Refinamento do Conjunto Inicial de Resultados baseado em Contexto para Recuperação Interativa de Imagens*

Luciano Araujo Dourado Filho[1], **Rodrigo Tripodi Calumby**[1]

[1]Universidade Estadual de Feira de Santana – Feira de Santana – BA – Brazil

lucianoadfilho@ecomp.uefs.br, rtcalumby@uefs.br

Abstract. *This paper proposes the exploitation of contextual information to refine the initial set of results in interactive image retrieval systems. This context-based method differs from traditional approaches given it considers the rich information present in the relationships between multiple images, rather than simply using traditional methods that computes pairwise distances between two images. In this work, an experimental evaluation of the effectiveness of this technique in different Web search scenarios was carried out. The results showed that the evaluated method was able to significantly improve the effectiveness of the retrieval, specially when applied in conjunction with similarity measures of greater distinctive power.*

Resumo. *Este trabalho propõe a exploração de informações contextuais para refinamento do conjunto inicial de resultados em sistemas de recuperação interativa de imagens. Este método baseado em contexto difere das abordagens tradicionais, pois considera a rica informação presente nas relações entre múltiplas imagens, ao invés de realizar o cômputo de similaridade apenas entre pares de imagens. Neste trabalho, realizou-se uma avaliação experimental da eficácia desta técnica em diferentes cenários de busca na Web. Os resultados mostraram que o método avaliado foi capaz aprimorar significativamente a eficácia da recuperação, especialmente quanto aplicado em conjunto com medidas de similaridade de maior poder distintivo.*

1. Introdução

O avanço das tecnologias de captura e armazenamento de conteúdo digital proporcionou o aumento da quantidade de dados como: imagens, vídeos, *e-mails*, documentos, entre outros. Isso gera demanda por técnicas para sua exploração de forma eficiente e eficaz tendo em vista sua quantidade e disponibilidade crescentes. Esses dados podem ser utilizados para fins diversos, como: medicina, análise forense, estudos da biodiversidade, redes sociais, bibliotecas digitais, entre outros. Assim, a qualidade das técnicas para sua exploração é de fundamental importância na disponibilização de acesso prático e eficaz.

Quando tratam-se de imagens por exemplo, os principais paradigmas de busca consideram informações textuais associadas (anotações, metadados e palavras-chave) para realizar a recuperação a partir de uma base de dados [Baeza-Yates and Ribeiro-Neto 2011]. Outra possibilidade é efetuar a representação das propriedades visuais das imagens, como cor, forma e textura por meio da extração de características baseadas por

*Este trabalho foi desenvolvido com o apoio do PIBIC/CNPq (processo nº 165871/2017-2).

exemplo em *pixels*, segmentos de borda, regiões ou objetos presentes nas imagens. Isso possibilita que, dada uma imagem de consulta, as imagens da base possam ser comparadas por similaridade por meio de uma função de distância e que sejam ranqueadas e retornadas para o usuário [Torres and Falcão 2006]. Para isso, utilizam-se os descritores de imagens[1].

A fim de melhor definir a similaridade entre as imagens de uma base de dados pode-se aplicar técnicas de re-ranqueamento cujo propósito é refinar as distâncias entre elas para torná-las mais adequadas. A motivação por trás disso é que ao melhorar a qualidade da medida de similaridade entre as imagens da base o processo de recuperação torna-se mais eficaz. Isso aumenta a possibilidade de satisfação na experiência de um usuário com o sistema, já que a quantidade de interações que ele precisará realizar para satisfazer suas necessidades tende a ser menor.

Uma abordagem para aprimoramento da eficácia de sistemas de recuperação de imagens é permitir que os usuários interajam com o sistema de modo a indicar a relevância dos itens exibidos nos resultados obtidos para uma consulta, processo chamado de Realimentação de Relevância. Assim, o usuário alimenta o sistema indicando a relevância dos itens encontrados em cada iteração, a partir disso o sistema aprende quais propriedades melhor caracterizam as imagens relevantes e adapta-se para atender às necessidades do usuário [Calumby et al. 2014]. Dessa forma, se o conjunto inicial de resultados apresenta imagens relevantes, a tendência é que o usuário possa fornecer informação de qualidade para o sistema e, assim, os resultados ao longo das iterações subsequentes sejam mais satisfatórios [Calumby et al. 2017]. Por outro lado, abordagens tradicionais de recuperação de imagens produzem resultados baseados no cômputo de similaridade apenas entre pares de imagens, deixando de explorar a informação existente nas relações entre elas [Pedronette et al. 2014]. Alternativamente, é possível combinar técnicas de re-ranqueamento com a realimentação de relevância, a fim de propagar as melhorias na eficácia dos *rankings* iniciais ao longo das iterações de *feedback*.

