



AGRUPAMENTO DE BANDAS EM IMAGENS HIPERESPECTRAIS

Luiz R. F. ALVES¹; Diego SAQUI²

RESUMO

Com os avanços tecnológicos da atualidade, surgiram oportunidades de implementação de métodos de reconhecimento de padrões, em específico os *clusterings*, os quais implementam técnicas computacionais que permitem realizar agrupamentos de dados com base em suas semelhanças. Os *clusterings* podem ser aplicados, inclusive, dentro do contexto de imagens, em específico imagens hiperespectrais com uma vasta gama de bandas. Assim, utilizando o algoritmo de *clustering* HDBSCAN por meio da plataforma Colab, junto a um dataset, o algoritmo retornou 9 grupos, dos quais um representa as bandas que não foram classificadas dentro dos demais. Após algumas análises, constatou-se que o algoritmo HDBSCAN é um bom candidato para realizar agrupamentos de bandas em imagens hiperespectrais.

Palavras-chave:

HDBSCAN; Dimensionalidade; *Clustering*.

1. INTRODUÇÃO

Os avanços tecnológicos da atualidade são enormes quando comparados aos de anos anteriores. Isso, também, ocorre nos sistemas de aquisição e armazenamento de dados que, junto à redução dos custos dos dispositivos, geram oportunidades para a implementação de métodos de reconhecimento de padrão. Exemplo disso são os *clusterings*, ou agrupamentos, os quais consistem na implementação de técnicas computacionais com o intuito de separar um determinado conjunto de dados em diferentes agrupamentos utilizando como base suas semelhanças, podendo agrupar diversos tipos de dados, como bandas de imagens hiperespectrais (IHs). Imagens, as quais passaram por desenvolvimentos significativos nas últimas três décadas, surgindo como ferramentas para diversas aplicações civis, ambientais e militares (KHAN et al., 2018).

As imagens são geradas a partir de processos específicos aplicados na energia eletromagnética capturada por sensores (LILLESAND; KIEFER; CHIPMAN, 2004). Tais imagens possuem informações que são importantes sobre as áreas e objetos analisados, e, mediante a isso, uma possível categorização é conforme o número de bandas da mesma, como RGB, multiespectral e hiperespectral (SAQUI; et al., 2016). As IHs têm diversas bandas adicionais em relação ao RGB, possuindo maior abrangência na faixa espectral, aumentando as possibilidades de uso. Exemplo de utilização dessas bandas adicionais é na melhora da precisão de classificação de objetos em imagens (KHAN et al., 2018). Com os avanços mencionados anteriormente, pode-se captar uma maior

¹Discente do Bacharel em Ciência da computação, IFSULDEMINAS – *Campus* Muzambinho. E-mail: luiz.ferlin@alunos.ifsuldeminas.edu.br.

²Orientador, IFSULDEMINAS – *Campus* Muzambinho. E-mail: diego.saqui@muz.ifsuldeminas.edu.br.

abrangência na faixa espectral, as quais são úteis para os mais diversos fins, e, em decorrência desse aumento, há uma maior dimensionalidade dos dados, o que leva à uma alta demanda de recursos computacionais, como armazenamento, transmissão e capacidade de processamento, como citado por Sohaib, Ulhaq e Mushtaq (2013).

Consequentemente, com esse aumento de demanda por recursos computacionais, diversas técnicas de redução de dimensionalidade (RD) surgiram para diminuir a dimensionalidade dos dados, estes que muitas vezes utilizam *clustering* em seus processos. E, embora cada ponto de dado possa ser composto por milhares de características, ainda assim é possível descrevê-lo como uma função de apenas alguns parâmetros subjacentes. Ou seja, os pontos de dados são amostras de uma variedade de baixa dimensão que está contida em um espaço de alta dimensão (CAYTON, 2005).

Exemplo de *clustering* é o *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN). O HDBSCAN é um algoritmo de *clustering* que estende o algoritmo DBSCAN, convertendo-o em um algoritmo de agrupamento hierárquico e, em seguida, usando uma técnica para extrair um *clustering* plano com base na estabilidade dos *clusters* (THE HDBSCAN Clustering Library, 2016). Os grupos advindos dos processos do HDBSCAN podem ser aplicados para diversas finalidades, como reconhecimento de objetos em imagens e técnicas de RD.

