

# Recuperação de Imagens na Web com Fusão Adaptativa de Credibilidade Baseada em Algoritmos Genéticos

Wanderson Bezerra da Silva  
Universidade Estadual de Feira de Santana  
Avenida Transnordestina, s/n - Novo Horizonte  
Feira de Santana, Bahia 44036-900  
wbsilva@ecompu.uefs.br

Rodrigo Tripodi Calumby  
Universidade Estadual de Feira de Santana  
Avenida Transnordestina, s/n - Novo Horizonte  
Feira de Santana, Bahia 44036-900  
rtcalumby@uefs.br

## ABSTRACT

Credibility information gives an indication of which users are most likely to share relevant images on social network feeds and consequently may help estimating the relevance of an image for retrieval purposes. Considering multiple credibility evidences has been shown as an effective method for image ranking. In order to select and combine multiple credibility descriptors, this work proposes a genetic algorithm-based automatic context-adaptive weight adjustment model. The experimental results show promising effectiveness when compared to the baseline.

## KEYWORDS

Machine Learning, Image Retrieval, Credibility Descriptors

## 1 INTRODUÇÃO

Dadas as necessidades dos usuários e a grande quantidade de informação disponível na web, é imprescindível o uso de técnicas eficientes e eficazes para exploração de grandes coleções de dados. Tratando-se de busca de imagens, por exemplo, a abordagem mais comum baseia-se na utilização de informações textuais (metadados, palavras-chaves, páginas web, etc.) [1]. Outro paradigma utiliza a descrição das propriedades visuais (cor, forma, textura, estrutura, etc.) das imagens para indexá-las [4]. Quando consideramos o compartilhamento de informação em mídias sociais, outro fator importante é a reputação das fontes de dados, que podem representar a qualidade e confiabilidade do conteúdo [8].

Neste contexto, informações de credibilidade considerando as tags que um usuário associa ao conteúdo das imagens que compartilha, buscam fornecer uma estimativa automática da qualidade global do conteúdo presente nas relações tag-imagem para as contribuições deste usuário [8]. A credibilidade de um usuário pode ser computada levando-se em consideração fatores como, e.g., a descrição textual da imagem, a quantidade de visualizações, frequência de coocorrência entre as tags de um usuário e entre diferentes usuários, proporção de imagens que apresentam faces humanas, similaridade entre as tags de localização das imagens que representam o mesmo local, entre outros. Estas informações são particularmente valiosas para explorar o contexto social dos dados. Elas fornecem uma indicação sobre quais usuários são mais propensos a compartilhar imagens com anotações relevantes/confiáveis. Neste processo, a

credibilidade do usuário é um fator de grande importância para o refinamento dos resultados obtidos por um sistema de busca. Contudo, um único critério de credibilidade pode não ser capaz de capturar adequadamente a reputação de um usuário e permitir que suas imagens tenham seu grau de relevância corretamente estimado para um determinado contexto de busca. De modo a atenuar este problema, descritores de credibilidade podem ser combinados para refinar os resultados da busca, e.g., eliminando-se itens fornecidos por usuários de baixa credibilidade. Esta não é uma tarefa fácil, dado que é preciso definir quais as melhores combinações entre os descritores de modo a prover o melhor resultado para um determinado cenário de busca.

O uso dos descritores de credibilidade tem atraído grande atenção de pesquisadores nos últimos anos. Neste sentido, podemos citar a MediaEval Retrieving Diverse Social Images Task [7], desafio que teve por cenário a simulação do uso de mecanismos de buscas em redes sociais por um turista que está interessado em obter uma visão geral sobre uma determinada localidade, por exemplo, pontos turísticos e regiões geográficas, antes de escolher seu destino. Nesta atividade, descritores de credibilidade podem ser utilizados, e.g., para ranquear, filtrar e/ou diversificar os resultados em sistemas de recuperação de imagens.

Considerando as limitações individuais dos descritores de credibilidade e a necessidade de modelos eficazes para sua combinação, este trabalho propõe uma técnica de fusão linear para descritores de credibilidade, cujos pesos são automaticamente ajustados por meio da meta-heurística de Algoritmos Genéticos (AG). Assim, estes algoritmos são utilizados para a descoberta da importância/contribuição de cada descritor, permitindo assim a seleção e fusão de múltiplas informações de credibilidade e, consequentemente, o aprimoramento dos resultados da busca.

