

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
DEPARTAMENTO DE SISTEMA E COMPUTAÇÃO

ÍTALO DE PONTES OLIVEIRA

**UMA ABORDAGEM PARA ANALISAR A ATENÇÃO DE
ESPECTADORES UTILIZANDO VISÃO COMPUTACIONAL**

Campina Grande
2016

ÍTALO DE PONTES OLIVEIRA

**UMA ABORDAGEM PARA ANALISAR A ATENÇÃO DE
ESPECTADORES UTILIZANDO VISÃO COMPUTACIONAL**

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Orientadores: Prof. Dr. Herman Martins Gomes
Prof. Dr. Eanes Torres Pereira

Campina Grande
2016

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

O48a

Oliveira, Ítalo de Pontes.

Uma abordagem para analisar a atenção de espectadores utilizando visão computacional / Ítalo de Pontes Oliveira. – Campina Grande, 2017.

121 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2016.

"Orientação: Prof. Dr. Herman Martins Gomes, Prof. Dr. Eanes Torres Pereira."

Referências.

1. Visão Computacional. 2. Sinalização Digital. 3. Análise da Atenção de Espectadores. 4. Estimação de Idade e de Gênero. 5. *Deep Learning*. I. Gomes, Herman Martins. II. Pereira, Eanes Torres. III. Título.

CDU 004.932(043)

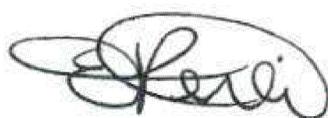
**"UMA ABORDAGEM PARA ANALISAR A ATENÇÃO DE ESPECTADORES
UTILIZANDO VISÃO COMPUTACIONAL"**

ÍTALO DE PONTES OLIVEIRA

DISSERTAÇÃO APROVADA EM 01/12/2016



HERMAN MARTINS GOMES, Ph.D, UFCG
Orientador(a)



EANES TORRES PEREIRA, D.Sc, UFCG
Orientador(a)



FREDERICO MOREIRA BUBLITZ, Dr., UFCG
Examinador(a)

RUDINEI GOULARTE, Dr., USP
Examinador(a)

CAMPINA GRANDE - PB

Resumo

A Sinalização Digital consiste na transmissão de conteúdo via mídia digital, sendo normalmente utilizada por meio de painéis informativos dispostos em locais públicos para despertar a atenção dos espectadores. A pesquisa descrita nesta dissertação, consistiu em criar um aparato para exibição de vídeos e gravação do ambiente, análise e modelagem da atenção dos espectadores quanto à exibição de vídeos em um cenário de sinalização digital. No estudo, foram utilizadas ferramentas de visão computacional para categorizar os espectadores por gênero e faixa etária. A partir do registro dos anúncios exibidos e da classificação dos espectadores foram automaticamente construídas Redes Bayesianas, as quais representaram corretamente o perfil dos espectadores com área sob a ROC de 0,82. Durante a realização da pesquisa, foram evidenciadas lacunas na validação de classificadores de idade utilizando imagens de faces. Para tratar tais lacunas, uma abordagem de estratificação de faixas etárias foi proposta para treinar uma rede neural que obteve desempenho superior ao estado da arte, que foi de 72% quando validado na base de faces MORPH. Para que fosse possível realizar esta pesquisa, devido à inexistência de uma base de vídeos disponível publicamente com rótulos categorizando seus conteúdos, foi criada uma base de vídeos contendo 152 vídeos rotulados manualmente em seis categorias. As análises dos vídeos dos espectadores utilizando as imagens de suas faces e a classificação de tais imagens em relação ao gênero e a idade indicou que indivíduos pertencentes a categorias diferentes dedicam diferentes graus de atenção a vídeos diferentes. Assim, o sistema desenvolvido pode ser utilizado na categorização dos espectadores, para exibição de anúncios de forma automática como forma de auxílio na alocação do tempo de exibição das empresas anunciantes, dentre outras aplicações.

Palavras-chave: Sinalização Digital. Análise da Atenção de Espectadores. Estimação de Idade e de Gênero. *Deep Learning*.

Abstract

Digital Signage consists in content transmission by digital media, very often using informative panels displayed in public places for audience attention arouse. The research described in this dissertation, consists in creating a method for video display, environment recording, and attention modeling in video display for digital signage scenarios. In this study, computer vision tools were used to categorize audience by gender and age group. Bayesian Networks were automatically built using the records obtained from transmitted ads, in which the correct audience profile representation had area under the curve (AUC) of 0.82. When performing this research, it were observed gaps in validation of age classifiers using faces. To address these deficiencies, an age stratification approach was proposed to train a neural network that achieved superior performance in comparison with the state of the art which was 72% when validated on the MORPH face dataset. To make it possible to carry out this research, due to the lack of a publicly available dataset of labeled videos, a video dataset was created containing 152 videos manually labeled in six categories. The analysis of the faces of the viewer and the classification of such images with respect to gender and age indicated that the individuals belong to different categories that dedicated varying degrees of attention to different videos. Thus, the system developed can be used in the categorization of the spectators, for automatically transmitted advertisements as a way to aid in the allocation of time for the transmission of advertising companies, among other applications.

Keywords: Digital Signage. Spectators Attention Analysis. Gender and Age Estimation. Deep Learning.

Agradecimentos

O autor agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Hewlett-Packard (HP) pelo auxílio financeiro.

Aos meus pais, José Wanderley de Oliveira e Eronice de Pontes Oliveira, a minha linda namorada, Francielly Gomes Araujo, pelo suporte e apoio.

Aos orientadores, Herman Martins Gomes e Eanes Torres Pereira, pela dedicação.

Aos colegas do Laboratório de Percepção Computacional (LPC), em destaque Oeslle Alexandre, João Lucas, Carlos Daniel e Tarciso Oliveira, pelas trocas de ideias e as ajudas na realização desta pesquisa.

Em memória aos meus avós: Inácio Camilo de Oliveira, Rita Fernandes de Oliveira, João Augustinho de Pontes e Isabel Avelina de Pontes.

Por fim e não menos importante, aos professores: Daniella Dias Cavalcante, Rhavy Maia Guedes e Carlos Danilo de Miranda Regis, que humildemente me concenderam cartas de recomendação para o ingresso neste mestrado.

Conteúdo

1	Introdução	1
1.1	Relevância	3
1.2	Objetivos	4
1.2.1	Objetivo Geral	4
1.2.2	Objetivos Específicos	4
1.3	Estrutura da Dissertação	6
2	Revisão Bibliográfica	7
2.1	Análise Anônima dos Espectadores	7
2.1.1	Considerações Demográficas e Contextuais	8
2.1.2	Considerações Temporais na Atenção dos Espectadores	11
2.2	Influência da Sinalização Digital	13
2.2.1	Influência no Comportamento Humano	14
2.2.2	Influência nos Estabelecimentos Comerciais	15
2.3	Considerações Finais	17
3	Materiais e Métodos	18
3.1	Materiais	18
3.1.1	Bases de Dados	20
3.1.2	<i>Hardware</i> s	27
3.1.3	<i>Software</i> s	28
3.2	Metodologia	32
3.3	Abordagens Propostas	34
3.3.1	Definição das Faixas Etárias	34

3.3.2	Módulo de Classificação de Idade e Gênero	37
3.3.3	Análise Automática dos Espectadores	44
3.4	Considerações Finais	47
4	Avaliação Experimental	49
4.1	Validação dos Classificadores de Idade e de Gênero	50
4.2	Relação entre o Conteúdo exibido, o Gênero e a Idade do Espectador	55
4.3	Estudo Sobre as Distribuições de Idade dos Espectadores	57
4.4	Relação entre o Conteúdo exibido com a Idade, o Gênero e a Distância do Espectador ao Monitor	61
4.5	Comparação com Outras Pesquisas	66
4.5.1	Pesquisa de Mercado Realizada pela Nielsen Company	66
4.5.2	Pesquisa de Mercado Realizada pela Ibope Inteligência	68
4.6	Considerações Finais	68
5	Considerações Finais	70
5.1	Conclusões	70
5.2	Pesquisas Futuras	72
5.2.1	Análise em Tempo Real	74
5.2.2	Pré-análise do Aspecto Tempo	76
A	Fundamentação Teórica	87
A.1	Teorema de Bayes	87
A.2	Redes Bayesianas	88
A.3	Algoritmo de Aprendizagem K2	89
B	Base de Vídeos Própria: Exibição	92
C	Amostra dos Vídeos Exibidos	105
D	Códigos Fonte	111
D.1	Sincronização da Exibição e Captura	111
D.2	Processamento dos Dados Utilizando o Weka	115
D.3	Conversão da Base de Dados para o Formato LMDB	116

D.4	Modificações na Definição da Arquitetura de Rede	118
E	Publicações	120

Lista de Siglas e Abreviaturas

AAM - *Active Appearance Models*

AIM - *Audience Impression Metric*

ANS - Agência Nacional de Saúde

ARFF - *Attribute-Relation File Format*

AVA - *Anonymous Viewer Analytics*

BSD - *Berkeley Software Distribution*

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

COES - Centro de Operações de Emergências em Saúde

CODEC - Codificador/Decodificador

DNN - *Deep Neural Network*

EUA - Estados Unidos da América

EPBEST - Encontro Paraibano de Estatística

FBI - *Federal Bureau of Investigation*

FDDDB - *Face Detection Data Set and Benchmark*

FN - Falso Negativo

FP - Falso Positivo

GAD - Grafos Acíclicos Dirigidos

GPL - Licença Pública Geral

GPU - *Graphics Processing Unit*

GUI - *Graphical User Interface*

HDMI - *High-Definition Multimedia Interface*

HOG - *Histograms of Oriented Gradients*

KNN - *k-Nearest Neighbors*

LBP - *Local Binary Pattern*

LCD - *Liquid Crystal Display*

LDP - *Local Directional Patterns*

LMDB - *Lightning Memory-Mapped Database*

LTS - *Long Term Support*

MAC - *Media Access Control*

MAC OS - *Macintosh Operating System*

MFOR - *Multi-feature Ordinal Ranking*

MORPH - *Craniofacial Longitudinal Morphological Face Database*

OPENCV - *Open Source Computer Vision*

PCA - *Principal Components Analysis*

PICO - *Pixel Intensity Comparisons Organized in Decision Trees*

POPAI - *Point of Purchase Advertising International*

RAM - *Random-Access Memory*

RITA - *Revista de Informática Teórica e Aplicada*

RMSE - *Root Mean Square Error*

RNA - *Redes Neurais Artificiais*

RNC - *Redes Neurais Convolucionais*

ROI - *Return On Investment*

SDK - *Software Development Kit*

SIBGRAPI - *Conference on Graphics, Patterns and Images*

SMS - *Short Message Service*

SURF - *Speeded Up Robust Features*

SVM - *Support Vector Machine*

STF - *Supremo Tribunal Federal*

UHF - *Ultra High Frequency*

TAN - *Tree Augmented Naïve Bayes*

TDC - *Transformada Discreta do Cosseno*

VN - *Verdadeiro Negativo*

VP - *Verdadeiro Positivo*

Lista de Figuras

3.1	Amostra das imagens da base de dados FDDB.	22
3.2	Amostra de imagens da base de dados MORPH.	23
3.3	Amostra de um vídeo sobre política.	24
3.4	Estrutura e cenário da realização do experimento de sinalização digital. . .	28
3.5	Disposição do <i>hardware</i>	29
3.6	Comparação do detector de faces PICO com outros na base de dados FDDB.	30
3.7	Representação do método proposto nesta dissertação.	33
3.8	Representação do método de classificação de gênero.	38
3.9	Representação do método de classificação de idade.	39
3.10	Exemplos de variações de face geradas sinteticamente.	39
3.11	Exemplo de face detectada próximo à margem da imagem.	40
3.12	Resultado dos pontos detectados enumerados.	41
3.13	Exemplos de variações de face geradas sinteticamente.	42
3.14	Diagrama de estados para o processamento em tempo de execução.	45
3.15	Diagrama de estados para o processamento <i>off-line</i>	46
4.1	Arquiteturas de rede: GoogLeNet - <i>inception</i> e tradicional.	51
4.2	Desempenho dos dois modelos: com e sem aumento de dados.	54
4.3	Comparação entre diferentes métodos de estimação de idade.	55
4.4	Rede bayesiana e tabelas de probabilidade geradas com base no processamento de todos os vídeos capturados dos espectadores.	56
4.5	Relação entre o conteúdo exibido com a faixa etária do espectador agrupada por conteúdo.	59

4.6	Relação entre o conteúdo exibido com a faixa etária do espectador agrupada por geração.	60
4.7	Heurística utilizada para avaliar a distância do espectador em relação ao monitor.	62
4.8	Estrutura da rede bayesiana.	63
4.9	Comparação de interesses para diferentes categorias de idade.	67
5.1	Diagrama de alto nível, representando o processamento em tempo de execução.	75
5.2	Estrutura da rede bayesiana considerando as variáveis tempo e distância.	77
A.1	Representação gráfica de uma rede bayesiana.	89
C.1	Amostra de um vídeo sobre diversão, em que são apresentados jogadores de futebol que se atrapalharam em suas jogadas.	106
C.2	Amostra de um vídeo sobre economia, em que são apresentadas orientações financeiras de como economizar em tempos de crise.	107
C.3	Amostra de um vídeo noticiário, em que são apresentadas novidades tecnológicas.	108
C.4	Amostra de um vídeo de tema policial, em que é noticiado o furto de um helicóptero que ocorreu em Mairinque (SP).	109
C.5	Amostra de um vídeo sobre saúde, em que são noticiados os cuidados no combate ao Zika vírus e a relação com o surto dos casos de microcefalia.	110

Lista de Tabelas

3.1	Probabilidades <i>a priori</i> de todos os conteúdos transmitidos neste experimento.	27
3.2	Modificações nos pontos fiduciais da mandíbula.	43
3.3	Modificações nos pontos fiduciais do nariz.	43
3.4	Modificações nos pontos fiduciais do queixo.	43
4.1	Ordem das preferências dos espectadores com base em cada geração e o seu respectivo gênero.	58
4.2	Relação entre o conteúdo saúde e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.	64
4.3	Relação entre o conteúdo política e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.	64
4.4	Relação entre o conteúdo tecnologia e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.	65
4.5	Relação entre o conteúdo entretenimento e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.	65
4.6	Relação entre o conteúdo economia e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.	66
4.7	Relação entre o conteúdo policial e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.	66
5.1	Tabela de probabilidades condicionadas incluindo a variável tempo nas análises.	78

Lista de Quadros

3.1	Comparação de <i>frameworks</i> populares de <i>Deep Learning</i>	32
3.2	Detalhamento das gerações de idades com base nas definições dos sociólogos.	36
3.3	Mapeamento das gerações com base nas faixas etárias definidas pela Agência Nacional de Saúde (ANS).	37
5.1	Levantamento das principais contribuições desta dissertação.	71
B.1	Relação completa dos vídeos rotulados manualmente e que foram exibidos aos espectadores.	93
E.1	Publicação na Revista de Informática Teórica e Aplicada (RITA Vol. 23). . .	121
E.2	Publicação no Encontro Paraibano de Estatística (EPBEST).	121
E.3	Publicação no evento <i>Conference on Graphics, Patterns and Images</i> (29 th SIBGRAPI 2016).	121

Lista de Códigos Fonte

D.1	Instalação das depências utilizadas para realizar a transmissão e captura sincronizadas.	112
D.2	Código utilizado para realizar a exibição e captura sincronizadas.	112
D.3	Código utilizado para estimar o tempo do vídeo de captura.	113
D.4	Código para construção da Rede <i>Bayesiana</i> utilizando a base de dados. . . .	116
D.5	Código para conversão da base de dados para o formato LMDB.	117

Capítulo 1

Introdução

A Sinalização Digital expõe informações como notícias urgentes, previsão do tempo, entretenimento e outros temas para captação da atenção das pessoas, usando painéis de visualização, tipicamente construídos nas tecnologias de *Liquid Crystal Display* (LCD) ou Plasma de alta resolução, com seus conteúdos atualizados em tempo real (BURKE, 2009). Tem sido utilizada em diferentes aplicações, como na promoção e lançamento de produtos e serviços, ou de utilidade pública como informação aos frequentadores de um determinado ambiente e até mesmo em ocasiões críticas, como foi o caso do terremoto no Japão em abril de 2009 em que os monitores foram úteis na evacuação de emergência (PALESTINI, 2013). Além disso, a difusão da Sinalização Digital tem sido alta, sendo utilizada em diversos lugares, como por exemplo, bancos, aeroportos, academias, igrejas, hospitais e feiras. Para gerar melhor satisfação dos usuários, novas tecnologias têm sido desenvolvidas, por exemplo, o *Audience Impression Metric Suite* (AIM) da Intel[®], que adquire informações indiretamente dos espectadores, por meio de câmeras de monitoramento. Além disso, estudos como o de Kinebuchi et al. (2009), fazem uso das informações coletadas pelas câmeras na seleção automática de anúncios. Tais abordagens permitem tornar a sinalização digital mais interativa com o usuário, quando a adaptação do conteúdo exibido se torna automática (RAVNIK; SOLINA, 2013b).

Alguns padrões na Sinalização Digital tem sido definidos, tais como o padrão definido pela *Point of Purchase Advertising International* POPAI (2014), sob a patente US 4670798 A que se trata de uma associação comercial global sem fins lucrativos dedicada a melhorar ambientes de varejo, tendo mais de 2 mil empresas associadas em todo o mundo, como

a Cisco® e a LG®. Essa padronização define níveis de tecnologias a serem empregadas, como forma de organizar as exibições de Sinalização Digital e sugerir padrões de qualidade a serem seguidos (e.g. na codificação de áudio e vídeo). Essa coletânea de informações visa maximizar o Retorno Sobre o Investimento (*Return On Investment* - ROI) das empresas com estudos que otimizem a audiência obtida. Assim, observa-se que conhecer o perfil dos consumidores, para uma seleção personalizada dos anúncios a serem entregues, mostra-se relevante. De modo semelhante aos anúncios exibidos via internet, que utilizam informações pessoais (e.g., *cookies* dos navegadores), que adaptam o conteúdo entregue ao contexto do usuário, evitando que cause um impacto negativo na entrega de anúncios indesejados ao usuário (BUBLITZ et al., 2014).

Nesse sentido, câmeras de monitoramento podem ser utilizadas para categorização dos espectadores e assim, utilizar informações sobre o perfil demográfico dos mesmos para associar ao conteúdo exibido de maneira anônima (RAVNIK; SOLINA, 2013a). A análise dos espectadores pode também servir como entrada para estratégias de cobrança pelo espaço publicitário, de tal modo que, as empresas que desejem alocar uma faixa de tempo para exibição de seus anúncios, teriam acesso a relatórios precisos sobre o público-alvo alcançado pelos anúncios. Por fim, sistemas com essa característica poderiam ser utilizados para adaptar a exibição dos anúncios automaticamente. Isto é, aprender as preferências de grupos demográficos e, ao detectar a presença de um grupo prevalente, exibir aquele anúncio que retém maior interesse desses espectadores ou exibir o anúncio da empresa que esteja buscando esse perfil de espectador específico para exibir os seus anúncios. Ou seja, a rede bayesiana modela as preferências dos grupos demográficos, e o anúncio recomendado é aquele de maior preferência entre os grupos predominantes no local.

Nesta dissertação foi desenvolvido um método para análise automática de espectadores na Sinalização Digital. O método proposto utiliza uma câmera para monitorar os espectadores. O núcleo da aplicação realiza análises nos vídeos capturados, coletando informações como o gênero e a idade dos espectadores. O método de classificação de idade proposto é competitivo com o estado-da-arte, fornecendo dados mais confiáveis do que outras abordagens de classificação para as análises dos espectadores. Além disso, foi realizado um estudo detalhado sobre a definição das faixas etárias dos espectadores e apresentada uma abordagem para especificá-las com base em estudos sociológicos, a exemplo de (HOPE, 2016)

que caracteriza os indivíduos nascidos na era da Internet e do acesso as redes sociais, como pertencentes de uma mesma geração. Foi realizado um estudo de caso para testar experimentalmente o sistema desenvolvido e analisar a atenção dos espectadores na exibição de vídeos com diferentes tipos de conteúdo. Verificou-se que os espectadores em função do gênero e da idade, prestam atenção de maneira diferenciada a depender do conteúdo do vídeo exibido, conforme observado por Yin et al. (2012). Mais de cem vídeos públicos na internet foram manualmente categorizados e disponibilizados para que outros autores possam utilizá-los em suas pesquisas.

Este capítulo está organizado como segue: a Seção 1.1 sintetiza a relevância desse estudo. Os objetivos são apresentados na Seção 1.2. Por fim, a estrutura de como a dissertação está organizada é apresentada na Seção 1.3.