Levando em conta todo esse contexto, percebe-se a importância do agrupamento de bandas em IHS de forma eficiente. Com base nisso, buscou-se avaliar se o algoritmo HDBSCAN é uma possibilidade quanto ao agrupamento de bandas em imagens hiperespectrais.

2. MATERIAL E MÉTODOS

Para atingir o objetivo proposto, utilizou-se o arquivo MATLAB (arquivos “.mat”, os quais possuem dados que incluem variáveis, funções, matrizes, dentre outras informações) “Indian Pines³” do Grupo de Inteligência Computacional (GIC). Esse arquivo é um conjunto de dados de segmentação de imagem hiperespectral, com dados que consistem em bandas hiperespectrais de uma única paisagem em Indiana, Estados Unidos, com 145 por 145 pixels. No qual, para cada pixel se tem 220 bandas de reflectância espectral as quais representam porções do espectro eletromagnético nas faixas 0,4-2,5.10-6. Também se utilizou recursos como o Colab da empresa Google, com o qual é possível executar códigos Python em nuvem, sem precisar ter o Python instalado localmente. Além das bibliotecas: pandas, matplotlib, scipy.io, numpy e HDBSCAN.

Dentro do Colab, foi criado um documento para realização dos agrupamentos, no qual o pandas realizou a leitura e estruturação dos dados, assim como foi utilizado na criação de uma nova “tabela” contendo as correlações entre as bandas presentes no arquivo (Figura 1). Essa tabela com

³ https://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Indian_Pines. Acesso em Junho de 2023

as correlações entre as bandas foi utilizada pelo HDBSCAN, o qual retornou agrupamentos baseados na mesma, tais agrupamentos foram utilizados para análise das relações entre bandas e para visualização com o auxílio do scipyy.io.

	0	1	2	3	4	5	6	7	8	9	...	210	211	212
0	1.000000	0.068589	0.077494	0.107796	0.079550	0.078562	0.074408	0.074151	0.069290	0.068191	...	0.000702	-0.001469	-0.000778
1	0.068589	1.000000	0.802970	0.828921	0.834285	0.836317	0.834296	0.831253	0.827655	0.825458	...	0.717857	0.716006	0.716493
2	0.077494	0.802970	1.000000	0.905989	0.925892	0.926113	0.926251	0.924491	0.922550	0.920246	...	0.808956	0.805503	0.808174
3	0.107796	0.828921	0.905989	1.000000	0.951380	0.957841	0.957111	0.956367	0.955261	0.953628	...	0.844471	0.840934	0.843210
4	0.079550	0.834285	0.925892	0.951380	1.000000	0.979337	0.982393	0.981323	0.980954	0.979785	...	0.873524	0.870575	0.872362
...
215	0.001684	0.711827	0.800200	0.834236	0.863362	0.873193	0.882789	0.888257	0.893862	0.897325	...	0.973513	0.972550	0.974418
216	0.004485	0.700086	0.786389	0.820534	0.848067	0.856798	0.867047	0.871481	0.877264	0.880388	...	0.955135	0.954014	0.955831
217	0.006011	0.635472	0.710705	0.736402	0.760972	0.767885	0.776851	0.781136	0.785663	0.788150	...	0.856263	0.855179	0.857815
218	0.002821	0.426339	0.470473	0.482379	0.489972	0.493193	0.496920	0.498921	0.501773	0.502054	...	0.546767	0.546772	0.547137
219	-0.000897	0.400691	0.435987	0.449159	0.458719	0.461754	0.466156	0.467593	0.469585	0.470627	...	0.513174	0.513852	0.513549

20 rows x 220 columns

Figura 1: Tabela de correlações

3. RESULTADOS E DISCUSSÃO

Após a realização do processo de agrupamento com o algoritmo HDBSCAN no arquivo MATLAB “Indian Pines”, chegou-se a 9 agrupamentos, ou grupos, (Figura 2), identificados de -1 a 7, dos quais o identificado por -1 refere-se ao grupo das bandas que não foram classificadas como similares aos demais grupos. Esse grupo em específico, por ser um grupo de objetos que não se encaixam nos demais, acaba por gerar resultados relativamente “estranhos”, exemplo disso é caso seja analisada a correlação de duas bandas dentro deste grupo, têm-se bandas com baixas correlações e outras com altas correlações.