## 2 TRABALHOS RELACIONADOS

Nesta seção, apresentamos uma breve revisão sobre trabalhos envolvendo o uso de descritores de credibilidade na etapa de filtragem ou re-ranqueamento dos resultados de sistemas de recuperação da informação baseados em recuperação de imagens em mídias sociais.

Em [6], os autores propõem a exploração de dados de credibilidade dos usuários no processo de recuperação de imagens. O objetivo foi demonstrar um sistema de recuperação da informação capaz de retornar para o usuário resultados relevantes e diversificados. Para isso, na etapa de ranqueamento, os descritores de credibilidade foram utilizados para aumentar a relevância dos resultados considerando imagens provenientes de usuários com boas estimativas de credibilidade. Seus resultados mostraram que os

In: XV Workshop de Trabalhos de Iniciação Científica (WTIC 2018), Salvador, Brasil. Anais do XXIV Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres. Porto Alegre: Sociedade Brasileira de Computação, 2018.

© 2018 SBC – Sociedade Brasileira de Computação.  
ISBN 978-85-7669-435-9.

dados de credibilidade são eficazes quando utilizados em complemento com dados textuais e/ou visuais no processo de recuperação de imagens. Os escores de credibilidade utilizados foram obtidos comparando-se as tags dos usuários com os conceitos disponíveis na ImageNet [5], sem qualquer adaptação ao conjunto de dados de avaliação. Mesmo nesse cenário, os resultados foram considerados satisfatórios, ratificando o potencial dos descritores de credibilidade.

O trabalho em [8] apresenta um coleção de imagens projetada para dar suporte ao desenvolvimento de sistemas de recuperação que apresentem resultados de busca relevantes e diversificados. O conjunto de dados, coletados do Flickr<sup>1</sup> e Wikipedia<sup>2</sup>, consiste em informações para 300 pontos turísticos em 35 países diferentes, e.g., museus, monumentos, edifícios, estradas, pontes, cavernas, entre outros. Para cada local, as seguintes informações são fornecidas: palavra-chave da localização (identificador textual exclusivo no conjunto de dados), número de localização (identificador numérico exclusivo) e coordenadas GPS (latitude e longitude em graus). Dentre os recursos presentes na base, há um conjunto de descritores de credibilidade, e.g.: o *faceProportion*, que indica a porcentagem de fotos de um usuário que apresentam faces humanas; o *tagSpecificity*, que é obtido calculando a especificidade média das tags de um usuário; e o *LocationSimilarity*, que dá ao usuário uma pontuação com base na similaridade entre as *tags* de localização das suas imagens e outras imagens que representam o mesmo local.

Em [3], dentre as abordagens apresentadas, é proposto o uso de um descritor de credibilidade para o ranqueamento dos resultados de buscas de imagens. Para isso, os autores utilizaram apenas o descritor *visualScore* para ordenar as imagens. Este descritor usa as tags das imagens do usuário como parâmetro de busca na coleção ImageNet [5]. Para cada tag do usuário idêntica a um conceito no ImageNet, é feito o comparativo entre as características visuais da imagem associada à tag e as características visuais das imagens recuperadas do ImageNet, obtendo um score de classificação para a tag. O score final é obtido através da média das scores obtidos para cada tag.

Para a combinação de múltiplos descritores de credibilidade, em [2], os autores propuseram uma abordagem multimodal com a utilização de metadados de imagens e informações de credibilidade do usuário. No arcabouço desenvolvido, os descritores de credibilidade são usados para ranqueamento de resultados de busca imagens em redes sociais. Os descritores foram integrados por meio de uma combinação linear simples, sendo os respectivos pesos fixados de forma empírica. Os resultados apresentaram melhor qualidade de ranking quando comparados com os originais obtidos pelo sistema de ranking da rede social, neste caso o Flickr.

