Fusão Adaptativa de Credibilidade para Recuperação de Imagens em Redes Sociais

Wanderson Bezerra da Silva¹; Rodrigo Tripodi Calumby²

1. Bolsista PROBIC, Graduando em Engenharia da Computação, UEFS, bsilva.wanderson@gmail.com

2. Orientador, DEXA, UEFS, rtcalumby@uefs.br

PALAVRAS-CHAVE: Fusão de dados, Recuperação de Imagens, Credibilidade

INTRODUÇÃO

Considerando o compartilhamento de informação em mídias sociais, um fator importante é a reputação das fontes de dados, que podem representar a qualidade e confiabilidade do conteúdo. 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 (Ionescu et al, 2015). 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, 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 prover um resultado mais adequado para cada determinado cenário de busca. Este trabalho propõe uma técnica da fusão linear para descritores de credibilidade, cujos pesos são automaticamente ajustados por meio da meta-heurística de Algoritmos Genéticos (AG).

METODOLOGIA

O método aqui proposto tem como objetivo fornecer para cada imagem 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 busca em um sistema de recuperação de imagens em rede social. Para obter um score a partir da combinação dos descritores de credibilidade, é utilizada a combinação ponderada (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 i . A descoberta dos melhores pesos (p_i) para a combinação dos descritores foi realizada por meio da técnica de algoritmos genéticos. No AG implementado, a solução (indivíduo no AG) foi modelada como um genótipo (vetor de pesos) formado por múltiplos genes (valores de ponto flutuante), que representam os pesos dos descritores

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

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. Os indivíduos selecionados são submetidos à etapa de reprodução, onde são submetidos aos operadores de cruzamento e mutação. O operador de mutação utilizado foi o *Creep* e o operador de cruzamento foi o *BLX Alpha*. Finalizado o ciclo de reprodução, a nova população é avaliada e os indivíduos menos aptos são descartados. Este processo é repetido até atingir a condição de parada. A avaliação da aptidão de cada indivíduo (*fitness*) 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 em avaliação. 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 dado como o *fitness* 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.

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.

Tabela 1. Parâmetros utilizados no algoritmo genético

Probabilidade de Cruzamento	Probabilidade de Mutação	Tamanho da População	Quantidade de Sobreviventes
50%	50%	200	10

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). 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 (Ionescu et al, 2015), totalizando 222 consultas. Esta base consiste em informações para pontos turísticos em 35 países diferentes, fornecendo para cada local cerca de 300 imagens e as seguintes informações: palavra-chave da localização, número de localização e coordenadas GPS (latitude e longitude). Dentre os recursos presentes na base, há um conjunto com 13 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 fornece ao usuário uma pontuação com base na similaridade entre as tags de localização das suas imagens em relação a outras imagens que representam o mesmo local.

Com a unificação dos conjuntos de dados, aplicou-se então a técnica da validação cruzada. A validação cruzada ajuda a avaliar a capacidade de generalização do AG, ou seja, estimar a eficácia da solução encontrada pelo AG quando submetida a um novo conjunto de

dados. O modelo de validação cruzada utilizado foi o *5-fold cross validation*. Desse modo, em cada etapa da validação cruzada aproximadamente 178 consultas foram utilizadas para descoberta dos pesos e 44 consultas foram utilizadas para avaliação.

Para a avaliar o *fitness* dos indivíduos gerados durante a etapa de treinamento do AG, foram utilizadas as medidas Mean Average Precision (MAP) e Precision ($P@N$), especificamente $P@10$, $P@20$ e $P@50$. Utilizando a mesma configuração de validação cruzada, os experimentos foram executados separadamente para cada uma destas funções de fitness. Para a etapa de teste, os pesos encontrados são utilizados na Equação 1 para obtenção do 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 curvas de $P@N$ e MAP.

RESULTADOS E DISCUSSÕES

Os rankings obtidos com o método aqui proposto foram comparados com os rankings originais obtidos do Flickr e com a utilização dos melhores pesos obtidos com o método em (Calumby et al, 2015). O gráfico da figura 1 apresenta um comparativo entre os resultados obtidos utilizando as medidas $P@10$, $P@20$, $P@50$ E MAP como fitness. Este comparativo foi feito utilizando uma curva de $P@N$.

figura 1: Comparativo entre os resultados obtidos com o AG.

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.

Por meio da Figura 1, 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 a encontrar uma solução adequada

Figura 2: Resultados obtidos usando P@20 como fitness.

A figura 2 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.

CONSIDERAÇÕES FINAIS

Os resultados 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.

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. 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.

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] Ionescu, B., Popescu, A., Lupu, M., Gînscă, A. L., Boteanu, B., & Müller, H. (2015, March). Div150cred: A social image retrieval result diversification with user tagging credibility dataset. In Proceedings of the 6th ACM Multimedia Systems Conference (pp. 207-212). ACM.
- [2] R. T. Calumby, I. B. A. do C. Araujo, V. P. Santana, J. A. V. Muñoz, O. A. B. Penatti, L. T. Li, J. Almeida, G. Chiachia, M. A. Gonçalves, and R. da S. Torres. Recod @ mediaeval 2015: Diverse social images retrieval. In Working Notes Proceedings of the MediaEval 2015 Workshop, Wurzen, Germany, September 14-15, 2015.