

Diversificação de Imagens baseada em Agrupamento Adaptativo por Consulta

José Solenir Lima Figuerêdo¹ e Rodrigo Tripodi Calumby²

1. Bolsista PIBIC/CNPq, Graduando em Engenharia da Computação, UEFS, e-mail: solenir.figueredo@gmail.com

2. Orientador, Departamento de Ciências Exatas, UEFS, e-mail: rcalumby@uefs.br

PALAVRAS-CHAVE: Agrupamento; Relevância; Diversificação.

INTRODUÇÃO

À medida que a tecnologia avança, grandes transformações acontecem, sejam em âmbito puramente social ou tecnológico. Nesse contexto, observa-se que as imagens têm impactado de forma direta em muitas dessas transformações. As imagens são utilizadas em contextos diversos, como em sistemas de medicina, de biodiversidade e bibliotecas digitais (Torres & Falcao, 2006). Desta forma, ao longo dos anos, muito tem sido feito para melhorar a eficácia com a qual essas imagens são recuperadas e analisadas. Uma dessas técnicas é a chamada recuperação de imagens por conteúdo (Veltkamp & Tanase, 2002). Essas técnicas, basicamente, tentam recuperar imagens semelhantes à uma especificação ou padrão definido pelo usuário (por exemplo, esboço de forma, uma imagem modelo) (Torres & Falcao, 2006).

O processo de recuperação de informação exige que determinados aspectos sejam observados com cautela, como ambiguidade, redundância, relevância e diversidade. Além disso, as imagens que devem ser exibidas para um determinado usuário são as imagens consideradas relevantes, aquelas que oferecem informações úteis. Contudo, embora a utilização da relevância como critério seja eficaz, tem-se observado que em certas situações o seu uso não satisfaz por completo as necessidades de consulta que demandam diversidade visual, por exemplo, Chang *et al* (2016); Chang & Wang (2016); Fan *et al* (2008). Uma solução comumente explorada para amenizar esse problema é a utilização de técnicas de agrupamento de dados, que visa encontrar grupos de objetos que tenham certa semelhança, sem levar em consideração informações prévias sobre os dados existentes. Todavia, muitos algoritmos demandam um valor de referência para determinar o número de grupos a serem gerados.

A determinação do número de grupos é uma tarefa que exige esforço, pois envolve um conjunto de propriedades e características das imagens. Trabalhos anteriores tentaram encontrar um número fixo de grupos independente da consulta a ser executada (Ferreira *et al*, 2016); (Tollari, 2016), ao invés de utilizar métodos adaptativos. Em outros trabalhos, mesmo simulando-se o número exato de clusters para cada consulta (com base no ground-truth), os resultados não foram satisfatórios (Araujo, 2016). Diante disso, este estudo buscou formular uma abordagem que auxiliasse na detecção automática do número de grupos, adaptável a cada consulta.

METODOLOGIA

Considerando um conjunto de descritores, buscamos identificar o número de grupos mais adequado a serem formados, utilizando um intervalo de possíveis valores de k . Este processo é feito para cada consulta, deste modo, o número de grupos é escolhido de modo específico para cada uma, em tempo de execução. Para efeitos de comparação, fez-se um outro agrupamento, mas agora utilizando um mesmo número de grupos para todas as consultas, assim como é comumente realizado na literatura. Com isso, pudemos comparar a eficácia das duas abordagens.

Antes de executar o agrupamento, realizou-se para cada consulta uma filtragem das imagens de entrada, baseada na medida de similaridade textual BM25. Para isso, a lista de entrada foi ranqueada considerando as imagens mais relevantes, e ao término deste processo, selecionou-se as 150 primeiras imagens. Com as imagens mais relevantes separadas, executou-se o algoritmo de agrupamento *KMedoids* (Jin & Han, 2011), variando-se o k entre 15 e 25. Escolheu-se este intervalo por cobrir variações positivas e negativas em relação ao número de grupos comumente utilizado na literatura. Após realizar o agrupamento para todas as possibilidades do intervalo, aplicou-se as medidas de avaliação, aqui chamadas de medidas de qualidade. A escolha do melhor número de grupo é feita utilizando cada medida de qualidade separadamente. Nos experimentos, várias medidas de qualidade foram avaliadas, sendo elas: índice de *Davies-Bouldin*, abreviado para DB, o coeficiente Silhueta, índice de *Dunn*, erro quadrático, e índice *Xie-Beni*, aqui chamado de índice XB.