### 3 MÉTODO PROPOSTO

O método aqui proposto tem como objetivo fornecer uma única informação de relevância por meio da combinação de múltiplos descritores de credibilidade. Esta informação é utilizada na etapa de re-ranqueamento dos resultados de uma determinada busca em um sistema de recuperação de imagens em rede social. Os descritores de credibilidade utilizados foram: *visualScore*, *faceProportion*,

*tagSpecificity*, *locationSimilarity*, *photoCount*, *uniqueTags*, *uploadFrequency*, *bulkProportion*, *meanPhotoViews*, *meanTitleWordCounts*, *meanTagsPerPhoto*, *meanTagRank* e *meanImageTagClarity* [8].

Para obter um score a partir da combinação dos descritores de credibilidade é utilizada a fusão linear (Equação 1), sendo  $d_i$  um elemento contido no conjunto dos descritores  $D$ , tal que  $D = \{d_1, d_2, \dots, d_n\}$  e  $p_i$  o peso associado ao descritor  $d_i$ .

$$credScore = \frac{\sum_{i=1}^n p_i d_i}{\sum_{i=1}^n p_i} \quad (1)$$

A descoberta dos melhores pesos ( $p_i$ ) para a combinação dos descritores foi realizada por meio da técnica de algoritmos genéticos, cuja estrutura geral é apresentada na Figura 1. No AG implementado, a solução (indivíduo no AG) foi modelada como um vetor de pesos (genótipo) formado por valores de ponto flutuante (genes), em que cada valor representa o peso de um descritor.

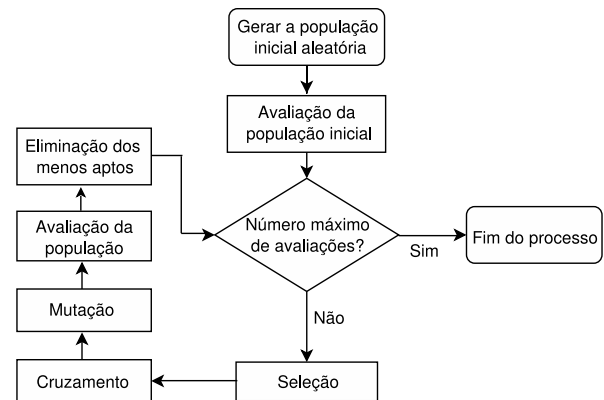


Figura 1: Fluxograma do algoritmo genético implementado.

A primeira etapa do algoritmo é a criação uma população aleatória de indivíduos, ou seja, vetores cujos pesos são inicializados com valores aleatórios. Em seguida, esses indivíduos são avaliados e submetidos ao operador de seleção por torneio binário. Feito isso, os indivíduos selecionados são submetidos à etapa de reprodução, onde são submetidos aos operadores de cruzamento e mutação. Finalizado o ciclo de reprodução, a nova população é avaliada e os indivíduos menos aptos são descartados. Este processo é repetido até atingir o número máximo de avaliações.

A avaliação da aptidão (*fitness*) de cada indivíduo é feita da seguinte forma: dado um ranking de entrada, a fusão linear dos descritores de credibilidade é realizada utilizando os pesos fornecidos pelo indivíduo avaliado. Em seguida, as imagens presentes no resultado original são re-ranqueadas utilizando o score de credibilidade obtido. Uma medida de avaliação é utilizada para avaliar a eficácia do novo ranking. O valor obtido com a medida de avaliação é utilizado como a aptidão do indivíduo.

Após a realização de todas as etapas do AG, o melhor indivíduo da última geração reproduzida é definido como a melhor configuração de pesos para os descritores.

<sup>1</sup><https://www.flickr.com>

<sup>2</sup><https://www.wikipedia.org>

#### 4 METODOLOGIA EXPERIMENTAL

O AG desenvolvido tem como base o modelo NSGA-II. Para o auxiliar no desenvolvimento, foi utilizado o framework JMetal<sup>3</sup>. O comportamento do AG depende da configuração de alguns parâmetros. A Tabela 1 apresenta a configuração utilizada no AG implementado. Os parâmetros de probabilidade de cruzamento e de mutação definem as chances que cada indivíduo tem de ser submetido a estes operadores, enquanto o tamanho da população e a quantidade de sobreviventes definem, respectivamente, a quantidade de indivíduos em uma população e o número de indivíduos que serão replicados para a próxima geração. A quantidade de avaliações define o número máximo de indivíduos que devem ser avaliados.

| Cruzamento | Mutação | População | Sobreviventes | Avaliações |
|------------|---------|-----------|---------------|------------|
| 90%        | 7%      | 200       | 10            | 5000       |

Tabela 1: Parâmetros utilizados no algoritmo genético.

