

Estimação de Idade de Telespectadores para Aplicações de Sugestão de Conteúdo em *Smart TVs*

Nicoli Pinheiro de Araújo, Elloá B. Guedes

¹ Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – Amazonas

{npda.eng, ebgcosta}@uea.edu.br

Abstract. *This meta-paper describes the style to be used in articles and short papers for SBC conferences. For papers in English, you should add just an abstract while for the papers in Portuguese, we also ask for an abstract in Portuguese (“resumo”). In both cases, abstracts should not have more than 10 lines and must be in the first page of the paper.*

Resumo. *Este meta-artigo descreve o estilo a ser usado na confecção de artigos e resumos de artigos para publicação nos anais das conferências organizadas pela SBC. É solicitada a escrita de resumo e abstract apenas para os artigos escritos em português. Artigos em inglês deverão apresentar apenas abstract. Nos dois casos, o autor deve tomar cuidado para que o resumo (e o abstract) não ultrapassem 10 linhas cada, sendo que ambos devem estar na primeira página do artigo.*

1. Introdução

2. Trabalhos Relacionados

A aplicação de redes neurais convolutivas em problemas de classificação e detecção de objetos em imagens têm obtido resultados significativamente positivos. Em (SIMONYAN; ZISSERMAN, 2014), (HE et al., 2016), (SZEGEDY et al., 2015), (REDMON et al., 2016), (LIU et al., 2016) e outros, são descritas arquiteturas robustas capazes de detectar dezenas de objetos em várias situações. Treinadas com conjuntos de dados visuais que contam com milhares de exemplos como a ImageNet, Pascal VOC e COCO, estas redes são conhecidas por seu bom desempenho. Algumas destas redes foram afinadas utilizando conjuntos de dados menores e especializados para a tarefa de estimação de idade aparente.

O trabalho de (ROTHER; TIMOFTE; GOOL, 2015) relata um método para estimação de idade aparente em imagens de faces imóveis utilizando *deep learning*. Propõe-se um conjunto de 20 redes neurais convolucionais classificadoras com arquiteturas VGG-16 pré-treinadas com a base de dados visuais ImageNet, e ajustadas utilizando imagens disponibilizadas pelo IMDB, Wikipedia, e o conjunto de dados *Looking At People*–LAP para anotação de idade aparente. Cada modelo tem como saída um número discreto entre 0 e 100, representando a idade prevista. A saída final do modelo consiste na média entre as idades previstas pelos 20 redes. A solução atingiu um MAE (*Mean Average Error*) de 3.221 na fase de testes.

Em (LIU et al., 2015) cria-se um estimador composto pela fusão de um modelo regressor e outro classificador. A rede neural convolucional profunda *GoogLeNet* (ou *Inception*) (SZEGEDY et al., 2015) sofreu modificações em sua arquitetura, como adição de normalização do batch, remoção de camadas de *dropout* e perda. O conjunto de modelos conseguiu prever idades com MAE de 3.3345.

Ademais, é possível encontrar resultados satisfatórios para a tarefa de aprendizado proposta utilizando modelos menos complexos. Com o objetivo de consolidar um método de classificação de idade e gênero, (LEVI; HASSNER, 2015) propõe uma rede neural convolucional de natureza mais simples, se comparada com (SZEGEDY et al., 2015), (SIMONYAN; ZISSERMAN, 2014) ou (HE et al., 2016). Sua arquitetura consiste em três camadas convolucionais com *dropout* e funções de ativação *ReLU*, seguidas por três camadas totalmente conectadas. A camada de saída tem como função de ativação a Soft-max. A escolha por um design de rede menor é motivado pelo desejo de reduzir o risco de *overfitting* e pela natureza do problema, que contém apenas 8 classes de idade. O modelo é treinado utilizando apenas o conjunto de referência *Adience*, composto por imagens não filtradas para classificação de idade e gênero. Considerando uma margem de erro de uma classe vizinha, a melhor rede obteve acurácia de $84.7\% \pm 2.2$ ao empregar a técnica de sobre-amostragem.

3. Objetivo Geral

O objetivo geral deste trabalho consiste em propor um estimador de idade para telespectadores de *Smart TVs*. Para alcançar esta meta, alguns objetivos específicos precisam ser contemplados, a citar:

- Formular um referencial teórico sobre redes neurais convolucionais, modelo de *machine learning* considerado, contemplando suas características, principais arquiteturas, métodos de treinamento e teste;
- Consolidar uma base de dados para a tarefa de *machine learning* proposta, contemplando exemplos realísticos;
- Identificar tecnologias adequadas para implementar o estimador proposto;
- Propor, treinar e testar diferentes arquiteturas de redes neurais convolucionais para a tarefa em questão;
- Avaliar comparativamente os estimadores propostos.

4. Justificativa

A realização de um trabalho de conclusão de curso desta natureza é justificada por várias razões. No contexto da interação entre telespectador e *SmartTV*, um estimador pode ser utilizado para facilitar a coleta de informações que contribuam para facilitar a coleta de informações que contribuam para melhor experiência de provimento de conteúdo e de configurações personalizadas. Em particular, a estimação de idade dos telespectadores pode ser especialmente para a implementação de um controle parental mais eficiente, protegendo crianças e adolescentes de conteúdos inadequados à sua faixa etária.

