preto e partes coloridas, e esses padrões detectados são então aprendidos. Outro conjunto de imagens, constituído apenas por imagens de saias (roupas inferiores), é utilizado como consultas (teste) a fim de recuperar a roupa superior. O resultado da recuperação é uma lista classificada das imagens ordenadas por relevância para uma consulta. Uma vez que existem vários algoritmos SFR/DFR propostos, os resultados da recuperação são múltiplas listas classificadas onde cada lista corresponde a uma saída de algoritmo diferente.

Em (SAXENA et al., 2013), os autores propuseram um sistema de recomendação baseado em imagens de sapato, a fim de ajudar a modernizar o setor de varejo. O sistema avalia um conjunto de dados de mais de 13.000 imagens de sapatos masculinos. A recomendação é realizada extraindo-se características de natureza visual das imagens, baseando-se em informações de forma, cor e textura associadas a cada um desses sapatos. Em seguida, são recomendados uma lista de dez sapatos que são semelhantes à imagem de entrada de um usuário com base nesses recursos.

Ambas as propostas utilizam apenas características visuais para realizar a recomendação, o que difere do trabalho proposto nesta dissertação que emprega características textuais, além das visuais, bem como combinações dessas, extraídas de formas variadas, para a tarefa de recomendação. Além disso, o trabalho proposto difere por não realizar a tarefa de predição e sim, recomendação *top-N*, bem como a não restrição do sistema à itens de um domínio específico (como por exemplo, apenas itens relacionados à moda ou à sapatos).

Em (LEMOS et al., 2012), os autores apresentaram um sistema de recomendação de imagens para dispositivos móveis que explora características de contexto, tanto do usuário quanto da foto, como um meio de melhorar a recomendação. Em sua execução, o método explora o contexto atual do usuário adquirido pelo seu dispositivo móvel, bem como o contexto das imagens, uma vez que a maioria das imagens são capturadas por dispositivos móveis que armazenam a localização, data, hora e outras informações contextuais. Com base nessas informações o sistema calcula uma similaridade entre contextos e recomenda para os usuários fotos criadas em contextos semelhantes ao contexto atual do usuário. Dessa forma, o sistema busca beneficiar a dois tipos de usuário, onde o primeiro tipo são aqueles que estão em um contexto incomum (por exemplo, visitando uma local turístico pela primeira vez) e eles podem então apreciar as fotos recomendadas para ter referências a atividades ou novos locais para explorar e, o segundo, usuários que já estiveram nesse contexto semelhante e que as fotos recomendadas podem dar uma nova visão e perspectiva da situação que eles encontram.

Alguns pontos diferenciam este trabalho do trabalho proposto. O primeiro deles é o fato de o trabalho proposto não ser projeto para dispositivos móveis e para uso *online*. Outro ponto é que o trabalho proposto utiliza características de natureza visual e textual

apenas para descrição dos itens, não utilizando informações de contexto. Além disso, não há características extraídas dos usuários, de modo que a relação entre os usuários e os itens se dá associando os itens favoritos de cada usuário. Uma terceira diferença ocorre no cálculo das similaridades onde, no trabalho proposto, esse cálculo é realizado apenas entre itens e, então, itens semelhantes a itens que são do agrado do usuário são recomendados, enquanto em (LEMOS et al., 2012), a similaridade é calculada entre os contextos dos usuários e os contextos dos itens.

Em (AARTHI; SAMPATH, 2013), os autores utilizaram filtragem colaborativa baseada em item para recomendação de imagens. Por meio da matriz usuário-item, que consiste na matriz que associa os usuários e os itens que são relevantes à ele, os autores identificaram as relações entre diferentes itens e, indiretamente, usaram esses relacionamentos para calcular as recomendações para os usuários. Para isso, utilizaram um fenômeno natural conhecido como difusão de calor (*heat diffusion*). O fenômeno mostra que o fluxo de calor ocorre a partir de uma posição com alta temperatura para outra de baixa temperatura. Segundo eles, o processo de influência entre pessoas é muito semelhante à esse fenômeno onde, em uma rede social por exemplo, os usuários inovadores e adotadores iniciais de um produto ou inovação atuam como fontes de alta temperatura e têm uma quantidade muito alta de calor à margem desta rede social; os demais usuários são tidos como posição de baixa temperatura e passam a adotar esse produto ou inovação da fonte de alta calor, ou seja, do usuário inovador. Assim, a recomendação de novos itens provenientes de usuários relevantes tende a possuir maior acurácia comparado à recomendações de novos itens provenientes dos demais usuários.

Já o trabalho proposto nesta dissertação utiliza a recomendação baseada em conteúdo, como técnica de recomendação, cujo foco principal é a forma de descrição dos itens e a similaridade entre os mesmos, de forma que os novos itens são recomendados à um usuário com base em certo um grau de semelhança entre os itens que esse usuário já demonstrou interesse, não utilizando a filtragem colaborativa e a relação direta entre os usuários. Trata-se de abordagens distintas utilizadas para um mesmo propósito: recomendar novos itens a um usuário, onde os itens são as imagens.

2.2 Recomendação Multimodal

Sistemas de Recomendação de natureza multimodal tem sido utilizados em diferentes contextos, e tem sido objetos de estudo na área devido a possibilidade de representar os itens de maneira mais precisa. Nessa seção são apresentados trabalhos de diferentes domínios que utilizam como fonte a extração de características multimodais.

Em (YANG et al., 2007), os autores utilizaram recomendação multimodal em um sistema *online* para recomendação de vídeos. O sistema dispõe de características textuais

(a) diretas: texto fornecido pelo próprio usuário, legendas ocultas (CC - *Closed Caption*), reconhecimento automático de fala (ASR) e reconhecimento óptico de caracteres (OCR), e (b) indiretas: categorias dos vídeos (por exemplo, praia, deserto, flor, música). Além disso, características visuais: histograma de cor, intensidade e frequência; e características de natureza auditivas do vídeo, levando-se em conta a velocidade média do som em cada janela de tempo, também foram utilizadas. Combinado às informações de natureza multimodal, os autores utilizaram dados de *feedback* dos usuários de modo a descobrir as preferências dos usuários em relação aos itens obtendo, assim, melhor acurácia nas recomendações.

Em (LI et al., 2014), os autores também realizaram recomendação multimodal de vídeos, dessa vez com o objetivo de prever a relevância de um item para um usuário. Os autores basearam-se em características de áudio (informações sonoras) e imagem (informações visuais, como cores). Além disse, cada vídeo foi avaliada de forma positiva ou negativa para o usuário, com base em avaliações e tempo de execução dos vídeos por parte dos usuários. Eles utilizaram o classificador SVM (*Support Vector Machine*) (VAPNIK, 1995), um algoritmo de aprendizado de máquina, para aprender as preferências dos usuários com base nas informações e avaliações dos vídeos, e modelaram o problema como um problema de classificação binária. Assim, um item pode ser classificado como relevante ou não-relevante para o usuário. Os autores puderam conferir a eficácia do método para problemas de recomendação de vídeos comparando-o à métodos tradicionais como utilizar o cosseno do ângulo para calcular a similaridade entre os itens.

Em (SCHEDL; SCHNITZER, 2014), os autores propuseram um sistema de recomendação multimodal de música que utiliza informação do conteúdo da música, aspectos do contexto da música e do contexto do usuário. As informações de conteúdo e contexto das músicas foram incorporadas usando extratores de características do estado-da-arte, enquanto as informações de contexto dos usuários foram abordadas levando-se em conta a preferência musical e dados geoespaciais do usuário, extraídos do conjunto de dados.

Em (ARYAFAR; SHOKOUFANDEH, 2014), os autores também utilizam uma abordagem multimodal de extração de características, dessa vez não para recomendar, mas para classificar e identificar um artista. Para isso, eles utilizam características de áudio (como a frequência) e de texto (como as letras das canções e *tags*). Os autores utilizaram o *TF-IDF* para representar os itens em forma vetorial, por meio das características, e esses vetores foram usadas para o treinamento de um classificador que se baseia nos k vizinhos mais próximos combinado ao algoritmo de recomendação *SLIM* para classificar os artistas.

Como pode ser observado, informações de natureza multimodal podem ser utilizadas de diversas formas em sistemas de recomendação. Porém, o foco desse trabalho, diferentemente dos mencionados acima, é aplicar abordagem multimodal em um domínio específico de imagens.

2.2.1 Recomendação Multimodal de Imagens

Nesta subseção apresenta-se trabalhos de natureza multimodal, porém no domínio específico de imagens, alvo dessa dissertação. Foram apresentados os trabalhos relacionados e as principais diferenças destes para o trabalho proposto.

Em (PUEYO, 2010), os autores utilizaram uma abordagem multimodal combinando características visuais, textuais e sociais para prever as fotos favoritas de um usuário no sítio *Web Flickr*. O trabalho foca na utilização das características multimodais para treinamento de um classificador baseado em árvore de decisão. Os autores mostraram que características textuais, visuais e sociais capturam efetivamente as necessidades de usuários mais ativos do *Flickr*, uma vez que cada usuário possui diferentes motivações para escolher uma foto. Nesta dissertação a tarefa de recomendação consiste em retornar uma lista de imagens ordenada pela relevância prevista pelo algoritmo de recomendação enquanto, em (PUEYO, 2010), os autores estão interessados em prever a nota que um usuário daria para uma foto.

Em (PEDRO; SIERSDORFER, 2009), os autores utilizaram características (a) visuais, como cor, contraste, brilho, saturação, e (b) textuais, como *tags* que foram manipuladas usando o TF e o TF-IDF, para classificar imagens em atrativas e não atrativas. Para as características textuais, foi ordenada uma lista de 50 termos, considerando-se várias categorias, para fotos já classificadas em atrativas e não-atrativas (onde as fotos marcadas mais de 5 vezes como favoritas foram tidas como atrativas e, do contrário, não-atrativas). Utilizando uma base de 12.000 imagens no total. A partir dessas listas, foram construídos os vetores textuais para as imagens a serem testadas. Os autores utilizaram o classificador SVM (*Support Vector Machine*). O trabalho proposto também utiliza características visuais e textuais mas para realizar uma recomendação top-N de imagens e não trabalha com a tarefa de agrupamento ou classificação.

Em (ZHU et al., 2013), os autores propuseram um novo método de recomendação multimodal denominado MSLIM (*Multimodal Sparse Linear Integration Method*), para sugestão de itens baseado em conteúdo. Para tal, utilizaram características visuais: HSV (histograma de cor), HOG (borda) e SIFT (descritor) e textuais - onde termos e palavras-chaves foram representadas em um vetor binário conforme sua frequência - para a descrição das imagens e o algoritmo de vizinhança KNN para encontrar os vizinhos de cada imagem. Os resultados da recomendação foram gerados, de acordo com as pontuações de similaridade dos vizinhos. A abordagem proposta também difere de todos os trabalhos mencionados quanto ao acréscimo dos redutores de dimensionalidade ao sistema de recomendação.

2.3 Métodos Lineares Esparsos

Como mencionado anteriormente, o quadro proposto neste trabalho é motivado por uma classe de métodos baseados em itens conhecidos como Métodos Lineares Esparsos, que foram desenvolvidos para a recomendação *top-N*.

O primeiro membro da classe é o algoritmo recomendador SLIM (NING; KARYPIS, 2011), que considera apenas a matriz de classificação como entrada para calcular as recomendações, ou seja, apenas a relação entre usuários e itens. O outro membro importante da classe de métodos lineares esparsos é conhecido como SSLIM (NING; KARYPIS, 2012), que propõe explorar, além da matriz de classificação, informações descritivas sobre os itens.

Esta classe de métodos de recomendação representa o estado da arte nos sistemas de recomendação *top-N*. Os métodos SLIM e SSLIM são detalhados na Seção 3.3.

Em (CREMONESI; KOREN; TURRIN, 2010), os autores avaliaram - através de métricas de precisão - o desempenho de diferentes algoritmos de filtragem colaborativa para tarefas de recomendação *top-N*, como os baseados em vizinhança e os baseados em fatores latentes. Com base nos experimentos realizados, os autores consideraram que o algoritmo *PureSVD* possui melhor desempenho, superando métodos baseados em fatores latentes mais sofisticados e detalhados.

O trabalho proposto nesta dissertação tem por base os algoritmos do estado-da-arte citados anteriormente, (NING; KARYPIS, 2011; NING; KARYPIS, 2012), e não utiliza decomposição do tipo SVD.

2.4 Recomendação com uso de Redução de Dimensionalidade

A maioria dos problemas de recomendação precisam lidar com dados em altas dimensões, tal fato motiva a utilização de redutores de dimensionalidade aliados aos algoritmos de recomendação. A representação de redução de dimensionalidade de usuários e itens é aplicada com sucesso em sistemas de recomendação.

Em (BARRILERO M. ; ALDUAN, 2011), os autores apresentaram um sistema de recomendação de imagens baseado em conteúdo de baixa dimensão implementado a nível de rede por meio de uma Rede de Conteúdo Consciente (CAN - *Content-Aware Network*). Eles observaram a afinidade entre itens preferidos dos usuários para recomendar os novos itens. Para isso, os autores utilizam características de baixo nível das imagens, provenientes dos parâmetros do *software* MPEG-7 (luminância, cor e textura), para descrevê-las. O processo de recomendação proposto pelos autores foi realizado por meio de três etapas: (i) análise de preferências dos usuários, (ii) processamento de imagens desconhecidas e (iii)

recomendação final. Na primeira etapa, foram extraídas as características de cada imagem via parâmetros do *software* MPEG-7 construindo, assim, um vetor de representação das mesmas. Com esses vetores foi construída uma matriz de representação dos itens e sobre essa matriz foi aplicado o PCA a fim de diminuir sua dimensionalidade. Na segunda etapa, são extraídas (usando-se os mesmos parâmetros da primeira etapa) as características das imagens desconhecidas (isto é, as imagens não avaliadas pelos usuários). Dessa forma, as novas imagens são comparadas às já conhecidas, para, assim, recomendar as que são mais semelhantes às preferidas dos usuários. Na terceira etapa, é realizado um cálculo de distância entre as imagens conhecidas do usuário e as desconhecidas, uma vez que ambos os vetores que as representam passaram pelo mesmo processo e, com base na distância de *Mahalanobis*, as imagens mais próximas são selecionadas e recomendadas ao usuário. Após realizado o processo de recomendação, os autores utilizaram um rede CAN para tornar a implementação do recomendador à nível de rede. Nesse cenário, o conteúdo das imagens é gerenciado por "Nodos de Conteúdo", onde seus metadados associados são armazenados em um Componente Multimídia que permite que a rede acesse e roteie informações da imagem e seus metadados, de forma que o conteúdo das imagens sejam atualizados constantemente, fornecendo uma descrição enriquecida das mesmas, de forma que o recomendador é alimentando constantemente melhorando, assim, sua performance.

Em (SYMEONIDIS, 2008), os autores propuseram um sistema de recomendação combinando as técnicas de filtragem colaborativa e informações de conteúdo do usuário, se concentrando em explorar uma representação de elemento modal único, onde o vetor de características que representa o usuário é determinado pela frequência de seus termos. Para reduzir a dimensionalidade da representação do usuário, os autores utilizam o modelo de Indexação Semântica Latente (LSI), a fim de revelar suas características dominantes. Após a redução, os k vizinhos mais próximos de cada usuário são recomendados, com base na similaridade calculada via cosseno.

O trabalho proposto nesta dissertação, concentra em aplicar técnicas de redução de dimensionalidade sobre os vetores de representação dos itens, que possuem informações de natureza multimodal, enquanto os trabalhos citados se concentram em uma única dimensão para a representação dos itens.

3 Fundamentação Teórica

Neste capítulo são introduzidos os principais conceitos e técnicas que fundamentam este trabalho. Na Seção 3.1, apresenta-se os conceitos de sistemas de recomendação de um modo geral, seus tipos, tarefas e tipos de *feedbacks*. Na Seção 3.2, apresenta-se os métodos de representação das imagens, abordando as características textuais, visuais e os métodos de redução de dimensionalidade. Por fim, na Seção 3.3, é apresentado o SSLIM - algoritmo de recomendação base do arcabouço proposto nesta dissertação.

3.1 Sistemas de Recomendação

Sistemas de Recomendação são ferramentas computacionais que têm como objetivo sugerir itens relevantes para usuários (JANNACH et al., 2010). Podem ser definidos como sistemas que procuram auxiliar indivíduos a identificar conteúdo relevante em um conjunto de opções. Esse conteúdo pode ser de diversos tipos como, por exemplo, fotos, livros, filmes, músicas, viagens, notícias, dentre outros.

Sistemas de Recomendação podem ser divididos em três tipos: sistemas baseados em filtragem colaborativa, sistemas baseados em conteúdo e sistemas híbridos (RICCI; ROKACH; SHAPIRA, 2011). Os sistemas colaborativos utilizam como base as preferências de um usuário comparada a um grupo de outros usuários. Os sistemas baseados em conteúdo identificam itens com características similares aos que o usuário já demonstrou interesse no passado. Ainda existe a possibilidade de combiná-los, gerando uma abordagem híbrida de recomendação. Tais abordagens serão explicadas mais detalhadamente a seguir.

3.1.1 Tipos de Sistemas de Recomendação

Sistemas de Recomendação podem ser classificados basicamente em três abordagens, a partir de como a recomendação é feita: filtragem colaborativa (*collaborative filtering*), baseado em conteúdo (*content-based*) e abordagem híbrida (*hybrid recommendation*) (RICCI; ROKACH; SHAPIRA, 2011).

Os métodos baseados em **filtragem colaborativa** baseiam-se na recomendação de itens que pessoas com gosto semelhante preferiram no passado. Nos sistemas de recomendação colaborativos, a essência está na troca de experiências entre as pessoas que possuem interesses comuns. Para cada usuário procura-se identificar um conjunto de “vizinhos próximos”, que são assim classificados por possuírem um comportamento semelhante (ADOMAVICIUS; TUZHILIN, 2005a). Analisa-se a vizinhança do usuário a partir

da regra: “Se um usuário gostou de A e de B, um outro usuário que gostou de A também pode gostar de B”.