1.1 Relevância

Diversas empresas têm investido em publicidade e na Sinalização Digital, mas ainda não há métodos bem definidos que permitam analisar o perfil dos seus clientes e/ou associar o seu gênero, a sua faixa etária e outras características que possam ser extraídas do vídeo capturado dos espectadores com o conteúdo exibido. A Sinalização Digital realizada sem tais métodos gera dificuldades para as empresas atingirem o seu público-alvo (MACEDO, 2009). Nesta dissertação, foi proposta e implementada uma metodologia capaz de identificar o perfil dos espectadores a partir de técnicas de visão computacional e aprendizagem de máquina, contribuindo com inferências sobre a atenção e o conteúdo exibido. Durante o desenvolvimento da pesquisa, o desempenho do classificador de idade foi aperfeiçoado ao ponto de superar o desempenho de estado-da-arte. Além disso, foi implementado um sistema completo que possibilitou validar estatisticamente a metodologia proposta. Em resumo os resultados desta dissertação contribuem positivamente para a área de sinalização digital, viabilizando maior efetividade no processo, solucionando problemas como o desempenho dos classificadores, fornecendo empacotamento dos dados processados para que outros autores possam replicar os experimentos e conceber inferências importantes que auxiliem às empresas de publicidade em atingir o público-alvo da Sinalização Digital com mais eficácia.

1.2 **Objetivos**

Esta seção apresenta os objetivos gerais e específicos desta dissertação.

1.2.1 **Objetivo Geral**

Brevemente discutido na seção anterior, esta pesquisa propõe uma metodologia para adquirir conhecimento e analisar a atenção dos espectadores diante a Sinalização Digital. Objetivou-se desenvolver uma metodologia que permitisse investigar, quais são os anúncios que chamam mais atenção de certo perfil dos espectadores e que permitisse determinar quais as preferências de um determinado gênero ou faixa etária. Nesse sentido, o intuito desta pesquisa é aprimorar a análise automática da atenção dos espectadores em sistemas de Sinalização Digital, de tal forma que o sistema seja capaz de analisar o perfil demográfico dos consumidores e exibir o vídeo com conteúdo de maior afinidade, além de fornecer estatísticas e dados relevantes para as empresas de publicidade.

Nesse cenário, foram empregadas técnicas de visão computacional para automaticamente detectar as faces dos espectadores e classificá-las quanto ao gênero e à faixa etária. O propósito central do sistema é extrair conclusões acerca da audiência do grupo de indivíduos em diferentes tipos de exibição.

1.2.2 **Objetivos Específicos**

Os seguintes objetivos específicos foram formulados:

1. Investigar as diferentes modos de analisar as preferências dos espectadores na Sinalização Digital.
2. Montar uma estrutura para a exibição da Sinalização Digital em um cenário real e gerar capturas dos espectadores.
3. Gerar definições de faixas etárias para agrupar os espectadores demograficamente.
4. Analisar as capturas automaticamente e gerar estatísticas acerca das preferências dos espectadores.

5. Montar uma base pública de vídeos de sinalização digital com o intuito de permitir que outros pesquisadores repliquem o experimento desenvolvido para melhor validação.

O problema científico abordado nesta dissertação é desenvolver um método que seja capaz de analisar o perfil dos espectadores e categorizá-los com base em suas preferências a diferentes tipos de anúncios exibidos.

Para alcançar esses objetivos, as seguintes atividades foram realizadas: (1) Foi realizado um levantamento bibliográfico (apresentado no Capítulo 2), o qual apresenta diversos estudos que investigaram a categorização dos espectadores da Sinalização Digital e que serviram de inspiração/motivação para a metodologia desenvolvida na presente dissertação; (2) Na Seção 3.1 são apresentados os materiais utilizados nesta dissertação, todos os algoritmos utilizados possuem código-fonte aberto, até mesmo as bases de dados utilizadas nos experimentos, como também, os principais *scripts* utilizados são disponibilizados no Apêndice D. De tal maneira, é possível que outros autores utilizem os materiais utilizados para replicação dos experimentos ou avaliações comparativas com outros métodos; (3) A estrutura montada para comportar os equipamentos da Sinalização Digital é detalhada na Seção 3.1.2. Com ela é possível realizar a experimentação prática para a gravação dos vídeos dos espectadores em um cenário real de Sinalização Digital; (4) Foi realizado um estudo para a definição das faixas etárias fundamentada em conceitos sociológicos, apresentado na Subseção 3.3.1. Com isso, é possível agrupar os espectadores demograficamente de maneira mais adequada; (5) Para análise dos espectadores, foram geradas redes bayesianas que são apresentadas no Capítulo 4, desse modo, é possível analisar isoladamente e conjuntamente os diferentes fatores envolvidos no experimento, tais como o gênero, a faixa etária e o conteúdo exibido; (6) Os vídeos utilizados nas exibições aos espectadores, foram devidamente catalogados e rotulados manualmente para que outros autores pudessem verificar a estrutura da base e assim selecionar novos vídeos para exibição com base nas informações contidas na relação de vídeos apresentados no Apêndice B. Por fim, alguns dos resultados parciais desta pesquisa foram publicados, conforme apresentado no Apêndice E.

1.3 Estrutura da Dissertação

Esta dissertação está organizada como segue. No Capítulo 2, são apresentados os trabalhos relacionados. Enquanto que, no Capítulo 3, são apresentados os materiais e métodos utilizados para a realização dos experimentos, descrevendo o desenvolvimento do sistema e as abordagens propostas. A avaliação dos módulos do sistema e a sua aplicação experimental em um cenário real são apresentados no Capítulo 4. As considerações finais e os trabalhos futuros são apresentados no Capítulo 5. Por fim, o Apêndice A apresenta a fundamentação teórica das redes bayesianas empregadas na análise da atenção dos espectadores. O Apêndice B, contém a relação de vídeos exibidos nesta dissertação. O Apêndice C, contém uma amostra de cada categoria de vídeo exibido. No Apêndice D são apresentados os códigos-fonte que foram desenvolvidos nesta dissertação. A relação de publicações é apresentada no Apêndice E.

Capítulo 2

Revisão Bibliográfica

Este capítulo apresenta um levantamento das pesquisas relacionadas ao estudo relatado nesta dissertação. São apresentadas pesquisas que abordam a análise da atenção dos espectadores na sinalização digital, que propuseram modelos de análise ou que realizaram experimentos sobre as preferências dos espectadores. Esses estudos foram coletados de relevantes portais de pesquisa como o IEEEExplore, ACM *digital library*, Google Scholar e Springer. Foram considerados na seleção, os artigos que possuem maior número de citações e publicados nos últimos cinco anos. Verificaram-se também as pesquisas citadas nos artigos encontrados e as palavras-chave utilizadas para realização de novas pesquisas.

O presente capítulo está organizado em três seções. Na Seção 2.1 são apresentados os estudos que analisaram a atenção dos espectadores de forma demográfica ou considerando o tempo de visualização, enquanto na Seção 2.2 são apresentadas as pesquisas que analisaram as influências da sinalização digital no comportamento humano ou nas vendas dos estabelecimentos comerciais. Por fim, a Seção 2.3 apresenta as considerações finais sobre o levantamento bibliográfico realizado.

2.1 Análise Anônima dos Espectadores

Diversas pesquisas têm sido realizadas sobre a análise automática dos espectadores. Sistemas de Análise Anônima dos Espectadores (do inglês, *Anonymous Viewer Analytics - AVA*) têm sido incentivados por manterem a privacidade dos espectadores. Sistemas AVA não armazenam vídeos dos espectadores nem realizam reconhecimento facial. Um exemplo desse

tipo de sistema foi desenvolvido pela Intel[®], chamado de *Audience Impression Metric Suite* (Intel[®] AIM Suite) (YIN et al., 2012), que é um serviço para nuvem com tempo de processamento próximo do real, que utiliza sensores ópticos colocados imediatamente acima (ou à esquerda/direita/abaixo) da tela digital para a detecção de faces, classificação de gênero e estimação de idade. Sistemas como esse auxiliam as empresas anunciantes a conhecer o público alvo de suas exibições e têm motivado a realização de pesquisas que utilizam recursos diversos na análise dos fatores que influenciam a atenção dos espectadores. Esta seção destaca as principais pesquisas nesse contexto da seguinte forma: Subseção 2.1.1 organiza as pesquisas que agruparam os espectadores demograficamente e contextualmente, enquanto na Subseção 2.1.2, são apresentadas as pesquisas que focaram no aspecto temporal ao analisar a atenção dos espectadores.

2.1.1 Considerações Demográficas e Contextuais

A pesquisa de Kinebuchi et al. (2009), apresentou um arcabouço para seleção de anúncios utilizando o AIM Suite. Esse estudo utilizou cinco sistemas de sinalização digital instalados em diferentes lugares. Os dados coletados foram analisados com base em duas abordagens diferentes de predição dos espectadores: (i) utilizando as informações demográficas (i.e., o gênero e a idade); e (ii) utilizando o contexto (i.e., conteúdo do vídeo, data, hora, localização do dispositivo, o clima e outras informações). Por fim, foram utilizadas árvores de decisão para estimar o vídeo exibido com base nas informações demográficas e de contexto. O modelo de recomendação com base em informações demográficas se mostrou mais eficiente do que o modelo com base no contexto. Entretanto, a forma que o experimento foi conduzido pode ter dificultado a abordagem de predição demográfica, pois o estudo foi realizado nos *campi* da Intel[®], em que a maioria dos funcionários são jovens adultos do sexo masculino (KINEBUCHI et al., 2009). Como consequência, torna-se mais complexo distinguir as preferências dos espectadores se eles têm um grupo demográfico prevalente, desse modo, se houvesse grupos demográficos mais distintos, a abordagem de categorização demográfica poderia fornecer resultados ainda melhores em relação a recomendação com base no contexto.

Outra pesquisa relevante foi realizada por Ravnik e Solina (2013a), que desenvolveram um sistema para análise de audiência em tempo real. A aplicação consiste em capturar

vídeos de espectadores da sinalização digital utilizando uma câmera posicionada abaixo do televisor que exibe os anúncios. No estudo de caso realizado em uma boutique na cidade de Ljubljana, capital da Eslovênia, a faixa etária e o gênero dos espectadores foram relacionados ao conteúdo exibido e ao tempo de permanência e de visualização. O processamento dos vídeos capturados consistiu em: (i) realizar segmentação do plano de fundo e, assim, detectar os indivíduos presentes no local e estimar o tempo de permanência; (ii) detectar as faces e rastreá-las para estimar o tempo de observação dos espectadores; e por fim, (iii) aplicar classificadores de gênero e de idade às imagens de faces detectadas. As análises indicaram que o gênero do espectador tem impacto significativo no tempo de permanência e de atenção, sendo os homens mais receptivos do que as mulheres independente do tipo de conteúdo exibido. Também foi observado que a faixa etária tem influência sobre o tempo de atenção, sendo os mais jovens (com idades entre 1 e 14 anos) os que assistiram aos anúncios por mais tempo, enquanto a faixa etária de 35 à 44 anos de idade foi a que menos assistiu aos anúncios. Por fim, a relação com o tipo de conteúdo exibido foi analisada, se é estático ou dinâmico (isto é, uma imagem ou um vídeo, respectivamente), constatou-se que os anúncios dinâmicos retiveram 43% mais tempo de atenção dos espectadores que os anúncios estáticos.

Na pesquisa de Ravnik e Solina (2013b) são feitas considerações sobre os tipos de câmeras utilizadas na captura de vídeos de espectadores em ambientes de sinalização digital. No experimento realizado, foram comparadas a utilização de uma câmera monocular e uma câmera Kinect¹. Foi observado que, para distâncias superiores a quatro metros, a utilização da câmera monocular se mostrou melhor por obter menor *Root Mean Square Error* (RMSE) na detecção de faces. Essas análises reforçam os resultados obtidos por Ravnik e Solina (2013a), visto que foi utilizado o mesmo modelo de câmera monocular em seus experimentos. Adicionalmente, a pesquisa realizada por Ravnik e Solina (2013b) apresenta a correlação entre as variáveis analisadas, por exemplo, observou-se que o gênero obteve baixa correlação com a faixa etária dos espectadores e com o tipo de conteúdo exibido, embora, o gênero e a faixa etária possuam alta correlação com o tempo de atenção dos espectadores. As evidências obtidas nessas pesquisas foram utilizadas como suporte para outro estudo de Ravnik e Solina (2014), que propuseram um modelo de distribuição de anúncios que considera as

¹Sensor de movimentos desenvolvido pela Microsoft, que utilizam câmeras para captura de informação de cor e de profundidade.

informações coletadas nos estudos anteriores, dando a possibilidade de adaptar a seleção dos anúncios de modo interativo dependendo da audiência local.

Entretanto, pode existir um paradoxo no desenvolvimento de sistemas inteligentes que reconhecem o perfil dos usuários e selecionam o melhor conteúdo a ser exibido, pois, em Bublitz et al. (2014) são feitas considerações sobre a seleção automática dos anúncios, e observa-se que no projeto de um sistema de sinalização digital, precisa-se ter o seguinte cuidado: se apenas os melhores anúncios são entregues aos consumidores, os objetivos dos anunciantes são, eventualmente, negligenciados. Por outro lado, ao se considerarem apenas os interesses dos anunciantes, a relevância dos anúncios para os consumidores pode diminuir consideravelmente. É importante selecionar os conteúdos que melhor se adequem aos dois contextos. Dessa maneira, algumas pesquisas têm evidenciado tendências em trazer sucesso para a empresa contratante quando a satisfação do receptor é levada em consideração na seleção dos anúncios, como destacado na pesquisa de Bublitz et al. (2014).

No contexto de seleção automática de anúncios, a pesquisa realizada por Bublitz et al. (2013), utiliza o perfil individual dos espectadores na seleção dos anúncios via *smartphones*. Naquela pesquisa foi gerada uma rede de Petri para selecionar os conteúdos a serem exibidos de maneira ideal, ou seja, que satisfizessem ao máximo os interesses dos consumidores e das empresas anunciantes. Foram utilizadas informações de localização fornecidas pelos *smartphones* via *bluetooth*, identificando cada pessoa individualmente pelo endereço *Media Access Control* (MAC). Os testes de validação do modelo se mostraram promissores, visto que, com base na opinião dos usuários se observou um aumento de 18% na relevância média dos anúncios entregues quando a abordagem proposta foi empregada. Isto indica que os anúncios entregues com base nas decisões tomadas utilizando o modelo proposto, em relação à entrega aleatória, são mais interessantes para os clientes e, conseqüentemente, para as empresas anunciantes. Evidenciou-se, assim, a necessidade de se conhecer o perfil dos espectadores da sinalização digital ao selecionar o melhor conteúdo a ser exibido.

No estudo de Bublitz et al. (2014) foi desenvolvido um modelo capaz de inferir informações demográficas (como o gênero e a idade), psicográfica (classe social) e informações comportamentais (preferências) com base em informações de localização providas pelos *smartphones*. O experimento aconteceu em Maceió - Brasil e contou com a participação de 112 voluntários com idade entre 16 e 55 anos. Essa abordagem teve base no fato de que

alguns lugares terem público bem definido, por exemplo, uma sofisticada loja de roupas é normalmente frequentada por pessoas de classe social mais elevada (BUBLITZ et al., 2014). Dessa maneira, foram utilizadas informações de localização, i.e., os lugares frequentados pelos usuários, para estimar características dos consumidores. Foi possível estimar a idade dos indivíduos com margem de erro de 5 anos com 79% de acerto. Enquanto o gênero foi determinado com acurácia de 95%. O modelo gerado foi aplicado em um cenário de sinalização digital objetivando a entrega de anúncios com maior acurácia. Como resultado, foi observado um aumento de cerca de 10% na acurácia dos anúncios recomendados com base no modelo proposto, quando comparado com a técnica de recomendação aleatória.

2.1.2 Considerações Temporais na Atenção dos Espectadores

A pesquisa de Yin et al. (2012) apresenta um modelo de análise dos espectadores, que foi obtido por meio da exibição de diferentes tipos de anúncios por um televisor enquanto uma câmera posicionada acima do televisor realizava a aquisição de vídeo. Após isso, foi realizada uma análise dos vídeos capturados, por meio de detecção de faces, olhos e a boca na região da face. Essas informações de posição foram utilizadas para estimar a pose da face, isto é, estimar se a face está ou não voltada para o monitor. No experimento, foram exibidos diferentes tipos de anúncios, tais como: noticiário internacional, noticiário de Taiwan, notícias sobre o clima e outros. O conteúdo exibido foi relacionado à quantidade de espectadores e ao tempo médio de visualização por espectador. Observou-se que o noticiário internacional chamou atenção de uma quantidade de espectadores aproximada das demais notícias exibidas, entretanto, pelo maior tempo de visualização. Por outro lado, outros anúncios como o noticiário de Taiwan, chamou atenção para um número maior de pessoas quando comparado com outros tipos de anúncios, entretanto, por um curto período de tempo. Pode-se observar que diferentes conteúdos chamam atenção dos espectadores de modo diferenciado. Entretanto, nesse estudo não foi realizada análise demográfica dos espectadores (e.g., considerações sobre o gênero e a idade), como foi realizado em Ravnik e Solina (2013b).

Outra maneira de se medir a audiência foi apresentada por Hyun et al. (2014), que implementou um ferramenta utilizando informação 3D provida pelo sensor Kinect para: (i) detectar e contabilizar a presença de espectadores; (ii) classificar a faixa etária e o gênero com base na face detectada; (iii) estimar a altura e a distância do espectador em relação ao

monitor; e (iv) estimar se o indivíduo está ou não assistindo o vídeo, com base na posição dos olhos e boca (semelhante ao que foi feito por Yin et al. (2012)). Entretanto, apenas o modelo foi proposto e nenhum estudo de campo foi realizado para validar a sua proposta. Além disso, o código fonte da aplicação não foi disponibilizado para que outros autores reproduzissem os resultados. Apesar das ressalvas apresentadas acima, as análises sobre a atenção dos espectadores realizadas na pesquisa de (YIN et al., 2012) foram pertinentes para o desenvolvimento desta dissertação.

Em Maeda et al. (2010), é utilizada uma abordagem menos intrusiva de medir audiência e sem câmeras ou *smartphones* nas análises, o método consiste em medir a interferência causada pela presença do ser humano próximo de uma antena *Ultra High Frequency* (UHF), para estimar a presença de um indivíduo, acionar a tela e assim, exibir os anúncios. Desse modo, a economia no consumo energético foi de aproximadamente 60% ao desligar o monitor nos momentos em que não há presença de pessoas, maximizando assim, o Retorno Sobre o Investimento (*Return On Investment* - ROI) sem incomodar as pessoas que transitam pelo local com mensagens de texto. Além disso, a abordagem proposta associou a quantidade de trafegantes ao horário do dia. Observando os momentos do dia que há maior passagem de indivíduos, permitindo analisar esses padrões e elaborar uma estratégia de exibição de anúncios mais robusta. Entretanto, essa abordagem não permite estimar com eficácia a audiência dos anúncios exibidos, pois, embora os indivíduos estejam presentes, as suas faces não estão voltadas para o monitor. Assim, as conclusões podem ser inconsistentes.

Alguns estudos exploraram fatores diferentes dos supracitados para investigar a audiência dos espectadores, a pesquisa de Beyer et al. (2011) explorou como o tipo de tela utilizado nas exibições influencia o tempo de visualização. As telas planas foram comparadas com telas cilíndricas e foi observado que este tipo de tela retém menos a atenção dos indivíduos, ou seja, a média de pessoas que pararam de trafegar pela frente da tela para assistir ao anúncio foi menor nas telas cilíndricas do que nas telas planas, o tempo médio de observação também foi menor e o custo mais elevado, fazendo dele uma opção menos interessante na retenção da atenção dos espectadores. Assim, essa evidência favorece a utilização da tela plana nesta dissertação.