Um outro aspecto que ressalta a importância da realização de um trabalho desta natureza é a prática e a proposição de soluções envolvendo *machine learning*. Esta é uma área de vanguarda na Computação e seu potencial para resolução de problemas práticos está em franco desenvolvimento. Ao considerar a elaboração do estimador proposto, será necessário dominar conhecimentos de ferramental tecnológico atual, o que pode colaborar na minimização da distância entre o profissional em formação e os anseios do mercado de trabalho da área.

Por fim, há que se mencionar a relação entre a área de pesquisa considerada neste trabalho de conclusão de curso e o Laboratório de Sistemas Inteligentes (LSI). Este trabalho alinha-se com os objetivos desta iniciativa do Núcleo de Computação (NUCOMP), motivando o desenvolvimento de uma solução inovadora que utiliza técnicas da Inteligência Artificial.

5. Metodologia

A metodologia para o desenvolvimento deste trabalho consiste na realização da *fundamentação teórica sobre machine learning*, em especial contemplando os conceitos relativos às redes neurais convolucionais. Para tanto, considerar-se-á a literatura desta área para que haja o entendimento das bases biológicas deste modelo computacional, como funcionam, quais as características e os modelos mais importantes. Neste estudo, além dos aspectos teóricos, serão considerados os ambientes de desenvolvimento, bibliotecas e outras tecnologias para implementação dos conceitos contemplados..

Os demais passos que compõem a metodologia deste trabalho baseiam-se no *fluxo de atividades de machine learning*, conforme (MARSLAND, 2015). Inicialmente, haverá a aquisição e o pré-processamento de imagens para *consolidar a base de dados* para esta tarefa de aprendizado. Nesta etapa, será considerada a literatura e, se possível, outras bases de dados já disponíveis e com licença livre de utilização.

A seguir, há a *proposição de diferentes modelos de redes neurais convolucionais* para a tarefa de aprendizado em questão. Nesta etapa, serão consideradas diferentes arquiteturas, parâmetros e hiperparâmetros de configuração. Estes procedimentos visam consolidar um espaço de busca de modelos que possam endereçar a tarefa de maneira mais eficiente.

O próximo estágio consiste no *treinamento das redes neurais convolucionais* para o problema em questão, considerando diferentes parâmetros de configuração. Durante este processo, uma parte da base de dados é apresentada para que haja o ajuste de pesos nas camadas que compõe o aprendizado de características necessárias para que os modelos propostos sejam capazes de realizar a tarefa de *machine learning* pretendida. O treinamento das redes ocorrerá utilizando computação em nuvem, tendo em vista a capacidade computacional necessária para realizar este procedimento.

Segue-se então o *teste das redes*, respeitando uma abordagem de validação cruzada e utilizando métricas de desempenho apropriadas. O objetivo desta fase consiste em aferir os modelos propostos e treinados quanto à sua capacidade de generalização.

Por fim, para identificação de um modelo mais adequado à esta tarefa, as *métricas de desempenho serão comparadas* e os melhores modelos elencados a partir destes valores, apontando assim um estimador apropriado para o problema inicialmente considerado.

Além destas atividades, há que se considerar a escrita da proposta e do projeto final do trabalho de conclusão de curso, bem como as defesas parcial e final, que precisam ser incorporadas no cronograma.

6. Cronograma

Uma visão geral do cronograma de atividades deste trabalho de conclusão de curso pode ser vista na Tabela 1. Elas possuem relação com a metodologia detalhada na seção 5, que detalha as regras e diligências estabelecidas para realizar este trabalho.

Referências

HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.

Tabela 1: Cronograma de atividades levando em consideração os dez meses (de 02/2018 a 12/2018) para a realização do TCC.

	2018											
	02	03	04	05	06	07	08	09	10	11	12	
Escrita da Proposta	X	X	X	X	X							
Fundamentação Teórica sobre Machine Learning	X	X	X	X								
Consolidação da Base de Dados		X	X									
Proposição de Modelos de Redes Neurais Convolucionais				X	X	X	X	X				
Defesa da Proposta					X							
Escrita do Trabalho Final						X	X	X	X	X	X	
Treinamento das Redes Neurais Convolucionais					X	X	X	X	X	X		
Teste das Redes Neurais Convolucionais					X	X	X	X	X	X	X	
Comparação de Metricas de Desempenho						X	X	X	X	X	X	
Defesa do Trabalho Final											X	

LEVI, G.; HASSNER, T. Age and gender classification using convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2015. p. 34–42.

LIU, W. et al. Ssd: Single shot multibox detector. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 21–37.

LIU, X. et al. Agenet: Deeply learned regressor and classifier for robust apparent age estimation. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. [S.l.: s.n.], 2015. p. 16–24.

MARSLAND, S. *Machine learning: an algorithmic perspective*. [S.l.]: CRC press, 2015.

REDMON, J. et al. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 779–788.

ROTHER, R.; TIMOFTE, R.; GOOL, L. V. Dex: Deep expectation of apparent age from a single image. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. [S.l.: s.n.], 2015. p. 10–15.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SZEGEDY, C. et al. Going deeper with convolutions. In: CVPR. [S.l.], 2015.