A Tabela 1 apresenta um exemplo de como a filtragem colaborativa funciona. Assim, se quisermos recomendar uma foto ao usuário Mauro procuraremos outros usuários com hábitos de consumo semelhantes. No caso, Paulo e João já gostaram de fotos que Mauro também gostou (Foto 2). Em seguida, recomendamos a Mauro fotos que estes dois outros usuários gostaram, mas que Mauro ainda não gostou, como a Foto 1 e a Foto 5. A decisão sobre a recomendação destes produtos baseia-se no histórico de avaliações comuns e no valor de predição calculado.

Usuários	**Foto 1**	**Foto 2**	**Foto 3**	**Foto 4**	**Foto 5**
João	x	x			
Rosana			x	x	x
Jéssica	x		x		
Paulo		x			x
Mauro	?	x			?

Tabela 1 – Recomendação baseada em Filtragem Colaborativa.

A Figura 4 representa um modelo de recomendação baseada em filtragem colaborativa. Note que a similaridade entre usuários determina os itens que serão recomendados.

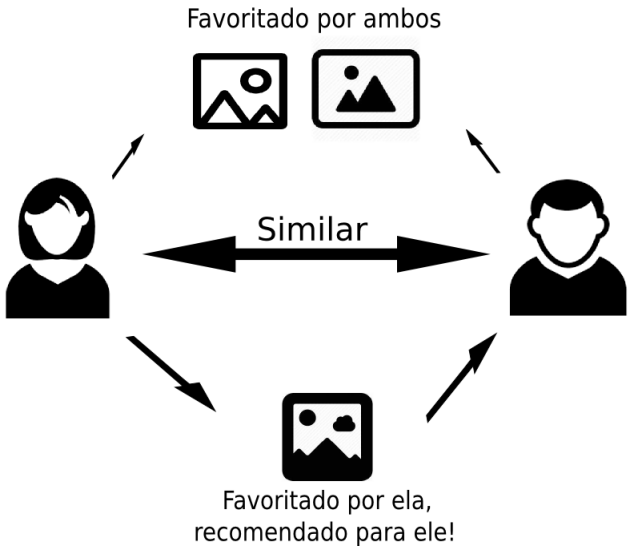

Figura 4 – Recomendação baseada em filtragem colaborativa. Na recomendação baseada em filtragem colaborativa a similaridade entre perfis de usuários é o fator determinante para se realizar a recomendação. Na imagem, a semelhança entre o consumo dos dois usuários faz com que um item que foi consumido apenas por um deles seja recomendado para o outro.

No entanto, a abordagem colaborativa enfrenta alguns desafios. Os usuários do sistema, geralmente, avaliam poucos itens, fazendo com que os dados disponíveis sobre

avaliações sejam esparsos. Outro ponto é que para um item ser recomendado é necessário que ele tenha sido avaliado pelos usuários similares ao usuário ativo. Desta forma, um item que nunca recebeu nenhuma avaliação não é recomendado. Da mesma forma, um item novo, recém inserido no sistema e que ainda não possui avaliação, também não é recomendado até que um número expressivo de usuários o avalie. Este problema é denominada como *cold-start* de item (ADOMAVICIUS; TUZHILIN, 2005b).

Já um sistema de recomendação **baseado em conteúdo** recomenda ao usuário ites que sejam semelhantes ao que ele preferiu no passado. Desta forma, a abordagem baseada em conteúdo parte do princípio de que os usuários tendem a interessar-se por itens similares aos que demonstraram interesse anteriormente (HERLOCKER, 2000). Em um cenário de recomendação de filmes, por exemplo, um usuário que, assiste e gosta do filme "*Matrix*" teria recomendações do gênero ação e ficção científica.

A Figura 5 representa um modelo de recomendação baseado em conteúdo. Note que é a similaridade entre itens o fator que define a recomendação, isto é, a semelhança entre um item que o usuário já gostou no passado e um item que ele ainda não conhece.

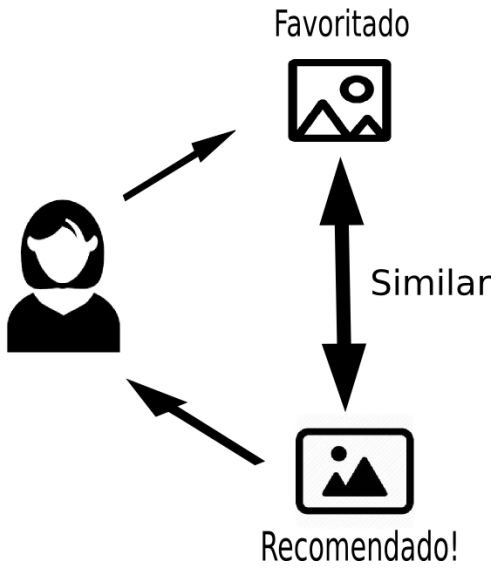

Figura 5 – Recomendação baseada em conteúdo. Na recomendação baseada em conteúdo apenas a similaridade entre os itens que o usuário não conhece e os itens que ele já gostou é levada em conta para se realizar a recomendação.

No entanto, a abordagem baseada em conteúdo sofre o problema da super especialização do sistema em tópicos frequentes no perfil do usuário. Dessa forma, ao aprender o perfil do usuário o sistema não inova em suas recomendações tendo em vista a pouca diversidade de itens avaliados pelo usuário (BEZERRA, 2004).

Por fim, um sistema **híbrido** consiste em combinar duas ou mais abordagens (como filtragem colaborativa e baseado em conteúdo, por exemplo) com o objetivo de oferecer

sugestões mais precisas e driblar os problemas individuais enfrentados por cada uma, isto é, *cold-start* de item e super especialização.

3.1.2 Tarefas de Recomendação

Um sistema de recomendação tem por finalidade predizer se um determinado usuário irá gostar de um item em particular - *tarefa de predição de nota* - ou identificar um conjunto de N itens que serão de interesse de um certo usuário - *recomendação top-N* (DESHPANDE; KARYPIS, 2004).

A tarefa de **predição de nota** consiste em predizer as notas que os usuários darão para itens ainda não avaliados por eles. Essa tarefa tem seu resultado comumente avaliado pela medição dos erros cometidos nas predições. Para tal, são utilizadas medidas como, por exemplo, MAPE (*Mean Absolute Percentage Error*) e RMSE (*Root Mean Squared Error*) (HYNDMAN; KOEHLER, 2006).

Um exemplo do que se deseja realizar em uma tarefa de predição se encontra na Tabela 2. As linhas da tabela correspondem aos usuários, as colunas aos itens e o seu preenchimento se dá com as notas dadas pelos usuários a alguns itens. A tarefa consiste em prever quais seriam os valores dados pelos usuários para itens ainda não avaliados, utilizando como base as notas já existentes.

Tabela 2 – Tarefa de predição de notas. O objetivo é encontrar o melhor valor que prevê a nota que o usuário daria para itens que ele ainda não avaliou.

Usuários	**Imagem 1**	**Imagem 2**	**Imagem 3**
Ted	4	2	1
Carol	?	3	4
Bob	?	4	?

Já a tarefa de **recomendação top-N** consiste em identificar os N melhores itens que satisfaçam o interesse de um determinado usuário. Neste caso, a nota dada pelo usuário para cada item não é o mais importante, mas sim a ordem em que eles aparecem na lista de itens de cada usuário. Deste modo, métricas de acurácia, tais como precisão e NDCG (*Normalized Discounted Cumulative Gain*), tornam-se mais adequadas para avaliação desta tarefa do que métricas de erro (CREMONESI; KOREN; TURRIN, 2010).

Com base na Tabela 2 usada anteriormente para ilustrar a tarefa de predição, com relação à tarefa de recomendação top-N não se deseja mais predizer a nota dada pelo usuário *Bob* ao *Item3*, mas sim gerar uma lista de tamanho N com os itens que melhor agradariam o usuário *Bob*, por exemplo. Note que a ordem dessa lista é de total relevância no processo, entendendo-se que o primeiro item da lista possui maior predição do que o n-ésimo item. A principal diferença nesse caso é o fato de que o objetivo da recomendação top-N não está relacionado à predição das notas em si, mas sim na predição

de uma ordenação de N elementos para cada usuário. O foco deste trabalho é a tarefa de recomendação top-N.

3.1.3 Tipos de Feedback

Em sistemas de recomendação, as preferências dos usuários são inferidas levando-se em consideração o *feedback* direto do usuário, que pode acontecer nas formas ímplicita ou explícita (PARRA et al., 2011).

Na forma **implícita**, as informações são obtidas ao se mensurar a interação do usuário com diferentes itens e pode ser aprendida com o comportamento do usuário com o tempo. Por exemplo, podemos citar o monitoramento de *clicks* em páginas de *e-commerce*, tempo de visita, *cookies* do *browser*, localidade geográfica, histórico de sites visitados, a quantidade de vezes que uma música foi executada, entre outros. Já na forma **explícita**, as informações são obtidas pela consulta direta ao usuário, como por exemplo notas dadas a um determinado item, uso de questionários, entre outros.

Nesta dissertação há um interesse particular em estratégias que exploram *feedback* explícito, já que as informações foram obtidas através dos *likes* (curtidas) realizados pelos usuários, para a coleção de dados do *Flicker*.

3.2 Métodos de Representação de Imagens

Nessa dissertação, os itens (imagens) foram representados através da extração de características de natureza textual e visual. As características visuais consistem em informações obtidas da própria imagem, utilizando-se para isso técnicas de visão computacional. Já as características textuais consistem em informações obtidas dos textos que acompanham as imagens. Também foram empregados métodos de redução de dimensionalidade do vetor de características. Tais técnicas são apresentadas nessa seção.

3.2.1 Características Textuais

As características textuais foram obtidas por meio da utilização de duas métricas: TF (*Term Frequency*) e TF-IDF (*Term Frequency–Inverse Document Frequency*).

O **TF**, como o próprio nome infere, captura a frequência dos termos em um mesmo documento, ou seja, ele conta o número de vezes que um determinado termo aparece no documento (YAMAMOTO; CHURCH, 2001).

Já o **TF-IDF** captura a frequência dos termos (TF) em um documento e pondera sua importância por meio da raridade do termo (IDF), de modo a diminuir o peso dos termos que ocorrem mais frequentemente no conjunto de textos selecionados (BAEZA-YATES; RIBEIRO-NETO et al., 1999). Na prática, o valor do TF-IDF consiste na multiplicação de

duas medidas calculadas de forma independente: a frequência do termo (TF) e o inverso da frequência do termo (IDF). A Equação 1 mostra como é calculado o valor do TF-IDF para cada termo i em um documento j.

$$tf * idf_{ij} = tf_{ij} * \log \frac{N}{df_i} \tag{1}$$

onde tf_{ij} corresponde à frequência do item i no documento j; N corresponde ao número total de documentos e df_i corresponde ao número de documentos que contem o termo i.

3.2.2 Características Visuais

As características visuais foram obtidas utilizando-se técnicas das áreas de Visão Computacional e Processamento de Imagens e são descritas nas próximas subseções.

3.2.2.1 Histograma de Cor

Cada camada de uma imagem pode ser decomposta em um ou mais canais de cores. O histograma utilizado decompõe a imagem em 3 canais: os canais RGB, R (*Red* - vermelho), G (*Green* - verde) e B (*Blue* - azul), onde cada canal suporta uma faixa de cores de 0 a 255 (valores inteiros), sendo que um *pixel* preto tem o valor 0 em todos os canais de cor, e um *pixel* branco tem o valor 255 em todos os canais.

A representação de cada imagem é feita por meio de um vetor, contendo os valores dos 3 canais para cada *pixel* da imagem.

3.2.2.2 SIFT

O *SIFT* (*Scale Invariant Feature Transform*) (LOWE, 1999) é um descritor de imagens que permite a detecção e extração de características locais, razoavelmente invariáveis a mudanças de iluminação, ruído, rotação, escala e perspectiva.

O método pode ser dividido em 4 etapas principais: (i) detecção dos extremos, (ii) localização dos pontos chave, (iii) definição da orientação e (iv) descritor dos pontos chave (RIBEIRO, 2015; SILVA, 2014).

Na primeira etapa são detectados os pontos chave (pontos de interesse). Isto é feito utilizando-se a função *Diferença das Gaussianas* (DoG) de modo a identificar os potenciais pontos de interesse invariáveis a escala e a rotação. Esta é a parte mais custosa do algoritmo.

A segunda etapa consiste em localizar os pontos detectados, rejeitando aqueles que são suscetíveis a ruído. Para cada extremo detectado, um modelo detalhado é ajustado

de modo a determinar sua localização e escala. Pontos chave são então selecionados baseando-se em medidas de estabilidade.

Na terceira etapa é definida a orientação de cada ponto chave por meio dos gradientes locais da imagem. Toda operação a partir de então será feita com relação a dados da imagem transformados em relação à orientação, escala e localização de cada ponto chave. Desta maneira se obtém invariância a essas transformações.

Por fim, é criado o descritor ao se medir gradientes locais em uma região vizinha a cada ponto de interesse. Estas medidas são então transformadas para uma representação que permite níveis significativos de distorção e mudança na iluminação. Para cada ponto de interesse, são definidas nxn regiões, com kxk *pixels* cada, ao redor da localização do ponto chave. Para cada região, tem-se um histograma em oito direções. Este histograma é construído com a magnitude dos *pixels* pertencentes a cada região. O descritor é então representado pelos histogramas das regiões.

Na Figura 6 é apresentado o processo de construção do descritor SIFT explicado na quarta etapa. Na primeira parte da figura, são ilustradas as regiões ao redor dos pontos chave, e na segunda parte, os histogramas pelos quais são feitas a descrição da imagem.

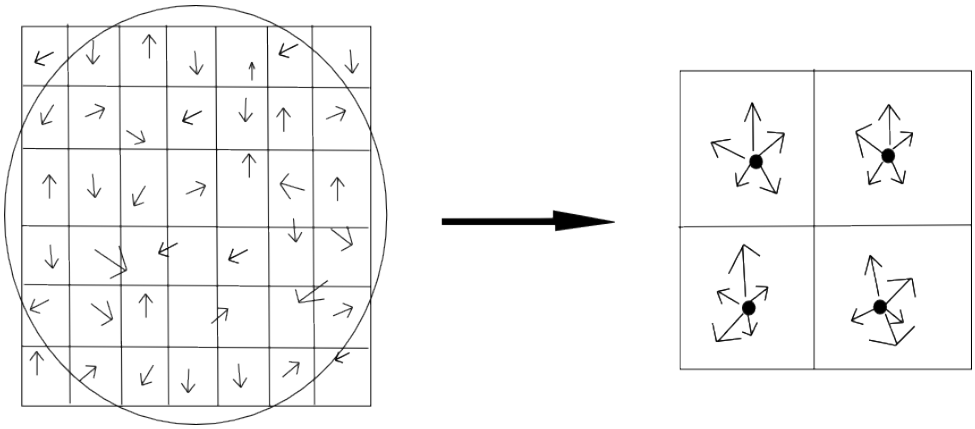

Figura 6 – Construção do descritor de pontos chave.

3.2.2.3 SURF

O *SURF* (*Speeded Up Robust Features*) (BAY; TUYTELAARS; GOOL, 2006) é um detector de características invariáveis a iluminação, escala e rotação em imagens. Consiste em uma versão do operador de diferença das Gaussianas (DoG), onde as transformadas de Haar (*Haar wavelets*) são usadas para calcular uma aproximação das derivadas de segunda ordem do núcleo Gaussiano.

Basicamente, a construção do descritor é dividido em 3 partes: criação da imagem integral, extração dos pontos de interesse e descrição dos pontos de interesse (BAY; TUYTELAARS; GOOL, 2006).

A extração dos pontos de interesse é feita com uso do determinante da matriz Hessiana H(x,σ), mostrada na Equação 2:

$$H(p,\sigma) = \begin{bmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{xy}(p,\sigma) & L_{yy}(p,\sigma) \end{bmatrix} \quad (2)$$

onde $L_{xx}(p,\sigma)$ é a convolução da derivada da Gaussiana de segunda ordem da imagem no ponto p, com as coordenadas (x,y) e σ é a escala da imagem. Para determinar a orientação é utilizada a transformada de Haar.

3.2.2.4 HOG

O *HOG* (*Histogram of Oriented Gradients*) (DALAL; TRIGGS, 2005) ou Histograma de Gradientes Orientados, é uma técnica que utiliza as orientações dos gradientes de uma imagem para obter seus descritores.

A ideia principal deste descritor é que a aparência e forma de objetos em uma imagem podem ser descritos através da distribuição dos gradientes de intensidade dos *pixels* ou pelas direções das bordas (GRITTI et al., 2008).

O processo para gerar o descritor pode ser dividido em quatro etapas: cálculo do gradiente em cada *pixel*, agrupamento dos *pixels* em células, agrupamento das células em blocos e obtenção do descritor.

Na primeira etapa, calcula-se o gradiente de cada *pixel* da imagem em relação a sua vizinhança, tanto no eixo vertical quanto no horizontal.

O passo seguinte é responsável por agrupar os *pixels* de uma determinada região, criando-se o que se chama de célula. Todas as células criadas na imagem possuem mesmo formato e tamanho. Cria-se também um histograma com orientação do vetor gradiente dos *pixels* que compõe essa célula, onde são computados os valores de magnitude de acordo com o ângulo do vetor. O histograma possui uma quantidade finita de divisões. Por apresentar melhor desempenho (GRITTI et al., 2008), utiliza-se o histograma com nove divisões.

Após a segunda etapa, os blocos são criados com o uso do agrupamento de células de uma certa região. Assim como as células, os blocos também sempre possuem o mesmo formato e tamanho em toda a imagem por motivos de padronização.

Na etapa final, cria-se o descritor, que é uma lista dos histogramas de todos os blocos.

3.2.3 Redução de Dimensionalidade

Métodos de redução de dimensionalidade consistem em uma forma de reduzir a dimensão dos dados de entrada, ou seja, diminuir o tamanho do vetor de representação das imagens, além de aliviar o problema da esparsidade na descrição dos itens. O problema da esparsidade ocorre, especialmente, na modalidade textual, uma vez que o vocabulário completo é infinitamente maior que o vocabulário das imagens favoritadas pelo usuário, tornando sua descrição extremamente esparsa. Nesse contexto, os métodos redutores de dimensionalidade tendem a aprender apenas informação que são de fato relevantes na descrição dos itens, de forma que a perda de informação natural da redução da dimensionalidade equivale a perda de conteúdos desnecessários (lixos).