Na pesquisa de Nelson e Webster (2016), a sinalização digital foi investigada na perspectiva dos usuários dos sites de notícias dos EUA, relacionando a quantidade de tempo que os

visitantes gastam nos sites e o número de visitantes únicos que um site recebe. Foram utilizados os dados providos pela comScore®, uma empresa de análise do tráfego *Web* que informa estimativas mensais de audiência de diversos sites. Esses dados também foram utilizados por outros autores nas análises de audiência on-line, por exemplo, em Taneja e Wu (2013) a audiência dos sites estimada com base no perfil dos seus espectadores, sendo observado que a linguagem e a dissimilaridade geográfica são os fatores mais influentes para a estimação de audiência. A empresa que forneceu os dados, a comScore, coleta seus dados a partir de cerca de 1 milhão de pessoas que carregam o *software* de monitoramento em seus computadores. O *software* rastreia a *Uniform Resource Locator* (URL) acessada pelo usuário e o tempo que passou visualizando cada endereço. Em setembro de 2015, a comScore registrou um público total nesses sites de cerca de 230 milhões de visitantes únicos. Desses, cerca de 165 milhões visitaram sites de notícias on-line. Nesse contexto, foi observado por Nelson e Webster (2016), que não há correlação significativa entre o tempo médio de acesso dos visitantes e o tempo médio de visualização da página. Isto é, nota-se que os usuários com tempo médio baixo também frequentam sites com tempo médio de acesso alto (e vice-versa). Ou seja, o conteúdo do site não está relacionado ao tempo de navegação dos usuários. O que pode ser contrastado no ambiente de sinalização digital via televisão, por exemplo. A pesquisa de Yin et al. (2012) observou que alguns conteúdos chamam atenção de um número alto/baixo de espectadores, embora o tempo médio de visualização seja baixo/alto, respectivamente.

2.2 Influência da Sinalização Digital

Esta seção apresenta como a utilização de sistemas de análise dos espectadores da sinalização digital tem influenciado o comportamento humano e/ou as vendas nos estabelecimentos comerciais. Dessa maneira, esta seção se divide em duas: a Subseção 2.2.1 destaca a influência da sinalização no comportamento humano, muitas vezes utilizada para melhorar a experiência do espectador, enquanto a Subseção 2.2.2 destaca a influência da sinalização no aumento do ROI.

2.2.1 Influência no Comportamento Humano

Na pesquisa de Santos, Gomes e Campos (2011) é apresentado um modelo de análise de espectadores, que consiste em realizar a aquisição de vídeo dos transeuntes, realizar a detecção facial e, com base nas faces, fazer a classificação das emoções (neutra, felicidade, tristeza, surpresa e nojo). Por fim, o sistema desenvolvido gera relatórios para a análise manual com estatísticas sobre a representatividade das emoções despertadas em diversos horários ao longo do dia. O sistema foi testado em um ambiente real durante um dia de captura, sendo observada a predominância de faces neutras ao longo do dia, e ao final do dia, um aumento significativo nas expressões de felicidade e tristeza, respectivamente, com diminuição da expressão neutra. Assim, a ferramenta pode ser útil para profissionais de *marketing* que desejam utilizar essas informações na escolha dos anúncios exibidos. Entretanto, em função da necessidade da utilização de humanos nas análises, essa ferramenta pode ser inaplicável em cenários reais por ser pouco escalável. Além disso, o autor se restringiu em suas análises a não correlacionar o conteúdo exibido com a emoção despertada. Outros autores correlacionaram os dados coletados dos vídeos com o conteúdo exibido, como, por exemplo, Yin et al. (2012) e Ravnik e Solina (2013b). Além disso, poderia ser investigada a razão pela qual houve aumento na porcentagem das faces com expressão de felicidade e tristeza no final do dia. A motivação poderia ser o conteúdo exibido ou outros fatores externos, como a expressão de felicidade ser motivada pelo final do expediente de trabalho e a de tristeza devido ao cansaço após um dia de trabalho e não necessariamente, devido ao conteúdo exibido.

A abordagem de Kramer, Guillory e Hancock (2014) relacionou o impacto de postagens dos espectadores com conteúdo afetivo (positivo/negativo) manipulado. Reforçando as observações realizadas na análise de atenção com base em questionários e análises indiretas. Nessa pesquisa, foi realizado um experimento com mais de 600 mil usuários do Facebook, no qual, mensagens positivas ou negativas (i.e., que possuem ao menos uma palavra positiva/negativa) foram enviadas para as páginas de notícias dos usuários e foram relacionadas ao conteúdo que os mesmos usuários postaram em seguida. Durante uma semana de análise, mais de três milhões de postagens foram realizadas por todos os usuários envolvidos no experimento. Foi observado que as postagens dos usuários estavam correlacionadas com os anúncios que recebiam em sua página inicial. Isto é, anúncios positivos/negativos geravam postagens positivas/negativas, respectivamente. Dessa maneira, pode-se observar a

interferência que os anúncios podem causar no comportamento humano.

Outra abordagem para analisar o efeito dos anúncios via computador, mas com base na opinião dos receptores, foi apresentada em Grobelny e Michalski (2015), esse estudo analisou as preferências de estudantes de duas faculdades diferentes. Foram exibidos anúncios com diferentes variações na cor do plano de fundo, no posicionamento do texto e no tipo da fonte. Foi observado que, a cor do anúncio não foi um fator relevante para as mulheres na escolha do anúncio de sua preferência, enquanto os homens tiveram preferência por anúncios de cor cinza (independente da sua instituição de origem). Assim, nota-se que fatores implícitos aos anúncios podem chamar a atenção dos espectadores de forma diferenciada, o que foi discutido em Feng et al. (2011) que estudou como homens e mulheres têm áreas do cérebro estimulados de forma diferente.

Por fim, a pesquisa realizada por Moritz (2016) utilizou questionários para analisar a satisfação dos espectadores, o experimento realizado envolveu a satisfação nas viagens de trens, focando na excelência da experiência pelos passageiros de um sistema de transporte público na Holanda. No experimento, foram exibidos diferentes tipos de conteúdo, tais como: informações de viagem, notícias, curiosidades sobre o percurso e o destino da viagem, condições climáticas, destaques do esporte e outros. Os 835 passageiros foram divididos em dois grupos, os que receberam e os que não receberam anúncios para testar os efeitos da sinalização digital na experiência emocional, com base em um questionário que foi aplicado posteriormente. Foi observado que a satisfação da viagem foi melhor para aqueles que receberam os anúncios, embora as propagandas tenham influenciado negativamente sobre alguns passageiros, que sugeriram não mostrá-las. Quando o conteúdo do anúncio exibido estava relacionado à viagem, foi observado um parecer favorável dos passageiros. Evidenciou-se, assim, que o conteúdo exibido exerce influência sobre o interesse dos espectadores e em sistemas fechados como esse, a aplicação de um sistema de reconhecimento demográfico automático pode selecionar os melhores anúncios com base nas opiniões contidas nos formulários.

2.2.2 Influência nos Estabelecimentos Comerciais

A pesquisa de Helander (2010), utilizou dados fornecidos por *smartphones* para categorização dos espectadores. Naquele estudo foi realizada uma investigação sobre o ROI. Para se

ter controle das informações, foi utilizada comunicação celular para confirmar o interesse do usuário em receber anúncios com cupons de descontos e assim efetuar o cálculo do lucro obtido com os anúncios. Os resultados mostraram que mais de 90% dos consumidores que autorizaram o envio dos cupons de desconto, efetuaram alguma compra na loja utilizando os cupons recebidos. Por outro lado, apesar do impacto positivo no ROI, a sinalização digital pode ser considerada um meio intrusivo que incomode os espectadores e que, por isso, deve ser utilizada com maior cuidado para garantir a satisfação do cliente, como é discutido em Lee e Yoon (2014).

Na pesquisa de Mikawa, Izumi e Tanaka (2011), foi desenvolvido um sistema de sinalização para recomendação de livros, que estima o gênero e a faixa etária do indivíduo que trafega pelo local com base em sua silhueta. Mesmo com acurácia baixa, 77,5% para o classificador de gênero e 37,5% para o grupo de idade utilizando o *Support Vector Machine* (SVM) na classificação, foi observado que 25% das pessoas que trafegaram pelo local se interessaram por adquirir a revista que estava sendo anunciada. Mostrou-se, assim, a eficiência dos sistemas de sinalização digital e a importância do processo de reconhecimento demográfico.

No contexto de vendas, a pesquisa de Roggeveen, Nordfalt e Grewal (2016) evidenciou o impacto da sinalização digital nas vendas de lojas no varejo. Naquela pesquisa se observou que, após a inserção da sinalização digital, as vendas tiveram relação direta com o tamanho do estabelecimento comercial. Isto é, as vendas subiram nos grandes estabelecimentos comerciais, como os hipermercados. Nos supermercados, não houve alteração nas vendas ao se incluir sinalização digital, enquanto nas lojas menores (como as lojas de conveniência), houve impacto negativo nas vendas. Embora os anúncios exibidos tenham sido os mesmos e as lojas tenham considerável similaridade demográfica, isto é, possuem público semelhante, ao se considerar as faixas etárias e a situação de emprego dos consumidores que a frequentaram, o efeito da sinalização digital foi diferente nos estabelecimentos comerciais. Esse fato reforça a necessidade de estudos experimentais que analisem a atenção dos espectadores, principalmente em espaços pequenos para que a sinalização digital favoreça as vendas. Como apresentado por Mikawa, Izumi e Tanaka (2011), a sinalização digital pode significar sucesso de vendas, se empregada de maneira correta em um estabelecimento comercial de pequeno porte.

2.3 Considerações Finais

Nesse levantamento bibliográfico, percebeu-se que a maioria dos autores que aplicaram abordagens inspiradas em visão computacional nos seus experimentos, como nas pesquisas de Ravnik e Solina (2014), Ravnik e Solina (2013a) e de Yin et al. (2012), utilizou implementações de código-fonte aberto disponível na biblioteca *Open Source Computer Vision* (OpenCV), entre elas, a implementação do detector de faces proposto por Viola e Jones (2001), e os classificadores de gênero e idade inspirados nas técnicas de Turk e Pentland (1991) e de Fisher (1936). Além disso, certos classificadores foram validados em pequenas bases de dados, como a FERET (PHILLIPS et al., 1997) que contém 3.737 imagens e foi utilizada no treinamento do modelo de Ravnik e Solina (2014).

Esta seção destacou as principais pesquisas envolvendo a análise da atenção dos espectadores e o impacto da sinalização digital. O impacto indireto da sinalização digital nas vendas e no comportamento humano foi analisado. A presente dissertação utiliza essas pesquisas como suporte para validar as inferências realizadas. A partir da análise da literatura, observa-se a necessidade de otimizar os sistemas de exibição automática de anúncios, o que pode ser feito de maneira menos intrusiva utilizando recursos de visão computacional. Além disso, existem lacunas de pesquisa na categorização dos espectadores, que foram tratadas pela abordagem proposta nesta dissertação com a geração de redes bayesianas na associação do perfil demográfico do espectador ao conteúdo exibido.

Capítulo 3

Materiais e Métodos

Neste Capítulo são apresentados os materiais e métodos empregados nesta dissertação. A Seção 3.1 apresenta os materiais na forma de uma descrição de itens que foram utilizados nos experimentos, enquanto a Seção 3.2 apresenta a metodologia adotada para consecução dos objetivos propostos. Na Seção 3.3, apresentam-se os métodos associados que fazem parte das abordagens propostas. Por fim, a Seção 3.4, apresenta as considerações finais deste capítulo.

3.1 Materiais

Com o intuito de facilitar a replicação dos experimentos realizados nesta dissertação, esta seção apresenta detalhadamente os materiais utilizados nos experimentos. A seguinte lista organiza esses materiais de forma resumida em três categorias: bases de dados, *hardwares* e *softwares*. As subseções seguintes detalham cada item para tornar mais clara a sua utilidade.

- Bases de Dados:
 - *Face Detection Data Set and Benchmark* (FDDB) (JAIN; LEARNED-MILLER, 2010): utilizada para treinar e validar o detector de faces.
 - *MORPH Database* (RICANEK; TESAFAYE, 2006): utilizada para treinar e validar o classificador de gênero e de idade.
 - Bases de vídeos próprias:

- * Exibição: catálogo de vídeos provenientes da Internet ou de canais abertos de televisão, que foram utilizados para serem exibidos aos espectadores enquanto eles eram gravados.
- * Captura: para cada vídeo de exibição, existe um vídeo sincronizado de captura dos espectadores, *i.e.*, o vídeo foi capturado simultaneamente com a exibição.

- *Hardwares:*

- Aparelho de televisão LG modelo 42LM3400 de 42 polegadas *full HD* (1920 × 1080) utilizado na exibição dos anúncios aos espectadores.
- Câmera modelo Logitech HD Pro C920 com resolução *full HD* codificação H.264 e taxa de captura de 30 quadros por segundo (*frames per second* - fps). Utilizada para capturar imagens dos espectadores enquanto assistiam aos anúncios exibidos na televisão.
- Computador Intel Core i7 3770 3,4GHz com 8GB de memória RAM: utilizado para sincronizar as exibições, exibidas na televisão via cabo *High-Definition Multimedia Interface* (HDMI), com as capturas recebidas da *webcam* via cabo *Universal Serial Bus* (USB).
- Teclado e *mouse*: utilizados para operar o computador realizando tarefas como: fornecer *login* e senha para inicializar o sistema operacional e acessar o terminal para executar os *scripts* na exibição de vídeos e captura de imagens dos espectadores.
- Monitor auxiliar: utilizado para visualizar as configurações do computador enquanto os vídeos são exibidos na televisão.
- Torre de suporte para televisão: estrutura que agrega os equipamentos utilizados na sinalização digital: a televisão, a *webcam* e o computador.
- Placa de vídeo Nvidia Titan X com 12GB de *Random-Access Memory* (RAM): utilizada para realizar o treinamento da rede convolucional do classificador de gênero e o de idade.

- *Softwares:*

- Detector de Faces: *Pixel Intensity Comparisons Organized in Decision Trees* (PICO) proposto por Markus et al. (2013). Utilizado no processamento dos vídeos capturados pela *webcam* durante a exibição dos anúncios.
 - * *Open Source Computer Vision library* (OpenCV): é uma biblioteca de visão computacional com código-fonte aberto utilizada pelo PICO.
- *Framework* para *Deep Learning* CAFFE, desenvolvido por Jia et al. (2014): utilizado no treinamento da rede convolucional para estimação da idade e a classificação do gênero.
- Arquitetura de Rede: GoogLeNet, desenvolvida por Szegedy et al. (2014). Especificação da rede utilizada no aprendizado das características faciais para distinção dos gêneros e das idades.
- Ferramenta para construção das redes bayesianas: WEKA, desenvolvida por Hall et al. (2009). Ferramenta utilizada no processamento dos dados para análise dos resultados obtidos.

3.1.1 Bases de Dados

Nos experimentos realizados nesta dissertação, foram utilizadas quatro bases de dados: a FDDB, que serviu para validar o detector de faces; a MORPH, que foi utilizada para treinar o classificador de gênero e idade; e as bases de vídeos próprias, (i) a de *exibição*, que se caracteriza como uma relação de vídeos públicos que foram selecionados para serem exibidos aos espectadores e (ii) de *captura*, que consiste em vídeos privados correspondentes aos espectadores assistindo aos vídeos exibidos. Mais detalhes sobre cada uma dessas bases são discutidos nas subseções seguintes.

FDDB

A base de dados *Face Detection Data Set and Benchmark* (FDDB) (JAIN; LEARNED-MILLER, 2010), contém 2.845 imagens sendo 5.171 faces rotuladas manualmente, isto é, são fornecidas as coordenadas das faces nas imagens, podendo aparecer mais de uma face por imagem. As faces estão dispostas em cenários reais, ou seja, os ambientes em que as fotografias foram registradas são diversificados e não há controle de pose ou de iluminação

das faces, há sobreposição e/ou oclusão das faces e algumas faces são pouco nítidas. Tais variações dificultam a obtenção de bons resultados por detectores de faces nessa base de dados. Alguns exemplos de imagens podem ser observadas na Figura 3.1, em que são apresentadas imagens com sobreposição, inclinação da cabeça, mais de uma face por imagem, entre outras variações. Também são apresentadas as marcações da região de face, que foram utilizadas para validar o detector de faces discutido na Seção 3.1.3. Esta base de dados foi selecionada pois apresenta uma vasta gama de testes de validação utilizando diferentes detectores de faces, facilitando a escolha do detector que melhor se aplica ao contexto.

MORPH

A base de dados MORPH (RICANEK; TESAFAYE, 2006), contém 55.134 imagens originais de mais de 13.618 indivíduos, fotografadas entre os anos de 2003 e 2007. As idades variam de 16 a 69 anos, com uma idade média de 33 anos. O número médio de imagens por indivíduo é 4 e o tempo médio entre as fotografias foi de 152 dias, sendo o mínimo de 1 dia e o máximo de 1681 dias. O desvio-padrão de dias entre imagens é 180. A Figura 3.2 apresenta uma amostra de faces dessa base de dados com indivíduos de diferentes idades e ambos os gêneros. Essa base foi utilizada para treinar o classificador de gênero e de idade discutido na Seção 3.1.3. Esta base de dados foi selecionada por apresentar o maior número de amostras de faces frontais encontrado na literatura, o que permite utilizar classificadores como Redes Neurais Convolucionais, que utilizam grande volumes de dados para treinamento e fornecem resultados estado-da-arte em classificação.

Base de Vídeos Própria: Exibição

A base de vídeos própria de *exibição* consiste em um coletânea de vídeos utilizados para exibição aos espectadores, os quais pertencem a uma destas seis categorias: diversão, economia, tecnologia, policial, política e saúde. Esses vídeos foram exibidos aos espectadores a partir da TV aberta ou de canais públicos da internet, como o Youtube. Essas categorias foram selecionadas com base nos temas mais frequentes nos canais de consulta citados. Os vídeos transmitidos na TV aberta foram selecionados com base na sua data de publicação, dando-se preferência aos vídeos mais recentemente publicados. Enquanto os vídeos oriundos da internet, deu-se preferência aos mais visualizados e com boa avaliação por usuários.



(a) Exemplos de imagens da base de dados FDDB.



(b) Exemplos de imagens da base de dados FDDB com as faces marcadas.

Figura 3.1: Amostra das imagens da base de dados FDDB.

Fonte: Jain e Learned-Miller (2010).



Figura 3.2: Amostra de imagens da base de dados MORPH.

Fonte: Ricanek e Tesafaye (2006).

Todos os vídeos foram manualmente rotulados de acordo com o seu conteúdo, uma vez que, a busca na literatura não encontrou uma base apropriada para exibição aos espectadores que fosse devidamente rotulada.

A Figura 3.3 apresenta uma amostra de um vídeo que pertence à categoria política. Nesta figura, observam-se 50 quadros de um vídeo com mesmo intervalo temporal entre cada quadro. No vídeo em questão, proveniente do programa Bom Dia Brasil exibido em rede aberta pela emissora Rede Globo, mostra-se uma discussão entre os deputados na votação de afastamento do então presidente da Câmara dos Deputados do Brasil, Eduardo Cunha. Com o intuito de sintetizar os conteúdos explorados em cada categoria de vídeo, os temas abordados foram descritos em tópicos para contextualizar o leitor.

A base de vídeos exibidos é formada por 152 vídeos, que foram exibidos intercaladamente, em dias alternados e no mesmo período do dia. Uma relação completa dos vídeos exibidos pode ser encontrada no Apêndice C. Os vídeos exibidos foram rotulados em uma das seguintes categorias de acordo com seu conteúdo: policial (14 vídeos), economia (19 vídeos), saúde (20 vídeos), política (25 vídeos), entretenimento (34 vídeos) e tecnologia (40 vídeos).

- **Saúde.** De acordo com o Centro de Operações de Emergências em Saúde (COES, 2016), o Brasil passava por um surto nos casos de microcefalia no mesmo período que ocorreu a aquisição dos dados. Conseqüentemente, uma porção considerável da



Figura 3.3: Amostra de um vídeo sobre política.

Fonte: Rede Globo de Televisão. Vídeo intitulado: "Conselho de Ética adia mais uma vez votação do processo contra Eduardo Cunha", disponível em: <http://globoplay.globo.com/v/4663527/>, acessado em 11 de setembro de 2016.

programação na TV aberta tratou sobre esse tema. Sendo assim, essa classe é representada, na sua maioria, por vídeos sobre os casos da microcefalia, a sua relação com

o Zika vírus¹ e o combate aos mosquitos transmissores. Além disso, outros temas também associados à saúde foram exibidos, por exemplo, dicas de alimentação e de exercícios físicos, fiscalização na segurança dos brinquedos comercializados no país e técnicas para reverter a queda de cabelos causada pela quimioterapia.