Para determinar a probabilidade de cruzamento e mutação do AG foram utilizados os valores padrões para o modelo NSGA-II implementado no framework JMetal.

O AG precisa passar por uma etapa de busca de soluções (o que garante a evolução dos indivíduos) e por uma etapa de testes (para avaliar quão boa foi a solução final). Assim, de modo a avaliar adequadamente a capacidade do algoritmo de encontrar boas soluções e que sejam eficazes para novos dados (consultas), foi definida a utilização de conjuntos de dados distintos para cada etapa. Com base nesta ideia e com o objetivo de ter um grande conjunto de dados, optou-se por unir os conjuntos de desenvolvimento e testes apresentados em [8], totalizando 222 consultas com cerca de 300 imagens no resultado cada uma. Com a unificação dos conjuntos de dados, optamos por aplicar a técnica da validação cruzada. A validação cruzada ajuda a avaliar a capacidade de generalização do AG, ou seja, a possibilidade de estimar a eficácia da solução encontrada pelo AG quando submetida a um novo conjunto de dados.

O modelo de validação cruzada utilizado foi o *k-fold cross validation*. Esta técnica particiona aleatoriamente o conjunto de dados em *k* subconjuntos mutuamente exclusivos de tamanhos iguais. Para o treinamento são utilizados *k* - 1 subconjuntos, enquanto o conjunto restante é utilizado para testes. Este processo é realizado *k* vezes alternando de forma circular o subconjunto de testes. Para os experimentos realizados foi definido *k* = 5. Desse modo, em cada etapa da validação cruzada aproximadamente 178 consultas foram utilizadas para descoberta dos pesos e 44 consultas foram utilizadas para avaliação.

Como função de adequação dos indivíduos gerados durante a etapa de treinamento do AG, foram avaliadas as medidas *Mean Average Precision* (MAP) e *Precision* (P@N), especificamente P@10, P@20 e P@50. A medida P@N indica a proporção de imagens relevantes dentre as N primeiras imagens retornadas (Equação 2). Neste trabalho, a função *Average Precision* (AP) foi computada como a média dos valores de *Precision* da curva interpolada de *Precision-Recall* de 11 pontos. A medida MAP é então computada como a média dos valores de AP considerando múltiplas consultas.

$$P@N = \frac{\{Itens\ relevantes\} \cap \{Itens\ retornados\}}{\{Itens\ retornados\}} \quad (2)$$

Utilizando a mesma configuração de validação cruzada, os experimentos foram executados separadamente para cada uma destas funções de *fitness*.

Por fim, a etapa de teste foi realizada, na qual os pesos encontrados pelo AG foram utilizados na Equação 1 para obter um score final de credibilidade. Este score é utilizado para re-ranquear os resultados para as consultas presentes no subconjunto de teste. O novo ranking gerado foi então avaliado utilizando a medida MAP e curvas de P@N. Os resultados são apresentados até a profundidade máxima de 50, cobrindo aquilo que, de modo geral, é apresentado nas primeiras páginas dos sistemas de busca.

#### 5 RESULTADOS E DISCUSSÕES

Os rankings obtidos com o método aqui proposto foram comparados com os rankings originais obtidos do Flickr, com o método proposto em [3] e com a utilização dos melhores pesos obtidos com o método em [2]. Primeiro foi feito um comparativo entre os resultados obtidos utilizando as medidas P@10, P@20, P@50 e MAP como função de *fitness*. Os resultados em termos de P@N são apresentados na Figura 2.