Os métodos de redução de dimensionalidade podem ser divididos em duas categorias: (i) não-supervisionados e (ii) supervisionados. Basicamente, a redução de dimensionalidade não-supervisionada visa otimizar a minimização da representação dos dados, enquanto os métodos supervisionados de redução de dimensionalidade visam maximizar a informação discriminante entre classes.

Uma vez que os sistemas de recomendação são, por definição, um problema de aprendizagem não-supervisionado, foram utilizados os seguintes métodos de redução de dimensionalidade não-supervisionados: (i) PCA, (ii) LSA e (iii) ISOMAP, descritos nas subseções seguintes.

3.2.3.1 PCA

O PCA (*Principal Component Analysis* – Análise de Componentes Principais) (HOTELLING, 1933) possui por finalidade reduzir a dimensionalidade. Esta redução se dá através da representação do conjunto de dados em um novo sistema de eixos denominados componentes principais (PC), que permitem a visualização da natureza multivariada dos dados em poucas dimensões (SOUZA; POPPI et al., 2012).

O PCA consiste, portanto, em uma manipulação da matriz de dados com objetivo de representa-lá através de um número menor de características, construindo-se um novo sistema de eixos (denominados fatores, componentes principais ou variáveis latentes) para representar as amostras. Assim, a natureza multivariada dos dados permite sua visualização em poucas dimensões (FERREIRA et al., 1999).

Nos dados originais, as amostras são pontos localizados em um espaço n-dimensional, sendo n igual ao número de variáveis. Com a redução de dimensionalidade proporcionada pelo PCA, as amostras passam a ser pontos localizados em um espaço de dimensões reduzidas definidos pelos componentes principais, por exemplo, bi- ou tridimensionais. Matematicamente, no PCA, a matriz X é decomposta por um produto de duas matrizes, denominadas escores (T) e pesos (P), além de uma matriz de erros (E), como mostrado na

Equação 3:

$$X = TP^T + E \tag{3}$$

Os **escores** representam as coordenadas das amostras no sistema de eixos formados pelos componentes principais. Cada componente principal é constituído pela combinação linear das variáveis originais e os coeficientes da combinação são denominados **pesos**. Matematicamente, os pesos são os cossenos dos ângulos entre as variáveis originais e os componentes principais (PC), representando, portanto, o quanto cada variável original contribui para um determinado componente principal. O primeiro componente principal (PC1) é traçado no sentido da maior variação no conjunto de dados; o segundo componente (PC2) é traçado ortogonalmente ao primeiro, com o intuito de descrever a maior porcentagem da variação não explicada pelo PC1 e assim por diante; enquanto os escores representam as relações de similaridade entre as amostras. A avaliação dos pesos permite entender quais variáveis contribuem mais para os agrupamentos observados no gráfico dos escores. Por meio da análise conjunta do gráfico de escores e pesos, é possível verificar quais variáveis são responsáveis pelas diferenças observadas entre as amostras (SOUZA; POPPI et al., 2012).

3.2.3.2 LSA

O LSA (Análise Semântica Latente) (LANDAUER; FOLTZ; LAHAM, 1998) é um método de redução de dimensionalidade baseado na decomposição de valor singular (SVD – *Singular Value Decomposition*) (GOLUB; KAHAN, 1965) para encontrar uma aproximação de baixa classificação dos dados originais, de forma a reduzir o número de dimensões em um espaço vetorial com a menor perda possível de informação. Trata-se de uma técnica matemática/estatística, totalmente automática, para extrair e inferir as relações de uso contextual esperadas das características dos dados. O LSA foi inicialmente proposto no contexto de representação de documentos textuais, assim, nesta dissertação, aborda-se o seu funcionamento nesse contexto.

A primeira etapa do processo de redução de dimensionalidade seguido pelo LSA consiste na criação de uma matriz esparsa A que informa a co-ocorrência de termos por documentos (ou frações desse documento, como por exemplo, frases e parágrafos). Ou seja, trata-se de uma matriz mxn, onde m equivale à cada termo, n representa cada documento (ou fração do documento) e cada célula contém a frequência do termo no documento (ou na fração) (COSTA; MARTINS, 2015).

Em seguida, as entradas das células são submetidas a uma transformação preliminar, em que a frequência das células é ponderada por uma função que expressa tanto a importância da palavra na frase quanto o grau de informação que a palavra carrega no

documento (LANDAUER; FOLTZ; LAHAM, 1998).

A próxima etapa consiste em aplicar o SVD à matriz A, com o objetivo de reduzir o número de linhas (i.e., os termos correlacionados são agrupados em conceitos), preservando a estrutura das colunas (i.e. dos documentos). Nesta fase, a matriz A é decomposta no produto de três novas matrizes, conforma mostra a Equação 4.

$$A = UDV^T, \tag{4}$$

onde $U = [u_{ij}]$ é uma matriz ortonormal mxn cujas colunas são chamadas de vetores singulares a esquerda; $D = diag(\sigma_1, \sigma_2, ..., \sigma_n)$ é uma matriz diagonal nxn cujos elementos são chamados de valores singulares não negativos, os quais aparecem ordenados de forma decrescente; e $V = [v_{ij}]$ é uma matriz ortonormal nxn cujas colunas são chamadas de vetores singulares a direita.

A dimensão destas matrizes é então reduzida, eliminando as linhas e colunas correspondentes aos menores valores singulares da matriz D assim como seus correspondestes nas matrizes U e V. Dado o $posto(A) = k$, o SVD pode ser interpretado como o mapeamento do espaço de A em um espaço conceito (reduzido) de k dimensões, as quais são linearmente independentes.

Ainda que $posto(A)$ seja maior que k, quando seleciona-se apenas os k maiores valores singulares $(\sigma_1, \sigma_2, ..., \sigma_k)$ e seus correspondentes vetores singulares em U e V, pode-se obter uma aproximação de posto k para a matriz A (GONG; LIU, 2001). Deste modo, pode-se tratar vetores de termos e documentos como um *espaço conceito* de dimensão reduzida k, formalizado na Equação 5.

$$A_k = U_k D_k {V^T}_k \tag{5}$$

Por fim, normalmente, é utilizado o cosseno para calcular a similaridade dos vetores no espaço dimensional reduzido, a fim de obter a proximidade entre eles e descobrir o quão relacionados eles são. O valor do cosseno permite descobrir a proximidade entre um termo e um documento (ou fração do documento), e entre documentos.

3.2.3.3 ISOMAP

O ISOMAP (Mapeamento Isométrico) (TENENBAUM; SILVA; LANGFORD, 2000) é uma técnica de redução de dimensionalidade não-linear que utiliza uma distância geodésica (ou curvilínea) induzida por um grafo de item de vizinhança e tem como vantagem o fato de melhor preservar a estrutura geométrica dos dados.

O método é composto por três etapas, sendo elas: (i) construir um grafo de vizinhança, (ii) calcular os caminhos mais curtos e (iii) construir um espaço fixo de dimensão d.

A primeira etapa consiste em determinar quais pontos são vizinhos, com base em uma distância $d_x(i, j)$ entre pares de pontos i e j do espaço de entrada. Duas formas simples são (i) conectar cada ponto a todos os outros pontos dentro de algum raio fixo, ou (ii) a todos os seus k vizinhos mais próximos. Essas relações de vizinhança são representadas como um grafo ponderado G sobre os pontos de dados, com peso $d_x(i, j)$ nas arestas que ligam pontos vizinhos.

Na segunda etapa, o ISOMAP estima as distâncias geodésicas (ou distâncias curvilíneas) $d_M(i, j)$ entre todos os pares de pontos do espaço de entrada, calculando as distâncias de caminho mais baixas $d_G(i, j)$ no grafo G. A distância geodésica é calculada inicializando-se $d_G(i, j) = d_x(i, j)$ se os nodos i e j do grafo G são conectados por uma aresta, do contrário $d_G(i, j) = \infty$. Então, para os valores de $k = 1, 2, ..., n$, substitui-se todos os valores de $d_G(i, j)$ pelo $min\{d_G(i, j), d_G(i, k) + d_G(k, j)\}$ (KUMAR et al., 1994). Assim, a matriz D de valores finais $D_G = \{d_G(i, j)\}$ irá conter as distâncias mais curtas do caminho entre todos os pares de pontos em G. Outra forma de se calcular a distância geodésica é utilizar o algoritmo do *Dijkstra* (CORMEN et al., 2001).

A etapa final consiste em aplicar o método MDS clássico (COX; COX, 2000) à matriz D, de forma a construir uma associação dos dados em um espaço euclideano Y de dimensão reduzida d que melhor preserve a geometria intrínseca dos dados originais. Os vetores de coordenadas y_i dos pontos em Y são selecionados para minimizar a função de custo, conforma a Equação 6.

$$E = \quad \|\tau(D_G) - \tau(D_Y)\|_{L^2}, \tag{6}$$

onde D_Y denota a matriz de distâncias euclideanas $d_Y = \|y_i - y_j\|$ e $\|A\|_{L^2}$ é a norma L^2 da matriz - $\sqrt{\sum_{i,j} A_{ij}^2}$. O operador τ converte distâncias para produtos internos, que caracterizam de forma única a geometria dos dados através de uma otimização eficiente. O mínimo global da função de custo é definido configurando as coordenadas y_i para os principais vetores próprios da matriz de produto interno $\tau(D_G)$ obtidos da matriz D.

3.3 SSLIM

O arcabouço de recomendação proposto nesta dissertação é baseado no algoritmo SSLIM (*Sparse Linear Methods with Side Information*) (NING; KARYPIS, 2012), que, por sua vez, é um método baseado em fatores latentes para sistemas de recomendação top-N e consiste em uma extensão do método SLIM (*Sparse Linear Methods*) (NING; KARYPIS, 2011).

O SLIM é um método que foca em construir recomendações top-N de alta qualidade e eficientemente. O método aborda o problema de recomendação como um problema de otimização descrito na Equação 8 a fim de aprender uma matriz de coeficientes esparsa

para os itens no sistema a partir dos perfis dos usuários. A matriz aprendida deve obedecer a Equação 7.

$$\tilde{A} \sim AW, \tag{7}$$

$$\begin{aligned} \min_{W} \quad & \|A - AW\|_F^2 + \tfrac{\beta}{2} \|W\|_F^2 + \lambda \|W\|_1 , \\ \text{sujeito a} \quad & \begin{cases} W \geq 0 \\ diag(W) = 0 \end{cases} \end{aligned} \tag{8}$$

onde $\tilde{A}$ é a matriz das pontuações estimadas para A e é encontrada conforme a Equação 9 para cada usuário/item de forma isolada; A é a matriz binária usuário-item e W é a matriz de agregação dos coeficientes.

$$\tilde{a}_{ij} = a_i^T w_j, \tag{9}$$

onde $\tilde{a}_{ij}$ é a pontuação estimada de um usuário u_i em relação a um item ainda não estimado i_j, enquanto a_i^T corresponde a itens já avaliados pelo usuário e w_j é um vetor coluna esparso de coeficientes de tamanho n.

O SSLIM (*Sparse Linear Methods with Side Information*) agrega informações adicionais ao problema de otimização endereçado pelo SLIM. O SSLIM é especialmente aplicável a problemas nos quais há informação adicional sobre os itens, uma vez que ele é capaz de utilizar essas informações adicionais de modo a alcançar recomendações com maior acurácia do que o SLIM (NING; KARYPIS, 2012).

A construção do SSLIM é baseado na utilização de dois tipos de informações: as relações usuário-item e as descrições dos itens. A relação usuário-item é representada por uma matriz binária A de tamanho $m \times n$, onde m corresponde ao número de usuários e n ao número de itens, na qual para cada usuário i é associado um item j, sendo que a entrada será 1 se o usuário avaliou aquele item, ou 0, caso contrário. Já as informações dos itens são representadas por uma matriz de fatores adicionais B de tamanho nxn, em que cada item i é associado a outro item j e a entrada corresponde ao valor da similaridade entre eles.

O SSLIM assume que existem correlações entre os usuários que “gostam” de dois itens, bem como a semelhança entre eles. Análogo ao SLIM, uma vez que utilizam a mesma base teórica, a abordagem utilizada pelo método consiste em construir uma matriz de fatores latentes W esparsa de tamanho nxn para, dessa vez, reproduzir A e B, de forma a obedecer a Equação 10:

$$\tilde{A} \sim AW \quad e \quad \tilde{B} \sim BW, \tag{10}$$

onde $\tilde{B}$ é a matriz das pontuações estimadas para B e B é a matriz de *Side Information* (informações sobre os itens).

A matriz W também é aprendida pela resolução de um problema linear de otimização como mostra a Equação 11, análogo ao SLIM:

$$\min_{W} \quad \|A - AW\|_F^2 + \tfrac{\alpha}{2}\|B - BW\|_F^2 + \tfrac{\beta}{2}\|W\|_F^2 + \lambda\|W\|_1, \quad \text{sujeito a} \quad \begin{cases} W \geq 0 \\ diag(W) = 0 \end{cases} \tag{11}$$

onde $\|B - BW\|_F^2$ mede o quão bem a matriz dos coeficientes de agregação W se ajusta às informações adicionais. O parâmetro α controla a importância de usar essas informações adicionais.

Uma vez que as colunas de W são independentes, o problema de otimização referido na Equação 8 (para o SLIM) e na Equação 11 (para o SSLIM), pode ser decomposto em um conjunto de problemas de otimização, como mostra a Equação 12 (para o SLIM) e Equação 13 (para o SSLIM), resolvendo-se todas as colunas da matriz W de forma paralela (sendo essa uma importante característica do método), aumentando assim a sua eficiência.

$$\min_{W} \quad \|a_j - Aw_j\|_F^2 + \tfrac{\beta}{2}\|w_j\|_F^2 + \lambda\|w_j\|_1. \quad \text{sujeito a} \quad \begin{cases} W \geq 0 \\ diag(W) = 0 \end{cases} \tag{12}$$

$$\min_{W} \quad \|a_j - Aw_j\|_F^2 + \tfrac{\alpha}{2}\|b_j - Bw_j\|_F^2 + \tfrac{\beta}{2}\|w_j\|_F^2 + \lambda\|w_j\|_1. \quad \text{sujeito a} \quad \begin{cases} W \geq 0 \\ diag(W) = 0 \end{cases} \tag{13}$$

Nesse capítulo apresentou-se os principais conceitos para entendimento da metodologia proposta nesta dissertação, a qual, será detalhada no próximo capítulo.

4 Arcabouço de Recomendação de Imagens LDSSLIM

Neste capítulo é descrito arcabouço proposto nesta dissertação para recomendação de imagens. Na Seção 4.1, apresenta-se a formulação do problema de recomendação de imagens. Na Seção 4.2, revisa-se o algoritmo SSLIM que é a referência para o arcabouço proposto neste trabalho. Na Seção 4.3, detalha-se como o algoritmo SSLIM pode ser utilizado para considerar representações em baixas dimensões. Finalmente, na Seção 4.4 é descrito o arcabouço proposto nesta dissertação: LDSSLIM (*Low-Dimensional Sparse Linear Methods with Side Information*).

4.1 Formulação do Problema

Nesta seção, o problema de recomendação de imagens abordado nesta dissertação é formalizado e também é apresentada a notação utilizada para detalhar o método. A formulação do problema está alicerçada no método do estado-da-arte SSLIM ((NING; KARYPIS, 2012)).

Seja U o conjunto de usuários e M o conjunto de imagens (i. e., itens). Seja $A \subseteq U \times M$ a matriz de avaliação onde $a_{ij} = 1$ se a imagem m_j é relevante para o usuário u_i, e 0 caso contrário. O tamanho dos conjuntos de usuários e itens são dados por $|U|$ e $|M|$, respectivamente.

A *i-ésima* linha da matriz A, denotada por a_i^T, representa a preferência do perfil do usuário u_i para todas as imagens, enquanto a *j-ésima* coluna de A representa o perfil da imagem m_j em relação à todos os usuários. A informação descritiva de todas as imagens compõe uma matriz B e é dada por $n \times |M|$, onde n é o número de características em uma determinada modalidade (por exemplo, o número de palavras usadas na descrição textual). A *j-ésima* coluna de B representa o vetor de característica de informação descritiva da imagem m_j, denotado por b_j.

O problema central consiste em identificar, com base nas preferências de u_i (inferidas a partir de A) e características descritivas associadas às imagens, as *top-N* imagens que melhor se ajustam aos interesses do usuário-alvo da recomendação. Esse problema é ser formalizado como:

[**Recomendação *Top-N***] Dadas as matrizes A e B, e um usuário alvo $u_i \in U$, sugerir uma lista ordenada de novos itens em M que corresponda às preferências de u_i.

4.2 SSLIM

O SSLIM (*Sparse Linear Methods with Side Information*) (Seção 3.3) é o algoritmo de recomendação no qual o arcabouço proposto nesta dissertação é motivado. Nessa seção detalha-se o algoritmo SSLIM.

O SSLIM possui como entrada duas matrizes: A e B e, como retorno, uma lista ordenada de tamanho N, contendo as *top-N* imagens recomendadas.

A matriz **A** é a matriz que descreve a relação usuário-item. É uma matriz binária, de tamanho mxn, onde m corresponde ao número de usuários e n ao número de imagens. As posições da matriz são preenchidas de acordo com a preferência dos usuários em relação às imagens. No problema proposto, essa relação foi estabelecida através das preferências do usuário, extraídas de informações da coleção de dados, onde foram considerados os *"likes"* (curtidas) de um usuário em uma imagem, portanto, se um usuário i curtiu uma imagem j, a posição a_{ij} da matriz foi setada em 1, caso contrário, em 0.

A matriz **B** consiste em uma matriz de informações adicionais e estabelece relação de similaridade entre as imagens. É uma matriz de tamanho nxn, onde n corresponde ao número de imagens. As posições da matriz são preenchidas de acordo com a similaridade entre as imagens. No problema proposto, a similaridade foi calculada baseando-se em três funções de similaridade, como explicado na Subseção 5.5.3, onde dada uma imagem i e outra imagem j, a posição b_{ij} da matriz contem o valor de similaridade entre os vetores que representam as imagens i e j.