- **Política.** No período do estudo, o Brasil passava por um momento político efervescente: o conselho de ética da câmara dos deputados se reunia para definir o andamento do processo *impeachment* contra Eduardo Cunha; o governador do estado de São Paulo, Geraldo Alckmin, anunciou o fechamento de escolas que afetariam mais de 300 mil estudantes; gravações mostram o plano de fuga de Nestor Cerveró que estava sendo investigado pela operação lava jato; prisão do ex-senador Delcídio do Amaral; decisões do Supremo Tribunal Federal (STF) sobre regras para o processo de *impeachment* da presidente recém-eleita, Dilma Rousseff; entre outros temas. Além disso, também houve várias exposições sobre a política internacional, tais como: combate entre Camarões e Nigéria; divulgação de estratégias terroristas do Estado Islâmico pela internet; e participação inédita das mulheres em eleições na Arábia Saudita.
- **Tecnologia.** Nesse tipo de conteúdo, foram agrupados os vídeos que são relacionados à tecnologia, tais como: pronunciamento da Sony sobre o lançamento de um novo *smartphone*; informativo sobre lista de aparelhos da Motorola que receberão atualização na versão de seu sistema operacional Android; apresentação de um novo modelo de motor V8 para carros esportivos; instruções de como proceder para usuários que tem seus serviços de internet prejudicados; discussão de mitos criados a respeito da bateria dos celulares; e demonstração de avanços na computação quântica envolvendo a Google e a Nasa. Esses foram os conteúdos de alguns vídeos exibidos nessa categoria.
- **Entretenimento.** Os vídeos agrupados nesta categoria informam opções de lazer ou proporcionam diversão ao espectador. Por exemplo, a divulgação de atrações culturais no estado, como a realização de um encontro nacional de música raiz em Campina Grande e a chegada do circo dos 7 anões em João Pessoa. Quanto aos vídeos que

¹Até então, apenas indícios desta relação eram noticiadas, o que foi posteriormente publicado pela médica campinense Adriana Melo (MELO et al., 2016).

proporcionam diversão, existem vários exemplos do canal Lista10² do Youtube, tais como: "Dez jogadores de futebol que se atrapalharam em suas jogadas", "Dez truques de mágica que terminaram de maneira inusitada" e "Dez cachorros dormindo de formas cômicas".

- **Economia.** Essa categoria foi representada por vídeos relacionados à crise econômica mundial que ocorre desde a queda do banco Lehman Brothers (NATIONS, 2013) e o impacto dela na economia brasileira (CARDOSO, 2013). Alguns exemplos de conteúdos abordados nos vídeos podem ser destacados: profissionais paraibanos que apostaram na criatividade para superar a crise; brasileiros que por conta da crise abandonaram as linhas de celular para utilizar o Whatsapp; explicação sobre a influência que o aumento do dólar têm sobre a economia brasileira; e o aumento nos preços dos combustíveis.
- **Policial.** Assuntos policiais são comumente noticiados nos telejornais, assim, os vídeos desta categoria foram representado pelas notícias jornalísticas que contenham envolvimento policial, por exemplo: aplicação de golpes por parte dos funcionários em churrascaria em João Pessoa; roubo de motos para colocar peças à venda (também em João Pessoa); e assassinato de diversos jovens pela polícia militar no Rio de Janeiro. Observar que as investigações policiais envolvendo políticos foram considerados temas *políticos*, para fins de organização.

Base de Vídeos Própria: Captura

A base de vídeos própria de *captura* consiste em uma base de vídeos que foram capturados dos espectadores enquanto assistiam aos anúncios exibidos, ou seja, são 164 vídeos de captura, cada um deles associado a um vídeo de exibição. As capturas ocorreram nos meses de Novembro e Dezembro de 2015, na Cantina de Dona Inês localizada na UFCG, em dias alternados e no mesmo horário (entre 11h30 e 13h00). Na Tabela 3.1, é apresentada a probabilidade *a priori* do vídeo capturado ter sido proveniente de uma exibição daquele tipo de conteúdo. A variável *causa* deste experimento é o tipo do conteúdo do vídeo exibido, objetivando conhecer o perfil demográfico dos espectadores. Estes valores foram calculados

²Disponível em: <https://www.youtube.com/user/lista10devolta>, acessado em 11 de setembro de 2016.

considerando a quantidade total de quadros processados para as diferentes categorias dos vídeos exibidos, por exemplo, os espectadores estavam assistindo aos vídeos políticos em 19,6% dos quadros processados. Mais detalhes sobre o método utilizado para gerar essa base são apresentados na Seção 3.2.

Tabela 3.1: Probabilidades *a priori* de todos os conteúdos transmitidos neste experimento.

Tipo do Conteúdo	Saúde	Política	Tecnologia	Entretenimento	Economia	Policial
Probabilidades à Priori (%)	20,2	19,6	17,9	15,2	14,3	12,8

Fonte: Próprio autor.

3.1.2 Hardwares

Para a realização das inferências sobre a atenção dos espectadores, foi realizado um estudo de caso em um estabelecimento comercial localizado no Edifício Prof. Shiva Prasad da Universidade Federal de Campina Grande (UFCG), que consistiu na exibição das seis categorias de vídeos aos espectadores que foram gravados enquanto prestavam atenção à exibição dos anúncios. O objetivo da pesquisa realizada nesta dissertação foi relacionar o conteúdo exibido com o gênero e a idade dos espectadores. Com o intuito de facilitar a replicação do experimento, no Apêndice D é disponibilizado o *script* utilizado para sincronizar as exibições dos vídeos com as capturas dos espectadores.

Na Figura 3.4, pode ser observado o equipamento utilizado nos experimentos e o ambiente em que foram realizadas as exibições de sinalização digital. Na Figura 3.4a, é exibida uma fotografia dos equipamentos utilizados nesta dissertação, de cima para baixo: (i) a televisão em LED utilizada na exibição dos anúncios; (ii) a *webcam*, posicionada logo abaixo do monitor de modo a capturar as faces frontais à tela; (iii) o computador, utilizado para executar os *scripts* de sincronização, exibindo os anúncios enquanto captura os espectadores via *webcam*; (iv) teclado e *mouse*, utilizados para operar o computador realizando tarefas como a inicialização do sistema operacional fornecendo o *login* e senha, como também, executando os *scripts* pelo terminal que realizam a exibição de vídeos e captura de imagens

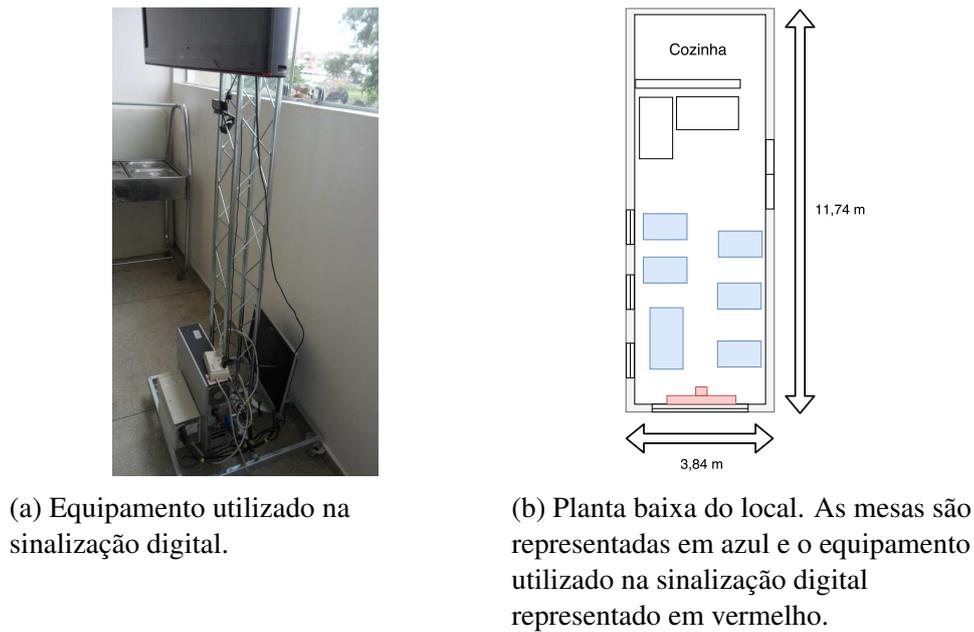


Figura 3.4: Estrutura e cenário da realização do experimento de sinalização digital.
Fonte: Próprio autor.

dos espectadores; (v) monitor auxiliar, foi utilizado para visualizar as configurações do computador, como o terminal que executa o *script* de exibição/captura de vídeos sincronizados enquanto os vídeos são exibidos apenas na televisão; e por fim, (iv) a torre de suporte, que acomodou todos os equipamentos durante o experimento. Na Figura 3.4b, é apresentada a planta baixa do estabelecimento comercial, ambiente em que foi realizado o experimento. Em que a televisão, voltada para os espectadores, foi representada de vermelho, enquanto que os espectadores estavam sentados à mesa, representadas em azul.

Também é ilustrado na Figura 3.5, um diagrama de conexão dos equipamentos, em que o cabo da esquerda ligando o computador ao monitor indica a conexão HDMI, que fornece vídeos da base própria de exibição com áudio e imagens digitais, enquanto que o cabo da direita conectando a *webcam* ao gabinete indica a conexão USB que coleta as capturas dos espectadores sincronizadas ao conteúdo exibido pelo computador. Estas imagens são úteis para ambientar o leitor ao cenário da experimentação.

3.1.3 Softwares

Esta seção referencia os *softwares* utilizados nesta dissertação, que estão organizados nas seguintes subseções: detector de faces, PICO, que foi utilizado para localizar as faces em

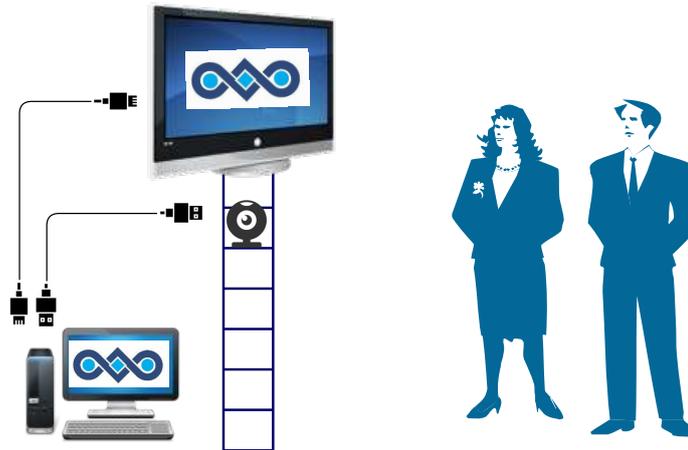


Figura 3.5: Disposição do *hardware*.

Fonte: Próprio autor.

todos os quadros de todos os vídeos da base de vídeos de captura; *framework* para *Deep Learning*, Caffe, que foi utilizado para realizar o treinamento do classificador de idade e gênero; definição da arquitetura de rede convolucional GoogLeNet; e por fim, a ferramenta para geração das redes bayesianas, Weka.

Detector de Faces

O detector de faces utilizado nesta dissertação é o *Pixel Intensity Comparisons Organized in Decision Trees* (PICO), desenvolvido por Markus et al. (2013), que é uma versão melhorada do algoritmo de Viola e Jones (2001). O PICO implementa algumas variações em relação ao algoritmo de Viola e Jones, como o uso do algoritmo de classificação GentleBoost (FRIEDMAN; HASTIE; TIBSHIRANI, 2000), mais robusto que o Adaboost (FREUND; SCHAPIRE, 1995) usado pelo algoritmo de Viola e Jones. A classificação ocorre em múltiplos estágios com aumento da complexidade em cada estágio. Cada etapa detecta quase todos os objetos de interesse (faces), rejeitando uma certa fração de não-objetos (não-faces). Assim, a maioria dos não-objetos é rejeitada nos estágios iniciais, isto é, com pouco tempo de processamento gasto. Ao se considerar o tempo de execução, foi observado que o PICO apresenta resultados satisfatórios, por exemplo, em relação ao detector de faces proposto por Li, Wang e Zhang (2011), o PICO foi 4,5 vezes mais rápido. Levando aproximadamente meio segundo para detectar todas as faces em uma imagem com resolução *full HD*.

Mais detalhes dos experimentos realizados pelo autor podem ser observados em Markus et al. (2013).

A Figura 3.6 ilustra o desempenho de diversos detectores de face aplicados à base de dados FDDB (JAIN; LEARNED-MILLER, 2010). São comparadas as métricas propostas por Jun e Kim (2012), Li, Wang e Zhang (2011), Jain e Learned-Miller (2011) e Viola e Jones (2001). Observa-se que o melhor resultado apresentado foi para o algoritmo de Jun e Kim (2012), entretanto, devido ao PICO ter o código-fonte aberto e ser escrito em C++, esse algoritmo foi selecionado para o desenvolvimento do experimento desta dissertação, pois, melhor se integra ao resto do sistema que possui outros módulos também escritos em C++, além de ser mais viável para o desenvolvimento de aplicações com tempo de resposta reduzido.

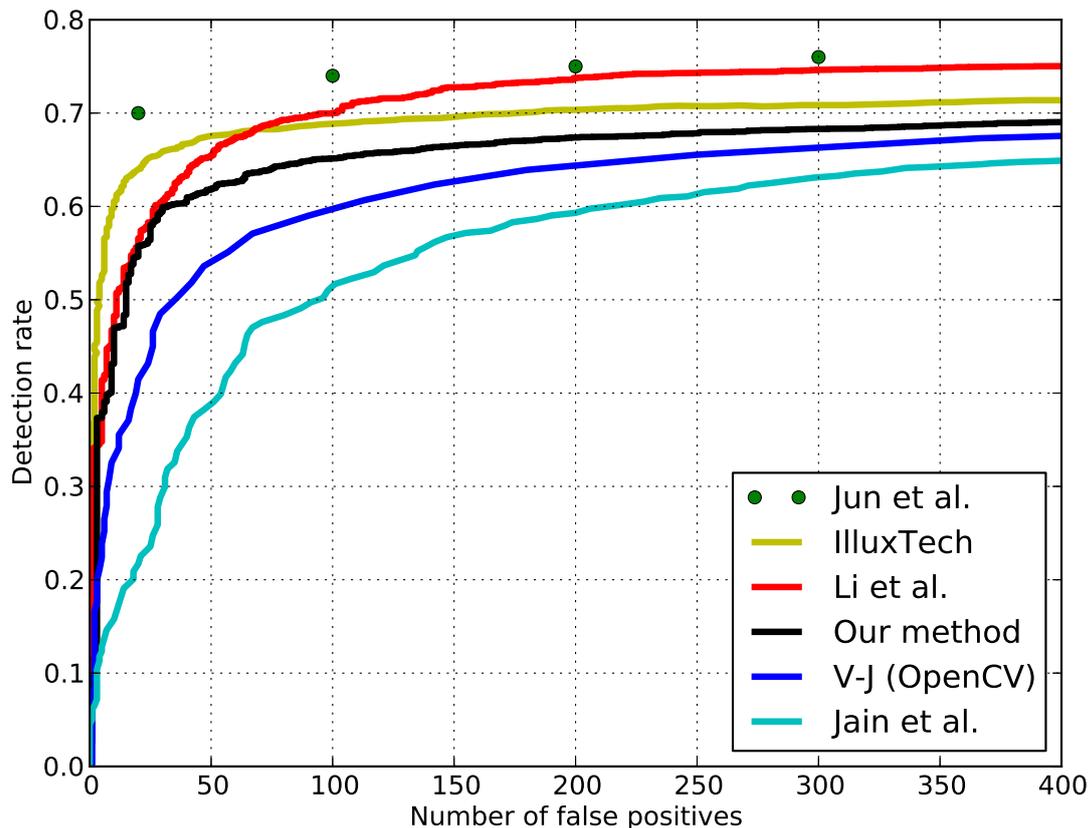


Figura 3.6: Comparação do detector de faces PICO com outros na base de dados FDDB. Fonte: Markus et al. (2013).

Nota: Observa-se no gráfico algumas informações em Inglês, que são elas: *Detection rate*: na ordenada do plano cartesiano representa a Taxa de Detecção; *Number of false positives* na abscissa do plano cartesiano representa o Número de Falsos Positivos; e *Our method*: representa o desempenho do algoritmo PICO proposto por (MARKUS et al., 2013).

O PICO utiliza a biblioteca de código-fonte aberto *Open Source Computer Vision* (OpenCV), que possui diversas funções básicas de processamento de imagem já implementadas. O OpenCV é distribuído sob a licença *Berkeley Software Distribution* (BSD), com interface em C/C++, Python e Java e com suporte para os sistemas operacionais Linux, Windows, MAC OS, iOS e Android. A versão do OpenCV 3.1.0 foi utilizada nesta dissertação.

Framework para Deep Learning: Caffe

Uma série de *frameworks* para execução de modelos de *Deep Learning* estão disponíveis ao público. O Quadro 3.1, identifica *frameworks* de particular interesse para a pesquisa desta dissertação, que são: o Caffe (JIA et al., 2014), o Decaf (DONAHUE et al., 2013), o OverFeat (SERMANET et al., 2013), o Theano/Pylearn (BERGSTRA et al., 2010) e o Torch (COLLOBERT; KAVUKCUOGLU; FARABET, 2011). Para compará-los foram utilizados os seguintes critérios: *framework* que seja de código-fonte aberto, escrito em C++ e que seja possível executá-lo tanto em CPU quanto em GPU. Assim, optou-se pelo Caffe, visto que, ele atende a essas quatro necessidades: (i) possui código-fonte aberto permitindo que seja possível realizar modificações necessárias para integrá-lo ao sistema apresentado nesta dissertação; (ii) que seja executável na CPU para que o sistema possa executar em máquinas comuns; (iii) que seja executável em GPU para realizar treinamentos com *big data* em menor tempo e (iv) é escrito em C++, o que além de ser computacionalmente mais eficiente, permite integrar de maneira mais fácil ao detector de faces também escrito em C++. Por fim, o Caffe apresenta um conjunto completo de funções para treinamento e teste de modelos de *Deep Learning*, como a implementação de redes convolucionais estado da arte, entre elas, a AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) e a GoogLeNet (SZEGEDY et al., 2014), além de possuir boa documentação e tutoriais disponíveis, contando também com uma larga comunidade de usuários no fórum, razões que se mostraram suficientes para usá-lo nesta dissertação.

Construção das Redes Bayesianas: WEKA

No processamento dos dados desta dissertação, foi utilizada a ferramenta Weka (WITTEN; FRANK, 2005) que é escrita em Java e disponível em código aberto sob uma licença GNU (*General Public License* - GPL). A sua utilização pode ser feita por meio de uma *Graphical*

Quadro 3.1: Comparação de *frameworks* populares de *Deep Learning*.

Framework	Linguagem	CPU	GPU	Código aberto
Caffe	C++	✓	✓	✓
Decaf	Python	✓	-	✓
OverFeat	Lua	✓	-	-
Theano/Pylearn	Lua	✓	✓	✓
Torch	Python	✓	✓	✓

Fonte: Jia et al. (2014).

User Interface (GUI) baseada em Swing ou pode ser acessada via linha de comando, sendo este modo adotado nesta pesquisa. Com o intuito de facilitar a replicação do experimento, os *scripts* gerados para processar os dados estão disponíveis no Apêndice D. Outra motivação para ter utilizado esta ferramenta se deu por ela disponibilizar várias rotinas já implementadas, como a construção automática de redes bayesianas, o que permitiu automatizar uma parte considerável do processamento dos dados.

3.2 Metodologia

Na Figura 3.7 é apresentado um diagrama das funcionalidades do método central proposto nesta dissertação, cujo objetivo é descobrir possíveis relacionamentos entre as variáveis do problema de sinalização digital, que são: o gênero (masculino ou feminino) e a faixa etária (definidas na Subseção 3.3.1) dos espectadores, a distância para o painel de sinalização (discutido na Seção 4.4) e o tipo de conteúdo exibido (apresentados na Subseção 3.1.1). O método recebe duas bases de dados: (i) base de vídeos a serem exibidos e (ii) a base de imagens de faces. A base de dados (i) contém uma coletânea de vídeos que serão exibidos para o espectador (*Base de Vídeos a Serem Exibidos*) paralelamente a captura de vídeos dos espectadores (*Base de Vídeos Capturados*). A base de dados (ii) (apresentada na Subseção 3.1.1), possui as faces utilizadas para treinar os classificadores de gênero e de idade, os quais, após treinados, fornecem modelos que são utilizados para estimar o gênero e a idade dos espectadores com base nas faces detectadas.