Os resultados do método proposto com avaliação de *fitness* baseada na medida P@10 sugerem que, como esta medida considera apenas o topo do ranking (as 10 primeiras posições), não foi sensível o suficiente para capturar variações no ranking e permitir ao AG avaliar adequadamente as soluções encontradas. Analisando-se as curvas para as medidas P@20 e P@50, percebe-se que com o aumento da profundidade de análise do ranking, melhores resultados são obtidos. Por fim, analisando-se a curva do resultado obtido com avaliação de *fitness* baseada na medida MAP, percebe-se uma degradação do resultado. Estima-se que, como a medida MAP leva em consideração todo o ranking, pequenas variações de posições na parte superior do ranking tem menor impacto no resultado computado, tornando a avaliação menos sensível a estas variações.

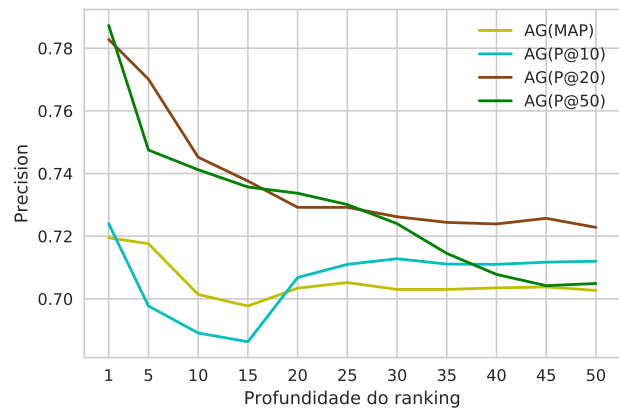


Figura 2: Comparativo dos resultados obtidos com o AG.

<sup>3</sup><http://jmetal.sourceforge.net>

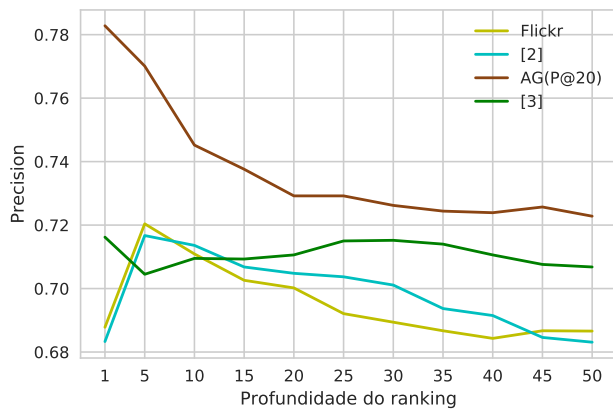


Figura 3: Comparativo do AG(P@20) com baselines.

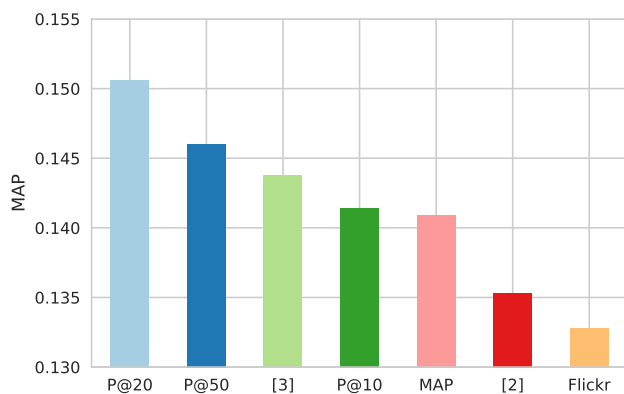


Figura 4: Avaliação comparativa dos resultados utilizando a medida MAP.

Por meio da Figura 2, percebe-se que, dentre as funções de *fitness* utilizadas, o melhor resultado foi obtido pela P@20, seguido de P@50. Estima-se que a utilização da medida P@20 teve vantagem em relação à P@50 por analisar uma parte do ranking que tende a ter uma maior proporção de itens relevantes em relação à quantidade de itens recuperados, ao mesmo tempo que avalia dados suficientes para ajudar o AG à encontrar uma solução adequada. Estes resultados são comparados com os obtidos utilizando os melhores pesos encontrados em [2], com os resultados obtidos utilizando o método proposto em [3] e com o ranking original do Flickr. A Figura 3 apresenta o comparativo entre os *baselines* e o resultado obtido utilizando P@20 como *fitness*. De forma geral, pode-se perceber que houve ganho expressivo em relação aos *baselines*.