O SSLIM possui três parâmetros de regularização, i.e., parâmetros que controlam a complexidade do modelo de recomendação: α, β e λ. Especificamente, os parâmetros β e λ foram utilizados em seus valores padrão (BIDART, 2015), por apresentarem bom desempenho, e os valores variados para teste não apresentarem impactos nos resultados dos experimentos. Já o parâmetro α, que acompanha a matriz B, foi selecionado entre [0.001, 0.01, 0.1, 0.5, 1.0, 2.0], utilizando-se α = 0.1, por apresentar o melhor desempenho referente aos testes aplicados em nossas coleções.

4.3 Do SSLIM ao LDSSLIM

Essa seção descreve as alterações realizadas no algoritmo do SSLIM de forma a gerar o método LDSSLIM proposto nessa dissertação.

O LDSSLIM assume que a matriz de coeficientes de agregação W do SSLIM pode ser representada como uma função de um espaço reduzido das características do item. Ou seja, a similaridade entre os itens é representada diretamente na matriz de informações adicionais, a partir de vetores descritivos reduzidos.

Isso foi feito criando-se a matriz B de semelhança entre os itens, explorando a representação de baixa dimensionalidade das imagens. É importante notar que usar uma representação de baixa dimensionalidade dos itens é altamente viável, uma vez que as descrições brutas do itens são muito esparsas (por exemplo, a representação por meio de *bag-of-words*, para modalidade textual).

A principal diferença entre o LDSSLIM e o SSLIM é sobre como a matriz W é construída em função dos dados de entrada. O algoritmo do LDSSLIM recebe como entrada as mesmas matrizes que o SSLIM, isto é, a matriz de classificação A e a matriz de descrição dos itens B. No entanto, ao invés de utilizar-se uma única matriz B de descrição, usamos um conjunto de matrizes B de descrição dos itens. Cada matriz $B \in \{\mathbf{B}\}$, fornece uma combinação das distintas representações modais dos itens. Por exemplo, se os itens são descritos usando informações textuais e visuais, temos $\{\mathbf{B}\} = \{B_{\text{textual}} + B_{\text{visual}}\}$, onde **B** é a concatenação das características obtidas pela representação textual (B_{textual}) e as características obtidas pela representação visual (B_{visual}). Sendo assim, o LDSSLIM explora as informações da descrição multimodal das imagens representando-as em um espaço latente comum por meio do uso de métodos de redução de dimensionalidade.

Além disso, o LDSSLIM é flexível no sentido de que se pode configurar como entrada os seguintes parâmetros: (i) o método de redução de dimensionalidade, (ii) a função de similaridade entre itens e (iii) a descrição multimodal utilizada para representar as imagens.

Os métodos SLIM e SSLIM originais não podem usar as melhorias propostas neste trabalho, a saber: (i) representação de itens de forma multimodais em um espaço latente de baixa dimensão, e (ii) representação explícita de semelhanças entre os itens. Especificamente, o método SLIM não utiliza informação descritiva dos itens, e o método SSLIM assume apenas uma matriz de semelhanças entre os itens.

4.4 LDSSLIM

Essa seção detalha o arcabouço LDSSLIM (*Low-Dimensional SSLIM*) proposto nesta dissertação. O arcabouço funciona utilizando diferentes representações multimodais de itens em alta dimensionalidade para um espaço latente de baixa dimensionalidade. Este recebe como entrada $|M|$ itens - no caso, imagens - representados por um conjunto de características x, que podem representar uma ou mais modalidades e são as informações descritivas com as quais o método trabalha. Assume-se que esses itens são representados por um pequeno número de variáveis descritivas ($|k| < |x|$), onde j refere-se ao conjunto de fatores latentes (informação adicional) do espaço de dimensão reduzida. Denomina-se essa representação como fatores latentes pelo fato de não poder ser observados diretamente no conjunto de dados dos usuários e dos itens (CUNNINGHAM; BYRON, 2014).

O LDSSLIM utiliza métodos de redução de dimensionalidade para descobrir e

extrair esses k fatores latentes dos dados de alta dimensionalidade de acordo com um métodos conhecidos (por ex., PCA). Ao mapear esses itens representados em um espaço latente de baixa dimensionalidade também é reduzida automaticamente a esparsidade da representação. Esta nova representação dos itens em um espaço de baixa dimensionalidade é então utilizada para calcular a semelhança entre pares de itens e essas semelhanças são passadas para um modelo linear a fim de sugerir itens relevantes a usuários com base em suas preferências expressas na matriz A. Uma vez que é utilizada uma representação de baixa dimensionalidade dos itens para regularizar o método SSLIM original, a abordagem proposta é denotada por LDSSLIM.

O algoritmo 1 detalha o arcabouço LDSSLIM proposto nesta dissertação.

Algoritmo 1: LDSSLIM

Input: Matrizes A e B, método de redução de dimensionalidade, função de similaridade.

for $i = 1, \ldots, |M|$ **do**
  $l_i \leftarrow$ low_dim$(\{b_i\})$ ▷ Representação do item em baixa dimensão
end
for $i = 1, \ldots, |M|$ **do**
  for $j = 1, \ldots, |M|$ **do**
    $L_{ij} \leftarrow$ sim(l_i, l_j) ▷ Cálculo da similaridade entre os itens
  end
end
for $i = 1, \ldots, |M|$ **do**
  ▷ Cálculo do coeficiente de agregação da matriz
  $W \leftarrow \underset{W}{\text{argmin}} \|\|a_i - Aw_i\|\|_2^2 + \frac{\alpha}{2}\|\|l_i - Lw_i\|\|_2^2 + \frac{\beta}{2}\|\|w_i\|\|_2 + \lambda\|\|w_i\|\|_1$
  subject to $W \geq 0$ and diag$(W) = 0$
end

O algoritmo funciona em três etapas sequenciais. No primeiro passo constroi-se a representação de baixa dimensão das imagens (primeiro laço). A única restrição nesse passo é que o comprimento da representação de baixa dimensionalidade precisa ser inferior à representação original da imagem i, isto é, $|l_i||b_i|$. O segundo passo identifica quão semelhantes são os pares de itens em termos da função de similaridade dada (segundo laço). Finalmente, o último passo calcula a matriz de coeficientes de agregação W. Aqui, segue-se o esquema de otimização do SSLIM (terceiro laço).

A Figura 7 ilustra os passos da recomendação de acordo com o arcabouço proposto nesta dissertação. Cabe ressaltar que embora o foco do trabalho tenha sido no uso de informação textual e visual, outras formas de descrição das imagens poderiam ser utilizadas. Por exemplo, características de redes neurais profundas (*Deep Neural Networks* poderiam ser utilizadas para alimentar o arcabouço.

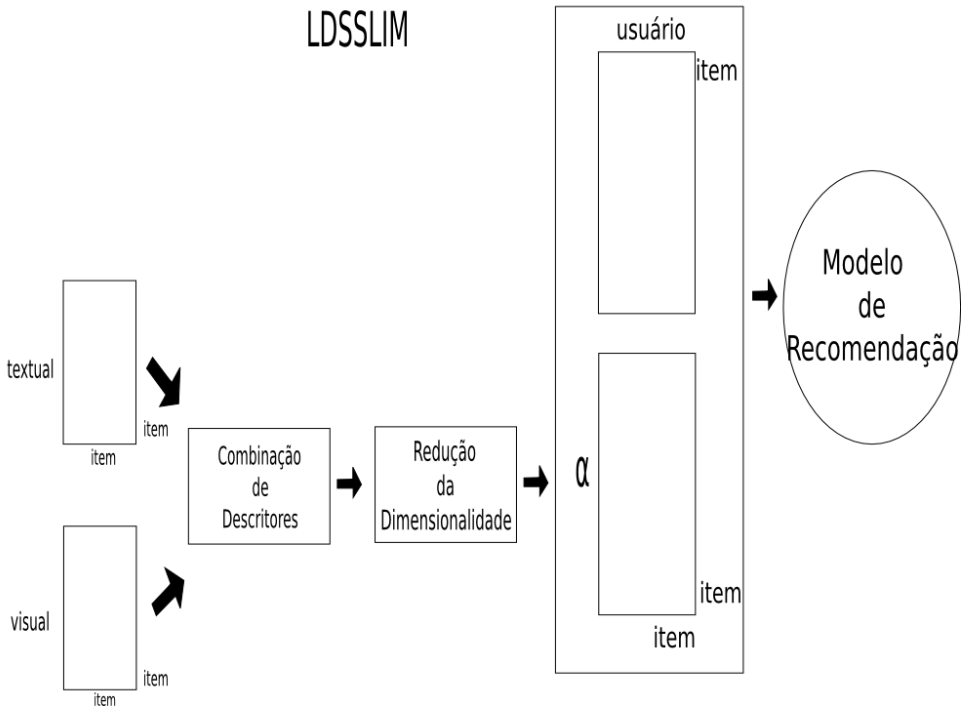

Figura 7 – Modelo de Recomendação. O modelo de recomendação é alimentado por duas matrizes: uma que relaciona os usuários e seus itens e outra que relaciona os itens, sendo essa uma combinação de descritores textuais e visuais em formato reduzido, influenciada por um parâmetro α que atribui peso à sua utilização.

5 Avaliação Experimental

Nesta seção, detalha-se os resultados da avaliação experimental do arcabouço LDSSLIM para recomendação de imagens multimodais proposto nesta dissertação. A seguir, lista-se as perguntas de pesquisa que guiam os experimentos e a discussão dos resultados:

Q1: O arcabouço proposto melhora a qualidade das imagens recomendadas em relação às estratégias estado-da-arte?

Q2: Qual o efeito dos métodos de redução de dimensionalidade na qualidade da recomendação?

Q3: Qual o efeito da variação do número de fatores latentes para a representação das imagens na qualidade da recomendação?

Q4: Qual o efeito da variação das métricas de similaridade na qualidade da recomendação?

Q5: Qual o efeito da variação no número de imagens sugeridas na qualidade da recomendação?

O restante deste capítulo está organizado da seguinte forma. Na Seção 5.1, detalha-se a base de dados utilizada nos experimentos. Na Seção 5.2, apresenta-se as métricas de qualidade da recomendação utilizadas para avaliar tanto o arcabouço proposto quanto os métodos estado-da-arte testados. Na Seção 5.3, detalha-se o protocolo de avaliação utilizado nos experimentos. Na Seção 5.5, detalha-se a escolha dos diferentes parâmetros utilizados no LDSSLIM. Na Seção 5.5.3, detalha-se as métricas de similaridade utilizadas para comparação das imagens. Por fim, na Seção 5.6, detalha-se os resultados obtidos em função das perguntas de pesquisa mencionadas acima.

5.1 Base de dados

Neste trabalho é utilizada uma coleção de dados real composta por imagens coletadas do sítio Web *Flickr*[1], denominado MIRFLICKR-1M (HUISKES; LEW, 2010). A base de dados MIRFLICKR-1M é composta por 1.000.000 de imagens da rede social de compartilhamento de imagens *Flickr* e 24.431 usuários.

Além das imagens e usuários, o MIRFLICKR-1M contém (i) *tags* e (ii) informações sobre as propriedades e configurações da câmara no momento em que as fotos foram tiradas (como: marca e fabricante da câmera, exposição, distância focal, resolução, data, hora,

[1] https://www.flickr.com/

localização, dentre outras) que não foram utilizadas nesse trabalho. O título e a descrição das imagens utilizados para extrair as características de natureza textual foram obtidas via API[2] (Application Programming Interface). Como avaliações (*ratings*) dos usuários em relação aos itens, foram consideradas as curtidas (*likes*) dadas pelo usuário à foto.

Sobre a base de dados original, foram realizados três cortes simultâneos: o primeiro, considerando-se apenas as imagens acompanhadas de no mínimo 2 palavras, de modo que fosse possível explorar as características textuais. No segundo, considerando-se a quantidade de *likes* (curtidas) realizados pelos usuários, por se tratar de um *feedback* explícito, como explicado no Capítulo 3, utilizado para estabelecer a relação entre os usuários e os itens. No segundo corte, foram considerados apenas os usuários que demonstraram preferência por, no mínimo, 5 imagens. Ainda foi realizado um terceiro corte, considerando apenas as 1.000 imagens que possuíam o maior número de palavras associadas a elas. A coleção de dados após o terceiro corte foi chamada MIRFLICKR-1k e sobre essa foram realizados os experimentos. A Tabela 3 contém os dados sem e após a realização dos cortes.

Tabela 3 – Dados: MIRFLICKR.

	MIRFLICKR-1M	MIRFLICKR-1k
#usuários	24.431	1.275
#imagens	1.000.000	1.000

5.2 Métricas de Avaliação

As métricas de avaliação utilizadas nessa dissertação são:

(a) **HR** (*Hit-Rate* – Taxa de Acerto): consiste na relação entre o número de usuários cujos itens recomendados mostraram-se relevantes e o número total de usuários, onde a validação é realizada sobre a base de teste. A Equação 14 descreve essa relação.

$$HR@N = \frac{\#acertos}{\#u}, \tag{14}$$

onde $\#acertos$ é o número de usuários para os quais o recomendador sugere itens relevantes na lista de recomendação de tamanho N e $\#u$ é o número total de usuários da base de dados.

[2] https://www.flickr.com/services/api/

(b) **ARHR:** (*Average-Reciprocal Hit Rate* – Média Recíproca da Taxa de Acerto) é uma versão ponderada do HR e mede o quão forte é a recomendação de um item, cujos pesos da ponderação são as respectivas posições dos acertos na lista de recomendação.

$$ARHR@N = \frac{1}{\#u} \sum_{i=1}^{\#acertos} \frac{1}{rank_i}, \tag{15}$$

onde $\#acertos$ é o número de usuários para os quais o recomendador sugere itens relevantes na lista de recomendação de tamanho N, $\#u$ é o número total de usuários da base de teste e $rank_i$ é a posição do item na lista de recomendação *top-N* considerando-se o *i-ésimo* acerto.

5.3 Protocolo de avaliação

No arcabouço foi aplicada uma metodologia de treinamento e teste denominada *k-fold cross-validation* para validar os resultados experimentais e garantir relevância estatística das comparações. Assim, a coleção de dados foi dividida, de forma aleatória, em cinco subconjuntos, cada um contendo 20% das avaliações. Quatro subconjuntos foram utilizados para treino dos métodos de recomendação e um subconjunto utilizado como teste para avaliação dos métodos. Os subconjuntos são comutados e o processo é repetido cinco vezes, cada um utilizando diferentes conjuntos para treino e teste. Ao final, a média dos resultados dos subconjuntos é considerada. Esta segmentação de treino e teste foi realizada apenas na matriz A que estabelece a relação dos usuários e imagens. Assim, as matrizes complementares de informação textual e visual (matriz B) não sofreram essa divisão, uma vez que o seu conteúdo é associado unicamente aos itens (i.e., imagens), não contendo informação de usuário.

5.4 Métodos de referência

Por meio do uso das mesmas métricas de avaliação, o LDSSLIM foi comparado a dois métodos do estado da arte para recomendações *top-N* que serviram de inspiração para esta dissertação: o SLIM e o SSLIM, onde para o SLLIM tem-se a mesma metodologia proposta, porém utilizando informações unimodais, isto é, exclusivamente textuais ou exclusivamente visuais, e multimodais sem a utilização dos métodos redutores de dimensionalidade.

Além dos métodos lineares esparsos acima apresentados (i.e., SLIM e SSLIM), o arcabouço proposto também foi comparado aos seguintes métodos: (i) *PureSVD*, que fatora a matriz de classificação utilizando uma técnica de fatorização semelhante ao método de Decomposição do Valor Singular e (ii) *WeightedItemKNN* (nomeado, nesta dissertação, por WI-KNN), que é um algoritmo de recomendação representativo dos métodos de vizinhança.

Os parâmetros do *PureSVD* referem-se ao número de fatores latentes, e os parâmetros do WI-KNN se referem ao número de vizinhos usado para prever a recomendação.

5.5 Parametrização dos componentes do LDSSLIM

Conforme discutido anteriormente, o arcabouço LDSSLIM pode ser instanciado com diferentes parâmetros. Assim, na Seção 5.5.1, apresenta-se o processo de representação das imagens, bem como as características textuais e visuais utilizadas. Na Seção 5.5.2, detalha-se as características dos métodos de redução de dimensionalidade utilizados nesta dissertação. Na Seção 5.5.3, discute-se as métricas de similaridade utilizadas nos experimentos.

5.5.1 Extração de Características

A etapa de extração de características foi realizada em duas modalidades: extração de características visuais e textuais. Cada imagem foi representada por vetores de ambas características, bem como suas combinações.

As características visuais consistem em informações obtidas da própria imagem, utilizando-se para isso técnicas de Visão Computacional. As características extraídas foram (i) histograma de cor, (ii) descritor HOG (forma), (iii) descritor SIFT e (iv) descritor SURF (pontos de interesse), onde as dimensões do vetor de representação de cada imagem equivalem a [765, 3.780, 1.000, 1.000], respectivamente. É válido ressaltar que para os descritores de pontos de interesse SIFT e SURF utilizou-se uma *Bag of Visual Words* a fim de determinar uma quantidade fixa em sua descrição, uma vez que o número de pontos-chaves (pontos de interesse) variam de imagem para imagem. Todas as características visuais estão detalhas na Subseção 3.2.2.

Já as características textuais consistem em informações obtidas dos textos que acompanham as imagens, sendo eles o título e a descrição, utilizando-se para isso vetores de características do tipo TF e TF-IDF, comumente utilizadas em Recuperação de Informação. As dimensões da representação de cada imagem para as características textuais foi de 13.564, que corresponde ao tamanho do vocabulário completo. As características textuais foram descritas na Subseção 3.2.1.

A Figura 8 ilustra o processo de representação das imagens por meio das etapas de extração das características e de redução de dimensionalidade, conforme descritas nas subseções seguintes.