A *Base de Vídeos Capturados* (comentada anteriormente na Subseção 3.1.1), possui uma coletânea de vídeos dos espectadores que assistiram aos anúncios exibidos. Sendo um vídeo de captura para cada vídeo exibido. Desta forma, é possível saber qual vídeo os especta-

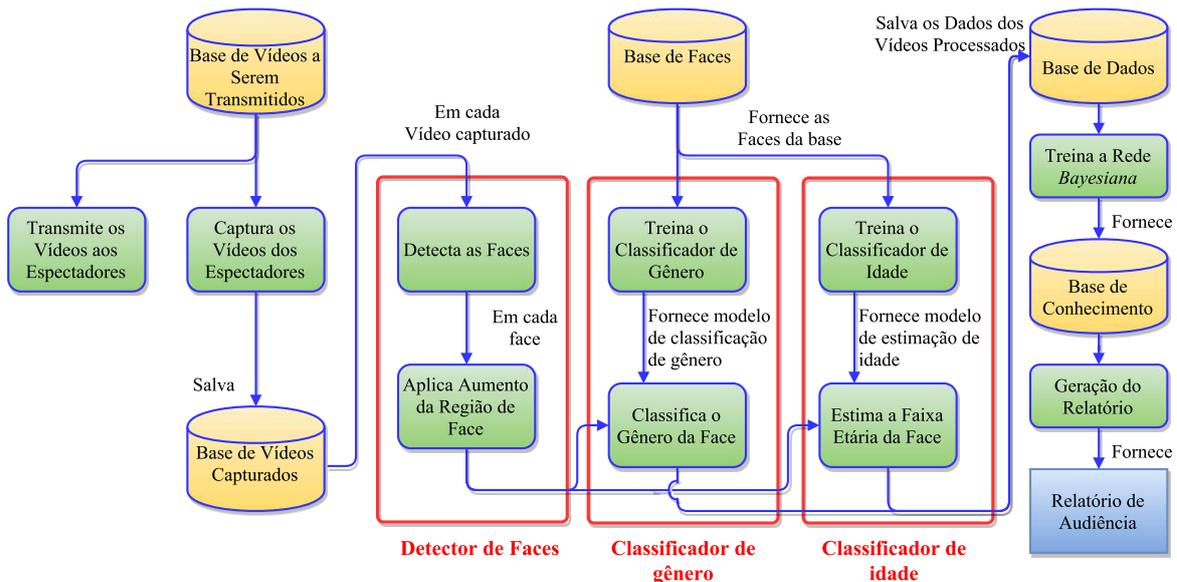


Figura 3.7: Representação do método proposto nesta dissertação.

Fonte: Próprio autor.

res de um determinado vídeo capturado estavam assistindo, conseqüentemente, é possível associar o vídeo capturado ao tipo do conteúdo do vídeo exibido com base em um rotulagem manual, que foi feita para os 164 vídeos exibidos. A lista completa dos vídeos exibidos com os seus respectivos rótulos é apresentada no Apêndice B.

Após a coleta da base de vídeos, inicia-se o processamento da base, o que consiste em realizar: (i) detecção de faces; (ii) classificação de gênero e (iii) classificação de idade. A detecção de faces consiste em analisar todos os quadros de todos os vídeos, detectando as faces frontais à câmera (conseqüentemente, também voltados ao monitor, visto que, se encontra imediatamente acima da câmera). Nesta dissertação não foi realizado o treinamento do detector, foi utilizado o treinamento realizado e disponibilizado pelo próprio autor (MARKUS et al., 2013). Após a detecção, é aplicado um aumento na região da face. Esse procedimento é aplicado para que a face recebida pelo classificador possua proporções compatíveis com o recorte aplicado pela rede neural que é detalhado na Seção 4.1.

Após a detecção de faces, os classificadores de gênero e de idade, previamente treinados, realizam a estimativa do gênero e da idade paralelamente. Todas as informações extraídas do vídeo processado são salvas na *Base de Dados*, que são: o gênero e a idade estimada pelo classificador, o conteúdo do vídeo exibido, além de informações secundárias como as coordenadas faciais e o número do quadro processado, que foi utilizado para fazer considerações

sobre o instante de visualização do espectador. Uma vez gerada a *Base de Dados*, é possível treinar com tais dados e gerar uma rede bayesiana que fornece uma *Base de Conhecimento* sobre a *Base de Vídeos de Captura*. Por fim, utilizando a *Base de Conhecimento* é possível realizar as análises e os comentários sobre o comportamento dos espectadores com a geração do *Relatório de Audiência*.

Os métodos que compõem a abordagem proposta são detalhadamente discutidos em quatro subseções: na Subseção 3.3.1, é explicado o método utilizado para definir as faixas etárias que serviram de parâmetro para a categorização dos espectadores; na Subseção 3.3.2, é apresentado como foi realizado o treinamento do classificador de gênero e de idade utilizando as faixas etárias definidas; enquanto que, a Subseção 3.3.3 inclui detalhes de mais baixo nível de como o processamento dos vídeos ocorreu; por fim, o Apêndice A, é incluída a fundamentação teórica para o Teorema de Bayes e as redes bayesianas, utilizados como base para a análise dos espectadores.

3.3 Abordagens Propostas

Esta seção apresenta as abordagens propostas desta pesquisa que foram agrupadas em três subseções: Subseção 3.3.1, apresenta o estudo realizado para definição das faixas etárias dos espectadores. Subseção 3.3.2, ilustra o procedimento realizado para classificação de gênero e estimação de idade pela Rede Neural Convolutacional. Por fim, na Subseção 3.3.3, é apresentado o funcionamento do sistema de captura de vídeo dos espectadores sincronizado à exibição dos vídeos de sinalização digital.

3.3.1 Definição das Faixas Etárias

Foi realizado um levantamento de como agrupar os espectadores em faixas etárias, para traçar o seu perfil e foi observado que diferentes autores têm desenvolvido análises em sinalização digital, escolhendo arbitrariamente as faixas etárias. Por exemplo, em (YIN et al., 2012) o autor agrupou os indivíduos como jovens, adultos e idosos definindo os intervalos de idade com critérios próprios. Para agrupar os espectadores em faixas etárias em (RAVNIK; SOLINA, 2013b), as faixas etárias fornecidas pela base de dados que treinou o seu classificador de idade foram utilizadas para categorizar o perfil de seu espectadores. Entre-

tanto, nesses trabalhos, não houve qualquer avaliação subjetiva de como definir os grupos etários. Portanto, nesta dissertação, utilizou-se da abordagem comumente utilizada por sociólogos (HOWE; STRAUSS, 2009), (FITCH, 1985) e (MILLER, 2011) para discriminar as faixas etárias. Eles utilizam o termo *gerações* para especificar os períodos de tempo que tenham em comum algum aspecto histórico, dessa maneira, permitindo que os indivíduos nascidos neste período se identifiquem com o coletivo, i.e, outras indivíduos que também nasceram nesse mesmo período. Para melhor esclarecer, as sete gerações comumente utilizadas pelos sociólogos são apresentadas no Quadro 3.2, tais gerações serviram de inspiração para categorizar os espectadores nesta dissertação.

Inspiradas nas definições das gerações, agrupando as faixas etárias contidas no Quadro 3.2, essas especificações foram utilizadas nesta dissertação e associadas à resolução da Agência Nacional de Saúde (ANS) (ANS, 2003), que define os limites a serem respeitados para a adoção de preço nos planos privados de saúde por faixa etária, para criar o nosso modelo de faixas etárias para maior organização na análise dos espectadores, uma vez que, não existem definições precisas dessas categorias. Pode ser observado, no Quadro 3.3, os termos utilizados nesse trabalho para categorizar os espectadores com base na idade estimada pelo classificador.

A MORPH (RICANEK; TESAFAYE, 2006), base de dados utilizada para treinar o classificador de idade, não dispõe de faces de indivíduos menores de 16 anos, uma vez que, tratam-se de indivíduos detidos pelo *Federal Bureau of Investigation* (FBI) nos Estados Unidos da América (EUA), ou seja, a base de dados contém apenas faces de indivíduos com maioridade penal nas leis regentes daquele país. Como também, a MORPH não dispõe de indivíduos de gerações mais avançadas, sendo 69 anos a idade máxima. Assim, uma limitação de se utilizar a rede treinada nesta dissertação em outros cenários, seria na estimação de idade de crianças e de pessoas com idade mais avançada, uma vez que, o conjunto de treinamento não possui exemplos de faces nessas faixas etárias. Entretanto, a aplicação desse modelo nesta dissertação, não se caracteriza como uma inconsistência nas análises. Pois, o ambiente em que foi realizado o experimento de sinalização digital não é (ou é pouco) frequentado por crianças, visto que, seu público predominante são jovens estudantes e adultos.

Quadro 3.2: Detalhamento das gerações de idades com base nas definições dos sociólogos.

Geração	Ano de Nascimento³	Aspecto histórico em comum
Z	1995-2012	Esse período se caracteriza pelo acesso à Internet e as redes sociais pelas crianças ao longo de uma grande recessão econômica da década de 2000 (HOPE, 2016).
Y	1986-1994	Crianças que cresceram em um período de grande desenvolvimento tecnológico, de maior prosperidade econômica e de maior facilidade ao acesso à informação (HOWE; STRAUSS, 2009).
X	1960-1985	As crianças nascidas nesse período, acompanharam uma grande mudança social e cultural em nível mundial (MILLER, 2011).
<i>Baby Boomers</i>	1943-1964	Este termo foi utilizado para referenciar o aumento significativo ("explosão") no número de nascimentos que ocorreu pós II Guerra Mundial (LUCHT, 2014).
Silenciosa	1923 - 1945	Pessoas que nasceram demasiadamente tarde para lutar a II Guerra Mundial e que viveram esse período na sua infância (CARLSON, 2008).
Grandiosa	1900-1924	São as pessoas que nasceram no período da I Guerra Mundial e que lutaram ou viveram a II Guerra Mundial (1939-1945) e a Guerra Fria (BROKAW, 2004).
Perdida	1882-1904	São os indivíduos que lutaram ou viveram a I Guerra Mundial (1914-1918) em sua juventude. Em sua fase adulta viveram a Grande Depressão (ou, a Crise de 1929) (FITCH, 1985).

Fonte: Especificadas na própria tabela.

Quadro 3.3: Mapeamento das gerações com base nas faixas etárias definidas pela Agência Nacional de Saúde (ANS).

Geração	Faixa Etária	Ano de Nascimento ⁴
Z	0-18	1997-2015
	19-23	1992-1996
	24-28	1987-1991
Y	29-33	1982-1986
	34-38	1977-1981
X	39-43	1972-1976
	44-48	1967-1971
<i>Baby Boomers</i>	49-53	1962-1966
	54-58	1957-1961
	59+	1956+

Fonte: Adaptado de ANS (2003).

3.3.2 Módulo de Classificação de Idade e Gênero

Esta subseção apresenta o módulo de classificação de gênero e de idade desenvolvido nesta dissertação. Sendo o classificador de idade publicado em Oliveira et al. (2016), o qual, apresenta uma metodologia de aumento de dados (chamada de: *data augmentation*), que forneceu resultados superiores ao estado da arte. O funcionamento desse classificador consiste em realizar quatro etapas, que são elas: (i) aumento sintético da base de dados, (ii) pré-processamento das faces, (iii) treinamento e (iv) teste. A diferença do classificador de idade para o de gênero, é que este último não possui a etapa (i) de aumento de dados, pois, por si só, forneceu boas taxas de acerto, conforme pode ser observado nas Figuras 3.8 e 3.9, que ilustram o funcionamento dos classificadores de gênero e de idade, respectivamente. As subseções seguintes detalham os procedimentos realizados em cada etapa de funcionamento dos classificadores, sendo identificadas as etapas que são comuns a ambos ou não.

Pré-processamento

O pré-processamento é aplicado a todas as imagens de entrada ao sistema, tanto na fase de treinamento quanto na fase de teste, é comum também a ambos os classificadores (o de gênero e o de idade). O pré-processamento consiste em realizar uma expansão da região da face detectada com o intuito de manter a razão de aspecto (em inglês *aspect ratio*) com o recorte aplicado pela rede convolucional adotada nos experimentos desta dissertação, isto é, a

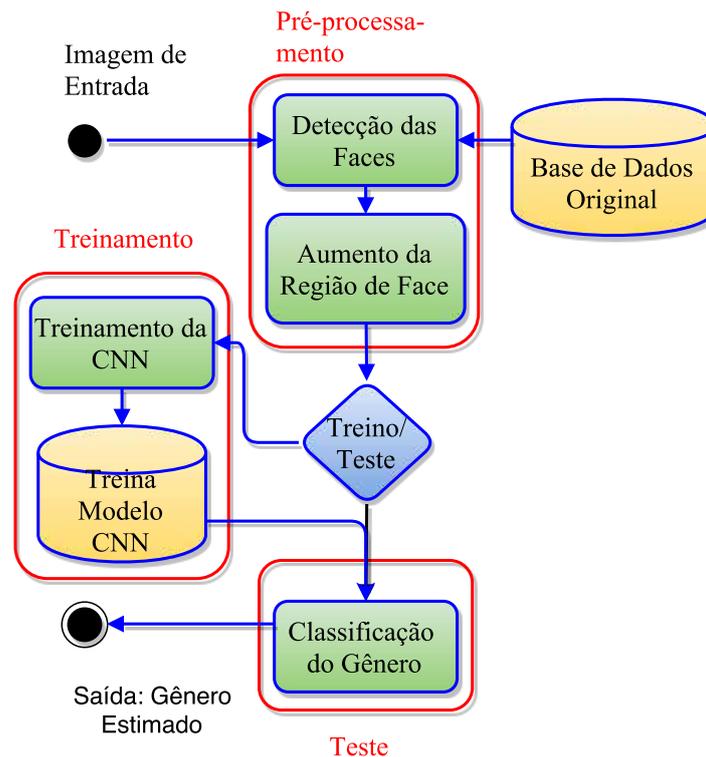


Figura 3.8: Representação do método de classificação de gênero.
 Fonte: Adaptado de Oliveira et al. (2016).

GoogLeNet recebe imagens com resolução padrão para classificação (256×256), na imagem são aplicados recortes aleatórios de resolução 224×224 no treinamento e um recorte central de mesma resolução na fase de teste. Ou seja, as faces detectadas tiveram uma expansão de aproximadamente 14% da região de face (o que corresponde ao aumento do recorte de 224×224 para a imagem de 256×256) e, por fim, a face detectada com o aumento da região de face é redimensionada para 256×256 , que é a resolução padrão de entrada para a GoogLeNet.

A Figura 3.13 ilustra todo o procedimento do aumento da região de face. Observar o recuo nas coordenadas iniciais do canto superior esquerdo da face na Subfigura 3.10a, em que o *Ponto A* indica a coordenada original e o *Ponto B* indica a nova coordenada após o aumento da região de face. Já a Subfigura 3.10b mostra o aumento nas dimensões da face, que, por se tratar de um quadrado detectado, o aumento é igual tanto para a largura quanto para a altura da face. Desta forma, a face detectada e a aumentada são concêntricas.

Em alguns casos, a face é detectada próximo à borda da imagem, consequentemente, o aumento da região de face pode passar dos limites da imagem. Assim, para evitar a perda

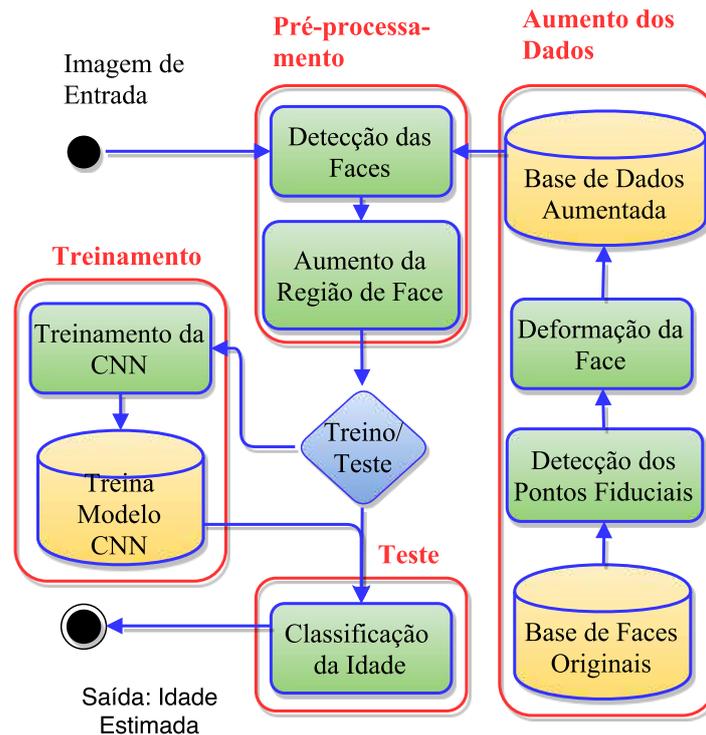
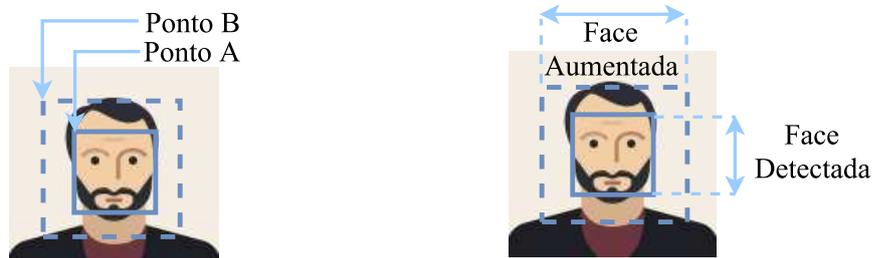


Figura 3.9: Representação do método de classificação de idade.

Fonte: Adaptado de Oliveira et al. (2016).



(a) Modificação nas coordenadas iniciais da face. (b) Modificação nas dimensões da imagem.

Figura 3.10: Exemplos de variações de face geradas sinteticamente.

Fonte: Adaptado de Oliveira et al. (2016).

do *aspect ratio*, foi utilizada a seguinte abordagem: espelhar a imagem em 8 direções e aplicar o recorte utilizando informações da imagem original com pedaços espelhados. A Figura 3.11 esclarece esse procedimento ilustrando um caso em que a face foi detectada próximo ao limite da imagem. Observar que a face recortada na direita, exibe um pedaço da testa da criança que foi espelhada na parte superior da imagem. Assim, o recorte central de resolução 224×224 pixels que a rede GoogLeNet irá aplicar na fase de teste, consistirá

apenas em região de face.

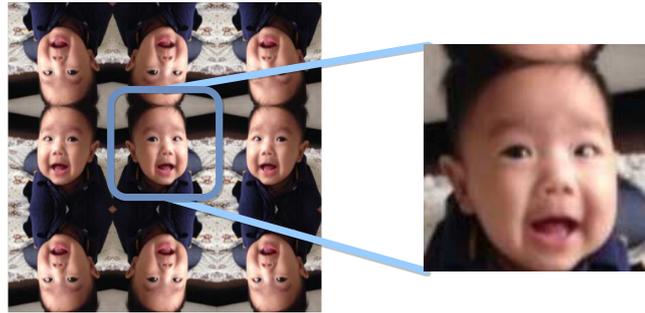


Figura 3.11: Exemplo de face detectada próximo à margem da imagem.

Fonte: Próprio autor utilizando uma imagem de Eidinger, Enbar e Hassner (2014).

Aumento dos Dados

O aumento no número de imagens da base de dados de treinamento, consistiu em fornecer mais amostras de imagens rotuladas para treinamento da rede convolucional, com o objetivo de maximizar o seu desempenho. Esse aumento de dados foi realizado apenas para o classificador de idade, que possui taxas de acerto inferiores às daquelas do classificador de gênero, com o intuito de melhorar as suas taxas, foi empregada tal técnica, o que resultou em um classificador com resultados competitivos com o estado da arte.

A técnica de aumento de dados consiste em duas etapas: (i) detectar os pontos fiduciais na face (com o aumento de região de face) e (ii) realizar a deformação em pontos específicos. A detecção dos pontos fiduciais foi realizada utilizando o algoritmo proposto por Schaefer, McPhail e Warren (2006), chamado *Active Appearance Models (AAM)*, que detecta 58 pontos em contornos de características faciais, tais como: as sobrancelhas, os olhos, o nariz, a boca e a mandíbula. Como pode ser observado na Figura 3.12, os 58 diferentes pontos fiduciais detectados em uma face de exemplo, numerados de acordo com a biblioteca utilizada, a AAM⁵. Os pontos marcados de 0 a 12 correspondem à mandíbula, os pontos de 13 a 20 correspondem ao queixo e os de 21 a 27 correspondem aos pontos do nariz, os quais, foram utilizados para realizar ligeiras deformações nas faces dos indivíduos e gerar faces sintéticas com variações sutis em relação à imagem original. As faces geradas mantiveram o rótulo da imagem original, ou seja, manteve a mesma idade.

⁵<https://github.com/greatyao/aamlibrary>

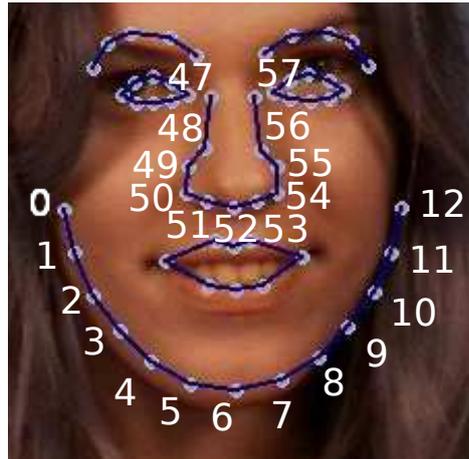


Figura 3.12: Resultado dos pontos detectados enumerados.

