

Classificação de Imagens por Similaridade

Ingrid Rosselis Sant Ana da Cunha¹, Pedro Olmo Stancioli Vaz De Melo¹

¹ Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
(UFMG)

Belo Horizonte – MG – Brasil

{irscunha,olmo}@dcc.ufmg.br

Tipo de Pesquisa: Científica

Abstract. *Divided into two parts, this course completion work aims to find correlations between movie posters, series and animations. These correlations will be found using image classification and clustering methods and techniques for extracting colors and features. In the first part, contemplated in this document, the datasets were extracted, the images were analyzed and the possible algorithms and techniques to be used in the second stage were studied. Among the methods considered, it was defined the use of a Convolutional Neural Network for classification by genre and by year and, possibly, of image segmentation tools and color descriptors for image clustering.*

Resumo. *Dividido em duas partes, este trabalho de conclusão de curso tem por objetivo final encontrar correlações entre pôsteres de filmes, séries e animações. Estas correlações serão encontradas utilizando métodos de classificação e clusterização de bases imagens, e técnicas para extração de cores e features. Na primeira parte, contemplada neste documento, foram extraídas as bases, analisadas as imagens e estudados os possíveis algoritmos e técnicas a serem usados na segunda etapa. Dentre os métodos vistos, ficou definido o uso de uma Rede Neural Convolucional para classificação por gênero e por ano e, possivelmente, de ferramentas segmentação de imagem e de descritores de cores para clusterização das imagens.*

1. Introdução

A classificação de imagens é o processo de aplicar rótulos utilizando-se de classes já existentes. Usando algoritmos de aprendizado supervisionado, o modelo aprende com instâncias previamente categorizadas e, baseado nas *features* da imagem e na classe da mesma, rotula novas instâncias.

Já a clusterização refere-se ao processo de agrupar instâncias dado similaridades que estas possuam. Por exemplo, em termos de pôsteres de filmes, grupos podem ser gerados com cores predominantes, quantidade de objetos, disposição do texto e características marcantes, como poses ou elementos em destaque.

Desde o uso para inteligência artificial em redes sociais, como *Facebook*, a aplicações de cunho social como acessibilidade para pessoas com deficiência visual, o estudo de técnicas para processamento de imagens tem ganhado cada vez mais espaço e reconhecimento no cenário tecnológico atual.

Entre outras aplicações de classificação e clusterização de imagens, podemos citar: reconhecimento de placas veiculares, análise de imagens médicas, organização automática de coletâneas de fotos, categorização automática de imagens e filtro de conteúdo.

Outro uso amplamente difundido é a identificação de aspectos em comum em um tipo específico de imagens – como nos pôsteres de filmes. Usando algoritmos de classificação e clusterização, pode-se encontrar grupos de imagens que possuem similaridades como mesmo estúdio de animação, diretor ou gênero.

Tendo em vista o grande sucesso das aplicações da classificação de imagens, este projeto tem como objetivo final encontrar relações entre pôsteres de filmes, séries e animações. A intenção é que sejam usados diversos tipos de algoritmos e técnicas que possam explicitar correlações não óbvias entre pôsteres de filmes, séries e animações de várias épocas e gêneros.

No escopo deste projeto, a análise é importante para identificar e entender como as formas de entretenimento se influenciam e se existem padrões que categorizam e definem imagens de mesmo grupo. De maneira geral, é também interessante pensar no uso dessas categorizações para selecionar imagens de referência, por exemplo, sejam elas por rótulos ou por similaridades.

Esta pesquisa foi estruturada como um projeto com duração de um ano, dividido nas disciplinas Projeto Orientado em Computação I e II. A primeira parte deste trabalho tem como objetivo a extração e tratamento dos dados, a análise exploratória e o estudo de métodos de classificação e clusterização de imagens. A segunda parte deste trabalho tratará da aplicação de alguns métodos às bases obtidas.

2. Referencial Teórico

Muitas soluções para o problema de classificação de imagens têm sido propostas nos últimos anos, principalmente para classificação de imagens. Nessa vertente, [Ivasic-Kos et al. 2015] utilizam um modelo algorítmico *multi-label* com três diferentes tipos de classificadores, comparando seus resultados em termos da classificação dos pôsteres de filmes por gênero. Um dos fatores mais interessantes como referência é a adaptação do método *Naive Bayes* para múltiplas classificações *single-label*.