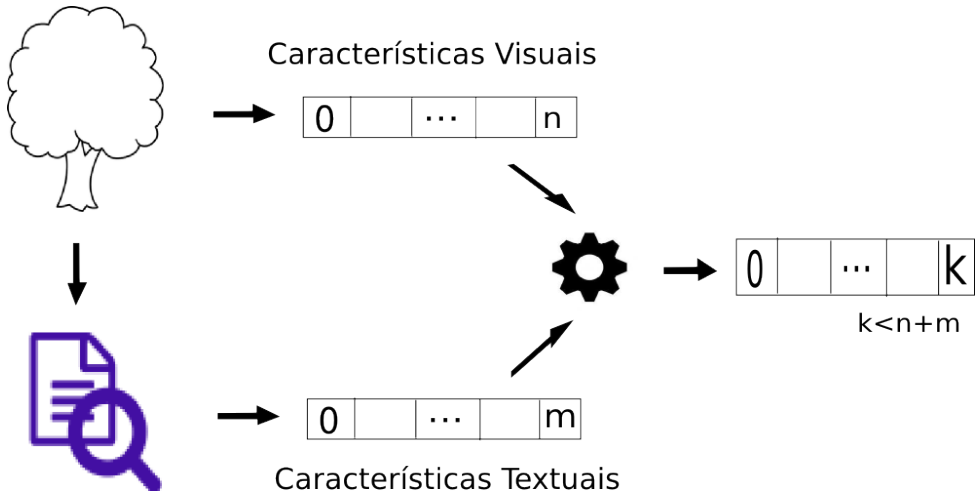

Figura 8 – Representação das Imagens. Após extraídas as características visuais, as imagens são representadas por um vetor de tamanho n, onde n pode ser o número de pontos chave, por exemplo, da mesma forma, após extraídas as características textuais, as imagens são representadas por um vetor de tamanho m, onde m pode ser o número de palavras do vocabulário, por exemplo; em seguida, é feita a redução da dimensionalidade, onde as imagens são representadas por um novo vetor de tamanho reduzido k.

5.5.2 Redução de Dimensionalidade

Após a representação das imagens pelos vetores de características, foi realizado um segundo passo: a redução de sua dimensionalidade. Tal etapa é importante pelo fato de diminuir os custos computacionais da execução, bem como o tempo de execução do algoritmo de recomendação.

Para a redução da dimensionalidade foram testadas três técnicas: PCA, LSA e ISOMAP, conforme descrito na Subseção 5.5.2.

Todas as técnicas recebem como entrada a combinação dos vetores de características visuais e textuais. Por exemplo, a junção do TF (textual) com o descritor SIFT (visual). Essa combinação possui um tamanho $n + m$ (onde n é o tamanho do vetor visual e m, o tamanho do vetor textual). Para o tamanho k (onde $k < n + m$) do novo vetor reduzido, foram testados os valores [50, 100, 200, 300, 400]. Foram utilizadas as implementações da biblioteca *Sickit-learn* em todos os métodos.

5.5.3 Similaridade entre as Imagens

As funções de similaridade são responsáveis por medir a semelhança entre as imagens. Os valores calculados são usados para compor a matriz de informações descritivas, que constitui uma das entradas do algoritmo de recomendação.

Nesse trabalho, foram utilizadas três métricas de similaridades diferentes: a função cosseno, a distância euclideana e a distância de *manhattan*.

(a) **Função Cosseno:** A similaridade do cosseno entre dois vetores é medida por meio do cálculo do cosseno do ângulo entre eles e é representada pela Equação 16:

$$\cos\theta = \frac{\vec{a}\vec{b}}{\|\vec{a}\| \left\|\vec{b}\right\|}. \tag{16}$$

O valor do cosseno calculado, que varia de 0 (zero) a 1 (um), indica a similaridade entre os itens. Quanto mais próximo a 1, mais similares são os itens, uma vez que o ângulo θ formado entre dois vetores iguais é igual a 0 (zero) e o $cos(0) = 1$, consequentemente, quanto mais próximo a 0, menos similares eles são.

(b) **Distância Euclidiana:** Para dois vetores num hiperplano de dimensão n, a distância Euclidiana é dada pela Equação 17:

$$E(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}. \tag{17}$$

O valor calculado é um número real no intervalo [0, ∞[e representa a similaridade entre os itens. Quanto mais próximo de 0 menor a distância entre os vetores e mais similares são os itens, consequentemente, quanto mais distante de 0 menos similares eles são.

(c) **Distância de Manhattan:** Para dois vetores, a distância de Manhattan é a distância média entre as dimensões e é dada pela Equação 18:

$$M(x, y) = \sum_{i=1}^{n} |x_i - y_i| . \tag{18}$$

Análogo à distância Euclidiana, o valor calculado é um número real no intervalo [0, ∞[e representa a similaridade entre os itens. Quanto mais próximo de 0 menor a distância entre os vetores e mais similares são os itens, consequentemente, quanto mais distante de 0 menos similares eles são.

Os valores retornados pelas funções de similaridade foram normalizados e, para a distância euclidiana e a distância de *manhattan*, foi utilizado o valor da distância subtraído de 1. Assim, por exemplo, duas imagens iguais possuem distância mínima (0) e similaridade máxima (1).

5.6 Experimentos e discussão

5.6.1 Comparação do arcabouço proposto e métodos estado-da-arte (Q1)

Os resultados aqui apresentados referem-se ao uso combinado de características multimodais (textuais - TF e TF-IDF - e visuais - SIFT, SURF, Histograma de Cor e HoG) (Subseção 5.5.1), de forma concatenada, sobre um corte de 1000 imagens da coleção de dados do *Flickr* (Seção 5.1). Sobre cada combinação multimodal, foram aplicados três métodos de redução de dimensionalidade: PCA, LSA e ISOMAP (Subseção 5.5.2) e os resultados então foram gerados utilizando-se as três métricas de similaridade: cosseno, distância euclidiana e distância de manhattan (Subseção 5.5.3).

Os resultados referem-se ao valor de α=0.1, por apresentar melhor acurácia, e o tamanho do ranking utilizado foi N=20. Por fim, a acurácia das recomendações foi medida pelas métricas de avaliação HR e ARHR (Seção 5.2).

A Tabela 4 apresenta a performance dos métodos de referência e do LDSSLIM em termos das métricas de avaliação HR e ARHR. Para os métodos de referência, os resultados relatados abrangem apenas a melhor combinação de representação em termos do tipo de característica da imagem considerando cada modalidade e medida de similaridade. É possível notar que o desempenho do *PureSVD* e do WI-KNN é inferior ao dos métodos lineares esparsos e ao resultado do arcabouço proposto. Também é apresentado o resultado do SSLIM ao concatenar informações textuais e visuais em uma única matriz, porém sem o uso das meodos redutores de dimensionalidade (SSLIM-concat).

Cabe ressaltar, que foi considerada apenas a primeira posição da lista recomendada (ou seja, N = 1) pelo fato de a maioria dos usuários da base de dados ter avaliado um número pequeno de imagens. Para todas as métricas, o arcabouço proposto supera os métodos de referência com ganhos entre 15% e 17% em termos das métricas HR e ARHR, respectivamente. É possível observar que os resultados do SLIM são realmente inferiores, pois o SLIM não considera informações descritivas das imagens, e a matriz de classificação da coleção é extremamente esparsa. Os melhores resultados, dentre os métodos de referência, tanto em HR@1 quanto em ARHR@1 foram obtidos ao usar ambos os descritores (i.e., textual e visual). A partir desses resultados, pode-se concluir que uma representação de informação multimodal de dimensões baixas melhora a qualidade da recomendação.

A Tabela 5 compara o arcabouço proposto LDSSLIM e o melhor método de referência (SSLIM-concat) variando o tamanho da lista retorna, uma vez que a Tabela 4 considera apenas a primeira posição da lista (N = 1). Novamente, são comparados os melhores resultados de cada método. É possível concluir que mesmo aumentando o número de itens retornados o arcabouço proposto continua com desempenho superior ao melhor método de referência.

	Método	Parâmetros	HR@1 (ganho)	Parâmetros	ARHR@1 (ganho)
Métodos de Referência	WI-KNN	100	0.0050	150	0.0062
	PureSVD	50	0.0065	50	0.0067
	SLIM	-	0.0090	-	0.0080
	SSLIM-textual	TFIDF:Euc	0.0314	TFIDF:Euc	0.0128
	SSLIM-visual	Hog:Euc	0.0011	Surf:Cos	0.0020
	SSLIM-concat.	TF+Surf:Euc	*0.0320*	TF+Surf:Euc	*0.0131*
	LDSSLIM	PCA:400 TF+HoG:Euc	0.0369 (0.1531)	LSA:300 TFIDF+Sift:Euc	0.0154 (0.1755)

Tabela 4 – Desempenho dos métodos de recomendação de imagem considerando $N = 1$. Os ganhos do arcabouço proposto são relativos ao melhor método dentre os *baselines* e são apresentados entre parênteses. O melhor desempenho dentre os *baselines* é representado em itálico.

	Método	Parâmetros	HR@1 (ganho)	Parâmetros	ARHR@1 (ganho)
N=5	SSLIM-concat.	TFIDF+Hog:Euc	0.0646	TF+Surf:Euc	0.0379
	LDSSLIM	LSA:50 TFIDF+HoG:Euc	0.0671 (0.0386)	LSA:300 TFIDF+Sift:Euc	0.0406 (0.0712)
N=10	SSLIM-concat.	TFIDF+Sift:Euc	0.0721	TF+Surf:Euc	0.0447
	LDSSLIM	PCA:50 TFIDF+Sift:Euc	0.0735 (0.0194)	LSA:300 TFIDF+Sift:Euc	0.0467 (0.0447)

Tabela 5 – Desempenho do arcabouço proposto - LDSSIM - e do melhor método de referência - SSLIM-concat. - considerando $N = 5$ e $N = 10$, ou seja, variando o tamanho da lista recomendada. Os ganhos do arcabouço proposto são apresentados entre parênteses.

5.6.2 Efeito dos métodos de redução de dimensionalidade (Q2)

Nesta Seção analisa-se método de redução da dimensionalidade e seu impacto na recomendação. A Tabela 6 mostra o impacto da variação dos métodos de representação das imagens em termos de HR@1 e ARHR@1, da distância euclideana enquanto função de similaridade e considerando o número de fatores latentes [300, 400]. Tais valores foram considerados por apresentarem melhor desempenhado, como mostrado acima. É possível notar que as combinações aliadas ao histograma de cor possuem desempenho bastante inferior aos demais, o que comprova que as características extraídas das imagens impactam o desempenho do recomendador e o histograma de cor consiste em uma descrição de baixa exatidão.

Como pode ser visto na Figura 9, o método de redução LSA atinge os melhores resultados sobre o melhor método de referência. Como o interesse é sugerir uma lista ranqueada de N imagens relevantes para os usuários, a Figura 9 também varia a posição da lista N. Como esperado, o desempenho de todos os recomendadores tende a aumentar à medida que aumenta-se o número de imagens recomendadas. Isso acontece porque é

		HR@1		ARHR@1	
		300	400	300	400
PCA	TF + Cor	0.0078	0.0081	0.0008	0.0009
	TF + HoG	0.0346	0.0369	0.0132	0.0146
	TF + SIFT	0.0243	0.0295	0.0084	0.0098
	TF + SURF	0.0284	0.0348	0.0103	0.0139
	TF-IDF + Cor	0.0065	0.0109	0.0006	0.0012
	TF-IDF + HoG	0.0269	0.0273	0.0098	0.0100
	TF-IDF + SIFT	0.0342	0.0256	0.0145	0.095
	TF-IDF + SURF	0.0296	0.0278	0.0107	0.0103
LSA	TF + Cor	0.0102	0.0120	0.0011	0.0014
	TF + HoG	0.0290	0.0283	0.0111	0.0100
	TF + SIFT	0.0298	0.0286	0.0104	0.0100
	TF + SURF	0.0328	0.0246	0.0120	0.0079
	TF-IDF + Cor	0.0089	0.0082	0.0010	0.0009
	TF-IDF + HoG	0.0301	0.0271	0.0111	0.0105
	TF-IDF + SIFT	0.0329	0.0246	0.0154	0.0095
	TF-IDF + SURF	0.0283	0.0237	0.0106	0.0088
ISOMAP	TF + Cor	0.0075	0.0072	0.0006	0.0005
	TF + HoG	0.0272	0.0250	0.0085	0.0083
	TF + SIFT	0.0263	0.0261	0.0087	0.0086
	TF + SURF	0.0280	0.0266	0.0094	0.0090
	TF-IDF + Cor	0.0064	0.0067	0.0005	0.0006
	TF-IDF + HoG	0.0262	0.0265	0.0087	0.0096
	TF-IDF + SIFT	0.0287	0.0287	0.0106	0.0088
	TF-IDF + SURF	0.0285	0.0345	0.0028	0.0030

Tabela 6 – Desempenho do arcabouço LDSSLIM variando-se os métodos de representação das imagens, considerando HR@1 e ARHR@1; a distância euclideana enquanto medida de de similaridade; [300, 400] enquanto o número de fatores latentes.

mais provável que o algoritmo recomende imagens relevantes a medida que o número total de imagens sugeridas aumenta.

5.6.3 Efeito da variação do número de fatores latentes (Q3)

A seguir, analisa-se o impacto da redução da dimensionalidade em relação à qualidade da recomendação. A Figura 10 mostra como o número de fatores latentes afeta o desempenho da recomendação. Os resultados mostram que o arcabouço proposto atinge os melhores resultados ao considerar mais de 300 fatores latentes para representar imagens. Assim, pode-se concluir que o número de fatores latentes deve ser grande o suficiente para discriminar as características das imagens.

5.6.4 Efeito da variação das métricas de similaridade (Q4)

O segundo parâmetro do LDSSLIM é a função de similaridade, que é usada para calcular como dois itens (isto é, duas imagens) são relacionados. Na Figura 11, apresenta-se o desempenho dos recomendadores de imagem em termos de ARHR@1 e dos diferentes conjuntos de funções de similaridade. Pode-se observar que a função de similaridade afeta o desempenho de cada método, incluindo o arcabouço proposto e o melhor método de referência. No geral, os melhores resultados foram obtidos com a distância euclidiana.

5.6.5 Efeito da variação do número de imagens recomendadas (Q5)

A Tabela 7 detalha o impacto do tamanho da lista retornada, ou seja, a quantidade de imagens recomendadas para os usuários; e o número de fatores latentes, como mostrado na Figura 9 e Figura 10. Os resultados são referentes à métrica de avaliação HR; e consideram a distância euclideana enquanto função de similaridade e o TF+HoG enquanto método de representação das imagens, por apresentar o melhor desempenho como mostra a Tabela 4.

		1	3	5	10	20
PCA	50	0.0286	0.0547	0.0651	0.0720	0.0738
	100	0.0294	0.0565	0.0634	0.0720	0.0738
	200	0.0291	0.0551	0.0651	0.0714	0.0732
	300	0.0346	0.0573	0.0638	0.0711	0.0730
	400	0.0369	0.0574	0.0651	0.0723	0.0743
LSA	50	0.0261	0.0563	0.0654	0.0722	0.0740
	100	0.0322	0.0577	0.0638	0.0722	0.0739
	200	0.0274	0.0568	0.0643	0.0724	0.0742
	300	0.0290	0.0571	0.0663	0.0731	0.0744
	400	0.0283	0.0537	0.0646	0.0722	0.0740
ISOMAP	50	0.0243	0.0482	0.0545	0.0678	0.0737
	100	0.0292	0.0472	0.0553	0.0680	0.0732
	200	0.0273	0.0451	0.0580	0.0707	0.0740
	300	0.0272	0.0483	0.0561	0.0713	0.0737
	400	0.0250	0.0485	0.0541	0.0701	0.0738

Tabela 7 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação HR; a distância euclideana enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

A Tabela 8 possui o mesmo foco da Tabela 7, porém em termos da métrica de avaliação ARHR e do TFIDF+Sift enquanto método de representação das imagens.

Nota-se, a partir de ambas as avaliações, que o desempenho do recomendador tende a melhoras à medida que aumenta-se o tamanho da lista, uma vez que é mais provável que o algoritmo recomende imagens relevantes a medida que o número total de imagens sugeridas aumenta. Outra conclusáo é que o arcabouço proposto tende a atingir

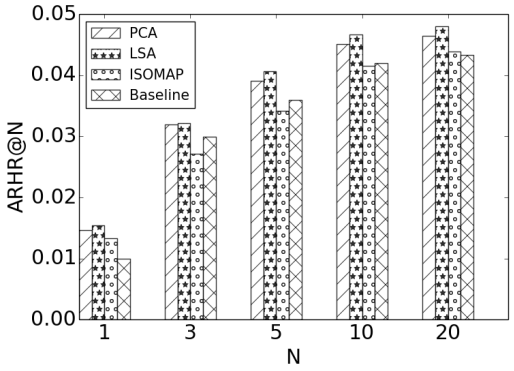

Figura 9 – Variação do tamanho da lista de recomendação *top*-N.

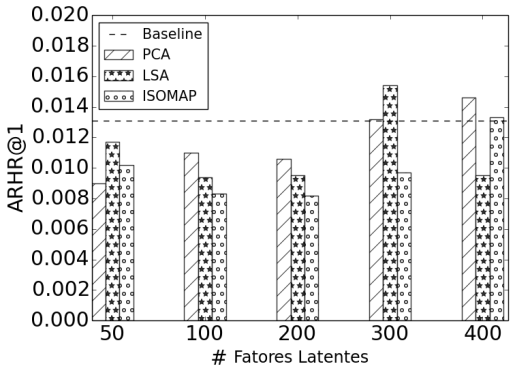

Figura 10 – Variação do número de fatores latentes.

