

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

Nicoli P. 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 work presents a proposal for estimating the age of viewers for content suggestion applications on Smart TVs using machine learning techniques. Such a tool can be used in a variety of ways, including to facilitate the collection of information that contributes to a better content delivery experience, to the creation and control of custom settings, and to the implementation of more efficient parental control.*

Resumo. *Este trabalho apresenta uma proposta para estimação de idade de telespectadores para aplicações de sugestão de conteúdo em Smart TVs utilizando técnicas de machine learning. Tal ferramenta pode ser utilizada de diversas maneiras, incluindo para facilitar a coleta de informações que contribuam para melhor experiência de provimento de conteúdo, para a criação e controle de configurações personalizadas e para a implementação de um controle parental mais eficiente.*

1. Introdução

As *Smart TVs* são o resultado da evolução tecnológica junto aos aparelhos de televisão domésticos. Possuem capacidades interativas ligadas à internet, acesso a conteúdo online, *e-commerce* de conteúdo televisivo, navegação web e acesso a redes sociais. Estes aparelhos podem ser equipados com câmeras e microfones embutidos e transmitem conteúdo 2D ou até mesmo 3D. Neste último caso, em particular, os telespectadores fazem uso de óculos especiais (NEWSROOM, 2011), (PERAKAKIS; GHINEA, 2015).

Segundo a Pesquisa Nacional por Amostra de Domicílios realizada pelo IBGE em 2015, foi observado um total de 103 milhões de aparelhos de televisões em residências e pontos comerciais, das quais 16 milhões são de *Smart TVs*. A pesquisa detalha que 94% destas *Smart TVs* foram adquiridas entre 2014 e 2015. Os números mostram um posterior aumento nas vendas de aparelhos televisores deste tipo, representando 68,2% do total de televisores vendidos no primeiro semestre de 2017 (IBGE, 2015).

Este aumento de vendas tem várias causas. Destacam-se os muitos benefícios resultantes do uso de *Smart TVs* quando comparadas aos aparelhos convencionais (SHIN; HWANG; CHOO, 2013), (BETWEEN, 2017). Em especial, cita-se o aumento da qualidade na transmissão, a utilização de aplicativos diversos e a possibilidade de acesso à conteúdo *online* e *on demand*, gratuitos ou mediante assinaturas. Além destes benefícios, cuja maioria é resultante da conectividade com a internet, outros fatores têm justificado o aumento das vendas e do interesse do público consumidor pelas *Smart TVs*, tais como o encerramento da transmissão de sinal analógico da televisão aberta, a Copa do Mundo 2018 e a tecnologia 4K (GUIMARÃES, 2017; BRAZILIENSE, 2018; CAPELAS, 2017).

Apesar da grande disponibilidade de conteúdo nas *Smart TVs* e sua grande difusão nos lares brasileiros, é imprescindível levar em conta as restrições e recomendações deste conteúdo para o público alvo a que se destina. Tendo isto em mente, o processo de classificação indicativa integra o sistema de garantias dos direitos da criança e do adolescente quanto a promover, defender e garantir o acesso a espetáculos e diversões públicas adequados à condição de seu desenvolvimento, mas reserva-se o direito final aos pais e responsáveis quanto à escolha do conteúdo adequado a estes (DEPUTADOS, 1995). No Brasil, a *Coordenação de Classificação Indicativa* (Cocind), vinculada ao Ministério da Justiça, é o órgão responsável pela classificação indicativa de obras destinadas à televisão e outros meios, incluindo até mesmo aplicativos.

Assim, buscando facilitar a interação entre telespectador e *Smart TV*, um estimador de idade pode ser utilizado para facilitar a coleta de informações que contribuam para melhor experiência de provimento de conteúdo e de configurações personalizadas, e pode ser especialmente útil para a implementação de um controle parental mais eficiente, protegendo crianças e adolescentes de conteúdos inadequados à sua faixa etária.