Fonte: Modificações em uma imagem da base de dados MUCT (MILBORROW; MORTEL; NICOLLS, 2010).

Uma vez que todos os pontos fiduciais das faces tenham sido detectados, as marcações foram utilizadas para gerar deformações na forma da mandíbula, do nariz e do queixo. Para isso, foram utilizadas as coordenadas de cada característica facial e aplicadas modificações em seus eixos x e y , deixando-os mais largos/finos ou com forma triangular/quadrada. Foi utilizada a biblioteca `imgwarp`⁶ para realizar a deformação após modificar as coordenadas das características faciais.

A Figura 3.13, ilustra todas essas deformações geradas. Foram utilizados os pesos apresentados nas Tabelas 3.2, 3.3 e 3.4, que representam as modificações nas coordenadas x e y de cada ponto fiducial modificado para a mandíbula, o nariz e o queixo, respectivamente. Após a geração da base de dados aumentada, a MORPH que possui aproximadamente 55 mil imagens originalmente, passou a conter mais de um milhão de imagens. Os experimentos de validação mostraram ganhos significativos quando a abordagem proposta de aumento de dados foi utilizada, o que é discutido no Capítulo 4.

Treinamento e Teste

O treinamento e o teste dos classificadores de gênero e de idade se deu utilizando a arquitetura de rede GoogLeNet, que é uma instanciação de uma rede convolucional inspirada no módulo *inception* proposta por Szegedy et al. (2014). Essa arquitetura de rede obteve o

⁶<https://github.com/cxcxcxcx/imgwarp-opencv>



Figura 3.13: Exemplos de variações de face geradas sinteticamente.

Fonte: Imagens da base de dados MORPH (RICANEK; TESAFAYE, 2006), com as deformações aplicadas em Oliveira et al. (2016).

primeiro lugar na competição de reconhecimento de objetos ImageNet *Large-Scale Visual Recognition Challenge* 2014 (ILSVRC2014) (RUSSAKOVSKY et al., 2014), que utiliza a

Tabela 3.2: Modificações nos pontos fiduciais da mandíbula.

Pontos Fiduciais	0	1	2	3	4	5	6	7	8	9	10	11	12
Coordenadas	(x, y)												
Largo	-1, 0	-1, 0	-1, 0	-1, 0	-1, 0	-1, 0	0, 1	1, 0	1, 0	1, 0	1, 0	1, 0	1, 0
Fino	1, 0	1, 0	1, 0	1, 0	1, 0	1, 0	0, 1	-1, 0	-1, 0	-1, 0	-1, 0	-1, 0	-1, 0

Fonte: Inspirado em Oliveira et al. (2016).

Tabela 3.3: Modificações nos pontos fiduciais do nariz.

Pontos Fiduciais	47	48	49	50	51	52	53	54	55	56	57
Coordenadas	(x, y)										
Largo	-1, 0	-1, 0	-1, 0	-1, 0	-1, 0	0, 0	1, 0	1, 0	1, 0	1, 0	1, 0
Fino	0, 0	0, -1	0, -1	0, 0	0, 1	0, 1	0, 1	0, 0	0, -1	0, -1	0, 0

Fonte: Inspirado em Oliveira et al. (2016).

Tabela 3.4: Modificações nos pontos fiduciais do queixo.

Pontos Fiduciais	4	5	6	7	8
Coordenadas	(x, y)				
Triangular	0, 2	0, 3	0, 4	0, 3	0, 2
Quadrático	-2, 2	-2, 2	0, 0	2, 2	2, 2

Fonte: Inspirado em Oliveira et al. (2016).

base de dados ImageNet com mais de um milhão de imagens no treinamento, sendo 1000 classes de objetos distintos.

A base de dados MORPH foi utilizada no treinamento dos classificadores. Inicialmente, todas as imagens da base de dados, que fazem parte do treinamento, são convertidas para um formato de suporte do Caffe, o *Lightning Memory-Mapped Database* (LMDB). O formato de dado LMDB armazena o seu conteúdo na forma chave-valor, em que, cada imagem da base é associada a um valor (representando o seu gênero, a sua idade ou a faixa etária). O *script* utilizado para realizar essa conversão consta no Apêndice D.

Na fase de teste dos classificadores, foi utilizado o módulo *Deep Neural Network* (DNN) do OpenCV. Assim, é possível carregar os pesos da rede treinada com o Caffe e realizar as predições diretamente nas imagens obtidas pelo detector de faces utilizando o módulo DNN/OpenCV, o que facilita a integração com detector de faces PICO, que também utiliza funções da biblioteca OpenCV.

3.3.3 Análise Automática dos Espectadores

O método de análise automática elaborado nesta dissertação, consiste em associar o perfil dos espectadores ao conteúdo dos anúncios exibidos. Para isso, foi aplicado em toda a base de vídeos de captura, o detector de faces seguido dos classificadores de idade e de gênero nas faces detectadas. Por fim, as informações de gênero e de idade de cada quadro foram relacionadas ao conteúdo exibido. Com base nessas informações, pode-se entender o funcionamento da aplicação de exibição síncrona à captura, conforme exibido na Figura 3.14, que ilustra o processo de geração da base de vídeos de captura.

O método proposto fornece uma base de vídeos capturados para processamento *off-line*, isto é, não é necessário que o processamento seja em tempo real (embora seja possível processá-lo em tempo real). Assim, como pode ser observado na Figura 3.14, o *script* de exibição/captura síncrona, calcula o tempo de execução do vídeo de exibição e inicia o processo de captura da *webcam* pelo mesmo tempo. O vídeo capturado mantém o identificador (ID) do vídeo exibido. Isso permite que no processamento da base de vídeos de captura, seja possível associar o conteúdo do vídeo que foi exibido aos espectadores enquanto eles assistiam. O Apêndice D apresenta o *script* utilizado para realizar o processo descrito na Figura 3.14.

Inicialmente, o *script* recebe dois parâmetros (i) o diretório contendo os vídeos à serem exibidos e (ii) o caminho para um diretório vazio, onde os vídeos capturados serão salvos. Cada vídeo do diretório dos vídeos de exibição é selecionado aleatoriamente, o *script* calcula o tempo de duração desse vídeo (t segundos) e captura simultaneamente da *webcam* (também por t segundos). Ao fim de cada captura, o vídeo é salvo com um identificador que é utilizado na fase de processamento dos dados e, então, o próximo vídeo da lista de exibição é selecionado aleatoriamente até que todos os vídeos tenham sido exibidos sem repetição. A Equação 3.1 ajuda a entender como o tempo de duração foi calculado.

$$q_c = \frac{q_e}{fps_e} * fps_c, \quad (3.1)$$

em que, a quantidade de quadros do vídeo de *captura* é representada por q_c e a quantidade de quadros do vídeo de *exibição*, é indicada por q_e . A taxa de quadros por segundo do vídeo *capturado* é representada por fps_c e a do vídeo *exibido* fps_e . Assim, essa equação possibilita

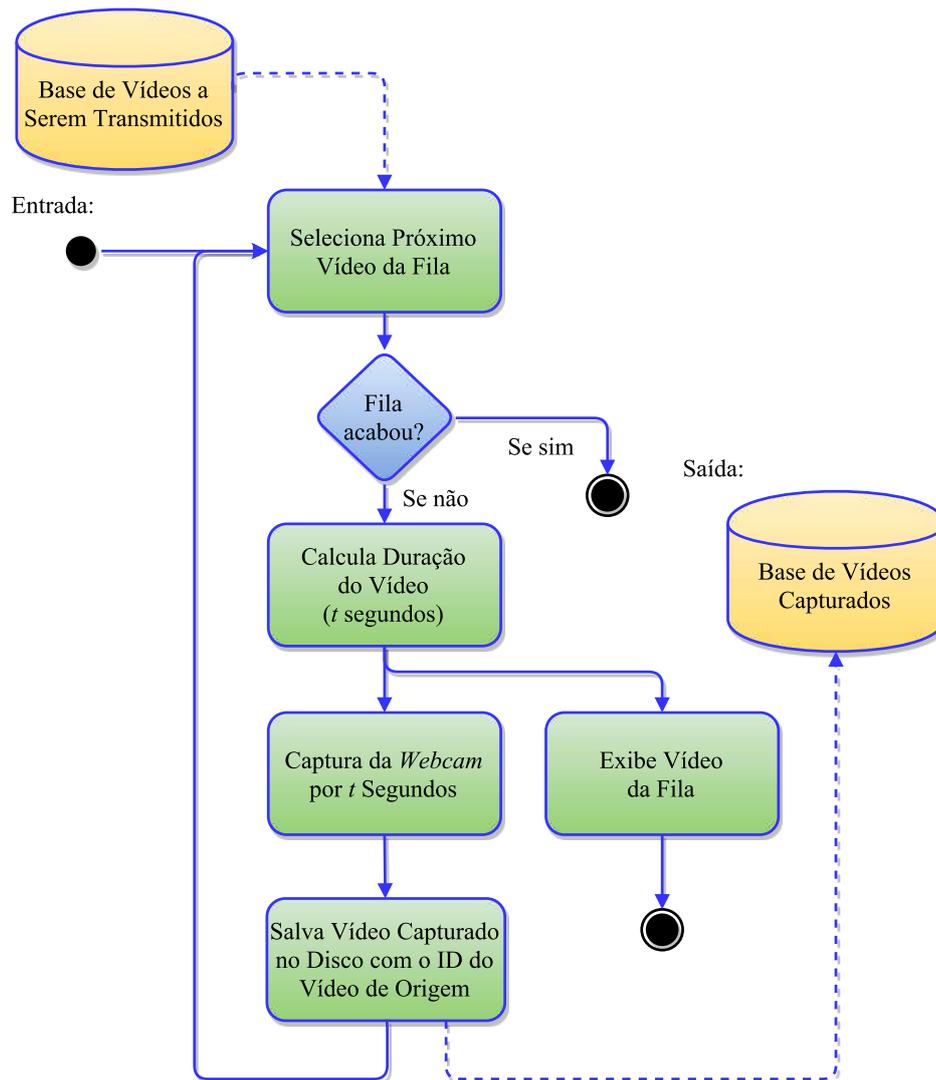


Figura 3.14: Diagrama de estados para o processamento em tempo de execução.
Fonte: Próprio autor.

a sincronização dos vídeos exibidos com os vídeos capturados. Sincronismo é garantido apenas quando se realizam os cálculos adequadamente e não apenas utilizando a quantidade de quadros do vídeo exibido como base para a captura, pois haverá dessincronização no tempo de duração do vídeo capturado em casos que a taxa de reprodução seja diferente, observando que, não é necessário que o vídeo de captura e o de exibição estejam com a mesma taxa de quadros por segundo (fps).

Uma vez capturada a base de vídeos dos espectadores, inicia-se o processo de análise demográfica. Tal processo consiste em aplicar o detector de faces (PICO) em todos os quadros de todos os vídeos da base capturada, após isso, utilizou-se o classificador de idade e

de gênero que classifica com base na face detectada. Por fim, essas informações são salvas em uma base de dados, mais especificamente, no formato de arquivo *Attribute-Relation File Format* (ARFF) (WITTEN; FRANK, 2005), que é o formato de entrada para a ferramenta Weka, a qual possui métodos para a construção de redes bayesianas empregadas na análise demográfica dos espectadores. A Figura 3.15 ilustra o método descrito.

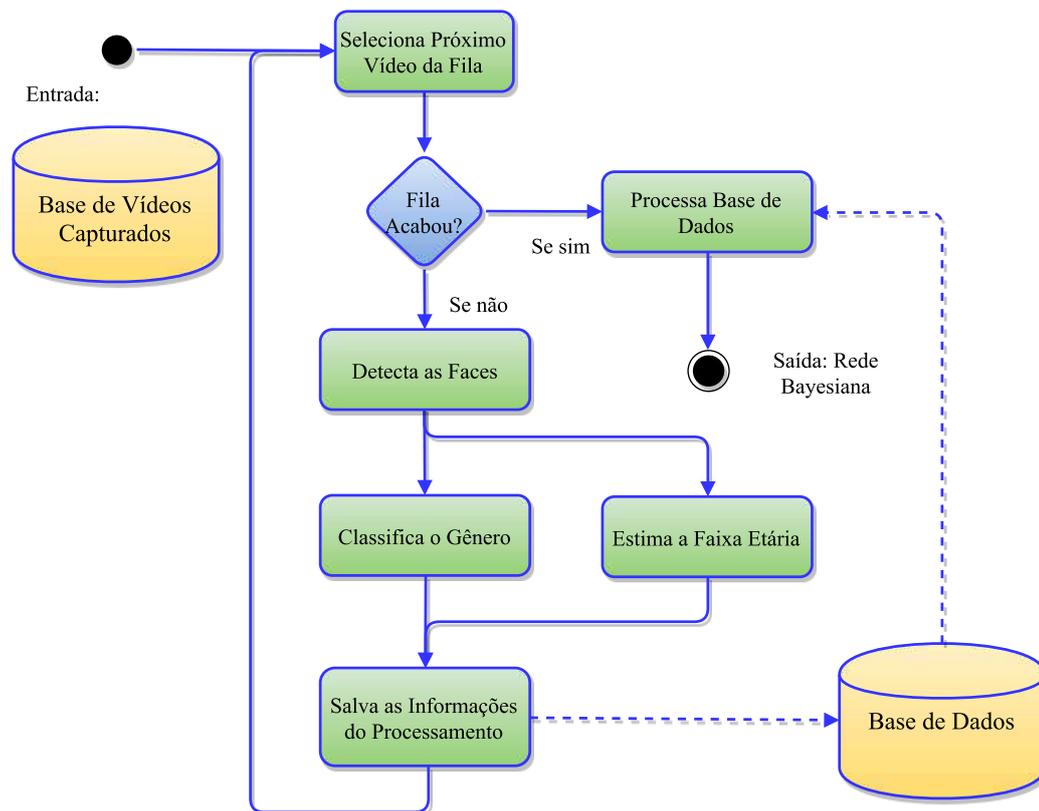


Figura 3.15: Diagrama de estados para o processamento *off-line*.

Fonte: Próprio autor.

Com isso, o sistema é capaz de identificar automaticamente uma série de informações e responder a diversas perguntas. Por exemplo, qual o público predominante dos vídeos políticos? Homens ou mulheres estiveram mais presentes? entre outras perguntas que foram respondidas no desenvolvimento de um estudo de caso nesta dissertação, conforme descrito no Capítulo 4. Essas informações são de interesse dos investidores e anunciantes em sinalização digital, visto que, constituem uma forma não intrusiva de selecionar anúncios a serem exibidos considerando o coletivo, permitindo o aumento do Retorno Sobre o Investimento (*Return On Investment* - ROI). Outras informações podem ser agregadas ao modelo para fazer análises mais detalhadas do perfil dos espectadores, por exemplo, incluir a hora do dia

que o anúncio foi exibido ou o dia do mês que as capturas ocorreram. Dessa maneira, as evidências levantadas nesta dissertação poderiam ser complementadas com outras variáveis a critério do cliente ou do pesquisador. O capítulo seguinte apresenta a análise dos resultados obtidos utilizando o sistema proposto em um cenário de captura real.

3.4 Considerações Finais

Neste capítulo, foram apresentados os materiais utilizados e os métodos desenvolvidos na pesquisa conduzida nesta dissertação. Os materiais consistem de base de dados, *hardwares* e *softwares*. As **base de dados** foram utilizadas para treinar/validar os detectores de face e classificadores de idade e de gênero, além disso, foram apresentadas duas bases próprias, a de vídeos exibidos e de vídeos capturados. Sendo a base de vídeos exibidos, uma base pública, disponível na internet e com rotulagem manual do conteúdo abordado em cada vídeo. Essa base foi utilizada para analisar o perfil demográfico dos espectadores em um cenário real de sinalização digital. Enquanto que a base de vídeos capturados consiste em uma sequência de vídeos capturados dos espectadores sincronizados com a exibição dos vídeos públicos. Os *hardwares* são os equipamentos utilizados no ambiente da sinalização digital. Consistem em: (i) o monitor, que exibiu os anúncios; (ii) a *webcam*, que realizou a captura dos espectadores; (iii) o computador, que executou o núcleo do sistema que realiza a exibição/-captura síncrona; (iv) a torre, estrutura onde todos os equipamentos foram montados; e por fim, (v) a placa de vídeo, que foi utilizada para diminuir o tempo de execução consumido no treinamento dos classificadores. A última categoria de materiais utilizados foram os *softwares*, que consistem em: implementação do detector de faces; *framework* de suporte para o classificador de gênero e de idade; ferramenta para geração da rede bayesiana.

Além disso, também foram apresentados nesse capítulo, os métodos desenvolvidos para analisar a atenção dos espectadores em um ambiente de sinalização digital. Os propósitos de tais métodos foram: analisar aspectos sociológicos e culturais na definição das faixas etárias e treinar o classificador de idade com base no estudo realizado. Além disso, utilizaram-se as faixas etárias definidas para agrupar os espectadores demograficamente e relacionar ao conteúdo exibido. Como parte do método, foi explicado como ocorre a análise automática dos espectadores, o que inclui a detecção de faces, a classificação de gênero e de idade e o pré-

processamento aplicado ao vídeo, além do aumento de dados aplicado na base de dados que aprimorou o desempenho do classificador de idade. Por fim, também foi contextualizado ao leitor o funcionamento do Teorema de Bayes, base para geração das redes bayesianas, que foi o método utilizado na análise da atenção dos espectadores. No próximo capítulo, são apresentados detalhes da validação do classificador de gênero e de idade, como também, os resultados do experimento desenvolvido sobre a análise da atenção dos espectadores, realizado em um ambiente real.

Capítulo 4

Avaliação Experimental

Este capítulo tem por objetivo descrever a avaliação experimental realizada nesta dissertação. Para isso, é apresentada a descrição da base de dados exibida aos espectadores, com o objetivo de relacionar o conteúdo exibido com informações demográficas (gênero e idade que foram estimados automaticamente). O capítulo também apresenta as redes bayesianas automaticamente construídas a partir dos dados coletados, juntamente com as tabelas de probabilidades, que serviram de base para análises estatísticas e validação experimental do estudo sobre a atenção dos espectadores.

Este capítulo está organizado como segue: na Seção 4.1, é apresentada a validação do classificador de idade e de gênero, sendo o método de aumento de dados aplicado ao classificar idade proposto nesta dissertação. A análise dos espectadores é descrita na Seção 4.2 (relação entre o gênero do espectador e o conteúdo exibido), na Seção 4.3 (relação entre faixa etária e conteúdo exibido) e Seção 4.4 (relação entre a idade, o gênero e a distância até o monitor com o conteúdo). Na Seção 4.5 são apresentadas relações entre as conclusões obtidas na avaliação experimental com resultados de pesquisa de mercado. Ou seja, embora a avaliação experimental conduzida tenha sido realizada em um único ambiente, os resultados de pesquisas globais destacam tendências de comportamento dos espectadores que, em alguns casos, são refletidas no cenário analisado, corroborando com as conclusões obtidas. A Seção 4.6 apresenta as considerações finais deste capítulo.