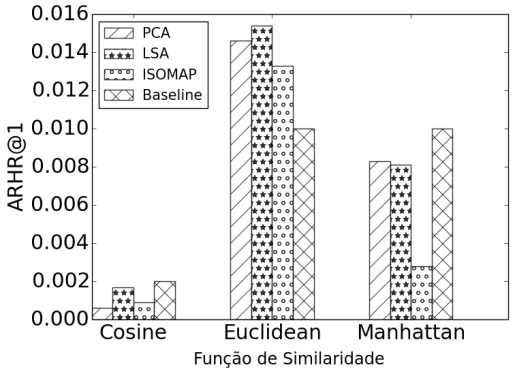

Figura 11 – Variação da medida de similaridade.

melhores resultados ao considerar mais de 300 fatores latentes para representar imagens, mas tal fator não representa tanta influência quanto o número de imagens recomendadas, uma vez que tal percepção não se mostra estável em todos os métodos de redução de

		1	3	5	10	20
PCA	50	0.0113	0.0289	0.0367	0.0441	0.0451
	100	0.0091	0.0267	0.0352	0.0421	0.0431
	200	0.0102	0.0265	0.0350	0.0413	0.0425
	300	0.0145	0.0321	0.0396	0.0445	0.0458
	400	0.0095	0.0267	0.0373	0.0436	0.0447
LSA	50	0.0117	0.0293	0.0389	0.0448	0.0461
	100	0.0094	0.0274	0.0357	0.0429	0.0444
	200	0.0095	0.0300	0.0360	0.0417	0.0431
	300	0.0154	0.0322	0.0406	0.0467	0.0480
	400	0.0095	0.0289	0.0381	0.0435	0.0446
ISOMAP	50	0.0063	0.0220	0.0302	0.0370	0.0392
	100	0.0069	0.0256	0.0324	0.0394	0.0413
	200	0.0084	0.0242	0.0336	0.0388	0.0412
	300	0.0106	0.0266	0.0330	0.0404	0.0429
	400	0.0088	0.0262	0.0332	0.0400	0.0413

Tabela 8 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação ARHR; a distância euclideana enquanto medida de de similaridade; e TFIDF+SIFT enquanto método de representação das imagens.

dimensionalidade e em toda variação do tamanho da lista.

Para ilustrar os tipos de resultados retornados pelo método, a Figura 12 mostra um exemplo de imagem recomendada para um usuário, dado suas imagens já avaliadas. As três imagens desses usuários estão de alguma forma relacionadas à arte (pinturas e vitrais), e a imagem sugerida traz uma escultura, o que é considerado uma boa recomendação.

Figura 12 – Exemplo ilustrativo de uma recomendação de imagem realizada pelo arcabouço proposto.

6 Conclusões e Trabalhos Futuros

O desenvolvimento de um arcabouço de recomendação que combine diferentes modos de representação de imagens em baixa dimensionalidade é uma tarefa desafiadora do ponto de vista científico. Um conjunto de técnicas diferentes deve ser empregado para que se possa obter boas descrições dos itens e das preferências dos usuários garantindo, assim, a qualidade das recomendações.

O trabalho de pesquisa apresentado nesta dissertação traz uma metodologia a ser aplicada nesse cenário, i.e., recomendação de imagens. Por meio do uso de técnicas de Aprendizado de Máquina, Visão Computacional e Recuperação de Informação, esta dissertação foca na representação multimodal de imagens, especificamente, textual e visual, em baixa dimensionalidade para melhorar a qualidade das recomendações.

Assim, esta dissertação propõe um arcabouço genérico para recomendação de imagens chamado LDSSLIM, o qual utiliza a representação de imagens em espaços de baixa dimensionalidade para sugerir recomendações relevantes aos usuários. A representação das imagens em baixa dimensionalidade alivia o problema da Maldição da Dimensionalidade (*Curse of Dimensionality*), que é um grande problema em sistemas de recomendação baseados em itens. Várias parametrizações do arcabouço proposto foram testadas, especificamente, (i) representação de imagem textual bruta, (ii) representação de imagem visual bruta, (iii) método de redução de dimensionalidade, (iv) número de fatores latentes e (v) função de similaridade.

Os resultados apresentados no Capítulo 5 mostraram que o arcabouço proposto apresentou ganhos sobre os algoritmos de recomendação baseados em itens estado-da-arte. Embora a lista de parâmetros investigada seja representativa, o espaço para outras possíveis escolhas é bastante grande. Em outras palavras, existem várias possibilidades para futuras investigações. Por exemplo, uma linha de pesquisa interessante é a representação das imagens por meio de características visuais utilizando Redes Neurais Convolutivas (*Convolutional Neural Networks*) (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Outra linha de investigação refere-se à construição da matriz de similaridade, eliminando valores muito baixos, que podem ser referir-se à ruídos na representação.

Aliada à essas possibilidades, ainda pode-se realizar, como trabalhos futuros, a identificação das melhores formas de representação, seja textual ou visual, das imagens testando-se outros métodos de representação, bem como as melhores combinações das mesmas, o teste de novos métodos de redução de dimensionalidade e medidas de similaridade, a fim de melhorar a qualidade da recomendação.

Referências

AARTHI, S.; SAMPATH, M. S. A heat diffusion method for mining web graphs for recommendations using recommendation algorithm. In: . [S.l.]: International Journal of Engineering Research Technology (IJERT), 2013. Citado na página 10.

ADOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. **IEEE Transactions on Knowledge and Data Engineering**, v. 17, n. 6, p. 734–749, June 2005. ISSN 1041-4347. Citado na página 15.

ADOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. **IEEE transactions on knowledge and data engineering**, IEEE, v. 17, n. 6, p. 734–749, 2005. Citado na página 17.

ARYAFAR, K.; SHOKOUFANDEH, A. Multimodal music and lyrics fusion classifier for artist identification. In: IEEE. **Machine Learning and Applications (ICMLA), 2014 13th International Conference on**. [S.l.], 2014. p. 506–509. Citado na página 11.

BAEZA-YATES, R.; RIBEIRO-NETO, B. et al. **Modern information retrieval**. [S.l.]: ACM press New York, 1999. v. 463. Citado na página 19.

BARRILERO M. ; ALDUAN, M. . S. F. . A. F. In-network content based image recommendation system for content-aware networks. In: . Shanghai: IEEE, 2011. p. 115 – 120. Citado 2 vezes nas páginas 5 e 13.

BAY, H.; TUYTELAARS, T.; GOOL, L. V. Surf: Speeded up robust features. In: SPRINGER. **European conference on computer vision**. [S.l.], 2006. p. 404–417. Citado na página 21.

BEZERRA, B. L. D. Uma solução em filtragem de informação para sistemas de recomendação baseada em análise de dados simbólicos. Universidade Federal de Pernambuco, 2004. Citado na página 17.

BIDART, R. **Toyslim - slim and sslim toy implementations.** 2015. Disponível em: <http://github.com/ruhan/toyslim>. Acesso em: 05 de outubro de 2015. Citado na página 30.

BOBADILLA, J. et al. Recommender systems survey. **Knowledge-based systems**, Elsevier, v. 46, p. 109–132, 2013. Citado na página 2.

BURKE, R. Hybrid recommender systems: Survey and experiments. **User Modeling and User-Adapted Interaction**, v. 12, n. 4, p. 331–370, November 2002. Citado na página 1.

CALLAN, J. e. a. Personalisation and recommender systems in digital libraries. **Joint NSF-EU DELOS Working Group Report**, 2002. Citado na página 2.

CORMEN, T. H. et al. **Introduction to algorithms second edition**. [S.l.]: The MIT Press, 2001. Citado na página 26.

COSTA, M. Â. A.; MARTINS, B. Uma comparaçao sistemática de diferentes abordagens para a sumarizaçao automática extrativa de textos em português. **Linguamática**, v. 7, n. 1, p. 23–40, 2015. Citado na página 24.

COX, T. F.; COX, M. A. **Multidimensional scaling**. [S.l.]: CRC press, 2000. Citado na página 26.

CREMONESI, P.; KOREN, Y.; TURRIN, R. Performance of recommender algorithms on top-n recommendation tasks. In: **Proceedings of the Fourth ACM Conference on Recommender Systems**. New York, NY, USA: ACM, 2010. (RecSys '10), p. 39–46. ISBN 978-1-60558-906-0. Disponível em: <http://doi.acm.org/10.1145/1864708.1864721>. Citado 3 vezes nas páginas 2, 13 e 18.

CUNNINGHAM, J. P.; BYRON, M. Y. Dimensionality reduction for large-scale neural recordings. **Nature neuroscience**, Nature Research, v. 17, n. 11, p. 1500–1509, 2014. Citado na página 31.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE. **2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)**. [S.l.], 2005. v. 1, p. 886–893. Citado na página 22.

DESHPANDE, M.; KARYPIS, G. Item-based top-n recommendation algorithms. **ACM Trans. Inf. Syst.**, ACM, New York, NY, USA, v. 22, n. 1, p. 143–177, jan. 2004. ISSN 1046-8188. Disponível em: <http://doi.acm.org/10.1145/963770.963776>. Citado 2 vezes nas páginas 2 e 18.

FERREIRA, M. et al. Chemometrics i: multivariate calibration, a tutorial. **Química Nova**, SciELO Brasil, v. 22, n. 5, p. 724–731, 1999. Citado na página 23.

GOLUB, G.; KAHAN, W. Calculating the singular values and pseudo-inverse of a matrix. **Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis**, SIAM, v. 2, n. 2, p. 205–224, 1965. Citado na página 24.

GONG, Y.; LIU, X. Creating generic text summaries. In: IEEE. **Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on**. [S.l.], 2001. p. 903–907. Citado na página 25.

GRITTI, T. et al. Local features based facial expression recognition with face registration errors. In: IEEE. **Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on**. [S.l.], 2008. p. 1–8. Citado na página 22.

HERLOCKER, J. L. **Understanding and improving automated collaborative filtering systems**. Tese (Doutorado), 2000. Citado na página 17.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of educational psychology**, Warwick & York, v. 24, n. 6, p. 417, 1933. Citado na página 23.

HUISKES, B. T. M. J.; LEW, M. S. New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative. In: **MIR '10: Proceedings of the 2010 ACM International Conference on Multimedia Information Retrieval**. New York, NY, USA: ACM, 2010. p. 527–536. Citado na página 34.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, n. 4, p. 679 – 688, 2006. ISSN 0169-2070. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0169207006000239>. Citado na página 18.

JANNACH, D. et al. **Recommender Systems: An Introduction**. 1st. ed. New York, NY, USA: Cambridge University Press, 2010. ISBN 0521493366, 9780521493369. Citado na página 15.

KABBUR, S.; , X. ; KARYPIS, G. Fism: factored item similarity models for top-n recommender systems. In: **ACM SIGKDD**. [S.l.: s.n.], 2013. p. 659–667. Citado na página 2.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105. Citado na página 46.

KUMAR, V. et al. **Introduction to parallel computing: design and analysis of algorithms**. [S.l.]: Benjamin/Cummings Redwood City, CA, 1994. v. 400. Citado na página 26.

LANDAUER, T. K.; FOLTZ, P. W.; LAHAM, D. An introduction to latent semantic analysis. **Discourse processes**, Taylor & Francis, v. 25, n. 2-3, p. 259–284, 1998. Citado 2 vezes nas páginas 24 e 25.

LEMOS, F. D. et al. Towards a context-aware photo recommender system. In: **Context-aware recommender system workshops**. [S.l.: s.n.], 2012. Citado 2 vezes nas páginas 9 e 10.

LI, Z. et al. Video recommendation based on multi-modal information and multiple kernel. **Multimedia Tools and Applications**, v. 74, n. 13, p. 4599–4616, 2014. ISSN 1380-7501. Disponível em: <http://link.springer.com/10.1007/s11042-013-1825-x>. Citado na página 11.

LOWE, D. G. Object recognition from local scale-invariant features. In: IEEE. **Computer vision, 1999. The proceedings of the seventh IEEE international conference on**. [S.l.], 1999. v. 2, p. 1150–1157. Citado na página 20.

LOWE, L. **125 Amazing Social Media Statistics You Should Know in 2016**. 2016. Disponível em: <https://socialpilot.co/blog/125-amazing-social-media-statistics-know-2016/>. Acesso em: 03 de junho de 2017. Citado na página 1.

MICHEL, F. **How many public photos are uploaded to Flickr every day, month, year?** 2016. Disponível em: <https://www.flickr.com/photos/franckmichel/6855169886>. Acesso em: 03 de junho de 2017. Citado na página 1.

NING, X.; KARYPIS, G. Slim: Sparse linear methods for top-n recommender systems. In: IEEE. **2011 IEEE 11th International Conference on Data Mining**. [S.l.], 2011. p. 497–506. Citado 3 vezes nas páginas 2, 13 e 26.

NING, X.; KARYPIS, G. Sparse linear methods with side information for top-n recommendations. In: ACM. **Proceedings of the sixth ACM conference on Recommender systems**. [S.l.], 2012. p. 155–162. Citado 5 vezes nas páginas 2, 13, 26, 27 e 29.

PARRA, D. et al. Implicit feedback recommendation via implicit-to-explicit ordinal logistic regression mapping. **Proceedings of the CARS-2011**, 2011. Citado na página 19.

PEDRO, J. S.; SIERSDORFER, S. Ranking and classifying attractiveness of photos in folksonomies. In: **Proceedings of the 18th International Conference on World Wide Web**. New York, NY, USA: ACM, 2009. (WWW '09), p. 771–780. ISBN 978-1-60558-487-4. Disponível em: <http://doi.acm.org/10.1145/1526709.1526813>. Citado na página 12.

PUEYO, R. van Z. A. R. L. G. Prediction of favourite photos using social, visual, and textual signals. In: **MM '10 Proceedings of the 18th ACM international conference on Multimedia**. New York, NY, USA: ACM, 2010. p. 1015–1018. Citado 2 vezes nas páginas 4 e 12.

RIBEIRO, J. P. M. P. **Método de correspondência para sistemas de visão multi-câmara**. Tese (Doutorado), 2015. Citado na página 20.

RICCI, F.; ROKACH, L.; SHAPIRA, B. Introduction to recommender systems handbook. In: **Recommender systems handbook**. [S.l.]: Springer, 2011. p. 1–35. Citado na página 15.

SAXENA, N. et al. **Building an image-based shoe recommendation system**. [S.l.]: Stanford University, 2013. Citado na página 9.

SCHAFFER J.B., K. J. A. R. J. E-commerce recommendation applications. **Data Mining and Knowledge Discovery. Kluwer Academic Publishers**, MA,USA, v. 5, n. 5, 2001. Citado na página 1.

SCHEDL, M.; SCHNITZER, D. Location-aware music artist recommendation. In: ____. **MultiMedia Modeling: 20th Anniversary International Conference, MMM 2014, Dublin, Ireland, January 6-10, 2014, Proceedings, Part II**. Cham: Springer International Publishing, 2014. p. 205–213. ISBN 978-3-319-04117-9. Disponível em: <http://dx.doi.org/10.1007/978-3-319-04117-9_19>. Citado na página 11.

SILVA, A. T. da. Descritores de imagem. Material de André Tavares sobre Descritores de Imagem. 2014. Citado na página 20.

SOUZA, A. M. d.; POPPI, R. J. et al. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte i. **Química Nova**, Sociedade Brasileira de Química, 2012. Citado 2 vezes nas páginas 23 e 24.

SUNDARESAN, V. J. R. P. A. B. W. D. N. Large scale visual recommendations from street fashion images. In: . New York, NY, USA: KDD '14, 2014. p. 1925–1934. Citado 2 vezes nas páginas 5 e 8.

SYMEONIDIS, P. Content-based dimensionality reduction for recommender systems. In: **Data Analysis, Machine Lear and Applications**. [S.l.]: Springer, 2008. p. 619–626. Citado na página 14.

TENENBAUM, J. B.; SILVA, V. D.; LANGFORD, J. C. A global geometric framework for nonlinear dimensionality reduction. **Science**, American Association for the Advancement of Science, v. 290, n. 5500, p. 2319–2323, 2000. Citado na página 25.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. Springer-Verlag New York, Inc., 1995. ISBN 0387945598. Disponível em: <http://portal.acm.org/citation.cfm?id=211359>. Citado na página 11.

XIE, P. et al. Mining user interests from personal photos. In: CITESEER. **AAAI**. [S.l.], 2015. p. 1896–1902. Citado na página 4.

YAMAMOTO, M.; CHURCH, K. W. Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus. **Comput. Linguist.**, MIT Press, Cambridge, MA, USA, v. 27, n. 1, p. 1–30, mar. 2001. ISSN 0891-2017. Disponível em: <http://dx.doi.org/10.1162/089120101300346787>. Citado na página 19.

YANG, B. et al. Online video recommendation based on multimodal fusion and relevance feedback. In: **Proceedings of the 6th ACM International Conference on Image and Video Retrieval**. New York, NY, USA: ACM, 2007. (CIVR '07), p. 73–80. ISBN 978-1-59593-733-9. Disponível em: <http://doi.acm.org/10.1145/1282280.1282290>. Citado na página 10.

ZHU, Q. et al. Multimodal sparse linear integration for content-based item recommendation. In: **Multimedia (ISM), 2013 IEEE International Symposium on**. [S.l.: s.n.], 2013. p. 187–194. Citado na página 12.

Apêndices

APÊNDICE A – Métricas de Avaliação: Precisão, Revocação, MAP e NDCG

Além das métricas de avaliação citadas na Seção 5.2, os experimentos realizados nesta dissertação também foram avaliadas segundo quatro outras métricas: (i) Precisão, (ii) Revocação (Recall), (iii) MAP e (iv) NDCG, descritas a seguir:

(a) **Precisão:** consiste na relação entre o número de acertos e o tamanho do conjunto de imagens recomendadas, onde o número de acertos é obtido verificando-se a existência das imagens retornadas na base de teste de cada usuário, como mostra a Equação 19.

$$precisao@N = \frac{\#acertos}{tamanho} \tag{19}$$

Onde $\#acertos$ é o número de acertos e $tamanho$ é o tamanho N da lista retornada.

(b) **Recall:** consiste na relação entre o número de acertos e o número de relevantes, onde o número de acertos é obtido verificando-se a existência das imagens retornadas na base de teste de cada usuário, como mostra a Equação 20.

$$recall@N = \frac{\#acertos}{\#relevantes} \tag{20}$$

Onde $\#acertos$ é o número de acertos e $\#relevantes$ é a quantidade de itens relevantes do usuário, ou seja, a quantidade de itens da base de teste de cada usuário.