Como base de comparação para o método proposto, fizemos a execução usando um mesmo número de grupos para todas as consultas. Escolheu-se $k = 20$, tendo em vista que este valor é tipicamente escolhido, correspondendo à quantidade comumente apresentada na primeira página de resultados de sistemas de busca. Os resultados foram avaliados usando medidas de *Precision* e *Cluster-Recall*. Para realização da análise, levou-se em consideração apenas as 50 primeiras imagens do *ranking*, cuja obtenção ocorreu via *round-robin*.

Os descritores e o algoritmo de agrupamento utilizado neste projeto foram avaliados a partir de uma coleção de imagens do *Retrieving Diverse Social Images Task* do MediaEval 2015 (Ionescu *et al*, 2014). A base utilizada nos experimentos é composta tanto por imagens do conjunto *Devset*, quanto do *Testset*, totalizando 222 localidades. Além disso, a base dispõe de outras informações adicionais, como descrição visual, meta dados e *ground-truth* em relação a diversidade e relevância. O gabarito em questão, utilizado para avaliar a qualidade do agrupamento, foi gerado mediante anotações de usuários especialistas reais (Ionescu *et al*, 2014), avaliando tanto o critério de relevância quanto de diversidade. Na condução dos experimentos foram utilizados 15 descritores, sendo textuais (Cosseto, BM25, Dice e Jaccard) e visuais (ACC, CN3x3, LUM, SCH, CEDD, FCTH, Gabor, JCD, PHOG, Tamura e CNN_AD).

RESULTADOS E DISCUSSÃO

Ao considerarmos o processo de diversificação, observou-se que a diferença de eficácia entre a otimização com as medidas de qualidade e a abordagem com um mesmo número de grupos para todas as consultas, não foram muito expressivas. Os valores alcançados pelas duas abordagens apresentaram variações marginais. Todavia, ao considerarmos a relevância das imagens recuperadas, notou-se que houve uma melhora nos resultados com a abordagem adaptativa. Análises mais detalhadas serão expostas nas seções seguintes. Em razão da limitação de espaço, exibiremos os resultados apenas dos descritores que apresentaram os melhores resultados.

Descritor ACC

Na Figura 1 podemos visualizar o resultado obtido utilizando-se o descritor ACC. Observa-se que os métodos de qualidade propostos obtiveram uma melhoria se comparados com a definição fixa do número de grupos. Isso pode ser fruto da particularidade que cada consulta possui, uma vez que o número de grupos tende a variar de acordo com a distribuição das imagens no espaço de características. Dos métodos de qualidade aplicados, o erro quadrático foi o que resultou em um melhor comportamento, superando os demais métodos em diferentes configurações.

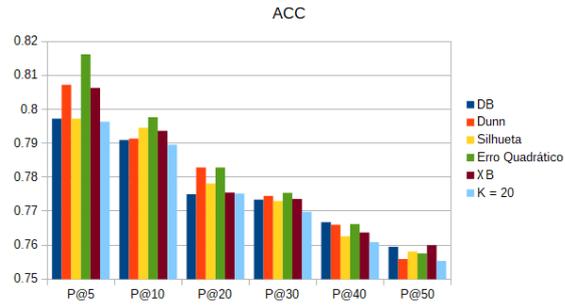


Figura 1. Precisão para o descritor ACC.

Descritor CN3x3

Ao analisar os resultados gerados com o descritor CN3x3, percebemos que a precisão para otimização com as medidas de qualidade foi superior. Isso pode ser verificado a partir da Figura 2. Em especial, para esse descritor, não houve um método predominantemente superior aos demais em todos os níveis do *ranking*. Percebe-se que o XB e o Erro quadrático foram os melhores em praticamente todos os níveis, excetuando-se o (P@5), cuja maior eficácia foi alcançada pelo DB.

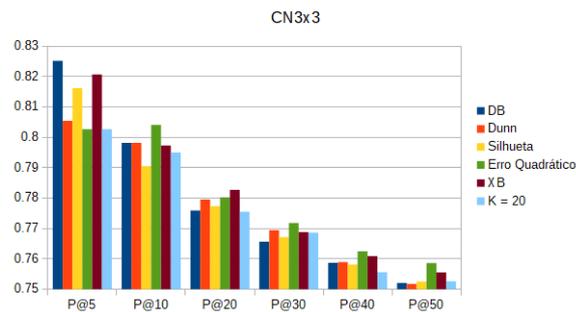


Figura 2. Precisão para o descritor CN3x3.