Neste contexto, este trabalho avalia experimentalmente um método baseado em espaços contextuais que explora as relações entre múltiplas imagens a fim de refinar o computo de distâncias entre elas. Para isso, considera-se uma base de dados heterogênea para avaliar a eficácia de sistemas de recuperação de imagens. O objetivo principal está em estimar o impacto dessa técnica sobre a qualidade do conjunto inicial de resultados considerando o cenário de buscas de imagens na *Web*.

2. Trabalhos Correlatos

Em diversas aplicações, utilizar apenas uma função de distância par-a-par para definir a similaridade entre duas imagens pode não ser uma abordagem eficaz, tendo em vista que deixa-se de explorar as relações existentes entre múltiplas imagens de um banco de dados. Dessa forma, em [Pedronette et al. 2014], os autores propõem uma alternativa para melhor computar as distâncias entre imagens por meio da exploração das informações nas múltiplas relações entre elas em um processo de refinamento iterativo baseado em kNN[2].

De forma simplificada, pode-se descrever o modelo de refinamento por meio de

[1]Par composto por uma função de extração de características de uma imagem e uma função para computar a distância entre duas imagens a partir das suas respectivas características.

[2]k vizinhos mais próximos, do inglês *k Nearest Neighbors*.

dois componentes, d_i e d_j, que representam, respectivamente: a distância entre duas imagens cuja similaridade deseja-se redefinir e a distância do contexto de uma imagem em relação à outra. A partir disso, o grau de similaridade entre as duas imagens é então redefinido conforme a Equação 4. Este processo é formalizado como segue:

A partir de conjunto de imagens $\mathcal{I} = (Img_1, Img_2, Img_3, ..., Img_m)$, define-se:

1. $k \in \mathbb{R}_+^*$, que define o tamanho do contexto (número de vizinhos).
2. $\mathcal{R}ank(Img_i, Img_j)$ como a posição de Img_j numa lista de imagens ordenada por similaridade em relação à Img_i, sendo $i \neq j$.
3. $kNN(Img_r) = \{\ Img_r \in \mathcal{I} \mid \mathcal{R}ank(Img_r, Img_s) < k\ \}$.
4. $kNN(Img_y)_i$ como a i-ésima imagem pertencente ao conjunto do $kNN(Img_y)$.
5. $\delta(Img_A, Img_B)$ como uma função de distância para imagens A e B.

Inicialmente determina-se o valor do componente d_i (Equação 1). Em seguida, calcula-se α (Equação 2) e então o componente d_j (Equação 3). Por fim, a nova distância $\delta_t^{'}(Img_A, Img_B)$ pode ser obtida (Equação 4). O processo é realizado de maneira iterativa com k variando de k_o a k_f para definir o número de vizinhos a ser considerado em cada iteração t, sendo as distâncias reajustadas a partir daquelas da iteração anterior. A intuição por trás disso é que a eficácia dos *rankings* aumente ao longo das iterações de modo que as imagens irrelevantes sejam afastadas das primeiras posições e que o valor de k seja incrementado para levar em consideração mais imagens [Pedronette et al. 2014].

$$d_i = \frac{\delta(Img_A, Img_B)}{k} \tag{1}$$

$$\alpha = \sum_{i=0}^{k} \delta(kNN(Img_A)_i, Img_B) \times (k - i) \tag{2}$$

$$d_j = \alpha / \frac{k \times (k-1)}{2} \tag{3}$$

$$\delta_t^{'}(Img_A, Img_B) = \sqrt{d_i^2 + d_j^2} \tag{4}$$

3. Metodologia Experimental

Para realizar os experimentos, utilizaram-se os mesmos parâmetros propostos em [Pedronette et al. 2014], onde se realizou um processo iterativo determinado pelos valores inicial e final de k (k_o, k_f). Dessa forma, os parâmetros utilizados para k foram: $(k_o, k_f) = \{(1, 10), (2, 10), (3, 10), (4, 10), (5, 10)\}$.