(c) **MAP:** *mean average precision* consiste na média das pontuações de precisão média para cada recomendação, como mostra a Equação 21.

$$MAP@N = \sum_{q=1}^{Q} \frac{AveP(Q)}{Q} \tag{21}$$

Onde Q é o número de recomendações e $AveP(Q) = \frac{\sum_{k=1}^{n} P(k) \times rel(k)}{\#relevantes}$, sendo $P(k)$ a precisão na posição k da lista retornada e $rel(k)$ é uma função de binária cujo valor é igual a 1 se o item na posição k é relevante e zero caso contrário.

(d) **NDCG:** *normalized discounted cumulative gain* consiste na normalização do DCG, cujo propósito é avaliar uma recomendação com base em sua posição na lista recomendada, por meio de uma lista ideal recomendada, conforme mostra a Equação 22.

$$NDCG@N = \frac{DCG_p}{IDCG_p} \tag{22}$$

Onde $DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{log_2 i}$, sendo rel uma função que indica a relevância da recomendação e p a posição da lista retornada e o $IDCG_p$ um valor ideal do DCG, que classifica as recomendações da lista segundo sua relevância.

Os resultados descritos no Capítulo 5 serão apresentados neste Apêndice, avaliados segundo as métricas acima descritas.

As Tabela 9 e Tabela 10 apresentam a performance dos métodos de referência e do LDSSSLIM em termos das métricas de avaliação Precisão e Recall; MAP e NDCG, respectivamente. Para os métodos de referência, os resultados relatados abrangem a melhor combinação de representação métrica em termos do tipo de característica da imagem para cada modalidade e da medida de similaridade. Também é apresentado os resultados do SSLIM ao concatenar todas as informações disponíveis em uma única matriz (SSLIM-concat). Foi considerada apenas a primeira posição da lista recomendada (ou seja, N = 1) pelo fato de a maioria dos usuários da coleção ter avaliado um número pequeno de imagens. Para todas as métricas, o arcabouço proposto supera os métodos de referência com ganhos de 5% em termos das métricas Precisão e NDCG, e 9% em termos das métricas Recall e MAP. É possível observar que os resultados do SLIM são realmente baixos, pois não consideram informações laterais, e a matriz de classificação da coleção é extremamente esparsa. A partir desses resultados, pode-se concluir que uma representação de informação bidimensional de dimensões baixas melhora a qualidade da recomendação.

	Método	Parâmetros	Precisão@1 (ganho)	Parâmetros	Recall@1 (ganho)
Métodos de Referência	SLIM	-	0.0015	-	0.0010
	SSLIM-textual	TF:Man	0.1324	TF:Man.	0.0982
	SSLIM-visual	Hog:Euc	0.0025	Hog:Euc	0.0013
	SSLIM-concat.	TF+Surf:Euc	*0.1396*	TF+Surf:Euc	*0.1069*
	LDSSLIM	PCA:400 TF+HoG:Euc	0.1470 (0.0530)	PCA:400 TF+Hog:Euc	0.1167 (0.0916)

Tabela 9 – Desempenho dos métodos de recomendação de imagem considerando Precisao@1 e Recall@1. Os ganhos do arcabouço proposto são relativos ao melhor método dentre os *baselines* e são apresentados entre parênteses. O melhor desempenho dentre os *baselines* é representado em itálico.

A Tabela 11 mostra o impacto da variação dos métodos de representação das imagens em termos de Precisão@1, Recall@1, MAP@1 e NDCG@1, da distância euclideana

	Método	Parâmetros	MAP@1 (ganho)	Parâmetros	NDCG@1 (ganho)
Métodos de Referência	SLIM	-	0.0010	-	0.0015
	SSLIM-textual	TF:Man	0.0982	TF:Man	0.1324
	SSLIM-visual	Hog:Euc	0.0013	Hog:Euc	0.0025
	SSLIM-concat.	TF+Surf:Euc	*0.1069*	TF+Surf:Euc	*0.1396*
	LDSSLIM	PCA:400 TF+HoG:Euc	0.1167 (0.0916)	PCA:400 TF+Hog:Euc	0.1470 (0.0530)

Tabela 10 – Desempenho dos métodos de recomendação de imagem considerando MAP@1 e NDCG@1. Os ganhos do arcabouço proposto são relativos ao melhor método dentre os *baselines* e são apresentados entre parênteses. O melhor desempenho dentre os *baselines* é representado em itálico.

enquanto função de similaridade e considerando o número de fatores latentes igual a 400. Tais fatores foram considerados por apresentarem melhor desempenhado, como mostrado nas Tabela 9 e Tabela 10. Com as demais métricas de avaliação, também é possível notar que as combinações aliadas ao histograma de cor possuem desempenho bastante inferior aos demais, o que comprova que as características extraídas das imagens impactam o desempenho do recomendador e o histograma de cor consiste em uma descrição de baixa exatidão.

As Tabela 12, Tabela 13, Tabela 14 e Tabela 15 detalham o impacto do tamanho da lista retornada, ou seja, a quantidade de imagens recomendadas ao usuário, e o número de fatores latentes. Os resultados são apresentados em termos da métrica de avaliação Precisão, Recall, MAP e NDCG - respectivamente, da distância euclideana enquanto função de similaridade e o TF+HoG enquanto método de representação das imagens, por apresentar o melhor desempenho como mostram as Tabela 9 e Tabela 10.

		Precisao@1	Recall@1	MAP@1	NDCG@1
		400	400		
PCA	TF + Cor	0.0336	0.0221	0.0221	0.0336
	TF + HoG	0.1470	0.1167	0.1167	0.1470
	TF + SIFT	0.1236	0.0945	0.0945	0.1236
	TF + SURF	0.1452	0.1112	0.1112	0.1452
	TF-IDF + Cor	0.0407	0.0276	0.0276	0.0407
	TF-IDF + HoG	0.1178	0.0920	0.0920	0.1178
	TF-IDF + SIFT	0.1210	0.0928	0.0928	0.1210
	TF-IDF + SURF	0.1237	0.0950	0.0950	0.1237
LSA	TF + Cor	0.0433	0.0304	0.0304	0.0433
	TF + HoG	0.1248	0.0929	0.1248	0.1248
	TF + SIFT	0.1271	0.0991	0.1271	0.1271
	TF + SURF	0.1092	0.0801	0.1092	0.1092
	TF-IDF + Cor	0.0358	0.0244	0.0244	0.0358
	TF-IDF + HoG	0.1254	0.0958	0.0958	0.1254
	TF-IDF + SIFT	0.1201	0.0916	0.0916	0.1201
	TF-IDF + SURF	0.1173	0.0870	0.0870	0.1173
ISOMAP	TF + Cor	0.0237	0.0147	0.0147	0.0237
	TF + HoG	0.1160	0.0902	0.0902	0.1160
	TF + SIFT	0.1125	0.0841	0.0841	0.1125
	TF + SURF	0.1182	0.0901	0.0901	0.1182
	TF-IDF + Cor	0.0259	0.0161	0.0161	0.0259
	TF-IDF + HoG	0.1147	0.0873	0.0873	0.1147
	TF-IDF + SIFT	0.1153	0.0861	0.0861	0.1153
	TF-IDF + SURF	0.1370	0.1062	0.1062	0.1370

Tabela 11 – Desempenho do arcabouço LDSSLIM variando-se os métodos de representação das imagens e as demais métricas de avaliação - Precisão@1, Recall@1, MAP@1 e NDCG@1, considerando a distância euclideana enquanto medida de de similaridade; e 400 enquanto o número de fatores latentes.

		1	3	5	10	20
PCA	50	0.1126	0.0945	0.0718	0.0428	0.0226
	100	0.1282	0.0938	0.0718	0.0429	0.0226
	200	0.1272	0.0931	0.0711	0.0424	0.0219
	300	0.1397	0.0946	0.0702	0.0421	0.0218
	400	0.1470	0.0939	0.0710	0.0427	0.0226
LSA	50	0.1102	0.0946	0.0735	0.0430	0.0225
	100	0.1260	0.0985	0.0719	0.0439	0.0233
	200	0.1265	0.0944	0.0706	0.0424	0.0223
	300	0.1282	0.0926	0.0712	0.0425	0.0222
	400	0.1248	0.0911	0.0717	0.0426	0.0224
ISOMAP	50	0.1076	0.0785	0.0598	0.0406	0.0223
	100	0.1166	0.0785	0.0604	0.0406	0.0223
	200	0.1168	0.0763	0.0626	0.0418	0.0226
	300	0.1127	0.0808	0.0616	0.0425	0.0226
	400	0.1160	0.0797	0.0614	0.0414	0.0224

Tabela 12 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação Precisão; a distância euclideana enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
PCA	50	0.0830	0.2070	0.2576	0.2949	0.3061
	100	0.1005	0.2113	0.2563	0.2952	0.3060
	200	0.0999	0.2092	0.2587	0.2948	0.3039
	300	0.1081	0.2134	0.2541	0.2935	0.3024
	400	0.1167	0.2121	0.2571	0.2938	0.3059
LSA	50	0.0786	0.2125	0.2641	0.2964	0.3066
	100	0.0979	0.2205	0.2548	0.2972	0.3044
	200	0.0953	0.2130	0.2556	0.2949	0.3047
	300	0.1004	0.2089	0.2579	0.2955	0.3038
	400	0.0929	0.2025	0.2592	0.2953	0.3065
ISOMAP	50	0.0787	0.1754	0.2116	0.2796	0.3038
	100	0.0896	0.1730	0.2159	0.2813	0.3035
	200	0.0888	0.1661	0.2238	0.2890	0.3055
	300	0.0842	0.1765	0.2173	0.2931	0.3053
	400	0.0902	0.1782	0.2170	0.2875	0.3052

Tabela 13 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação Recall; a distância euclideana enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
PCA	50	0.0830	0.0600	0.0435	0.0087	0.0009
	100	0.1005	0.0629	0.0457	0.0077	0.0000
	200	0.0999	0.0591	0.0371	0.0099	0.0000
	300	0.1081	0.0625	0.0410	0.0040	0.0000
	400	0.1167	0.0599	0.0441	0.0099	0.0069
LSA	50	0.0786	0.0686	0.0471	0.0102	0.0041
	100	0.0979	0.0649	0.0442	0.0135	0.0025
	200	0.0953	0.0662	0.0390	0.0005	0.0003
	300	0.1004	0.0620	0.0407	0.0083	0.0000
	400	0.0929	0.0589	0.0440	0.0059	0.0009
ISOMAP	50	0.0787	0.0533	0.0371	0.0072	0.0000
	100	0.0896	0.0502	0.0343	0.0129	0.0041
	200	0.0888	0.0506	0.0342	0.0167	0.0041
	300	0.0842	0.0496	0.0360	0.0150	0.0047
	400	0.0902	0.0533	0.0366	0.0168	0.0028

Tabela 14 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação MAP; a distância euclideana enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
PCA	50	0.1126	0.1821	0.1980	0.2089	0.2099
	100	0.1282	0.1900	0.2025	0.2140	0.2158
	200	0.1272	0.1872	0.2044	0.2147	0.2154
	300	0.1397	0.1931	0.2053	0.2178	0.2181
	400	0.1470	0.1985	0.2124	0.2227	0.2236
LSA	50	0.1102	0.1817	0.1974	0.2060	0.2074
	100	0.1260	0.1906	0.2032	0.2128	0.2141
	200	0.1265	0.1908	0.2013	0.2132	0.2143
	300	0.1282	0.1891	0.2054	0.2145	0.2145
	400	0.1248	0.1837	0.2040	0.2132	0.2142
ISOMAP	50	0.1076	0.1603	0.1727	0.1881	0.1910
	100	0.1166	0.1656	0.1793	0.1943	0.1959
	200	0.1168	0.1613	0.1791	0.1954	0.1966
	300	0.1127	0.1641	0.1772	0.1955	0.1965
	400	0.1160	0.1654	0.1779	0.1975	0.1982

Tabela 15 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação NDCG; a distância euclideana enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

APÊNDICE B – Medidas de Similaridade: Cosseno e Manhattan

Os resultados descritos no Capítulo 5 e Apêndice A utilizam a distância euclideana como medida de similaridade por apresentar melhor desempenho na maior parte dos experimentos. O arcabouço também foi projeto para o uso de outras métricas, e os experimentos apresentados também foram realizado utilizando-se o cosseno e a distância de manhattan como métricas de similaridade cujos resultados são apresentados neste Apêndice.

As Tabela 16 e Tabela 17 mostram o impacto da variação dos métodos de representação das imagens em termos de HR@1 e ARHR@1; Precisão@1, Recall@1, MAP@1 e NDCG@1 - respectivamente, o cosseno enquanto função de similaridade e considerando o número de fatores latentes igual a 300/400. Tais fatores foram considerados por apresentarem melhor desempenhado, como mostrado nas Tabela 4, Tabela 9 e Tabela 10. Com as demais métricas de avaliação, também é possível notar que as combinações aliadas ao histograma de cor possuem desempenho bastante inferior aos demais, o que comprova que as características extraídas das imagens impactam o desempenho do recomendador e o histograma de cor consiste em uma descrição de baixa exatidão.

As tabelas Tabela 18 e Tabela 19 possui o mesmo foco das tabelas acima, porém considerando a distância de manhattan como métrica de similaridade.

As Tabela 20, Tabela 21, Tabela 22, Tabela 23, Tabela 24, e Tabela 25 detalham o impacto do tamanho da lista retornada, ou seja, a quantidade de imagens recomendadas ao usuário, e o número de fatores latentes. Os resultados são apresentados em termos da métrica de avaliação HR, ARHR Precisão, Recall, MAP e NDCG - respectivamente, do cosseno enquanto função de similaridade e o TFIDF+Sift enquanto método de representação das imagens para o ARHR e TF+HoG para as demais métricas de avaliação, por apresentar o melhor desempenho como mostram as Tabela 4, Tabela 9 e Tabela 10.

As tabelas Tabela 26, Tabela 27, Tabela 28, Tabela 29, Tabela 30 e Tabela 31 possui o mesmo foco das tabelas acima, porém considerando a distância de manhattan como métrica de similaridade.

		HR@1		ARHR@1	
		300	400	300	400
PCA	TF + Cor	0.0099	0.0082	0.0008	0.0007
	TF + HoG	0.0079	0.0061	0.0007	0.0006
	TF + SIFT	0.0115	0.0092	0.0018	0.0012
	TF + SURF	0.0103	0.0094	0.0015	0.0012
	TF-IDF + Cor	0.0115	0.0091	0.0009	0.0008
	TF-IDF + HoG	0.0043	0.0019	0.0004	0.0002
	TF-IDF + SIFT	0.0055	0.0036	0.0004	0.0003
	TF-IDF + SURF	0.0057	0.0044	0.0005	0.0003
LSA	TF + Cor	0.0057	0.0058	0.0003	0.0003
	TF + HoG	0.0063	0.0064	0.0006	0.0007
	TF + SIFT	0.0057	0.0065	0.0007	0.0008
	TF + SURF	0.0046	0.0065	0.0006	0.0009
	TF-IDF + Cor	0.0057	0.0058	0.0003	0.0003
	TF-IDF + HoG	0.0074	0.0034	0.0011	0.0005
	TF-IDF + SIFT	0.0100	0.0157	0.0017	0.0030
	TF-IDF + SURF	0.0095	0.0176	0.0016	0.0021
ISOMAP	TF + Cor	0.0070	0.0128	0.0013	0.0015
	TF + HoG	0.0155	0.0182	0.0027	0.0035
	TF + SIFT	0.0184	0.0179	0.0038	0.0034
	TF + SURF	0.0205	0.0179	0.0043	0.0035
	TF-IDF + Cor	0.0059	0.0135	0.0011	0.0014
	TF-IDF + HoG	0.0050	0.0076	0.0007	0.0009
	TF-IDF + SIFT	0.0060	0.0085	0.0005	0.0009
	TF-IDF + SURF	0.0071	0.0061	0.0005	0.0009

Tabela 16 – Desempenho do arcabouço LDSSLIM variando-se os métodos de representação das imagens, considerando HR@1 e ARHR@1; o cosseno enquanto medida de de similaridade; [300, 400] enquanto o número de fatores latentes.

		Precisao@1	Recall@1	MAP@1	NDCG@1
		400	400		
PCA	TF + Cor	0.0247	0.0131	0.0131	0.0247
	TF + HoG	0.0289	0.0127	0.0127	0.0289
	TF + SIFT	0.0386	0.0272	0.0272	0.0386
	TF + SURF	0.0369	0.0264	0.0264	0.0369
	TF-IDF + Cor	0.0274	0.0156	0.0156	0.0274
	TF-IDF + HoG	0.0119	0.0089	0.0089	0.0119
	TF-IDF + SIFT	0.0156	0.0094	0.0094	0.0156
	TF-IDF + SURF	0.0179	0.0116	0.0116	0.0179
LSA	TF + Cor	0.0135	0.0086	0.0086	0.0135
	TF + HoG	0.0246	0.0169	0.0169	0.0246
	TF + SIFT	0.0314	0.0272	0.0272	0.0314
	TF + SURF	0.0320	0.0269	0.0269	0.0320
	TF-IDF + Cor	0.0137	0.0087	0.0087	0.0137
	TF-IDF + HoG	0.0186	0.0163	0.0163	0.0186
	TF-IDF + SIFT	0.0536	0.0382	0.0382	0.0536
	TF-IDF + SURF	0.0463	0.0315	0.0315	0.0463
ISOMAP	TF + Cor	0.0360	0.0161	0.0161	0.0360
	TF + HoG	0.0744	0.0479	0.0479	0.0744
	TF + SIFT	0.0709	0.0432	0.0432	0.0709
	TF + SURF	0.0731	0.0469	0.0469	0.0731
	TF-IDF + Cor	0.0348	0.0154	0.0154	0.0348
	TF-IDF + HoG	0.0307	0.0238	0.0238	0.0307
	TF-IDF + SIFT	0.0342	0.0243	0.0243	0.0342
	TF-IDF + SURF	0.0335	0.0227	0.0227	0.0335

Tabela 17 – Desempenho do arcabouço LDSSLIM variando-se os métodos de representação das imagens e as demais métricas de avaliação - Precisão@1, Recall@1, MAP@1 e NDCG@1, considerando o cosseno enquanto medida de de similaridade; e 400 enquanto o número de fatores latentes.

