



Data Science Academy

www.datascienceacademy.com.br

Visão Computacional

Mini-Projeto 3
Construindo um Sistema de Busca de
Imagens com Base em Conteúdo
Parte 1

Talvez você esteja se perguntando: o que exatamente é um sistema de busca de imagens?

Nós estamos familiarizados com os motores de busca baseados em texto como Google e Bing - você simplesmente insere algumas palavras-chave relacionadas ao conteúdo que deseja encontrar (ou seja, sua "consulta") e, em seguida, seus resultados são retornados para você. Mas para os motores de busca de imagens, as coisas funcionam um pouco diferente - você não está usando texto como sua consulta, em vez disso, está usando uma imagem.

E como você quantifica o conteúdo de uma imagem para torná-la “pesquisável”? Já iremos responder a esta pergunta, mas antes vamos aprender um pouco mais sobre os motores de busca de imagens.

Em geral, existem três tipos de motores de busca de imagens: busca por metadados, busca por exemplo e uma abordagem híbrida dos dois. Vejamos cada um deles.

Busca Por Metadados

A pesquisa por meta-dados é apenas marginalmente diferente dos motores de busca padrão baseados em palavras-chave mencionados acima. A pesquisa por sistemas de metadados raramente examina o conteúdo da própria imagem. Em vez disso, eles dependem de pistas textuais, como (1) anotações manuais e marcação realizados por humanos juntamente com (2) dicas contextuais automatizadas, como o texto que aparece perto da imagem em uma página da Web.

Quando um usuário realiza uma pesquisa em um sistema de busca por meta-dados, ele fornece uma consulta, assim como em um mecanismo de pesquisa de texto tradicional, e as imagens que possuem tags ou anotações similares são retornadas. Ao utilizar uma pesquisa por sistema de metadados, a imagem real em si é raramente examinada.

Um excelente exemplo de um mecanismo de pesquisa de imagens por metadados é o Flickr. Depois de carregar uma imagem para o Flickr, você recebe um campo de texto para inserir as tags que descrevem o conteúdo das imagens que você enviou. O Flickr armazena essas palavras-chave, indexa-as e as utiliza para encontrar e recomendar outras imagens relevantes.

Busca Por Exemplo

A pesquisa por sistemas de exemplo, por outro lado, depende unicamente do conteúdo da imagem - nenhuma palavra-chave é fornecida. A imagem é analisada, quantificada e armazenada para que imagens semelhantes sejam retornadas pelo sistema durante uma pesquisa.

Os motores de busca de imagens que quantificam o conteúdo de uma imagem são chamados sistemas de recuperação de imagem baseada em conteúdo (ou Content-Based Image Retrieval - CBIR). O termo CBIR é comumente usado na literatura acadêmica, mas, na realidade, é simplesmente uma maneira mais amável de dizer "motor de busca de imagens" com a facilidade de que o mecanismo de pesquisa está confiando estritamente nos conteúdos da imagem e não há anotações textuais associadas com a imagem.

Um ótimo exemplo de um sistema de Pesquisa por Exemplo é o TinEye (<https://www.tineye.com>). TinEye é, na verdade, um mecanismo de busca de imagem inversa onde você fornece uma imagem de consulta e, em seguida, o TinEye retorna combinações próximas da mesma imagem, juntamente com a página da Web em que a imagem original apareceu.

Então, considere isso: você vai rotular manualmente cada uma das seis bilhões de imagens no TinEye? Claro que não. Isso iria requerer um exército de funcionários e seria extremamente caro.

Em vez disso, você utiliza algum tipo de algoritmo para extrair "recursos" (ou seja, uma lista de números para quantificar e representar de forma abstrata a imagem) a partir da própria imagem. Então, quando um usuário envia uma imagem de consulta, você extrai recursos da imagem da consulta, compara-os ao seu banco de dados de recursos e tenta encontrar imagens semelhantes.

É importante reforçar o ponto em que os sistemas de Pesquisa por Exemplo dependem estritamente do conteúdo da imagem. Esses tipos de sistemas tendem a ser extremamente difíceis de construir e dimensionar, mas permitem um algoritmo totalmente automatizado para governar a pesquisa - nenhuma intervenção humana é necessária.

