UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA
EEL7513 - Introdução ao Aprendizado de Máquina
Professor: Danilo Silva

Proposta de Projeto Final

Aplicação de CNNs para a classificação multi-label de peças de roupa

Equipe
Kaue Cano Souza - 18204680
Ruan Cardoso Comelli - 201901993

Florianópolis, outubro de 2019

Tipo de tarefa

Neste projeto, a máquina deve aprender a classificar peças de roupa segundo diversos atributos. Esse problema é uma generalização da classificação *single-label*, em que, dada uma entrada, o modelo deve identificar a qual classe ela pertence. Um exemplo de conjunto de dados projetado para a classificação *single-label* é o FashionMNIST. Em contraste, em uma classificação de atributos de uma imagem, o modelo deve extrair características adicionais do produto como cor, cortes, estilo e peças adicionais contidas na foto. Para facilitar a visualização da problemática com modelos de categorização simples, a Figura 1 ilustra imagens de roupas em e-commerces de moda.

Figura 1: Exemplos de imagens de roupa em e-commerces de moda.

Como pode ser visto, marcas de grife raramente expõem produtos por si só nos seus websites. Ainda mais, a maneira que as peças de roupas são demonstradas varia drasticamente de marca para marca e até entre categorias de uma própria marca. Tem-se desde *fullbody shots* de um *look* neutro, dando bastante destaque para o produto ofertado, até *looks* completos integrados por outras peças de roupas não ofertadas, ou mesmo fotos cortadas de apenas uma área do produto.

Portanto fica claro que o dataset FashionMNIST, representado na Figura 2, é inadequado para aplicações no e-commerce em virtude da simplicidade das imagens.

Figura 2: Exemplos do FashionMNIST

Tendo isso em vista, este projeto visa construir uma CNN com elevado desempenho na categorização de imagens produtos encontrados em grandes marcas de moda feminina.

O aprendizado empregado será supervisionado, utilizando datasets públicos resultantes de esforços científico-comunitários. Adicionalmente, diferentes modelos serão comparados, já que o problema da categorização pode ser abordado de diferentes maneiras, como detalhado ao decorrer desta proposta.

Formato dos dados

O conjunto de dados consistirá de m exemplos, cada um sendo um par (x,y), em que x é uma imagem colorida de tamanho ainda não especificado. O vetor y depende da forma com que o problema será resolvido, e y é um vetor de K componentes, sendo K o número de classes a que cada imagem pode pertencer. A k-ésima componente de y é y_k=1, caso a imagem pertença à classe k, ou y_k=0, caso contrário. Por ser um problema multi-label, o vetor y pode possuir um número variável de componentes nulas.

Do dataset a ser selecionado, serão extraídos conjuntos de treinamento, de validação e de teste. O primeiro conjunto é aquele sobre o qual os parâmetros do modelo serão aprendidos. Por sua vez, o conjunto de validação será usado para se otimizarem hiperparâmetros. O conjunto de teste, por fim, será utilizado para se avaliar o desempenho do modelo e compará-lo com outros classificadores. Dados próprios, compilados de sites de moda, também serão utilizados como conjunto de teste.

Conjuntos de dados

Após uma revisão dos datasets disponíveis gratuitamente, foram levantadas as seguintes possibilidades.

DeepFashion

Liu *et al.* (2016): Powering Robust Clothes Recognition and Retrieval with Rich Annotations

Figura 3: Exemplos de imagens contidas no conjunto DeepFashion

Conjunto de dados extraídos principalmente dos sites da Forever21 e Mogujie, bem como resultados de buscas no Google Images. Ao total, tem-se 800.000 imagens classificadas em 50 categorias (humanamente adquiridas) e 1.000 atributos descritivos. Cada imagem possui de 4 a 8 atributos. Esse dataset introduz benchmarks também. Em cima desse dataset, a equipe por trás do projeto desenvolveu o FashionNet, um grupo de 3 CNNs aplicadas para recomendação de looks. Ademais, as imagens têm tamanhos diferentes. O lado maior de cada imagem foi fixado em 300pt, mantendo a razão de aspecto da fonte. O subset próprio para a aplicação descrita neste documento é o DeepFashion: Attribute Prediction, o qual contém 289.222 imagens de roupa com um total de 1.000 atributos.

Fashion550k

Inoue *et al.* (2017): Inoue et al. (2017): Multi-Label Fashion Image Classification with Minimal Human Supervision

Figura 4: Exemplos de imagens contidas no conjunto Fashion550k

Este conjunto é composto por 550.661 exemplos, cada um contendo uma imagem e um conjunto de rótulos associados. Esses dados foram extraídos de postagens encontradas na internet provenientes de diversas partes do mundo, sendo assim irrespectivas da sua localização. Existem 66 classes de produtos possíveis, sendo que uma imagem pode pertencer a mais do que uma classe, e os rótulos do conjunto de dados foram extraídos automaticamente das publicações, contendo um ruído significativo. Por isso, os autores rotularam manualmente um subconjunto dos dados, e treinaram um modelo nesse subconjunto para aprender a remover o ruído. O conjunto de dados final foi então obtido por meio da aplicação desse modelo ao restante do conjunto.