		HR@1		ARHR@1	
		300	400	300	400
PCA	TF + Cor	0.0094	0.0073	0.0010	0.0007
	TF + HoG	0.0231	0.0273	0.0070	0.0083
	TF + SIFT	0.0205	0.0187	0.0068	0.0059
	TF + SURF	0.0239	0.0190	0.0085	0.0058
	TF-IDF + Cor	0.0094	0.0073	0.0010	0.0007
	TF-IDF + HoG	0.0229	0.0196	0.0083	0.0058
	TF-IDF + SIFT	0.0202	0.0228	0.0060	0.0059
	TF-IDF + SURF	0.0189	0.0213	0.0061	0.0064
LSA	TF + Cor	0.0083	0.0086	0.0008	0.0009
	TF + HoG	0.0282	0.0251	0.0087	0.0092
	TF + SIFT	0.0235	0.0205	0.0080	0.0063
	TF + SURF	0.0201	0.0238	0.0072	0.0077
	TF-IDF + Cor	0.0083	0.0086	0.0008	0.0009
	TF-IDF + HoG	0.0252	0.0253	0.0089	0.0096
	TF-IDF + SIFT	0.0236	0.0197	0.0081	0.0053
	TF-IDF + SURF	0.0255	0.0193	0.0089	0.0067
ISOMAP	TF + Cor	0.0070	0.0100	0.0005	0.0010
	TF + HoG	0.0150	0.0164	0.0027	0.0033
	TF + SIFT	0.0222	0.0271	0.0057	0.0076
	TF + SURF	0.0222	0.0239	0.0053	0.0060
	TF-IDF + Cor	0.0070	0.0100	0.0005	0.0010
	TF-IDF + HoG	0.0161	0.0124	0.0034	0.0021
	TF-IDF + SIFT	0.0142	0.0138	0.0028	0.0026
	TF-IDF + SURF	0.0160	0.0151	0.0030	0.0028

Tabela 18 – Desempenho do arcabouço LDSSLIM variando-se os métodos de representação das imagens, considerando HR@1 e ARHR@1; a distância de manhattan enquanto medida de de similaridade; [300, 400] enquanto o número de fatores latentes.

		Precisao@1	Recall@1	MAP@1	NDCG@1
		400	400		
PCA	TF + Cor	0.0268	0.0147	0.0147	0.0268
	TF + HoG	0.1016	0.0741	0.0741	0.0741
	TF + SIFT	0.0924	0.0657	0.0657	0.0924
	TF + SURF	0.0923	0.0692	0.0692	0.0923
	TF-IDF + Cor	0.0268	0.0147	0.0147	0.0268
	TF-IDF + HoG	0.0892	0.0622	0.0622	0.0892
	TF-IDF + SIFT	0.0903	0.0647	0.0647	0.0903
	TF-IDF + SURF	0.0955	0.0677	0.0677	0.0955
LSA	TF + Cor	0.0330	0.0194	0.0194	0.0330
	TF + HoG	0.1141	0.0845	0.0845	0.1141
	TF + SIFT	0.0954	0.0706	0.0706	0.0954
	TF + SURF	0.1037	0.0789	0.0789	0.1037
	TF-IDF + Cor	0.0330	0.0194	0.0194	0.0330
	TF-IDF + HoG	0.1113	0.0802	0.0802	0.1113
	TF-IDF + SIFT	0.0878	0.0650	0.0650	0.0878
	TF-IDF + SURF	0.0934	0.0671	0.0671	0.0934
ISOMAP	TF + Cor	0.0332	0.0145	0.0145	0.0332
	TF + HoG	0.0693	0.0450	0.0450	0.0693
	TF + SIFT	0.0983	0.0700	0.0700	0.0983
	TF + SURF	0.0945	0.0642	0.0642	0.0945
	TF-IDF + Cor	0.0332	0.0145	0.0145	0.0332
	TF-IDF + HoG	0.0533	0.0356	0.0356	0.0533
	TF-IDF + SIFT	0.0603	0.0385	0.0385	0.0603
	TF-IDF + SURF	0.0665	0.0442	0.0442	0.0665

Tabela 19 – Desempenho do arcabouço LDSSLIM variando-se os métodos de representação das imagens e as demais métricas de avaliação - Precisão@1, Recall@1, MAP@1 e NDCG@1, considerando a distância de manhattan enquanto medida de de similaridade; e 400 enquanto o número de fatores latentes.

		1	3	5	10	20
	50	0.0137	0.0181	0.0218	0.0298	0.0458
	100	0.0125	0.0172	0.0198	0.0272	0.0396
PCA	200	0.0074	0.0176	0.0204	0.0252	0.0381
	300	0.0079	0.0189	0.0217	0.0275	0.0367
	400	0.0061	0.0174	0.0208	0.0242	0.0323
	50	0.0066	0.0164	0.0189	0.0286	0.0353
	100	0.0066	0.0148	0.0170	0.0268	0.0368
LSA	200	0.0071	0.0128	0.0180	0.0270	0.0336
	300	0.0063	0.0137	0.0170	0.0236	0.0324
	400	0.0064	0.0119	0.0154	0.0250	0.0326
	50	0.0177	0.0325	0.0393	0.0468	0.0514
	100	0.0190	0.0306	0.0381	0.0453	0.0506
ISOMAP	200	0.0177	0.0289	0.0371	0.0452	0.0505
	300	0.0155	0.0262	0.0381	0.0469	0.0521
	400	0.0182	0.0286	0.0364	0.0459	0.0517

Tabela 20 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação HR; o cosseno enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
	50	0.0040	0.0059	0.0064	0.0069	0.0093
	100	0.0020	0.0036	0.0048	0.0057	0.0065
PCA	200	0.0012	0.0034	0.0039	0.0048	0.0057
	300	0.0004	0.0015	0.0020	0.0029	0.0031
	400	0.0003	0.0011	0.0014	0.0019	0.0021
	50	0.0017	0.0047	0.0064	0.0074	0.0083
	100	0.0014	0.0041	0.0048	0.0067	0.0071
LSA	200	0.0015	0.0036	0.0062	0.0075	0.0083
	300	0.0017	0.0044	0.0066	0.0084	0.0084
	400	0.0030	0.0066	0.0070	0.0101	0.0111
	50	0.0013	0.0025	0.0033	0.0040	0.0055
	100	0.0008	0.0017	0.0026	0.0033	0.0044
ISOMAP	200	0.0007	0.0013	0.0021	0.0030	0.0043
	300	0.0005	0.0017	0.0022	0.0030	0.0048
	400	0.0009	0.0022	0.0029	0.0041	0.0070

Tabela 21 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação ARHR; o cosseno enquanto medida de de similaridade; e TFIDF+SIFT enquanto método de representação das imagens.

		1	3	5	10	20
	50	0.0517	0.0268	0.0199	0.0138	0.0098
	100	0.0513	0.0261	0.0193	0.0133	0.0093
PCA	200	0.0397	0.0242	0.0185	0.0124	0.0087
	300	0.0305	0.0251	0.0185	0.0131	0.0091
	400	0.0289	0.0237	0.0197	0.0135	0.0104
	50	0.0215	0.0236	0.0171	0.0141	0.0098
	100	0.0276	0.0218	0.0169	0.0144	0.0101
LSA	200	0.0248	0.0191	0.0173	0.0151	0.0105
	300	0.0241	0.0183	0.0165	0.0132	0.0093
	400	0.0246	0.0173	0.0156	0.0144	0.0096
	50	0.0725	0.0503	0.0372	0.0235	0.0136
	100	0.0741	0.0479	0.0367	0.0225	0.0131
ISOMAP	200	0.0724	0.0465	0.0358	0.0224	0.0131
	300	0.0647	0.0441	0.0349	0.0225	0.0134
	400	0.0744	0.0463	0.0339	0.0224	0.0130

Tabela 22 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação Precisão; o cosseno enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
	50	0.0138	0.0464	0.0566	0.0941	0.1250
	100	0.0202	0.0440	0.0530	0.0895	0.1188
PCA	200	0.0152	0.0336	0.0511	0.0864	0.1120
	300	0.0165	0.0350	0.0513	0.0825	0.1141
	400	0.0169	0.0322	0.0468	0.0902	0.1169
	50	0.0138	0.0464	0.0566	0.0941	0.1250
	100	0.0202	0.0440	0.0530	0.0895	0.1188
LSA	200	0.0152	0.0336	0.0511	0.0864	0.1120
	300	0.0165	0.0350	0.0513	0.0825	0.1141
	400	0.0169	0.0322	0.0468	0.0902	0.1169
	50	0.0469	0.0982	0.1222	0.1499	0.1700
	100	0.0466	0.0923	0.1199	0.1467	0.1702
ISOMAP	200	0.0449	0.0860	0.1148	0.1466	0.1687
	300	0.0387	0.0819	0.1129	0.1450	0.1724
	400	0.0479	0.0884	0.1101	0.1466	0.1702

Tabela 23 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação Recall; o cosseno enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
	50	0.0320	0.0039	0.0091	0.0027	0.0023
	100	0.0328	0.0036	0.0082	0.0048	0.0020
PCA	200	0.0212	0.0074	0.0036	0.0012	0.0011
	300	0.0145	0.0101	0.0015	0.0014	0.0030
	400	0.0127	0.0125	0.0028	0.0012	0.0027
	50	0.0138	0.0141	0.0088	0.0007	0.0044
	100	0.0202	0.0128	0.0081	0.0043	0.0000
LSA	200	0.0152	0.0099	0.0074	0.0039	0.0026
	300	0.0165	0.0102	0.0036	0.0006	0.0017
	400	0.0169	0.0055	0.0032	0.0041	0.0028
	50	0.0469	0.0257	0.0212	0.0092	0.0026
	100	0.0466	0.0269	0.0155	0.0031	0.0006
ISOMAP	200	0.0449	0.0222	0.0139	0.0050	0.0004
	300	0.0387	0.0232	0.0213	0.0013	0.0007
	400	0.0479	0.0225	0.0142	0.0022	0.0004

Tabela 24 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação MAP; o cosseno enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
	50	0.0517	0.0631	0.0682	0.0775	0.0873
	100	0.0513	0.0617	0.0656	0.0742	0.0820
PCA	200	0.0397	0.0580	0.0626	0.0669	0.0777
	300	0.0305	0.0534	0.0597	0.0633	0.0684
	400	0.0289	0.0488	0.0540	0.0563	0.0571
	50	0.0215	0.0465	0.0505	0.0595	0.0649
	100	0.0276	0.0470	0.0488	0.0579	0.0663
LSA	200	0.0248	0.0372	0.0448	0.0559	0.0630
	300	0.0241	0.0382	0.0456	0.0537	0.0619
	400	0.0246	0.0373	0.0446	0.0570	0.0637
	50	0.0725	0.1029	0.1113	0.1191	0.1229
	100	0.0741	0.0998	0.1114	0.1178	0.1223
ISOMAP	200	0.0724	0.0959	0.1075	0.1167	0.1197
	300	0.0647	0.0891	0.1041	0.1125	0.1159
	400	0.0744	0.0965	0.1057	0.1177	0.1181

Tabela 25 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação NDCG; o cosseno enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
PCA	50	0.0272	0.0500	0.0577	0.0691	0.0720
	100	0.0241	0.0484	0.0547	0.0659	0.0706
	200	0.0203	0.0415	0.0491	0.0630	0.0702
	300	0.0231	0.0401	0.0481	0.0623	0.0710
	400	0.0273	0.0397	0.0451	0.0636	0.0726
LSA	50	0.0298	0.0544	0.0632	0.0711	0.0728
	100	0.0317	0.0548	0.0620	0.0697	0.0721
	200	0.0283	0.0518	0.0603	0.0702	0.0721
	300	0.0282	0.0520	0.0611	0.0695	0.0719
	400	0.0251	0.0466	0.0598	0.0692	0.0713
ISOMAP	50	0.0235	0.0445	0.0509	0.0657	0.0704
	100	0.0209	0.0394	0.0512	0.0662	0.0735
	200	0.0218	0.0372	0.0460	0.0631	0.0741
	300	0.0150	0.0327	0.0431	0.0610	0.0715
	400	0.0164	0.0341	0.0454	0.0589	0.0716

Tabela 26 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação HR; a distância de manhattan enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
PCA	50	0.0120	0.0260	0.0329	0.0396	0.0418
	100	0.0090	0.0226	0.0296	0.0376	0.0402
	200	0.0101	0.0210	0.0268	0.0323	0.0393
	300	0.0060	0.0182	0.0220	0.0304	0.0346
	400	0.0059	0.0156	0.0180	0.0265	0.0317
LSA	50	0.0100	0.0240	0.0330	0.0401	0.0417
	100	0.0077	0.0206	0.0264	0.0337	0.0373
	200	0.0086	0.0206	0.0244	0.0320	0.0375
	300	0.0081	0.0177	0.0221	0.0293	0.0344
	400	0.0053	0.0163	0.0188	0.0263	0.0324
ISOMAP	50	0.0077	0.0184	0.0260	0.0329	0.0356
	100	0.0072	0.0170	0.0219	0.0287	0.0341
	200	0.0026	0.0115	0.0143	0.0207	0.0257
	300	0.0028	0.0090	0.0134	0.0181	0.0247
	400	0.0026	0.0086	0.0133	0.0182	0.0248

Tabela 27 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação ARHR; a distância de manhattan enquanto medida de de similaridade; e TFIDF+SIFT enquanto método de representação das imagens.

		1	3	5	10	20
PCA	50	0.1191	0.0838	0.0639	0.0400	0.0223
	100	0.1115	0.0793	0.0600	0.0391	0.0224
	200	0.0939	0.0707	0.0552	0.0353	0.0207
	300	0.0962	0.0700	0.0551	0.0353	0.0210
	400	0.1016	0.0693	0.0537	0.0377	0.0231
LSA	50	0.1221	0.0864	0.0677	0.0411	0.0224
	100	0.1269	0.0861	0.0666	0.0405	0.0221
	200	0.1180	0.0843	0.0651	0.0406	0.0220
	300	0.1130	0.0840	0.0637	0.0401	0.0218
	400	0.1141	0.0805	0.0643	0.0394	0.0213
ISOMAP	50	0.0960	0.0736	0.0558	0.0372	0.0208
	100	0.0919	0.0678	0.0559	0.0364	0.0215
	200	0.0810	0.0620	0.0498	0.0353	0.0212
	300	0.0639	0.0531	0.0458	0.0333	0.0203
	400	0.0693	0.0531	0.0479	0.0336	0.0209

Tabela 28 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação Precisão; a distância de manhattan enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
PCA	50	0.0931	0.1950	0.2436	0.2855	0.2994
	100	0.0972	0.1935	0.2397	0.2831	0.2975
	200	0.0869	0.1883	0.2361	0.2847	0.2975
	300	0.0826	0.1909	0.2319	0.2810	0.2941
	400	0.0845	0.1795	0.2347	0.2801	0.2960
LSA	50	0.0931	0.1950	0.2436	0.2855	0.2994
	100	0.0972	0.1935	0.2397	0.2831	0.2975
	200	0.0869	0.1883	0.2361	0.2847	0.2975
	300	0.0826	0.1909	0.2319	0.2810	0.2941
	400	0.0845	0.1795	0.2347	0.2801	0.2960
ISOMAP	50	0.0718	0.1627	0.1973	0.2640	0.2922
	100	0.0666	0.1441	0.1954	0.2559	0.2951
	200	0.0554	0.1238	0.1712	0.2474	0.2918
	300	0.0416	0.1091	0.1566	0.2328	0.2839
	400	0.0450	0.1089	0.1638	0.2331	0.2863

Tabela 29 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação Recall; a distância de manhattan enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
	50	0.0905	0.0573	0.0362	0.0054	0.0049
	100	0.0829	0.0499	0.0338	0.0074	0.0037
PCA	200	0.0677	0.0495	0.0300	0.0127	0.0003
	300	0.0679	0.0471	0.0329	0.0144	0.0027
	400	0.0741	0.0439	0.0302	0.0250	0.0063
	50	0.0931	0.0587	0.0446	0.0047	0.0015
	100	0.0972	0.0553	0.0426	0.0087	0.0022
LSA	200	0.0869	0.0575	0.0357	0.0126	0.0022
	300	0.0826	0.0583	0.0402	0.0181	0.0035
	400	0.0845	0.0553	0.0386	0.0130	0.0000
	50	0.0718	0.0493	0.0360	0.0176	0.0031
	100	0.0666	0.0461	0.0336	0.0125	0.0009
ISOMAP	200	0.0554	0.0380	0.0291	0.0174	0.0033
	300	0.0416	0.0308	0.0241	0.0191	0.0046
	400	0.0450	0.0349	0.0305	0.0178	0.0058

Tabela 30 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação MAP; a distância de manhattan enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.

		1	3	5	10	20
	50	0.1191	0.1732	0.1858	0.2007	0.1968
	100	0.1115	0.1617	0.1727	0.1872	0.1903
PCA	200	0.0939	0.1438	0.1556	0.1723	0.1782
	300	0.0962	0.1424	0.1580	0.1679	0.1745
	400	0.1016	0.1365	0.1491	0.1684	0.1761
	50	0.1221	0.1781	0.1941	0.2071	0.2038
	100	0.1269	0.1809	0.1962	0.2079	0.2033
LSA	200	0.1180	0.1707	0.1871	0.1989	0.1976
	300	0.1130	0.1702	0.1840	0.1954	0.1946
	400	0.1141	0.1667	0.1851	0.1970	0.1971
	50	0.0960	0.1488	0.1588	0.1779	0.1806
	100	0.0919	0.1343	0.1506	0.1658	0.1721
ISOMAP	200	0.0810	0.1182	0.1362	0.1551	0.1626
	300	0.0639	0.1058	0.1215	0.1434	0.1528
	400	0.0693	0.1055	0.1259	0.1441	0.1541

Tabela 31 – Desempenho do arcabouço LDSSLIM variando-se o tamanho da lista recomendada e o número de fatores latentes, considerando a métrica de avaliação NDCG; a distância de manhattan enquanto medida de de similaridade; e TF+HoG enquanto método de representação das imagens.