2. Trabalhos Relacionados

Segundo (FU; GUO; HUANG, 2010), a idade pode ser inferida a partir de padrões distintos que emergem através da aparência da face. Técnicas comuns para a estimação da idade envolvem a dedução de modelos matemáticos a partir do estudo do crescimento de medidas da face e do crânio (KWON; LOBO, 1999), da textura do rosto (LANITIS; TAYLOR; COOTES, 2002), da captura de tendências de envelhecimento a partir de várias imagens de indivíduos de mesma idade (FU; XU; HUANG, 2007) e a extração de características específicas relacionadas à idade (SUO et al., 2008), (LOU et al., 2018). Modelos de *machine learning* também são utilizados para a tarefa, em especial as redes neurais artificiais, K-vizinhos mais próximos e máquinas de vetores de suporte.

Recentemente, a aplicação de redes neurais convolucionais 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* (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 *Smart TV*, um estimador de idade pode ser utilizado 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 uma 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.

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

Referências

BETWEEN, D. *Difference between Smart TV and Normal TV*. 2017. <<http://www.differencebetween.info/difference-between-smart-tv-and-normal-tv>>. Acessado em 21 de Março de 2018.

BRAZILIENSE, C. *Copa e novas tecnologias prometem aumentar venda de TVs no Brasil em 2018*. 2018. <http://www.correiobraziliense.com.br/app/noticia/economia/2018/01/23/internas_economia,654966/copa-e-novas-tecnologias-prometem-aumentar-venda-de-tvs-no-brasil.shtml>. Acessado em 21 de Março de 2018.

CAPELAS, B. *Explosão no consumo de vídeos online coloca em xeque o futuro da televisão*. 2017. O Estado de S. Paulo. Acessado em 20 de Março de 2018. Disponível em: <<http://link.estadao.com.br/noticias/geral,explosao-no-consumo-de-videos-online-coloca-em-xeque-o-futuro-da-televisao,70001695828>>.

DEPUTADOS, C. dos. *Estatuto da Criança e do Adolescente*. BRASIL: [s.n.], 1995.

FU, Y.; GUO, G.; HUANG, T. S. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 32, n. 11, p. 1955–1976, 2010.

FU, Y.; XU, Y.; HUANG, T. S. Estimating human age by manifold analysis of face pictures and regression on aging features. In: IEEE. *Multimedia and Expo, 2007 IEEE International Conference on*. [S.l.], 2007. p. 1383–1386.

- GUIMARÃES, N. *Com fim do sinal analógico, busca por smart TVs cresce 11%*. 2017. <<http://www.leiaja.com/tecnologia/2017/07/17/com-fim-do-sinal-analogico-busca-por-smart-tvs-cresce-11/>>. Acessado em 22 de Março de 2018.
- 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.
- IBGE. *Pesquisa Nacional por Amostra de Domicílios: Acesso à Internet e à Televisão e Posse de Telefone Móvel Celular para Uso Pessoal*. 2015. <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv99054.pdf>>. Acessado em 16 de Março de 2018.
- KWON, Y. H.; LOBO, N. da V. Age classification from facial images. *Computer vision and image understanding*, Elsevier, v. 74, n. 1, p. 1–21, 1999.
- LANITIS, A.; TAYLOR, C. J.; COOTES, T. F. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 24, n. 4, p. 442–455, 2002.
- 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. Agetnet: 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.
- LOU, Z. et al. Expression-invariant age estimation using structured learning. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 40, n. 2, p. 365–375, 2018.
- MARSLAND, S. *Machine learning: an algorithmic perspective*. [S.l.]: CRC press, 2015.
- NEWSROOM, S. *Smart TV: Piece by Piece*. 2011. <<https://news.samsung.com/global/smart-tv-piece-by-piece>>. Acessado em 15 de Março de 2018.
- PERAKAKIS, E.; GHINEA, G. A proposed model for cross-platform web 3d applications on smart tv systems. In: ACM. *Proceedings of the 20th International Conference on 3D Web Technology*. [S.l.], 2015. p. 165–166.
- 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.
- SHIN, D.-H.; HWANG, Y.; CHOO, H. Smart tv: are they really smart in interacting with people? understanding the interactivity of korean smart tv. *Behaviour & information technology*, Taylor & Francis, v. 32, n. 2, p. 156–172, 2013.

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

SUO, J. et al. Design sparse features for age estimation using hierarchical face model. In: IEEE. *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*. [S.l.], 2008. p. 1–6.

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