Seguindo uma linha muito próxima à do artigo citado acima, [Chu and Guo 2017] usam uma rede neural convolucional para extrair a representação visual nos pôsteres e um *framework* de detecção de objetos nas mesmas imagens. Os dois resultados são tratados como entrada para uma rede neural profunda de classificação *multi-label* que retorna as probabilidades de um dado pôster pertencer a um gênero de filmes.

As redes neurais convolucionais têm se destacado como uma das soluções mais utilizadas para classificações de imagens de vários tipos. [Krizhevsky et al. 2012] utilizam-se de uma rede neural convolucional profunda para executar a tradicional tarefa de classificar imagens de diferentes tipos. É interessante notar que a rede foi capaz de trabalhar com grandes volumes de imagens e classes – em torno de um milhão de imagens e mil categorias, e gerar resultados melhores que o estado da arte na época.

[Chen and Wang 2004] trabalham em um escopo um pouco diferente: utilizam-se de um algoritmo de aprendizado por instâncias múltiplas para categorizar imagens por região. As regiões nada mais são que instâncias correspondentes às segmentações por

cores dominantes de uma mesma imagem. Segmentação por cores, além da forma como é utilizada no *framework* proposto neste artigo, é também uma ótima abordagem para clusterização de imagens.

Em relação à clusterização de imagens, [Schroff et al. 2015] apresentam um método de verificação e reconhecimento de rostos mapeando as faces em um espaço euclidiano onde as distâncias correspondem à similaridade entre os rostos. O *FaceNet* também utiliza uma rede neural convolucional profunda com o objetivo de otimizar a incorporação de métodos tradicionais como vetores de *features*.

3. Contribuição

As atividades feitas para a primeira metade do projeto foram divididas em três grandes grupos anteriormente citados: o trabalho com dados, a análise exploratória e o estudo de técnicas. No trabalho com dados, as bases foram coletadas e tratadas, na análise exploratória fez-se um estudo geral sobre os dados obtidos e, por último, no estudo de métodos aplicáveis, avaliou-se possíveis algoritmos para a classificação e clusterização das imagens. As partes serão detalhadas de maneira aprofundada nas próximas seções.

3.1. Dados

O dados escolhidos para este projeto são pôsteres de filmes, série e animações retirados dos sites IMDB ¹ e MyAnimeList ². A obtenção do dados foi feita via *web scraping* usando um script em *Python* para facilitar possíveis obtenções futuras de mais instâncias e *features*, se, por exemplo, o número de instâncias atual não for suficiente ou estiver desproporcional para alguma base ou categoria.

Os dados foram divididos em duas bases, uma para cada site. No entanto, as *features* extraídas são as mesmas para as duas bases, assim como o número de instâncias extraídas é muito próximo. Dessa forma, os algoritmos podem ser treinados com as bases de forma junta ou separada sem que o peso de uma classe influencie os modelos implementados.

Ao total, 9 categorias foram selecionadas, sendo: comédia, ação, aventura, terror, romance, *sci-fi*, drama, mistério, fantasia. Também por motivos de balanceamento, foram escolhidas categorias mais populares e que existem com grande quantidade de dados em ambas as bases.

As figuras em 3.1 são dois exemplos de pôsteres retirados dos sites – o filme *Toy Story 4* foi retirado do site IMDB e a animação *Magia Record: Mahou Shoujo Madoka★Magica Gaiden* foi retirada do site MyAnimeList. Além das imagens, também foram extraídos os títulos, os anos de lançamento e a lista de gêneros para cada instância.

O tratamento dos dados foi igual para ambas as bases: foram retirados quaisquer instâncias que não possuíssem os 4 atributos base, foram ajustadas as codificações e acentos nos títulos e todas as instâncias duplicadas por base foram removidas.