4.1 Validação dos Classificadores de Idade e de Gênero

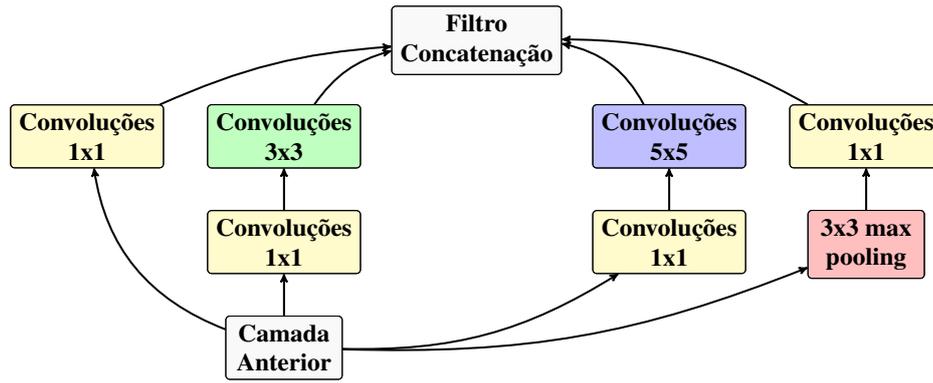
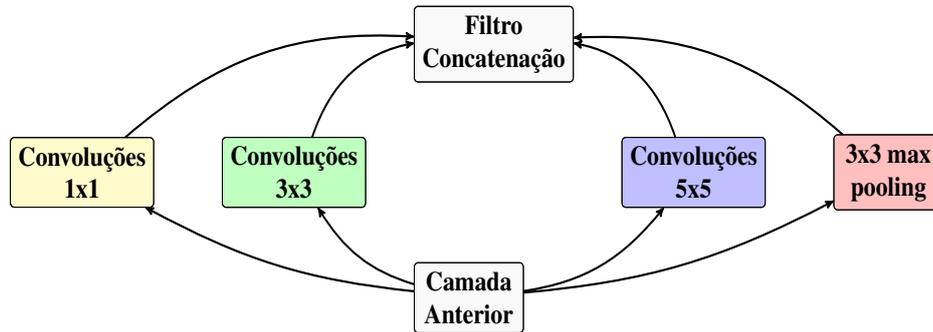
A arquitetura de rede convolucional adotada nestes experimentos foi a GoogLeNet, campeã na competição de reconhecimento de objetos ImageNet 2014 (RUSSAKOVSKY et al., 2014). A GoogLeNet é uma instanciação do modelo *inception* de arquitetura de rede. A Figura 4.1 ilustra o módulo *inception* de rede convolucional. Na Figura 4.1a é possível observar como as camadas convolucionais se interconectam de modo diferenciado das camadas tradicionais, visto na Figura 4.1b. A utilização de camadas múltiplas de convolução diminuem o tempo de processamento da rede, devido ao número de operações que cada camada executa (WANG; XU, 2015). Assim, a GoogLeNet requer menos memória e menor tempo de processamento para realização dos treinamentos e dos testes quando comparado a outros modelos de Redes Neurais Convolucionais (RNC). Além disso, o treinamento de grandes conjuntos de dados pode ser realizado com suporte de programação para GPU utilizando o *framework* Caffe, enquanto que os módulos de testes podem ser realizados utilizando apenas a CPU. Assim, o treinamento pode ser feito em um laboratório utilizando a GPU e levado para teste em um cenário real utilizando apenas a CPU, por exemplo.

Observa-se que a arquitetura *inception* tem duas vantagens principais em relação à abordagem tradicional: (i) através do uso de filtros de tamanhos diferentes em cada camada a rede retém informação espacial mais precisas; (ii) reduz significativamente o número de parâmetros livres na rede, tornando-se menos propensa a *overfitting* e permitindo que o aprendizado seja mais detalhado do que as arquiteturas tradicionais (CASTELLUCCIO et al., 2015).

Os classificadores de idade e de gênero foram validados com a base MORPH utilizando validação cruzada de cinco partes (*5-fold-cross-validation*).

Métricas Estatísticas

Esta subseção se destina a descrever as métricas estatísticas utilizadas para validar o classificador de idade e de gênero. O *F1-score* foi utilizado para avaliar o desempenho de classificadores em conjuntos de dados desbalanceados (SOKOLOVA; LAPALME, 2009). O *F-measure* é a média harmônica entre a precisão e a revocação (em inglês *recall*). Nesta dissertação, foi utilizada uma derivação da *F-measure*, chamada de *F1-score*, para avaliar o desempenho dos classificadores de gênero e de idade. O *F-measure* é dado por:

(a) Módulo *inception* da GoogLeNet.

(b) Módulo tradicional.

Figura 4.1: Arquiteturas de rede: GoogLeNet - *inception* e tradicional.
 Fonte: Adaptado de Szegedy et al. (2014).

$$F\text{-measure} = \frac{(\beta^2 + 1) \text{ Precisão} \times \text{Revocação}}{\beta^2 \times \text{Precisão} + \text{Revocação}}, \quad (4.1)$$

em que, considerando-se o balanceamento entre a precisão e a revocação, utiliza-se $\beta = 1$ (ZHANG; ZHANG, 2009), sendo assim, rescrevendo a Equação 4.1, têm-se:

$$F1\text{-score} = \frac{2 \times \text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}. \quad (4.2)$$

Enquanto que, a precisão é definida por:

$$\text{Precisão} = \frac{\sum_{i=1}^n \frac{VP_i}{VP_i + FP_i}}{n}. \quad (4.3)$$

E a Revocação é dada por:

$$\text{Revocação} = \frac{\sum_{i=1}^n \frac{VP_i}{VP_i + FN_i}}{n}. \quad (4.4)$$

Nas equações anteriores, n é o número de classes analisadas, VP o número de Verdadeiros Positivos, VN é o número de Verdadeiros Negativos, FP são os Falsos Positivos e o FN são os Falsos Negativos. Para o classificador de gênero, determina-se uma das classes para ser a classe positiva, por exemplo, o gênero masculino, então, têm-se: (VP) a face foi classificada como masculina quando se tratava de uma face masculina; (VN) a face foi classificada como feminina quando de fato se tratava de feminina; (FP) a face foi classificada como masculina quando a face era feminina; (FN) a face foi classificada como feminina quando era masculina.

Outra estatística utilizada para comparar o desempenho do método proposto para classificação de idade com os de outros autores foi a Pontuação Acumulada (PA), definida por:

$$PA = \frac{VP}{VP + VN + FP + FN}. \quad (4.5)$$

A pontuação acumulada foi aplicada para avaliar o desempenho do método proposto e comparar com outros autores (HUERTA et al., 2015) e (FERNÁNDEZ; HUERTA; PRATI, 2015) que não disponibilizaram os códigos-fonte de suas abordagens, mas que, disponibilizaram os resultados de avaliação de seus classificadores utilizando essa métrica. Além disso, foi utilizada a mesma base de dados e a mesma quantidade de partes na validação cruzada.

Avaliação de Desempenho

O experimento de validação da arquitetura de rede GoogLeNet como classificador de idade, consistiu em duas análises: (i) comparar o desempenho do classificador treinado sem utilizar a técnica de aumento de dados proposta nesta dissertação com o desempenho do classificador treinado utilizando a métrica de aumento de dados proposta; (ii) comparar o desempenho do classificador proposto com os desempenhos de classificadores de outras pesquisas. A principal contribuição dessa parte da pesquisa foi a proposição de uma metodologia de aumento de dados que eleva a taxa de acerto do classificador testado.

A comparação foi realizada utilizando a métrica *F1-score* utilizando a rede neural Goo-

gLeNet treinada com e sem a base aumentada, representados por GoogLeNet Com Aumento de Dados (GN-CAD) e GoogLeNet Sem Aumento de Dados (GN-SAD), respectivamente. Ambos os modelos treinados consideraram cada idade como um rótulo diferente, considerando que a base de dados MORPH possui uma variação de idade dos 16 aos 69 anos, então, o classificador considerou 54 classes diferentes. Para verificar o desempenho do classificador, as idades foram agrupadas em diferentes faixas etárias. Os resultados de *F1-score* para a média dos 5 grupos de validação cruzada, são representados na Figura 4.2. Em que, o aumento de dados foi aplicado apenas nos grupos de treinamento, mantendo-se as mesmas imagens nos grupos de teste. O valor de *F1-score* mais elevado, indica o classificador com melhor desempenho (SOKOLOVA; LAPALME, 2009). Os dois métodos de avaliação são os seguintes:

- **Método 1:** A faixa etária estimada é aquela que possui a idade mais provável.
- **Método 2:** A faixa etária estimada é aquela cujo somatório das probabilidades das idades do intervalo é maior.

Analisando a Figura 4.2, observa-se que o treinamento realizado utilizando a GN-CAD foi superior ao desempenho obtido ao se utilizar a GN-SAD. Além disso, conclui-se que o Método 1 apresenta resultados superiores quando existem muitas faixas etárias e utiliza aumento de dados. Ao considerar poucas faixas etárias (cinco ou menos), ambos os métodos são equivalentes. Assim, o GN-CAD aplicando o Método 1 de classificação foi utilizado nos experimentos práticos desta dissertação por apresentar o melhor desempenho. Sendo o resultado no conjunto de teste, para as quatro faixas etárias conforme apresentadas na Seção 3.3.1, de 72% de *F1-score*. Para o classificador de gênero, o *F1-score* obtido foi de 92%.

Comparando o classificador proposto (GN-CAD com o método de classificação do tipo 1, ou simplesmente GN-CAD), com outras pesquisas existentes, observa-se que o desempenho do classificador proposto é superior aos resultados apresentados nos artigos de Huerta et al. (2015) e de Fernández, Huerta e Prati (2015), que possuem desempenho estado-da-arte em classificação de idade. As métricas testadas com melhores desempenho foram: o *Fusion*, o *Multi-feature Ordinal Ranking* (MFOR) (WENG et al., 2013), o *Speeded Up Robust Features* (SURF) (BAY et al., 2008) e o *Local Binary Pattern* (LBP) (TOPI et al., 2000). O MFOR

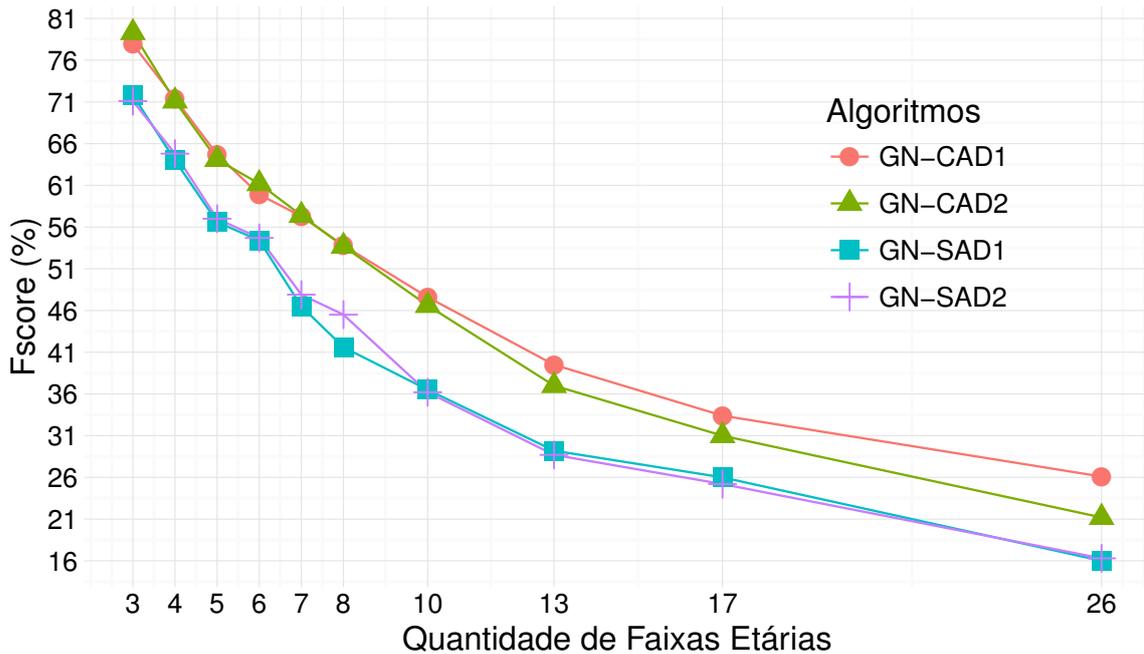


Figura 4.2: Desempenho dos dois modelos: com e sem aumento de dados.
Fonte: Adaptado de Oliveira et al. (2016).

é uma técnica de classificação híbrida que combina características obtidas pelos algoritmos *Principal Components Analysis* (PCA), o LBP e o *Bio-Inspired Features* (BIF) (MU et al., 2009). O PCA foi utilizado na redução de dimensão das imagens da base, o histograma LBP como descritor e o BIF como as características utilizadas. O SURF é um detector de pontos de interesse e descritor invariante à rotação e a escala. A abordagem *Fusion* consiste em combinar diferentes descritores: o SURF, o LBP e o *Histogram of Oriented Gradients* (HOG) (DALAL; TRIGGS, 2005). Foi utilizada a Pontuação Acumulada (PA) para avaliar o desempenho dos classificadores aplicados na base de dados MORPH. Na Figura 4.3, é apresentada a relação entre a PA e o nível de erro do classificador, na ordenada e abcissa do gráfico, respectivamente. O nível de erro corresponde à diferença entre a idade real e a estimada. Considerando que, quanto mais elevada a curva do classificador, melhor o seu desempenho (FAWCETT, 2006), pode-se observar que o melhor desempenho foi obtido para o classificador GN-CAD.

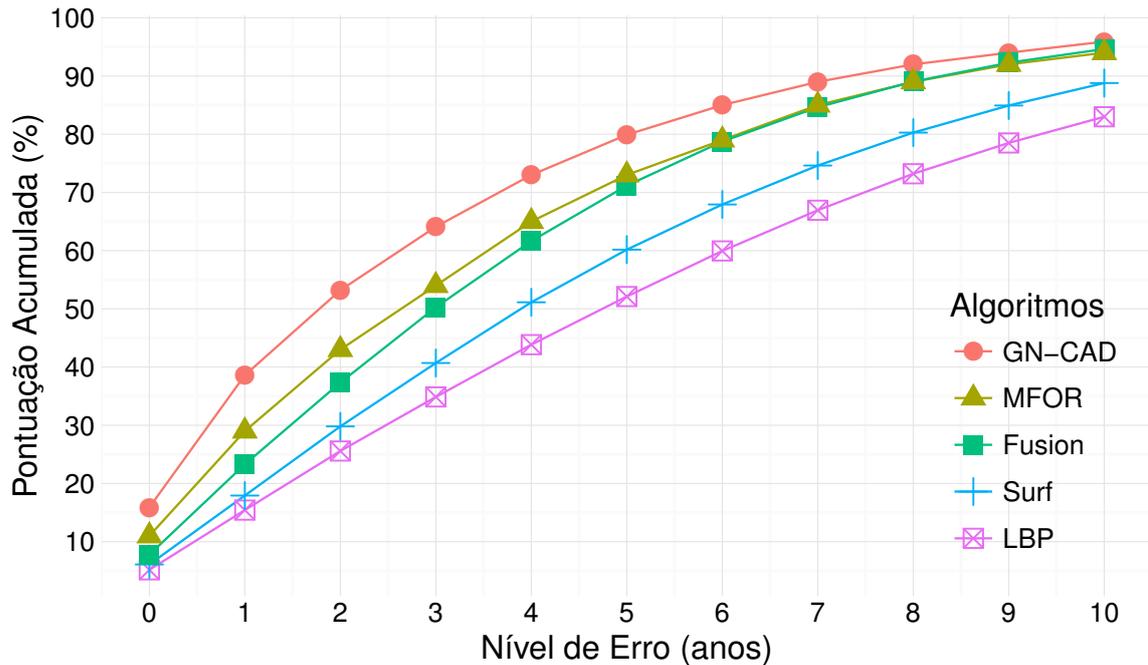


Figura 4.3: Comparação entre diferentes métodos de estimação de idade.
 Fonte: Adaptado de Oliveira et al. (2016).

4.2 Relação entre o Conteúdo exibido, o Gênero e a Idade do Espectador

Esta seção tem como objetivo apresentar a análise das relações entre o conteúdo exibido e o gênero dos espectadores. Na Figura 4.4, pode-se observar o diagrama da rede bayesiana gerada utilizando o algoritmo K2 (COOPER; HERSKOVITS, 1992), implementado na Ferramenta Weka (HALL et al., 2009), com as suas respectivas tabelas de probabilidades para cada variável, considerando os dados que foram processados nesta pesquisa. Com base na estrutura do diagrama, podemos evidenciar o efeito da variável *causa* (no caso, o conteúdo do vídeo exibido) sobre as variáveis *efeito* (o gênero e a idade do espectador). Sendo assim, observa-se que: (i) o gênero do espectador recebe influência apenas do conteúdo do vídeo exibido; enquanto que, (ii) a idade do espectador depende tanto do seu gênero quanto do conteúdo.

Com base nas tabelas de probabilidades exibidas na Figura 4.4, foi analisado o efeito causado pelo conteúdo exibido na atenção dos espectadores. Entretanto, o perfil do espectador, representado nesta dissertação pelo gênero e faixa etária a qual ele pertence, pode possuir

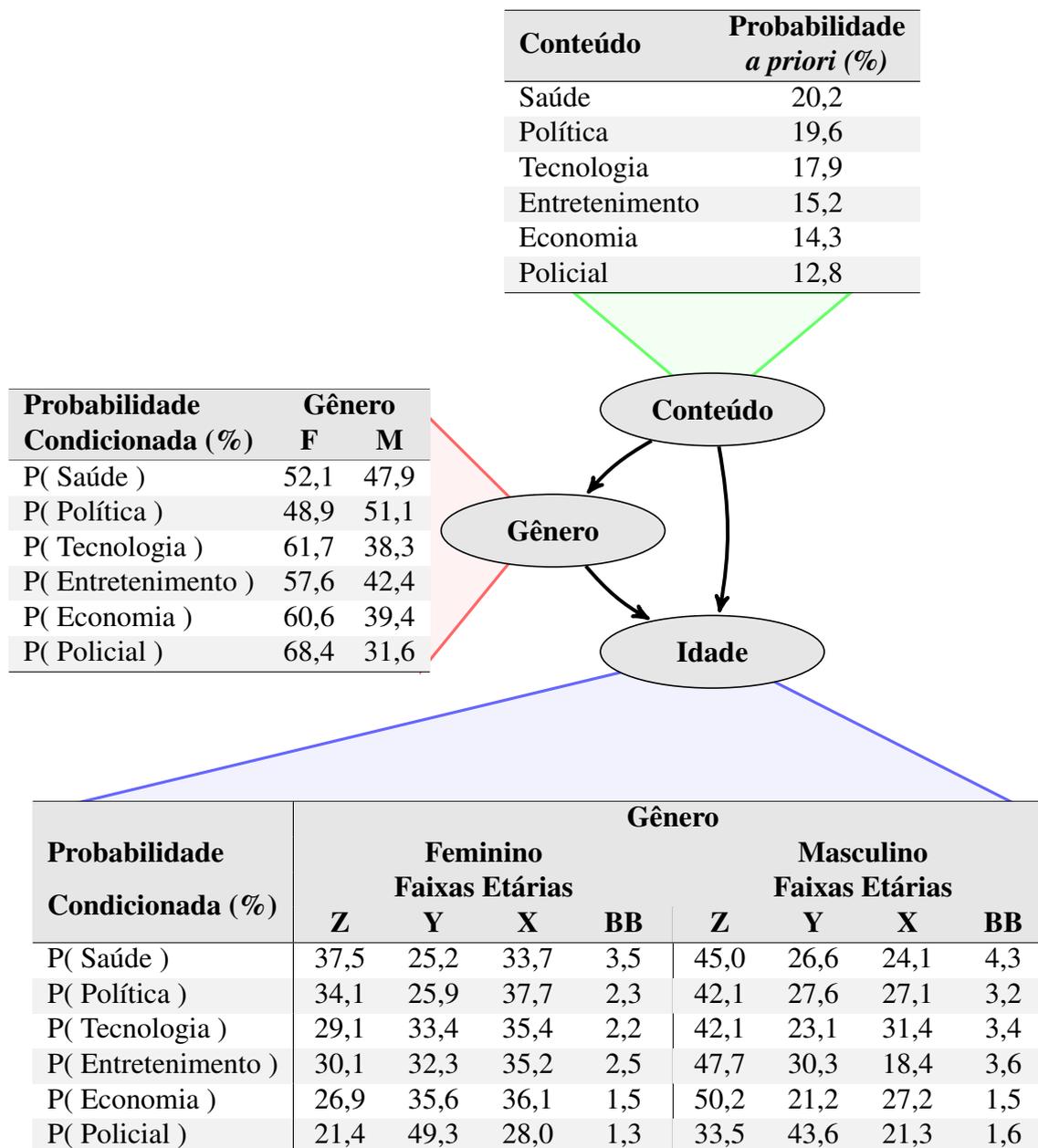


Figura 4.4: Rede bayesiana e tabelas de probabilidade geradas com base no processamento de todos os vídeos capturados dos espectadores.

Fonte: Próprio autor.

probabilidade *a priori* alta (ou baixa) demais e afetar as probabilidades condicionadas. Ou seja, se haviam muitos indivíduos de um determinado gênero (ou faixa etária), as probabilidades condicionadas para esse perfil será maior (ou menor), dependendo da probabilidade *a priori*. Assim, para evitar deduções errôneas¹ foi realizada uma análise intraclasse, ou seja, as probabilidades condicionadas foram comparadas para o mesmo perfil do espectador ao variar o conteúdo exibido.

Dessa maneira, percebemos que os espectadores variam a sua atenção a depender do conteúdo exibido. Por exemplo, considerando apenas o gênero do espectador, percebe-se um grande contraste na atenção dada pelos espectadores aos vídeos policiais. Esses vídeos foram os preferidos das mulheres, acompanhados dos vídeos sobre tecnologia e economia. Por outro lado, os vídeos políticos foram os de menor interesse. Como as probabilidades condicionadas são complementares, isto é, a soma dos homens com as mulheres corresponde a 100%, a lógica inversa é aplicável aos homens. Ou seja, o tema preferido do público masculino foi política, enquanto que, as categorias de menor interesse foram: economia, tecnologia e policial, respectivamente.

