ORGANIZAÇÃO DE BASE DE IMAGENS PARA USO EM TRABALHOS DE PESQUISA PARA FINS DE CLASSIFICAÇÃO

Carlos Masashi Kanda (PIBIC/CNPq/FA/Uem), Yandre Maldonado e Gomes da Costa (Orientador), e-mail: yandre@din.uem.br.

Universidade Estadual de Maringá / Centro de Tecnologia/Maringá, PR.

Ciências Exatas e da Terra / Ciência da Computação

Palavras-chave: base de imagens, classificação de imagens, rede neural convolucional.

Resumo:

Este projeto tem como proposta a criação de uma base de imagens para utilização em pesquisas de classificação de imagens. Para isso se utilizou o conjunto de imagens ImageNet e criou-se uma base com 59 classes de comidas totalizando mais de 59 mil imagens considerando critérios como: quantidade mínima de imagens por classe, existência do prato em nosso país, entre outros. Adicionalmente, testes com redes neurais convolucionais foram realizados sobre a base.

Introdução

Este projeto está ligado ao projeto de pesquisa “Uma Plataforma Avançada para Recomendação Automática de Pontos de Interesse em Cidades Brasileiras”. Nele, está prevista a ampliação das funcionalidades de uma plataforma para recomendação de pontos de interesse.
Uma possibilidade para isto é a utilização de classificadores de imagens baseados em técnicas de inteligência computacional. Para isto, é muito importante que se tenha recursos organizados a fim de viabilizar a proposição e experimentação de abordagens sobre os dados. Mas, é muito comum que surjam dificuldades na obtenção de uma base de dados adequada voltada ou ao domínio de aplicação específico em que se pretende trabalhar. É frequente problemas relacionados à quantidade de amostras presentes e a falta de balanceamento entre as diferentes classes.
Neste projeto, tem-se como objetivo principal a organização de uma base de imagens com categorias especificamente identificadas e voltadas para viabilizar a exploração de imagens encontradas no perfil de usuários de redes sociais. Também poderão ser propostos modelos de classificadores baseados em *deep learning* que extraiam informações úteis para a categorização das imagens tendo em vista melhorar a qualidade da resposta emitida por um sistema de recomendação.

Materiais e métodos

Formulação da base

Primeiramente, identificou-se conjuntos de imagens em potencial disponíveis para uso. Obteve-se os conjuntos em potencial: ImageNet, Food101, UEC FOOD 256, Tiny Images Dataset, USC-SIPI Image Database e o Flickr.
O conjunto mais relevante, o ImageNet (Deng et al., 2009), possui mais de 14 milhões de imagens, e é organizado com base na hierarquia da WordNet, possui validação das imagens (feita através do Amazon Turkey Machine).
Em seguida, foi realizada uma filtragem dos conjuntos ao se analisar o que poderia se aproveitar de cada um, e considerando que esse projeto em específico foi desenvolvido com o intuito de formar uma base para posterior utilização na classificação de comidas encontradas em estabelecimentos alimentícios no Brasil e possivelmente a utilização de modelos de redes neurais convolucionais para tal. Buscou-se a construção de uma base com o máximo possível de imagens distribuídas de forma equilibrada entre as classes. Escolheu-se a utilização dos recursos da ImageNet devido as características mencionadas dela. Dos motivos para se desconsiderar as outras bases estão: poucas imagens, não possuir imagens de comida, baixa resolução, falta de verificação manual e já estar pronta para uso em classificadores.
A ImageNet busca disponibilizar em torno de mil imagens por conjunto, tal valor foi estabelecido como quantidade mínima de imagens por classe. Dentro da estrutura hierárquica em árvore do ImageNet foi realizada uma busca manual sobre todos os filhos de Misc e sob todos os descendentes do nó *"food, nutrient"*, filho de *Misc*. Baseado nos critérios levantados, foram selecionados conjuntos considerados visualmente não semelhantes a outros já selecionados.

Rede Neural

As imagens foram estendidas para *227 x 227 pixels*, e mil de cada classe foram separadas em dados de treino e de teste 70% e 30%, respectivamente. Realizaram-se 7 testes com variações de rede neurais relativamente semelhantes ou de quantidade de classes de imagens.
O primeiro teste foi realizado com uma rede de 4 camadas convolucionais, todas com 64 filtros de dimensão 4x4 e *ReLu*. Cada uma seguida de uma camada de *max pooling*, exceto a última. Também, 2 camadas fortemente conectadas. A primeira com 256 neurônios, *ReLu* e um *dropout* de 50% e a segunda com 62 neurônios (um para cada classe de comida) com a operação de *softmax* para obter a predição.
O segundo, sexto e sétimo teste foram realizados com uma rede semelhante a primeira, mas as camadas convolucionais possuíam dimensão 5x5. 512 foi a quantidade de neurônios da primeira camada fortemente conectada e a última teve 16, 32 e 45 neurônios, respectivamente. O terceiro teste utilizou-se de uma rede neural idêntica a segunda, mas com uma camada

convolucional extra, com *max pooling*. O quarto teste possui uma rede semelhante ao segundo, mas com apenas 2 camadas convolucionais, E, a primeira camada fortemente conectada com 256. O quinto teste foi realizado sobre uma rede semelhante à do quarto teste, mas com uma camada convolucional idêntica extra, sem *max pooling*. Para todos foi utilizada a técnica de *data augmentation* nas imagens (zoom de até 20%, inclinação da imagem em até 0,2 radianos, giro horizontal e reescala sobre a cor). O primeiro teste foi executado com 50 épocas, o último com 60 épocas e os restantes com 90 épocas cada. A acurácia da parte de teste parou de evoluir muito antes da última época em todos os testes.