¹<https://www.imdb.com/>

²<https://myanimelist.net/>



Figura 1. Toy Story 4



Figura 2. Magia Record: Mahou Shoujo Madoka Magica Gaiden

Figura 3. Exemplos de pôsteres contidos nas bases de imagens

3.2. Análise Exploratória

Nesta seção serão apresentados gráficos que geraram análises iniciais sobre as imagens extraídas. O objetivo dessas análises é obter um melhor entendimento sobre os dados extraídos antes de aplicar algoritmos, conferindo a necessidade de usar outras técnicas nos dados antes de colocá-los em modelos preditivos, e achar curiosidades para as quais não é necessário nenhum tipo de ferramenta mais complexa.

Para fins comparativos, os mesmos gráficos e análises foram replicados em ambas bases. Os gráficos iterativos podem ser encontrados neste link ³.

3.2.1. IMDB

No total, a base do site IMDB traz 4943 instâncias. Foram selecionados os 1000 programas de TV mais populares de cada uma das categorias, incluindo as instâncias duplicadas – que foram removidas em um passo posterior. Esta base é uma coletânea de filmes e séries de 1922 a 2022.

A figura 4 é uma análise da quantidade de filmes e séries lançados a cada ano. Comparando com o gráfico para a base do MyAnimeList, figura 7 da próxima subseção, podemos notar que a produção de programas de TV é mais esparsa que a de animações, observando-se a discrepância entre as quantidades de um ano para outro.

Outra curiosidade é a quantidade de filmes e séries lançados no ano de 2018: foi um número recorde encontrado em ambas as bases, com um total de 495 instâncias. Apesar do motivo para o aumento de programas no ano não ser explícito, é interessante saber que só a franquia *Marvel*, responsável por uma série de filmes de super-heróis que

³<https://irscunha.github.io/pocl/>

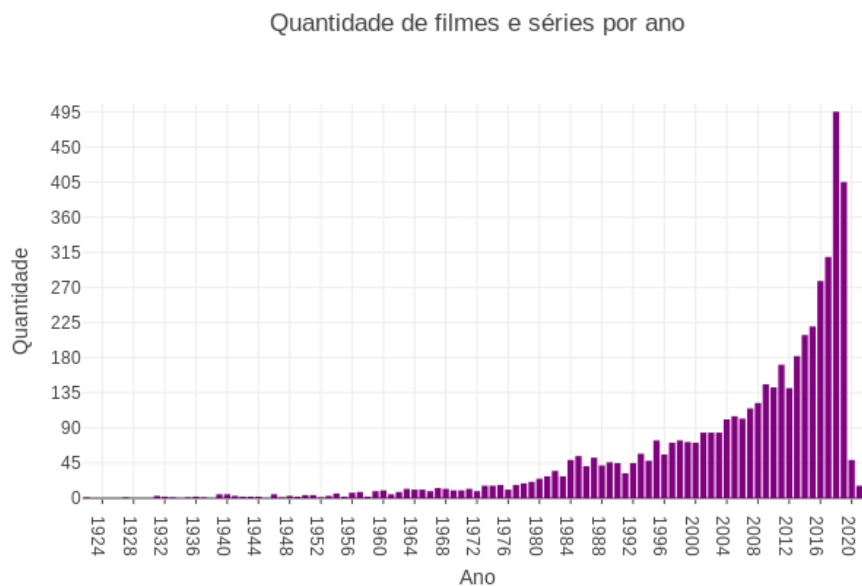


Figura 4. Análise da quantidade de filmes e séries lançados por ano

compartilham o mesmo universo e que quebram recordes de bilheteria, lançou 9 filmes em 2018.