A base de imagens utilizada foi a da *ImageCLEF Photographic Retrieval Task* [Arni et al. 2009], composta por 20.000 imagens de diversos locais ao redor do mundo contendo paisagens, cidades, animais, pessoas, dentre outros. A base inclui 39 consultas e *ground-truth* para avaliação. Nesse contexto, para avaliar o impacto nos resultados, diversas medidas de eficácia foram consideradas: *Precision@20* (P20), *Recall@20* (R20), *Mean Average Precision* (MAP), MAP@20 (MAP20), *Normalized Discount Cumulative*

Gain@20 (NDCG20), *Binary Preference* (BPREF), *Geometric MAP* (GMAP) e *Mean Reciprocal Rank* (MRR). P20 representa a fração de imagens relevantes retornadas em relação ao total de imagens retornadas; R20 representa a fração de imagens relevantes retornadas em relação ao total de imagens relevantes existentes; *Average Precision* (AP) é a média dos valores de *Precision* da curva *Precision*x*Recall* e MAP é o valor médio de AP para múltiplas consultas; BPREF baseia-se numa relação de preferência, penalizando a ocorrência de imagens irrelevantes acima das relevantes; MRR é dado pela média do inverso da posição da primeira imagem relevante no *ranking* de cada consulta.

Os experimentos foram conduzidos a partir do arcabouço de realimentação de relevância apresentado em [Calumby et al. 2014], para isso avaliou-se os *rankings* nas primeiras 20 posições dentre as 1000 retornadas como proposto pelo desafio em [Arni et al. 2009] além do ranking total. Os experimentos foram realizados apenas na modalidade textual, sendo que o texto utilizado descreve o conteúdo semântico das imagens. Foram utilizadas as seguintes medidas de similaridade: Jaccard [Lewis et al. 2006], Cosine [Baeza-Yates and Ribeiro-Neto 2011], BM25 [Lewis et al. 2006], TF-IDF-Sum [Santos et al. 2009], e Dice [Lewis et al. 2006]. Os experimentos foram avaliados com o gabarito fornecido junto à coleção e comparados com o *baseline*, baseado no *ranking* gerado sem a otimização por contexto e dado pela menor distância entre cada imagem do padrão de consulta e as imagem da base.

4. Resultados e Discussões

A qualidade dos *rankings* obtidos com o método proposto foi comparada em relação aos *rankings* obtidos pelos descritores originais sem ajustes das distâncias. Os resultados são apresentados nas Tabelas 1, 2, 3, 4 e 5 (melhores resultados estão em negrito e ganhos são computados do melhor contexto sobre o *baseline*). As tabelas mostram que os descritores utilizados não apresentaram ganhos em termos de GMAP, e que com as medidas JACCARD, DICE e TF-IDF de modo geral, também não houveram ganhos expressivos. Acredita-se que estes resultados estejam associados à simplicidade destas medidas, tendo em vista que, de modo geral, consideram apenas a ocorrência ou não dos termos de consulta nas descrições das imagens para computar a similaridade entre elas.

Diferentemente, as medidas COSINE e BM25, baseadas respectivamente em modelos mais complexos de frequência de termos e abordagem probabilística, apresentaram ganhos superiores de forma consistente para todas as medidas (exceto GMAP). Em termos de P20, o descritor BM25 apresentou resultado inferior quando comparado ao COSINE com ganho relativo de 1,3%, contra 9,3%. Já em termos de R20, ambos apresentaram resultados similares, com ganhos de 7,3% em relação ao baseline. Em termos de BPREF os ganhos foram bastante expressivos, sendo 15,8% (BM25) e 27,6% (COSINE). Considerando MRR, os resultados foram de 18,7% (BM25) e 26,4% (COSINE). As medidas MAP, MAP20 E NDCG20 também indicaram resultados satisfatórios em relação ao baseline, e.g., com MAP20 indicando ganho de 7,4% com COSINE e 13,8% com BM25.

