

# EXPLORANDO MÉTODOS NÃO SUPERVISIONADOS PARA DIVERSIFICAÇÃO VISUAL

**Nilson Augusto Magalhães Moraes Neto<sup>1</sup>; Rodrigo Tripodi Calumby<sup>2</sup>**

1. Bolsista PROBIC/UEFS, Graduando em Engenharia de Computação, UEFS, e-mail: [nilson.uefs@gmail.com](mailto:nilson.uefs@gmail.com)

2. Orientador, Departamento de Ciências Exatas, UEFS, e-mail: [rtcalumby@ecomp.uefs.br](mailto:rtcalumby@ecomp.uefs.br)

**PALAVRAS-CHAVE:** Diversidade; Aprendizado não supervisionado; Recuperação de Imagens.

## INTRODUÇÃO

Tradicionalmente, sistemas de recuperação de imagem por conteúdo estão sendo aprimorados utilizando mecanismos de aprendizado de máquina aliados à realimentação de relevância (Calumby et al. 2014), por meio da qual o usuário julga os resultados apresentados, permitindo com o que o sistema aprenda e apresente melhores resultados a cada interação.

Em vários cenários, tão importante quanto a relevância dos itens é a diversidade do resultado. No processo de recuperação por conteúdo, dadas as intenções de busca do usuário, não necessariamente os itens relevantes são aqueles que possuem maior valor de relevância em relação à consulta. Informações relevantes podem pertencer a itens com conceitos diferentes (Sanderson et al. 2009). A diversidade no resultado é essencial para que o sistema aprenda de forma mais precisa o que é realmente relevante na busca, pois permite ao usuário uma melhor visão dos possíveis itens obtidos como resultado. Porém, em contrapartida, pode também mostrar itens irrelevantes. Métodos de reranqueamento (Carbonell & Goldstein, 1998) e técnicas de agrupamento (Calumby et al. 2014) têm sido empregados com intuito de promover esta diversidade.

Em recuperação de informação, a técnica Floresta de Caminhos Ótimos (OPF, do inglês *Optimum-Path Forest*) (Papa, 2009) tem sido aplicada tanto em realimentação de relevância (supervisionada)(da Silva et al. 2012) quanto em agrupamento de dados (Rocha et al. 2009)(não supervisionada), mas não com o foco em promoção de diversidade. Esta técnica, baseada na construção de árvores de caminhos ótimos, visa conectar itens, criando grupos conceitualmente homogêneos. Consequentemente, estes grupos podem ser associadas às diferentes intenções de busca de um usuário ou diferentes interpretações que uma mesma consulta possa ter.

O objetivo deste trabalho é avaliar métodos não supervisionados em busca interativa com promoção de diversidade, comparando e analisando seus resultados em diferentes modalidades de busca.

## METODOLOGIA

Os métodos implementados e analisados, OPF (não supervisionado), Maximal Marginal Relevance (MMR) (Carbonell & Goldstein, 1998) e k-Medoids (k-MED) (Kaufman & Rousseeuw, 1987), foram avaliados por meio de experimentos baseados em um conjunto de imagens heterogêneas, que simula uma atividade de busca de imagens na web. Para isso, usamos a coleção de imagens do ImageCLEF Photographic Retrieval Task (Thomas et al. 2009) com um total de 20.000 imagens. Cada imagem da

coleção é acompanhada de um documento que contém um identificador único, título, descrição em texto livre do conteúdo semântico e visual, notas adicionais, fornecedor, local e data. Toda a avaliação foi realizada utilizando as métricas Recall, Cluster Recall e F1.

Na OPF (não supervisionada), dado um grafo ponderado, é calculado o peso de um nó através dos nós vizinhos. O nó que possuir maior peso em determinada região deste grafo é definido como a raiz de uma árvore. As raízes possuem uma zona de influência na qual “conquistam” os nós não-raízes afim de se obter uma árvore de caminhos ótimos. Esta zona de influência é delimitada pela quantidade máxima de nós ( $k$ ) que uma árvore pode ter. As árvores obtidas são denominadas grupos e a partir destes grupos é realizada diversificação do resultado por meio da seleção de imagens representativas de cada grupo.

No método de partição k-MED, inicialmente é definido um número de agrupamentos ( $n$ ), e arbitrariamente são selecionadas imagens nas quais representarão cada aglomerado, e a partir destas imagens é calculada a similaridade entre as demais imagens para a construção dos grupos. A partir destes grupos é contruído o resultado diversificado final selecionando as imagens representativas.