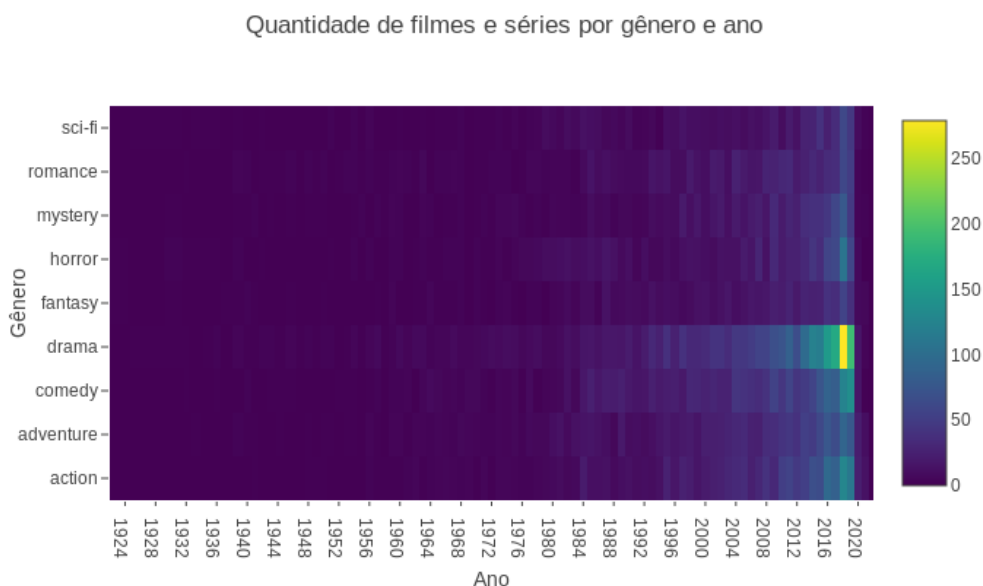


Figura 5. Análise da quantidade de filmes e séries lançados por ano, separado por gênero

A figura 5 traz outra visão com as mesmas informações do último gráfico: a quantidade de filmes e séries por ano, agora divididos por gênero. Também sem muitas explicações claras do porquê, o gênero drama se sobressai em relação aos demais com