De modo geral, observou-se que, no que diz respeito ao *ranking* inicial, os resultados com utilização do refinamento baseado em contexto alcançaram ganhos relativos médios em termos de *Precision@20* de aproximadamente de 2,8%, 3,0% em *Recall@20*, 4,0% em MAP, e 2,7% em NDCG, em comparação aos resultados sem otimização por contexto. Já em relação ao posicionamento das imagens relevantes ao longo dos *ran-*

kings (BPREF), observou-se um ganho médio relativo aproximado de 10,3% em relação ao baseline.

Tabela 1. Resultados obtidos para JACCARD.

JACCARD								
	P20	R20	MAP20	NDCG20	BPREF	MAP	GMAP	MRR
Baseline	0.1758	0.0706	0.0555	0.1883	0.1362	0.1216	0.0081	0.2687
Contexto: 1-10	0.1758	**0.0706**	**0.0555**	0.1883	0.1362	0.1216	**0.0081**	0.2687
Contexto: 2-10	**0.1795**	0.0677	0.0548	**0.1894**	0.1343	0.1168	0.0040	0.2480
Contexto: 3-10	0.1718	0.0639	0.0531	0.1868	0.1382	0.1195	0.0042	0.2575
Contexto: 4-10	0.1731	0.0657	0.0538	0.1866	0.1431	**0.1232**	0.0049	**0.2707**
Contexto: 5-10	0.1731	0.0653	0.0501	0.1786	**0.1442**	0.1223	0.0052	0.2335
Ganho Relativo	**2.1%**	0.0%	0.0%	**0.6%**	**5.9%**	**1.3%**	0.0%	**0.7%**

Tabela 2. Resultados obtidos para TF-IDF.

TF-IDF								
	P20	R20	MAP20	NDCG20	BPREF	MAP	GMAP	MRR
Baseline	0.1833	0.0807	0.0556	0.1939	0.1514	0.1332	0.0196	0.3240
Contexto: 1-10	**0.1897**	**0.0835**	**0.0572**	**0.1984**	0.1559	**0.1347**	**0.0196**	**0.3240**
Contexto: 2-10	0.1808	0.0793	0.0509	0.1825	0.1553	0.1243	0.0129	0.2803
Contexto: 3-10	0.1718	0.0720	0.0399	0.1703	**0.1608**	0.1239	0.0111	0.2683
Contexto: 4-10	0.1654	0.0689	0.0403	0.1703	0.1607	0.1313	0.0105	0.2938
Contexto: 5-10	0.1641	0.0664	0.0342	0.1609	0.1562	0.1272	0.0105	0.2473
Ganho Relativo	**3.5%**	**3.5%**	**2.9%**	**2.3%**	**6.2%**	**1.1%**	0.0%	0.0%

Tabela 3. Resultados obtidos para DICE.

DICE								
	P20	R20	MAP20	NDCG20	BPREF	MAP	GMAP	MRR
Baseline	0.1756	0.0706	0.0560	0.1890	0.1359	0.1229	0.0084	0.2690
Contexto: 1-10	0.1756	**0.0706**	**0.0560**	**0.1890**	0.1380	0.1229	**0.0084**	**0.2732**
Contexto: 2-10	0.1756	0.0661	0.0548	0.1885	0.1352	0.1185	0.0042	0.2606
Contexto: 3-10	0.1731	0.0646	0.0536	0.1871	0.1385	0.1204	0.0044	0.2555
Contexto: 4-10	**0.1769**	0.0668	0.0539	0.1879	**0.1442**	0.1241	0.0050	0.2713
Contexto: 5-10	0.1718	0.0647	0.0507	0.1804	0.1458	**0.1244**	0.0054	0.2507
Ganho Relativo	**0.7%**	0.0%	0.0%	0.0%	**6.1%**	**1.2%**	0.0%	**1.6%**

Tabela 4. Resultados obtidos para BM25.