Trabalhos Relacionados

Por meio de buscas acadêmicas, as seguintes referências foram compiladas por sua relevância para a aplicação proposta:

Wang *et al.* (2018): Wang et al. (2018): Attentive Fashion Grammar Network for Fashion Landmark Detection and Clothing Category Classification

Nesse artigo, é proposto uma técnica mais moderna em termos de arquitetura de rede para a classificação multiclasse, com foco em características concretamente visuais, e pouco baseadas em estilo em si. Os princípios das LSTMs é expandido para empregar também blocos de *attention* e *transformers*. Apesar de ser uma abordagem interessante para a solução do problema, podendo trazer resultados promissores, adotar os passos descritos neste artigo elevariam muito a necessidade computacional da rede, bem como acaba sendo um passo grande demais para o escopo deste projeto.

Lao *et al.* (2014): Lao e Jagadeesh (2016): Convolutional Neural Networks for Fashion Classification and Object Detection

Um dos primeiros artigos dedicado à classificação mais completa de peças de roupa. Escrito por alunos de uma disciplina similar em Stanford, o documento foca mais na visão macro dos problemas enfrentados e, em diversos tópicos como datasets, arquitetura de rede e infraestrutura empregada, este artigo já está ultrapassado.

Bossard *et al.* (2012): Apparel Classification with Style

Artigo publicado na Computer Vision – ACCV 2012 (Asian Conference on Computer Vision), foi um dos pioneiros na introdução de atributos múltiplos.

O núcleo do método consiste em um classificador multiclasse baseado em uma Random Forest que usa learners fortemente discriminativos como nós de decisão. Para tornar o pipeline o mais automático possível, também integra dados de treinamento rastreados automaticamente da Web no processo de aprendizado. Para avaliação, definiram-se 15 classes de roupas e introduziu-se um conjunto de dados de referência para a tarefa de classificação de roupas que consiste em mais de 80.000 imagens, disponibilizadas ao público.

Para cada imagem, o algoritmo seleciona uma janela de interesse que contenha a parte da roupa que está sendo classificada.

Wei *et al.* (2014): CNN: Single-label to Multi-label

Neste artigo, os autores propõem uma solução para o problema de classificação *multi-label* utilizando a solução para a *single-label*. Para isso, é aplicada a técnica BING para a detecção de objetos nas imagens. Cada objeto corresponde a uma *hipótese* que se faz quanto às classes presentes na imagem. Cada objeto é então classificado em um problema multiclasse por uma rede convolucional, e a solução do problema *multi-label* é a união de todas as classificações *single-label*.

Inoue *et al.* (2017): Multi-Label Fashion Image Classification with Minimal Human Supervision

Para a geração do conjunto de dados já mencionado (Fashion550k), os autores desse trabalho desenvolveram uma arquitetura utilizaram a arquitetura ResNet-50, disponível na biblioteca Keras e pré-treinada no conjunto ImageNet, finalizada com uma camada de 66 unidades sigmoidais, cada uma capaz de prever a probabilidade de uma das 66 classes do problema *multi-label*.

Técnicas auxiliares foram aplicadas para se filtrarem ruídos nos rótulos das imagens no conjunto de treinamento, mas isso está fora do escopo deste trabalho.

Ferreira *et al.* (2018): A Unified Model with Structured Output for Fashion Images Classification

Publicação do ISR (Instituto Superior Técnico, Universidade de Lisboa, Portugal) em parceria com a marca de luxo portuguesa Farfetch.

O conjunto de dados utilizado corresponde a um banco de dados pertencente à empresa Farfetch contendo imagens de e-commerce classificadas com cinco níveis de categorização por produto. Não foram utilizados os conjuntos Fashion550k e DeepFashion por conta da limitação ao uso comercial. Por isso, os modelos foram pré-treinados no ImageNet.

Os autores desenvolveram a CNN representada na Figura 4. Essa arquitetura inicia-se com uma ResNet-50 disponível na biblioteca Keras e pré-treinada no conjunto ImageNet. Em seguida, o fluxo de informação se divide em três, cada um representando um nível de categorização. Em cada linha, tem-se um perceptron multi-camadas. Para os níveis de categoria e sub-categoria, a rede culmina em uma ativação *softmax*, já que esses níveis correspondem à classificação multiclasse. Por outro lado, o nível de atributos considera um conjunto de ativações sigmoidais, o que corresponde a uma classificação *multi-label*.