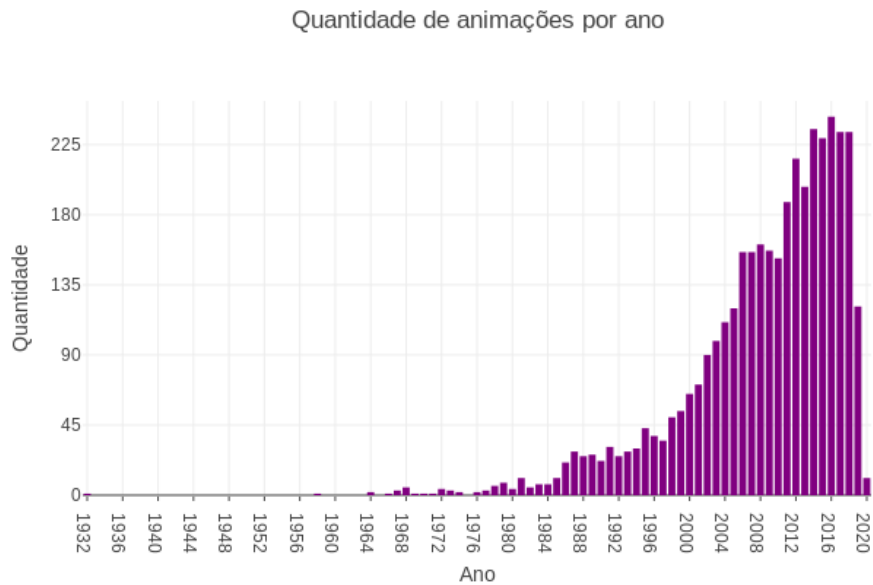


Figura 7. Análise da quantidade de animações lançadas por ano

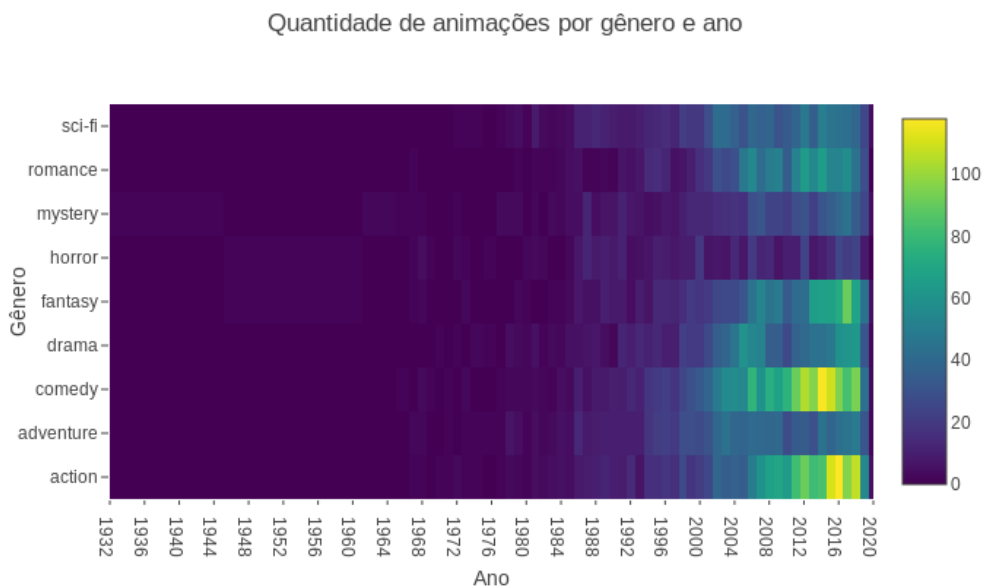


Figura 8. Análise da quantidade de animações lançados por ano, separado por gênero

Por fim, no *Wordcloud* da figura 9 existe um grande contraste em relação ao visto na figura 6. É notável a divisão das animações entre os filmes, *OVA*s – especiais que geralmente são exclusivos de mídias físicas, e episódios especiais. Também existe um grande destaque para partículas da língua japonesa, como “ga”, “ni” e “wa”, e para as sequências, como em “2nd Season”.

tentar categorizar pôsteres por palavras contidas nos títulos à partir das imagens.

Além da classificação de imagens, algumas abordagens interessantes podem ser feitas utilizando descritores de cores, métodos que capturam a distribuição espacial de cores em uma imagem, e ferramentas para extração de cores dominantes. Das aplicações vistas, pode-se pensar na classificação das imagens por *image segmentation* e no estudo da relação entre matrizes de cores e gêneros ou intervalos de tempo.

Por último, para a clusterização de imagens seria ideal aplicar e comparar métodos amplamente difundidos, como *K-Means* e *Transfer Learning*, com uma rede neural convolucional para treinamento não supervisionado.

4. Conclusão

Este trabalho de conclusão de curso tem como objetivo final fazer um estudo sobre relações que podem ser encontradas em diferentes bases de imagens. Ele foi dividido em duas partes; na primeira delas, o objetivo é extrair e analisar as imagens e fazer um estudo sobre possíveis algoritmos e técnicas que podem ser aplicados à esses dados.

Foram coletados aproximadamente 9000 pôsteres de filmes, séries e animações, juntamente com informações sobre seus títulos, gêneros e anos de lançamento. A análise desses dados permitiu averiguar a distribuição dos dados por gênero e por ano.

Como estudo, conclui-se que para a tarefa de classificação das imagens uma rede neural convolucional deverá ser usada, mas existem diversas possibilidades de rótulos diferentes, como gênero, ano e cores. Para clusterização de imagens, existem vários algoritmos que são estado da arte nesta tarefa e que podem ser facilmente comparados; ainda, existe a possibilidade de uso das redes neurais convolucionais para aprendizado não supervisionado. Por fim, pretende-se fazer um estudo de cores, que pode gerar insumos tanto para classificação quanto para clusterização das imagens.

Para segunda parte do projeto, deseja-se definir quais algoritmos serão usados e aplicar os conhecimentos práticos em uma análise aprofundada sobre os dados que explicita correlações entre os diferentes tipo de pôsteres.

Referências

- [Chen and Wang 2004] Chen, Y. and Wang, J. Z. (2004). Image categorization by learning and reasoning with regions. *J. Mach. Learn. Res.*, 5:913–939.
- [Chu and Guo 2017] Chu, W.-T. and Guo, H.-J. (2017). Movie genre classification based on poster images with deep neural networks. In *Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes*, MUSA2 '17, pages 39–45, New York, NY, USA. ACM.
- [Ivasic-Kos et al. 2015] Ivasic-Kos, M., Pobar, M., and Ipsic, I. (2015). Automatic movie posters classification into genres. *Advances in Intelligent Systems and Computing*, 311:319–328.
- [Krizhevsky et al. 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

[Schroff et al. 2015] Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823.