BM25								
Método	P20	R20	MAP20	NDCG20	BPREF	MAP	GMAP	MRR
	0.2051	0.0719	0.0426	0.1869	0.1517	0.1291	0.0244	0.2965
Contexto: 1-10	0.2051	0.0763	0.0464	0.2054	0.1517	0.1384	0.0244	0.3494
Contexto: 2-10	**0.2077**	0.0807	0.0482	0.2118	0.1560	0.1346	0.0167	0.3464
Contexto: 3-10	0.1897	**0.0819**	0.0506	0.1953	0.1595	0.1349	0.0191	0.3369
Contexto: 4-10	0.1821	0.0801	0.0511	0.1922	0.1688	0.1412	0.0231	0.3620
Contexto: 5-10	0.1808	0.0790	**0.0528**	**0.2118**	**0.1757**	**0.1485**	**0.0244**	**0.3620**
Ganho Relativo	**1.3%**	**7.3%**	**13.8%**	**3.1%**	**15.8%**	**7.3%**	0.0%	**18.7%**

Tabela 5. Resultados obtidos para COSINE.

COSINE								
	P20	R20	MAP20	NDCG20	BPREF	MAP	GMAP	MRR
Baseline	0.1654	0.0654	0.0379	0.1701	0.1240	0.1157	0.0280	0.2727
Contexto: 1-10	**0.1808**	0.0701	0.0406	0.1870	0.1485	**0.1205**	**0.0280**	0.3133
Contexto: 2-10	0.1731	**0.0702**	**0.0407**	0.1770	0.1457	0.1153	0.0240	0.3094
Contexto: 3-10	0.1769	0.0667	0.0381	0.1852	**0.1582**	0.1155	0.0235	0.3069
Contexto: 4-10	0.1782	0.0687	0.0376	**0.1878**	0.1568	0.1153	0.0225	0.3319
Contexto: 5-10	0.1667	0.0642	0.0333	0.1776	0.1489	0.1097	0.0222	**0.3446**
Ganho Relativo	**9.3%**	**7.3%**	**7.4%**	**10.4%**	**27.6%**	**4.1%**	0.0%	**26.4%**

5. Conclusões e Trabalhos Futuros

Este trabalho realizou uma análise da eficácia do método de Informações Contextuais sobre o conjunto inicial de resultados em cenários de busca de imagens na *Web*. Os resultados encontrados indicaram ganhos satisfatórios quando comparados ao *baseline*. Além disso, foi possível identificar o impacto direto da eficácia das medidas de similaridade textual em capturar as relações de similaridade entre as imagens, sobre o refinamento baseado em contexto. Considerando a aplicação com sucesso do método com ganhos de eficácia nos resultados iniciais de busca, acredita-se que isto gere impacto direto na qualidade da busca de sistemas interativos. Maior relevância nos primeiros resultados exibidos ao usuário permitem melhor troca de informação com o sistema e consequentemente uma melhor captura das intenções de busca do usuário. Neste sentido, trabalhos futuros podem ser realizados para validação experimental do impacto em múltiplas interações.

Referências

Arni, T., Clough, P., Sanderson, M., and Grubinger, M. (2009). Overview of the imageclefphoto 2008 photographic retrieval task. In *Evaluating Systems for Multilingual and Multimodal Information Access*, pages 500–511. Springer Berlin Heidelberg.

Baeza-Yates, R. and Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology Behind Search*. Addison-Wesley, USA, 2nd edition.

Calumby, R. T., da S. Torres, R., and Gonçalves, M. A. (2014). Multimodal retrieval with relevance feedback based on genetic programming. *Multimed Tools Appl*, (69):991–1019.

Calumby, R. T., da S. Torres, R., and Gonçalves, M. A. (2017). Diversity-based interactive learning meets multimodality. *Neurocomputing*, (259):159–175.

Lewis, J., Ossowski, S., Hicks, J., Errami, M., and Garner, H. R. (2006). Text similarity: an alternative way to search MEDLINE. *Bioinformatics*, 22(18):2298–2304.

Pedronette, D. C. G., da S. Torres, R., and Calumby, R. T. (2014). Using contextual spaces for image re-ranking and rank aggregation. *Multimed Tools Appl*, 69(3):689–716.

Santos, K., de Almeida, H. M., Gonçalves, M. A., and da Silva Torres, R. (2009). Recuperação de imagens da web utilizando múltiplas evidências textuais e programação genética. In *XXIV SBBD*, pages 91–105.

Torres, R. and Falcão, A. X. (2006). Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, (161-185):13(2).