A Figura 4 apresenta os resultados obtidos com as quatro medidas de *fitness* em comparação aos *baselines*, utilizando MAP como medida de avaliação. Dentre as medidas de aptidão utilizadas, mais

uma vez pode-se destacar os resultados obtidos utilizando P@20 e P@50. Ambas conseguiram resultados superiores em comparação aos *baselines*.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, apresentamos um modelo de re-ranqueamento dos resultados de um sistema de recuperação de imagens em redes sociais, utilizando a combinação de descritores de credibilidade através da fusão linear, cujos pesos são definidos por um algoritmo genético. Os resultados preliminares apresentam-se superiores aos *rankings* originais do sistema de busca da rede social e ao melhor *baseline*. Além disso, pode-se observar que a escolha da função de aptidão tem impacto direto na descoberta dos pesos e consequentemente na eficácia do método proposto.

Vale destacar que apenas uma configuração paramétrica foi avaliada para as etapas do algoritmo genético e portanto uma nova etapa de experimentação será realizada para melhor avaliar o método proposto. O ajuste destes parâmetros poderá permitir a obtenção de melhores soluções provenientes do AG, pois sintonizando seus parâmetros, o AG tende a se adequar às particularidades do problema. Outra etapa de grande importância é a experimentação de novos operadores para seleção, mutação e cruzamento, que podem influenciar de forma direta o processo de busca dos melhores pesos para os descritores de credibilidade.

O método proposto foi avaliado apenas em termos do aprimoramento do ranking, mas seus resultados podem ser utilizados para potencialização de métodos de diversificação visual que naturalmente são beneficiados por rankings de maior qualidade. De modo adaptativo, o método proposto pode ainda ser utilizado para descoberta de soluções para outros cenários de busca, diferentes bases de dados, ou tipos de objetos digitais (textos, vídeos, etc.).

## REFERÊNCIAS

- [1] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 2011. *Modern Information Retrieval - the concepts and technology behind search, Second edition*. Pearson Education Ltd., Harlow, England.
- [2] Rodrigo Tripodi Calumby, Iago B. A. do C. Araujo, Vinicius Pereira Santana, Javier A. V. Muñoz, Otávio Augusto Bizetto Penatti, Lin Tzy Li, Jurandy Almeida, Giovanni Chiachia, Marcos André Gonçalves, and Ricardo da Silva Torres. 2015. Recod @ MediaEval 2015: Diverse Social Images Retrieval. In *Working Notes Proceedings of the MediaEval Workshop*.
- [3] Xi Chen, Haokun Liu, Zhi-Hong Deng, and Yunlun Yang. 2015. Retrieving Relevant and Diverse Image from Social Media Images.. In *Working Notes Proceedings of the MediaEval Workshop*.
- [4] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Ze Wang. 2008. Image retrieval: Ideas, influences, and trends of the new age. *Comput. Surveys* 40, 2 (2008).
- [5] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.
- [6] Alexandru Lucian Ginsca, Adrian Popescu, Bogdan Ionescu, Anil Armagan, and Ioannis Kanellos. 2014. Toward an estimation of user tagging credibility for social image retrieval. In *Proceedings of the 22nd ACM International Conference on Multimedia*. ACM, 1021–1024.
- [7] Bogdan Ionescu, Alexandru-Lucian Ginsca, Bogdan Boteanu, Adrian Popescu, Mihai Lupu, and Henning Müller. 2014. Retrieving Diverse Social Images at MediaEval 2014: Challenge, Dataset and Evaluation.. In *Working Notes Proceedings of the MediaEval Workshop*.
- [8] Bogdan Ionescu, Adrian Popescu, Mihai Lupu, Alexandru Lucian Ginscă, Bogdan Boteanu, and Henning Müller. 2015. Div150cred: A social image retrieval result diversification with user tagging credibility dataset. In *Proceedings of the 6th ACM Multimedia Systems Conference*. ACM, 207–212.