O MMR constitui-se em um re-ranqueamento dos objetos levando em conta um índice entre a relevância e diversidade ( $\lambda$ ). Dado um conjunto de objetos inicialmente recuperado, o re-ranqueamento é feito a partir de uma função na qual se relaciona os objetos já recuperados com os demais objetos da base, assim computa-se um score combinado entre a relevância de uma nova imagem em relação à consulta e a diversidade em relação às imagens previamente selecionadas.

Para a comparação dos métodos foram simuladas 39 consultas, cada uma com 20 iterações do usuário, sendo mostradas 20 imagens para o usuário a cada iteração. A consulta se constitui por imagens e/ou textos utilizados na busca de um usuário. Abaixo na Figura 1, podemos observar a organização da máquina, desde quando o usuário faz a consulta (passo 1). Em seguida, é mostrado um conjunto de imagens no qual ele vai definir qual é relevante ou irrelevante (passo 2 e 3). Com essas informações o sistema treina, classifica (passo 4) as imagens em dois tipos (relevantes ou irrelevantes) utilizando a técnica da OPF supervisionada, organiza os resultados (passos 5 e 6), ranqueia as relevantes (passo 7), diversifica os resultados com os métodos analisados (passo 8) e por fim, apresenta novamente outro conjunto de imagens.

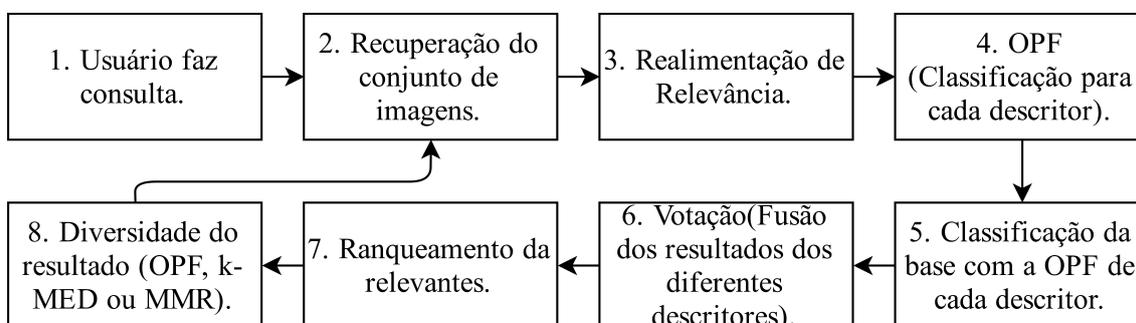


Figura 1. Funcionamento da máquina de busca interativa com diversidade.

Para a avaliação, os resultados obtidos com os métodos foram comparados entre si. As avaliações foram feitas à partir da técnica de full freezing, que é a concatenação

das listas de imagens mostradas ao usuário nas suas iterações. Utilizamos as 200 primeiras imagens mostradas ao usuário como resultado para qualificar os métodos.

Neste trabalho, utilizamos oito medidas de similaridade que também foram utilizados em Calumby (2014). As medidas textuais foram a Cosseno e BM25. Os descritores visuais foram para cor: BIC, GCH, JAC e para textura: CCOM, LAS e QCCH.

## RESULTADOS E DISCUSSÃO

O objetivo da tarefa de busca em estudo é obter um resultado que seja tanto diverso quanto preciso. Os experimentos foram divididos em três modalidades de busca, as quais são definidas como: somente texto (TXT), somente visual (VIS) e a busca multimodal (MM), que leva em consideração tanto o textual quanto o visual, para que pudéssemos observar a eficiência de cada método para as diferentes modalidades.

Além da variação da modalidade, para cada método, utilizou-se variações de parâmetros para comparação dos resultados. Para o método k-MED, utilizou-se o número de agrupamentos (n) em 20 ou 25 grupos. Para o MMR, o fator de importância da diversidade ( $\lambda$ ), que é um índice entre a relevância e diversidade do resultado, foi definido em 0,3; 0,5 e 0,7. Para a OPF fixou-se o número máximo de nós da árvore (k) em 3. Em todos os métodos, os níveis de profundidade de reranqueamento/agrupamento (d) utilizados foram 1,5; 2 e 3, que correspondem a 30, 40 e 60 imagens. Feita a análise dos resultados, constatou-se que os métodos avaliados apresentaram melhores resultados na modalidade MM.

Para objetivar a análise, os resultados apresentados levam em conta as melhores combinações de parâmetros por modalidade de cada método, para Recall@200: OPF (TXT: k=3, d=1,5; VIS: k=3, d=1,5; MM: k=3, d=1,5); k-MED (TXT: n=25, d=3; VIS: n=20, d=1,5; MM: n=25, d=3); MMR (TXT:  $\lambda=0,3$ , d=3; VIS:  $\lambda=0,3$ , d=3; MM:  $\lambda=0,3$ , d=3). Para CR@200: OPF (TXT: k=3, d=1,5; VIS: k=3, d=2; MM: k=3, d=2); k-MED (TXT: n=20, d=3; VIS: n=25, d=1,5; MM: n=25, d=3); MMR (TXT:  $\lambda=0,7$ , d=3; VIS:  $\lambda=0,7$ , d=3; MM:  $\lambda=0,5$ , d=1,5).