Abordagem Híbrida

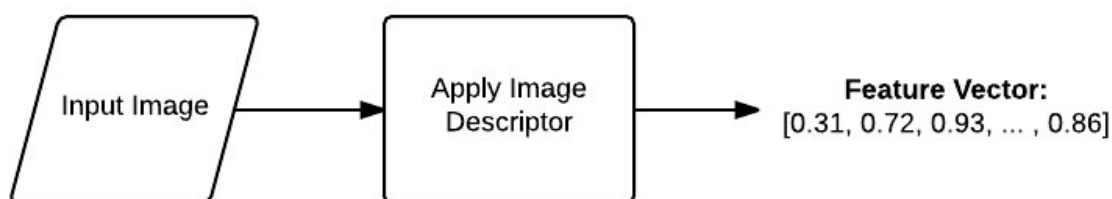
Claro, há um termo intermediário entre os dois - considere o Twitter, por exemplo. No Twitter, você pode fazer upload de fotos para acompanhar seus tweets. Uma abordagem híbrida seria correlacionar os recursos extraídos da imagem com o texto do tweet. Usando essa abordagem, você pode criar um mecanismo de busca de imagens que pode receber ambas dicas contextuais junto com uma estratégia de Pesquisa por Exemplo.

Revisando Alguns Conceitos Importantes

Vamos revisar alguns termos importantes. O primeiro termo que precisamos revisar é a extração de recursos. O processo de extração de recursos governa as regras, algoritmos e metodologias que usamos para quantificar de forma abstrata o conteúdo de uma imagem usando apenas uma lista de números, chamado vetor de recursos.

As características, por outro lado, são a saída do nosso descritor de imagem ou descritor de recursos. Quando você coloca uma imagem em um descritor de imagem, você obterá recursos do outro lado.

Nos termos mais básicos, os recursos (ou vetores de recursos) são apenas uma lista de números usados para representar e quantificar imagens de maneira abstrata. Dê uma olhada no exemplo abaixo:



Aqui, apresentamos uma imagem de entrada e aplicamos nosso descritor de imagem. E então nossa saída é uma lista de recursos usados para quantificar a imagem. Muitas vezes, usamos os termos recursos e vetores de recursos de forma intercambiável.

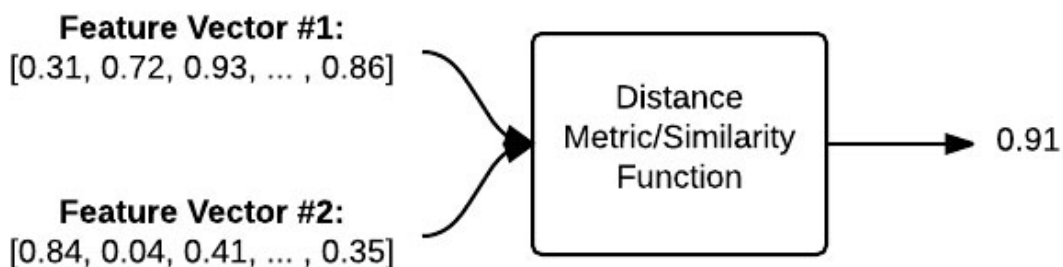
Extrair recursos de imagens é inerentemente uma tarefa paralelizada. Dado um conjunto de dados de imagens, nosso objetivo é aplicar exatamente o mesmo descritor de imagem a cada imagem no conjunto de dados. Assim, podemos aproveitar múltiplos núcleos do nosso sistema para tornar o estágio de extração de recursos executado, consideravelmente

mais rápido. Podemos até chegar a utilizar ferramentas de Big Data, como Hadoop, para acelerar ainda mais a extração de recursos.

Independentemente de extrair recursos de imagens em série, usando vários processos ou em um cluster Hadoop, ao criar um mecanismo de pesquisa de imagens, muitas vezes temos que indexar nosso conjunto de dados de recursos. Aqui nós recebemos os recursos extraídos do nosso conjunto de dados e armazenamos os recursos de forma eficiente para que eles possam ser comparados quanto à relevância.

Em alguns casos, indexar nosso conjunto de dados é tão simples como armazenar nossos vetores de recursos em um arquivo .csv. Mas, para os motores de busca de imagens em grande escala, usamos estruturas e algoritmos de dados especializados para fazer pesquisas em um espaço N-dimensional executado no tempo sub-linear.