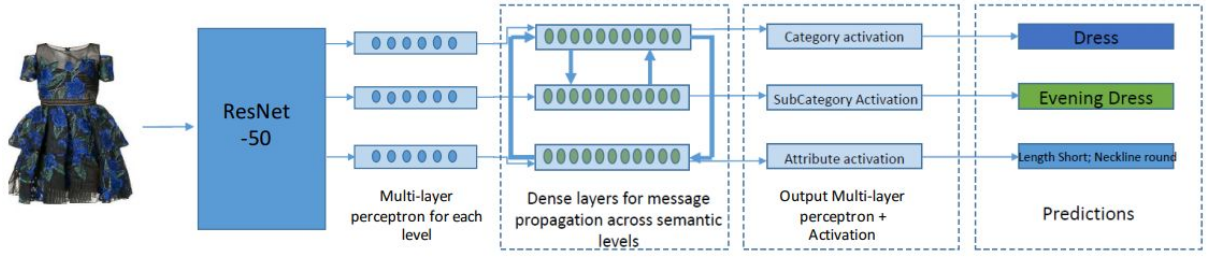

Figura 4: Representação da arquitetura desenvolvida.

Já que pode haver uma dependência entre os diferentes níveis de categorização (por exemplo, uma peça só pode pertencer à sub-categoria *vestido para o dia-a-dia* caso corresponder à categoria *vestido*), os autores incluíram um mecanismo de transporte de mensagem nas camadas densas.

Mercado

As seguintes empresas oferecem serviços de API para uso de modelos classificadores prontos para peças de roupa: Markable, Imagga e Ximilar.

A intenção da equipe é utilizar as amostras grátis destes serviços para

comparação em nível de mercado do modelo entregue por este projeto. Com isso, é possível estimar um valor aproximado para o nosso modelo final. Esta etapa é dependente da resposta hábil das empresas acima citadas, então não é algo certo.

Objetivos

Tendo em vista as diversas aplicações relacionadas anteriormente, a equipe chegou às seguintes propostas de projetos que seriam relevantes para a comunidade científica:

1. Construir uma CNN ResNet-50 considerando os conjuntos DeepFashion e Fashion550k, que não foram estudados em conjunto/contraste pelos autores previamente citados.
2. Expandir o mesmo trabalho usando diferentes redes pré-treinadas no lugar da ResNet-50, como VGG e SVM + Random Forest. Pretende-se aplicar as demais redes disponíveis na biblioteca Keras.

É interessante, também, dividir o problema em partes para um melhor fluxo de desenvolvimento e testes. Inicialmente, a equipe focará na categorização multi-classe de imagens de visuais completos vistos no Fashion550k, ou seja, detectar múltiplas peças de roupa em uma mesma imagem. A partir daí, com um modelo que cumpra as intenções da primeira parte, a equipe estenderá o projeto para a categorização de atributos, por exemplo, classificando um *look* como "calça, camiseta, vermelho, preto, mickey, couro". Nessa segunda etapa, o conjunto de dados utilizado será o DeepFashion: Attribute Prediction.

Os artigos listados anteriormente já introduziram métricas para o desempenho dos seus modelos. Portanto é possível fazer a comparação entre os modelos desenvolvidos neste trabalho e aqueles construídos na literatura.

Recursos computacionais

Em um primeiro momento, pretende-se utilizar a plataforma Google Colab para o desenvolvimento do código Python. Caso se mostre necessário, o programa também pode ser executado na plataforma IBM Cloud, à qual um dos autores tem acesso.

Para trabalhos futuros com redes mais complexas, maior poder computacional se torna imprescindível. Muito provavelmente, a plataforma IBM Cloud será aplicada para solucionar esse problema.

Full disclosure

Este projeto será parte do TCC do integrante Kauê Cano, orientado pelo Professor Danilo Silva. Os modelos desenvolvidos também servirão de base para entregas da Elint, empresa de um dos integrantes da equipe, para clientes do ramo de fabricação e varejo de roupas femininas.

Apenas os alunos aqui citados estão envolvidos no projeto definido por esta proposta.

Referências Bibliográficas

1. Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X., 2016. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1096-1104).
2. Kiapour, M.H., Yamaguchi, K., Berg, A.C. and Berg, T.L., 2014, September. Hipster wars: Discovering elements of fashion styles. In European conference on computer vision (pp. 472-488). Springer, Cham.
3. **Inoue, N., Simo-Serra, E., Yamasaki, T. and Ishikawa, H., 2017. Multi-label fashion image classification with minimal human supervision. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2261-2267).**
4. **Ferreira, B.Q., Baía, L., Faria, J. and Sousa, R.G., 2018. A unified model with structured output for fashion images classification. arXiv preprint arXiv:1806.09445.**
5. Wang, W., Xu, Y., Shen, J. and Zhu, S.C., 2018. Attentive fashion grammar network for fashion landmark detection and clothing category classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4271-4280).
6. Lao, B. and Jagadeesh, K., 2016. Convolutional neural networks for fashion classification and object detection. CCCV 2015: Computer Vision, pp.120-129.
7. Bossard, L., Dantone, M., Leistner, C., Wengert, C., Quack, T. and Van Gool, L., 2012, November. Apparel classification with style. In Asian conference on computer vision (pp. 321-335). Springer, Berlin, Heidelberg.
8. Wei, Y., Xia, W., Huang, J., Ni, B., Dong, J., Zhao, Y. and Yan, S., 2014. Cnn: Single-label to multi-label. arXiv preprint arXiv:1406.5726.