Na Tabela 1 podemos observar os resultados obtidos por cada método. Os números em negrito significam o melhor resultado. A Recall dá-se pela quantidade de imagens relevantes encontradas em um determinado resultado em relação ao total de imagens relevantes existentes na coleção. A CR é obtida através do número de sub-consultas cobertas nos primeiros N itens, ou seja, a diversidade do resultado. Por sua vez, a F1 é a média harmônica das duas medidas anteriores.

	TXT			VIS			MM		
	OPF	k-MED	MMR	OPF	k-MED	MMR	OPF	k-MED	MMR
Recall	0,1688	<b>0,2193</b>	0,2088	0,1892	0,2005	<b>0,2541</b>	0,2418	0,2504	<b>0,3063</b>
CR	0,4328	<b>0,4877</b>	0,4611	0,4149	0,4104	<b>0,4563</b>	0,5081	0,5706	<b>0,5766</b>
F1	0,2429	<b>0,3025</b>	0,2874	0,2600	0,2694	<b>0,3264</b>	0,3277	0,3481	<b>0,4001</b>

Tabela 1. Resultados dos melhores métodos por modalidade.

Contatou-se que o método de reranqueamento foi, para as modalidades VIS e MM, superior aos métodos baseados em agrupamento, o que pode estar associado à capacidade do método em melhor utilizar os critérios de similaridade e a dificuldade em

se construir grupos coerentes utilizando a informação conceitual de alto nível presente no texto.

Por outro lado, observamos que para a modalidade textual o método de agrupamento k-MED obteve os melhores resultados. O que sugere uma melhor capacidade de criação de grupos utilizando a informação textual. Entretanto, o método OPF não teve desempenho satisfatório e novos experimentos estão previstos como descrito na próxima seção.

## CONSIDERAÇÕES FINAIS

O método MMR se mostrou superior aos demais na base de dados utilizadas. Além disso, observamos também que a multimodalidade na busca é crucial para melhores resultados, independente do método proposto. O projeto continua em andamento, com perspectivas otimistas e propostas de otimização no processo e nos resultados, incluindo os testes em uma nova base de dados e um melhor ajuste ou ajuste automático no parâmetro da OPF. Uma avaliação mais rigorosa será realizada, incluindo a execução de testes de significância estatística.

## REFERÊNCIAS

- SANDERSON, M. et al. 2009. What else is there? Search diversity examined. In Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, ECIR '09, pages 562–569, Berlin, Heidelberg. Springer-Verlag.
- CALUMBY, R.T.; GONÇALVES, M. A. e TORRES, R. da S. 2014. Diversity-driven Learning for Multimodal Image Retrieval with Relevance feedback. In IEEE ICIP, pages 2197-2201, Paris, France.
- CARBONELL, J. & GOLDSTEIN, J. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In SIGIR, pages 335–336.
- CALUMBY, R.T. et al. 2014. Recod @ MediaEval 2014: Diverse Social Images Retrieval. Diverse Social Images Retrieval. Working notes papers.
- PAPA, J. P. et al. 2009. Supervised pattern classification based on optimum-path forest. *Int. J. Imaging Syst. Technol.* 19(2), 120.131.
- DA SILVA, A. T. et al. 2012. Incorporating multiple distance spaces in optimum-path forest classification to improve feedback-based learning. *Comput. Vis. Image Underst.* 116(4), 510-523.
- ROCHA, L. M. et al. 2009. Data clustering as an optimum-path forest problem with applications in image analysis. *Int. J. Imaging Syst. Technol.* 19(2), 50-68.
- THOMAS, A.; PAUL, C.; SANDERSON, M. and GRUBINGER, M. 2009. Overview of the ImageCLEFphoto 2008 photographic retrieval task. In *Evaluating Systems for Multilingual and Multimodal Information Access*, volume 5706 of Lecture Notes in Computer Science, pages 500 – 511. Springer Berlin / Heidelberg.
- CALUMBY, R. T. et al. 2015 RECOD @ Placing Task of MediaEval 2015. In *Proc. MediaEval*.
- KAUFMAN, L.; ROUSSEEUW P. J. 1987. Clustering by means of Medoids, in *Statistical Data Analysis Based on the L1-Norm and Related Methods*. North-Holland, páginas 405–416.