Agora que temos nossos vetores de recursos extraídos e armazenados, podemos comparar pares para relevância e similaridade usando uma métrica de distância ou função de similaridade. As métricas de distância e as funções de similaridade recebem dois vetores de recursos como entradas e, em seguida, exibem um número que representa como "semelhantes" os dois vetores de características. A figura abaixo demonstra o processo de comparação de duas imagens:



A saída da função de distância é um único valor de ponto flutuante usado para representar a semelhança entre as duas imagens.

Uma vez que temos a nossa métrica de distância ou função de similaridade, podemos realizar uma pesquisa real. A imagem que realmente enviamos ao nosso sistema CBIR (como digitar palavras-chave de texto no Google) é chamada nossa imagem de consulta ou simplesmente consulta.

Nosso sistema CBIR recebe a imagem de consulta, extrai recursos dele e usa nosso conjunto de dados e métrica de distância indexados para comparar imagens de similaridade estritamente, utilizando os vetores de recursos extraídos.

O conjunto de resultados, ou o conjunto de imagens relevantes para a consulta, são então classificados e apresentados ao usuário.

Se isso parece muito trabalho, na verdade, é muito trabalho mesmo. Dito isto, vejamos quais são as 4 etapas de construção de qualquer sistema CBIR.

Os 4 Passos de Qualquer Sistema CBIR

Independentemente do Sistema de Busca de Imagem Baseado em Conteúdo que você está construindo, todos podem ser resumidos em 4 etapas distintas:

1. **Definindo o seu descritor de imagem:** nesta fase, você precisa decidir qual o aspecto da imagem que deseja descrever. Você está interessado na cor da imagem? A forma de um objeto na imagem? Ou você quer caracterizar a textura?
2. **Extração de recursos e indexação de seu conjunto de dados:** agora que você tem seu descritor de imagem definido, seu trabalho é aplicar este descritor de imagem a cada imagem no seu conjunto de dados, extrair recursos dessas imagens e escrever os recursos no armazenamento (ex. Arquivo CSV, RDBMS, etc.) para que eles possam ser comparados posteriormente por similaridade. Além disso, você deve considerar se as estruturas de dados especializadas serão usadas para facilitar a busca mais rápida.
3. **Definindo sua métrica de similaridade:** agora temos uma coleção de vetores de recursos. Mas como você vai compará-los por similaridade? As escolhas populares incluem a distância euclidiana, a distância do Coseno e a distância do qui-quadrado, mas a escolha real é altamente dependente de (1) seu conjunto de dados e (2) os tipos de recursos que você extraiu.
4. **Pesquisa:** o passo final é realizar uma pesquisa real. Um usuário enviará uma imagem de consulta ao seu sistema (de um formulário de upload ou por meio de um aplicativo móvel, por exemplo) e seu trabalho será (1) extrair recursos dessa imagem de consulta e, em seguida, (2) aplicar sua função de similaridade para comparar os recursos de consulta para os recursos já indexados. A partir daí você simplesmente retorna os resultados mais relevantes de acordo com sua função de similaridade.

Estes são os quatro passos mais básicos de qualquer sistema CBIR. À medida que se tornam mais complexos e utilizam diferentes representações de recursos, o número de etapas cresce e você adicionará um número significativo de sub-etapas para cada etapa mencionada acima. Mas, por enquanto, vamos manter as coisas simples e utilizar apenas essas quatro etapas.

Vamos construir um Sistema de Busca de Imagens Baseado em Conteúdo? Então acompanhe o próximo item de aprendizagem.

Referências:

Recent Advance in Content-based Image Retrieval: A Literature Survey

<https://arxiv.org/abs/1706.06064>

Content-based Image Retrieval

<http://www.leeds.ac.uk/educol/documents/00001240.htm>

Content-Based Image Retrieval

<https://homes.cs.washington.edu/~shapiro/cbir.html>

Aigrain, P et al (1996) "Content-based representation and retrieval of visual media - a state-of-the-art review" Multimedia Tools and Applications 3(3), 179-202

Alsuth, P et al (1998) "On video retrieval: content analysis by ImageMiner" in Storage and Retrieval for Image and Video Databases VI, Proc SPIE 3312, 236-247

Androutsas, D et al (1998) "Image retrieval using directional detail histograms" in Storage and Retrieval for Image and Video Databases VI, Proc SPIE 3312, 129-137

Ardizzone, E and La Cascia, M (1997) "Automatic video database indexing and retrieval" Multimedia Tools and Applications 4, 29-56