Descritor COSSENO

Para o caso do COSSENO, Figura 3, o comportamento segue semelhante aos demais descritores. A principal diferença notada é a de que todas as medidas de qualidade, para todos os níveis considerado, apresentaram melhores resultados que a abordagem que usa um mesmo número de grupos para todas as consultas. Para este descritor, o índice DB foi o que apresentou melhores resultados para todos os níveis.

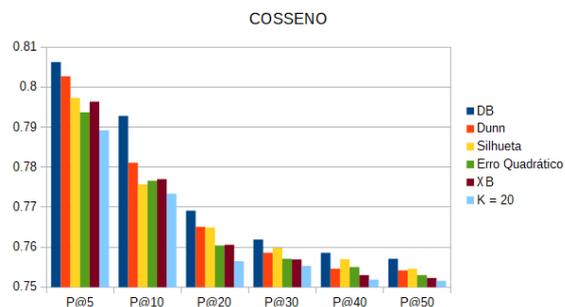


Figura 3. Precisão para o descritor COSSENO.

CONSIDERAÇÕES FINAIS

Considerando os experimentos realizados, as variações em termos de diversidade não foram expressivas, diferentemente do esperado. Esse comportamento pode ser fruto de inúmeras causas, entre elas a heterogeneidade da base de dados e o Gap Semântico. No caso de bases de dados heterogêneas, por se tratarem de imagens ricas em detalhes, e com grande variabilidade de conteúdo e o alto nível conceitual dos grupos esperados, incorporá-las em grupos homogêneos torna-se uma atividade difícil.

Já com o Gap Semântico entre características de baixo nível, extraídas das imagens, e o nível conceitual presente na imagem, pode afetar diretamente a formação de grupos. Ademais, estabelecer a relação entre as características da imagem e a opinião (gabarito) do usuário é um problema recorrente, uma vez que a interpretação humana é carregada de subjetividade, o que tende a impactar na determinação de grupos.

Os resultados mostraram que a nova abordagem pode proporcionar melhores resultados em termos de relevância, se comparamos com a abordagem que usa um mesmo número de grupos para todas as consultas, a qual comumente é utilizada. Além disso, evidenciamos que as medidas de qualidade – que compõem a abordagem desenvolvida – exibem um comportamento diferente dependendo do descritor, uma vez que a característica visual também influencia na formação dos grupos.

Outros resultados apontam, em relação a diversificação, que não houve uma diferença expressiva entre as abordagens. Diante disso, pretendemos realizar novos experimentos, usando novas configurações experimentais, como a utilização de outros descritores, outros métodos de ranking inicial, utilização de outros algoritmos de agrupamento, combinação de medidas de qualidade e combinação de descritores.

REFERÊNCIAS

- ARAUJO, Iago B. A. C.; CALUMBY, Rodrigo Tripodi. Features Fusion for Diversity Gap Reduction. In: SBBD. 2016. p. 175-180.
- CHANG, Wei-Yi; YEH, Yi-Ren; WANG, Yu-Chiang Frank. Style-oriented landmark retrieval and summarization. In: Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific. IEEE, 2016. p. 1-4.
- CHANG, Wei-Yi; WANG, Yu-Chiang Frank. Style-centric image summarization from photographic views of a city. In: Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. IEEE, 2016. p. 2787-2791.
- FAN, Jianping et al. A novel approach to enable semantic and visual image summarization for exploratory image search. In: Proceedings of the 1st ACM international conference on Multimedia information retrieval. ACM, 2008. p. 358-365.
- FERREIRA, Cristiano D. et al. Recod@ MediaEval 2016: Diverse Social Images Retrieval. In: MediaEval. 2016.
- JIN, Xin; HAN, Jiawei. K-medoids clustering. In: Encyclopedia of Machine Learning. Springer, Boston, MA, 2011. p. 564-565.
- IONESCU, Bogdan et al. Retrieving Diverse Social Images at MediaEval 2014: Challenge, Dataset and Evaluation. In: MediaEval. 2014.
- TOLLARI, Sabrina. UPMC at mediaeval 2016 retrieving diverse social images task. In: MediaEval 2016 Workshop. 2016.
- TORRES, Ricardo S.; FALCAO, Alexandre X. Content-based image retrieval: theory and applications. RITA, v. 13, n. 2, p. 161-185, 2006.
- VELTKAMP, Remco C.; TANASE, Mirela. Content-based image retrieval systems: A survey. Department of Computing Science, Utrecht University, p. 1-62, 2002.