Resultados e Discussão

A base C59 então ficou formada pelas seguintes *labels* (rótulos): *Baguet, Baked potato, Barbecue, Barbecued spareribs, Beef stew, Beef wellington, Boiled egg, Breadstick, Brie, Brioche, Burrito, Cannelloni, Chicken soup, Chocolate fondue, Chow mein, Club sandwich, Cornbread, Creme caramel, Crescent roll, Croquette, Cross bun, Curry, French fries, Fried egg, Fruit salad, Gelatin, Hamburger, Hotdog, Ice cream, Jam, Kabob, Lasagna, Mashed potato, Meatball, Mousse, Muffin, Omelet, Pea soup, Pizza, Poached egg, Popcorn, Pudding, Ratatouille, Risotto, Sashimi, Schnitzel, Shrimp cocktail, Soft pretzel, Souffle, Spaghetti, Stuffed tomato, Sweet corn, Taco, Tempura, Teriyaki, Toast, Truffle, White rice, Yogurt*. Atualmente, o menor conjunto possuí 1024 imagens, o maior 1522 e a média é de 1239.
Além disso, criou-se alguns metadados que podem ser úteis a quem for utilizar a base. Tais são: nomes das classes, mapeamento da base relativa a hierarquia do ImageNet, mapeamento das classes para seus identificadores no ImageNet, quantidade de imagens por classe, rótulos de cada classe, informações gerais sobre a base, termos de uso, e separador de imagens em treino e teste para mil imagens escolhidas aleatoriamente para cada classe, 70% e 30%, respectivamente.

Tabela 1

Experimento	Quantidade de classes	Épocas	Acurácia no treino	Acurácia no teste
Teste 1	62	50	34,51%	32,43%
Teste 2	16	90	85,78%	54,63%
Teste 3	16	90	83,11%	54,15%
Teste 4	16	90	79,67%	47,96%
Teste 5	16	90	78,79%	49,27%
Teste 6	32	90	63,72%	41,83%
Teste 7	45	60	53,09%	36,07%

Observando os resultados dispostos na Tabela 1 podemos observar que, possivelmente, os resultados dos testes 2, 6 e 7 indicam que para maiores

volumes de dados redes mais robustas são necessárias. Nesta sequência, os testes 7, 6 e 2 demonstram resultados progressivamente melhores, onde a única mudança é o tamanho da base. O teste 1 não utiliza a mesma rede, mas possui uma robustez semelhante à dos anteriores e demonstra resultados ainda piores que o teste 7, reforçando a hipótese.
Os testes 3, 4 e 5 são todos realizados sobre as mesmas 16 classes como o teste 2. Desta vez ocorrem variações sobre o modelo da rede neural. A rede do terceiro teste possuía uma camada convolucional a mais, mas não demonstrou ganhos sobre a acurácia. Enquanto que as redes dos testes 4 e 5 eram mais simples comparadas a rede do teste 2 e demonstraram uma acurácia, em média, 6,5% e 6,01% menor para treino e teste, respectivamente.

Conclusões

Ao se trabalhar com pesquisas de classificação utilizando-se de inteligência computacional, encontrar uma base de dados organizada e adequada para pronto uso pode se mostrar difícil. Neste trabalho organizou-se uma base, nomeada C59, com o intuito de futuramente ser utilizada para ajudar o projeto mencionado na introdução, assim como outras pesquisas.
Os experimentos demonstraram que a robustez da rede de um classificador adequado para a base proposta deve ser superior às utilizadas nos experimentos deste projeto, pois os resultados foram ruins se comparados a resultados de outros trabalhos, como a rede neural *AlexNet* definida por Krizhevsky, Sutskever e Hinton (2012), cuja taxa de erro é de apenas 15,3%.

Agradecimentos

Agradeço à Fundação Araucária, ao CNPQ e à UEM pela bolsa e oportunidade. Ao meu orientador Yandre pela oportunidade e experiência.

Referências

DENG J. et al. **ImageNet: A Large-Scale Hierarchical Image Database.** IEEE Computer Vision and Pattern Recognition (CVPR), 2009. Disponível em: <**https://ieeexplore.ieee.org/document/5206848/**>. Acesso em: 31 jul. 2018.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. **Imagenet classification with deep convolutional neural networks**. Pereira, F.; Burges, C. J. C.; Bottou, L.; Weinberger, K. Q., eds. Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012. P. 1097- 1105. Disponível em: <**https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pd**f>. Acesso em: 31 jul. 2018.