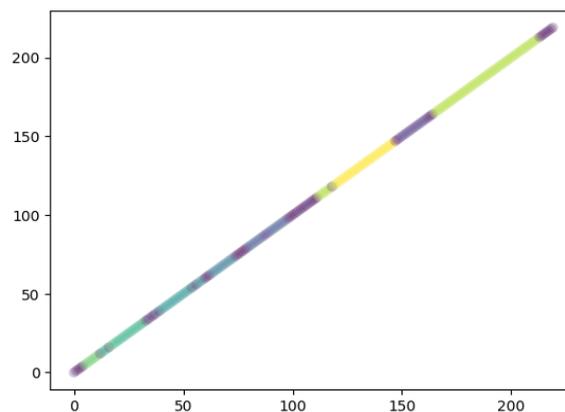


Figura 2: Representação visual dos grupos gerados pelo HDBSCAN

Ao se analisar a Figura 2, percebe-se que há uma dispersão de alguns grupos ao longo dos dados, enquanto outros se concentram em regiões mais delimitadas, como acontece com o grupo 7, que aparece representado pela cor amarela. Pegando esse grupo para análise, pode observar que as correlações entre os elementos deste mesmo grupo são elevadas (não necessariamente entre todas as

demais bandas), e, justamente por elas terem uma alta correlação, caso se queira remover bandas que possuem muitos dados “repetidos” ou muito similares, esse poderia ser um bom indicativo para um processo de RD e até seleção de bandas. Já um indicativo de que se pode preservar determinada banda é o que acontece no grupo -1, onde as bandas não possuem um padrão tão claro para o algoritmo, o que as fez ficarem em grupos diferentes.

Após se analisar todos os grupos, percebe-se que, com exceção do grupo -1, o que ocorreu com o grupo 7 se repete nos demais grupos. Embora, em alguns grupos, a maioria dos valores das bandas do mesmo não tenham valores tão elevados quanto as de outros grupos. Exemplo disso é o grupo 0, onde algumas bandas apresentam algumas correlações abaixo de “0.5”, mesmo tendo, nessa mesma banda, outras correlações superiores a “0.9”. Embora, nesse caso, o grupo 0 esteja mais disperso ao longo das bandas, o que pode explicar essa variação.

4. CONCLUSÃO

O objetivo deste trabalho foi verificar se o algoritmo de clustering HDBSCAN funcionaria no agrupamento de bandas em imagens hiperespectrais. Desta forma, após se realizar a aplicação desse algoritmo de *clustering* com o Indian Pines, pode-se dizer que o HDBSCAN é uma boa opção quando caso se queira realizar agrupamentos de bandas em imagens hiperespectrais, os quais podem ser utilizados para processos como processos de RD. E, junto a isso, deve-se ressaltar que todos os resultados obtidos e coletados por intermédio deste trabalho irão contribuir para a realização de um trabalho de seleção de bandas em imagens hiperespectrais utilizando o HDBSCAN.

REFERÊNCIAS

CAYTON, Lawrence. Algorithms for manifold learning. 15 jun. 2005. Disponível em: <https://cseweb.ucsd.edu/~lcayton/resexam.pdf>.

KHAN, Muhammad; KHAN, Hamid; YOUSAF, Adeel; KHURSHID, Khurram; ABBAS, Asad. Modern Trends in Hyperspectral Image Analysis: A Review. IEEE Xplore, [S. l.], 12 mar. 2018. Disponível em: <https://ieeexplore.ieee.org/document/8314827>.

LILLESAND, T.; KIEFER, R.; CHIPMAN, J. Remote sensing and image interpretation. Wiley, 2004. ISBN 9780471152279. Disponível em: <https://books.google.com.br/books?id=XiydAAAAMAAJ>.

THE HDBSCAN Clustering Library. [S. l.], 2016. Disponível em: <https://hdbscan.readthedocs.io/>.

SAQUI, D.; SAITO, J. H.; C., J. L. A. D.; J., F. E.; C., L. D.; P., H. J. Methodology for band selection of hyperspectral images using genetic algorithms and gaussian maximum likelihood classifier. p. 733–738, Dec 2016.

SOHAIB, M.; HAQ, I. U.; MUSHTAQ, Q. Dimensional reduction of hyperspectral image data using band clustering and selection based on statistical characteristics of band images.