Com base nas tabelas de probabilidades apresentadas na Figura 4.4, a subseção seguinte é reservada para apresentar uma análise sobre as tabelas de probabilidade, ao se considerar apenas as faixas etárias dos espectadores nas análises.

4.3 Estudo Sobre as Distribuições de Idade dos Espectadores

Nesta seção, é apresentada a relação entre os conteúdos dos vídeos exibidos e as faixas etárias dos espectadores. As Figuras 4.5 e 4.6 foram geradas para tornar mais clara a visualização dessas relações. A Figura 4.5, mostra a representatividade dos espectadores que cada categoria de vídeo possui. Por exemplo, metade dos espectadores de gênero feminino, que assistiram aos vídeos policiais, são da geração Y e metade dos espectadores do gênero masculino que assistiram aos vídeos sobre economia são da geração Z, sendo esta geração

¹Por exemplo, concluir que os indivíduos da geração BB prestam pouca atenção porque as suas probabilidades condicionadas são baixas, quando na verdade, haviam poucos indivíduos com esse perfil e por isso as suas probabilidades condicionadas foram mais baixas que as demais gerações.

a predominante do público masculino. Este gráfico permite estimar demograficamente o público presente por faixa etária. Uma análise mais detalhada é apresentada a seguir.

Na Figura 4.6, a distribuição de atenção é agrupada por faixa etária. Assim, é possível analisar as preferências e aversões que cada geração possui com maior facilidade. Por exemplo, as mulheres da geração Y tiveram maior interesse pelos vídeos policiais e menor interesse pelos vídeos de saúde e política. Nota-se também a baixa expressividade dos espectadores da geração BB e o desbalanceamento no grupo masculino com tendência de queda no avanço das gerações, isto é, quanto mais elevada a faixa etária, menor a expressividade. Além disso, com exceção do conteúdo policial, os demais vídeos foram mais representados pelos indivíduos da geração Z no grupo masculino.

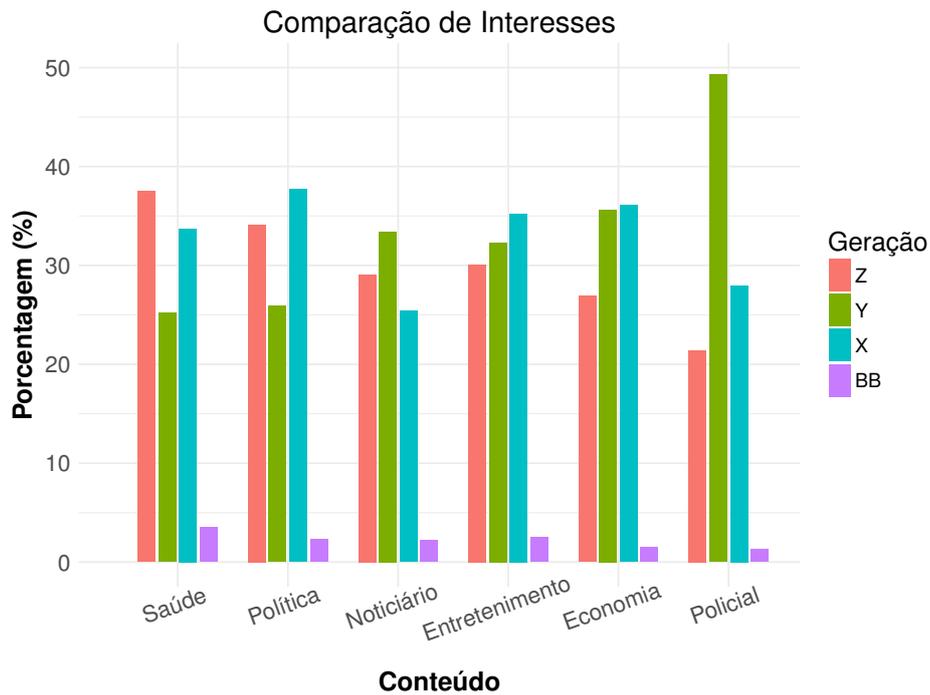
A Tabela 4.1 lista a ordem de preferência de cada geração por categoria de vídeo. A referida tabela serviu de base para argumentação sobre as análises da atenção dos espectadores, que são discutidas nos tópicos a seguir:

Tabela 4.1: Ordem das preferências dos espectadores com base em cada geração e o seu respectivo gênero.

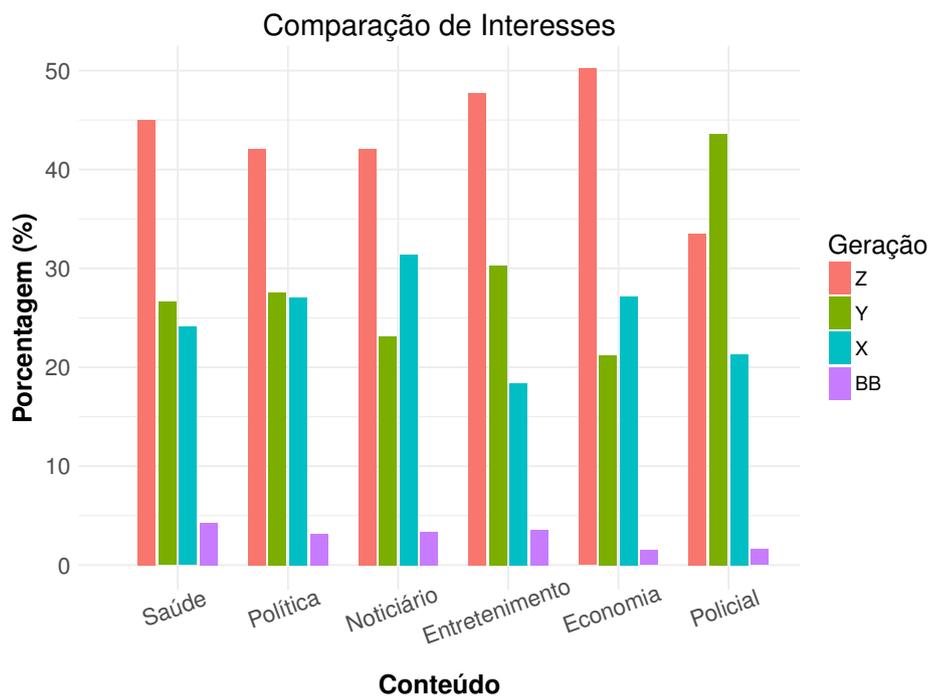
Conteúdo do vídeo	Gênero							
	Feminino Faixas Etárias				Masculino Faixas Etárias			
	Z	Y	X	BB	Z	Y	X	BB
Saúde	1º	6º	5º	1º	3º	4º	4º	1º
Política	2º	5º	1º	3º	4º	3º	3º	4º
Tecnologia	4º	3º	3º	4º	4º	5º	1º	3º
Entretenimento	3º	4º	4º	2º	2º	2º	2º	2º
Economia	5º	2º	2º	5º	1º	6º	2º	6º
Policial	6º	1º	6º	6º	6º	1º	5º	5º

Fonte: Próprio autor.

- **Saúde.** Foi o conteúdo preferido para os espectadores da geração BB (independente do gênero) e para as mulheres da geração Z. Enquanto que as mulheres das gerações intermediárias (X e Y) despertaram pouco interesse.
- **Política.** Existe um contraste entre os gêneros na geração Z, isto é, esse conteúdo foi um dos que mais chamou atenção das mulheres dessa geração, enquanto que os homens desse mesmo grupo etário prestaram pouca atenção. O tema político também



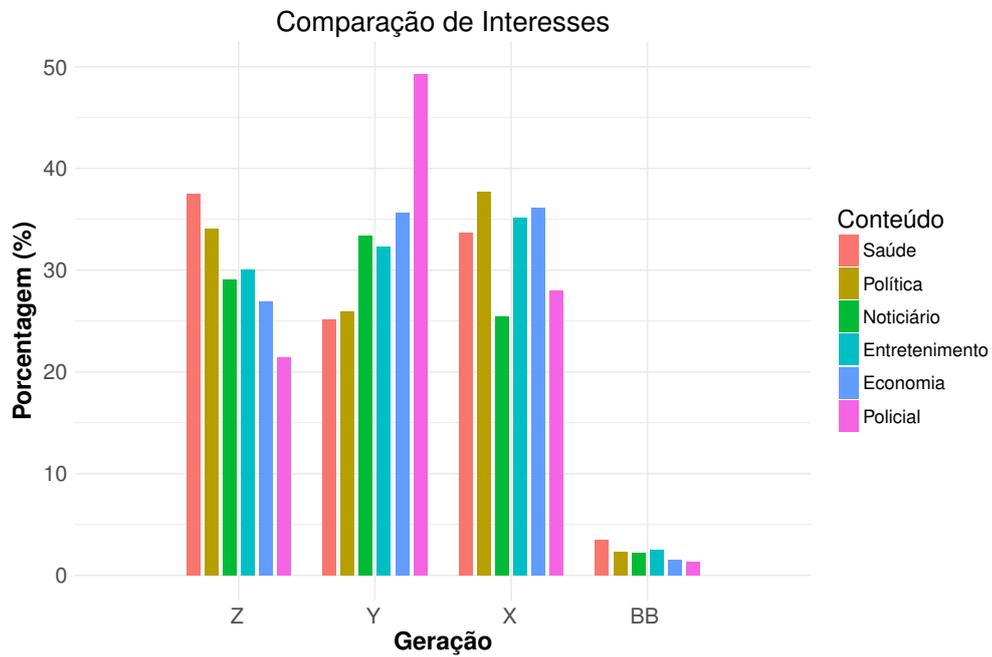
(a) Espectadores do gênero feminino.



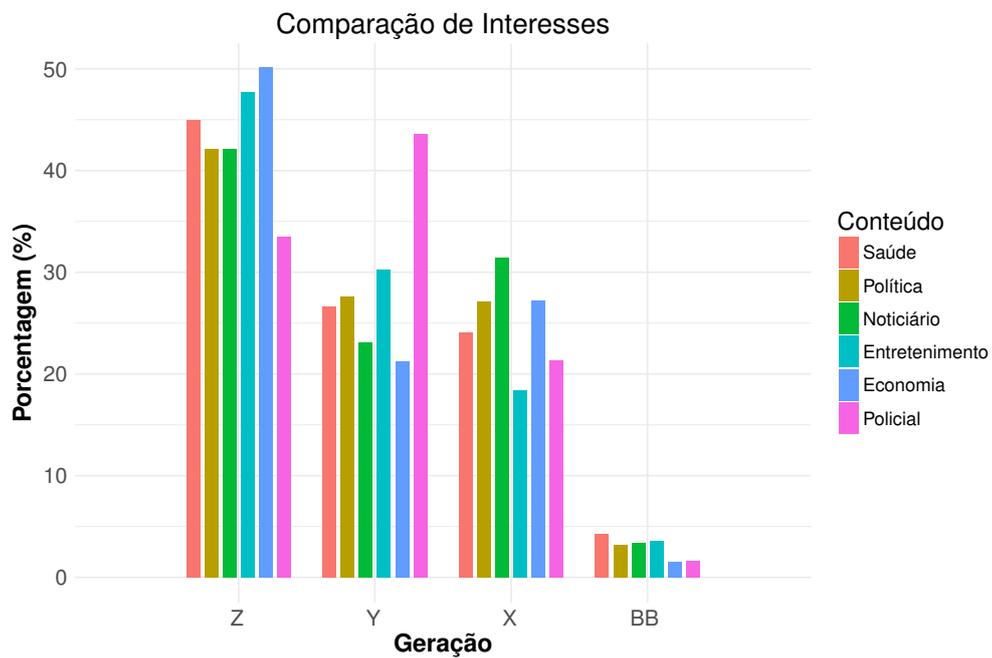
(b) Espectadores do gênero masculino.

Figura 4.5: Relação entre o conteúdo exibido com a faixa etária do espectador agrupada por conteúdo.

Fonte: Próprio autor.



(a) Espectadores do gênero feminino.



(b) Espectadores do gênero masculino.

Figura 4.6: Relação entre o conteúdo exibido com a faixa etária do espectador agrupada por geração.

Fonte: Próprio autor.

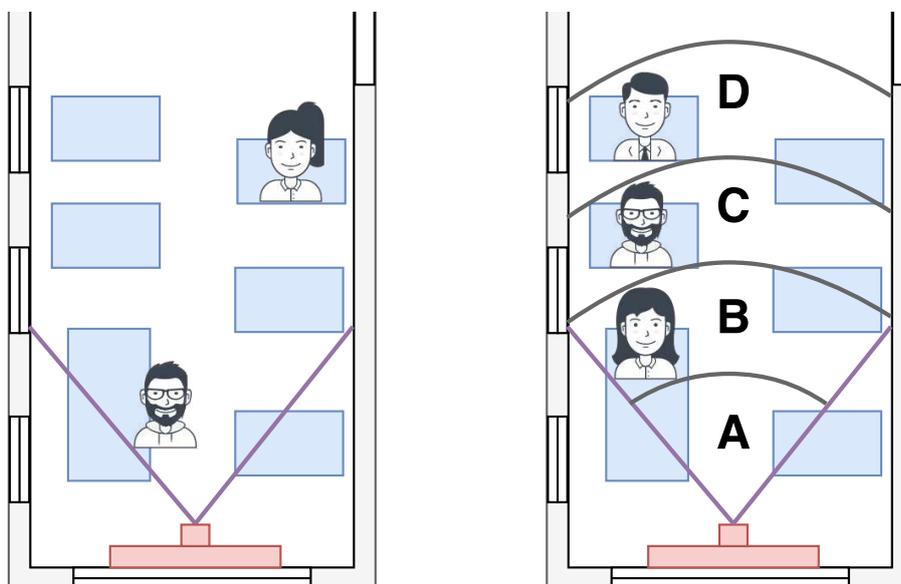
foi um dos de maior interesse das mulheres da geração X, já as mulheres da geração Y contrastam com as demais gerações por terem prestado pouca atenção.

- **Tecnologia.** Esse tema foi indiferente para as mulheres. Entretanto, os homens das gerações mais jovens (Z e Y) prestaram pouca atenção, enquanto que, esse tema foi um dos preferidos dos homens da geração X.
- **Entretenimento.** O público prevalente desse tema foram os homens, com exceção daqueles que são da geração X, pois prestaram pouca atenção ao tema. Além disso, esse tema foi um dos que chamaram mais atenção para as mulheres da geração BB.
- **Economia.** Esse tema chamou atenção dos espectadores de maneira diversificada. Nas gerações mais jovens (Z e Y), os homens prestaram atenção em proporção inversa às mulheres. Isto é, esse tema foi uma das preferências dos homens da geração Z e o de menor interesse da geração Y, por outro lado, as mulheres da geração Z despertaram pouco interesse, enquanto que as mulheres da geração Y tiveram grande interesse. Com as idades mais avançadas (X e BB), nota-se que os homens e as mulheres passam a convergir com as suas preferências. Ou seja, os espectadores da geração X de ambos os gêneros tiveram pouco interesse, enquanto os espectadores da geração BB prestaram mais atenção.
- **Policial.** A atenção para ambos os gêneros teve alta correlação, isto é, independente do gênero do espectador, a variação de sua atenção foi a mesma para as diferentes gerações de idade. O tema policial foi um dos temas de menor interesse dos espectadores (como dito acima, é invariante ao gênero), valendo observar um comportamento divergente na geração Y, que foi um dos temas de maior interesse dessa faixa etária.

4.4 Relação entre o Conteúdo exibido com a Idade, o Gênero e a Distância do Espectador ao Monitor

De forma a avaliar a relação entre o gênero e a idade do espectador com a proximidade ao visor, foi construída uma nova rede bayesiana com mais uma variável inclusa, a distância em relação à tela. A heurística utilizada para estimar a distância da face em relação ao monitor,

tem base na largura da face do espectador. Isto é, quanto mais larga (em *pixels*), mais próximo do visor o espectador está. A representação na Figura 4.7a mostra como a proximidade da face ao monitor interfere na largura que ela ocupa no campo de visão da câmera. Ou seja, a face do homem na imagem está próxima ao monitor e ocupa aproximadamente a metade do campo de visão da câmera (representado na figura como a linha lilás que parte da câmera), logo, essa face está próxima do monitor. Por outro lado, a face da mulher, possui o mesmo tamanho físico que a face do homem, está mais distante da câmera (e conseqüentemente, do monitor), por isto ocupa um espaço bem menor no campo de visão da câmera. Logo, a largura em *pixels* pode ser utilizada para indicar a proximidade da face em relação ao monitor, para estimar a distância do espectador em relação à tela e relacionar ao conteúdo do vídeo exibido.



(a) Ocupação do campo de visão pelas faces. (b) Classificação das distâncias das faces.

Figura 4.7: Heurística utilizada para avaliar a distância do espectador em relação ao monitor.

Fonte: Próprio autor.

Dessa maneira, as larguras das faces foram agrupadas em quatro conjuntos: com (A) 201-270 *pixels* de largura; (B) 151-200; (C) 101-150; e (D) 50-100. Sendo as faces do grupo "A", as mais próximas e "D", as mais distantes (quanto mais distante, menor a largura). Uma ilustração desse agrupamento é apresentada na Figura 4.7b, em que as faces com menor largura em relação ao campo de visão da câmera estão mais distantes (grupo "D"), enquanto

que as faces com maior largura estão mais próximas (grupo "A"). Assim, as menores faces têm 50 *pixels* de largura, enquanto que as maiores têm 270 *pixels*. Isto é, os tamanhos mínimos e máximos de todas as faces que foram detectadas.

A Figura 4.8, representa a estrutura da rede bayesiana automaticamente construída utilizando o algoritmo K2 (COOPER; HERSKOVITS, 1992) e implementada na ferramenta Weka (HALL et al., 2009), considerando o acréscimo da variável distância. Nela, pode-se observar que o gênero e a faixa etária do espectador dependem de sua distância em relação ao monitor. Foram geradas tabelas de probabilidades para todos os conteúdos de vídeos exibidos: Tabela 4.2 (Saúde), Tabela 4.3 (Política), Tabela 4.4 (Tecnologia), Tabela 4.5 (Entretenimento), Tabela 4.6 (Economia) e Tabela 4.7 (Policial). As maiores probabilidades estão destacadas em negrito.

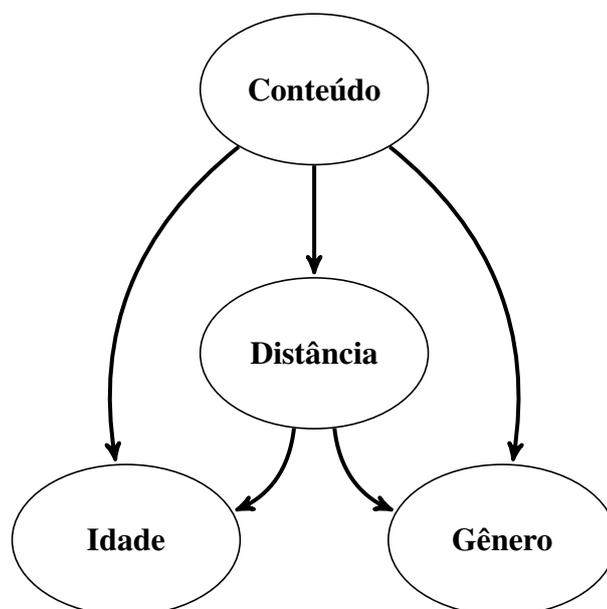


Figura 4.8: Estrutura da rede bayesiana.

Fonte: Próprio autor.

As análises indicaram que, por unanimidade, as mulheres ficaram mais próximas da TV do que os homens. Além disso, foi observado que:

- **Geração Z:** posicionaram-se mais afastados do monitor, com exceção de quando ocorreu a exibição dos vídeos sobre saúde e política. Os quais, foram os vídeos preferidos das mulheres.

- **Geração Y:** posicionaram-se mais próximos, com exceção na exibição dos vídeos políticos e policiais. Nesse caso, sentar-se afastado do monitor não significa aversão ao conteúdo exibido, pois, embora os espectadores dessa geração tenham se sentado mais afastados do monitor durante a exibição dos vídeos políticos e policiais, esse tema foi um dos preferidos dos homens dessa faixa etária. Entretanto, nota-se uma certa aversão das mulheres dessa faixa etária na exibição dos vídeos políticos.
- **Geração X:** se posicionaram em frente ao monitor de forma balanceada, isto é, sem preferência predominante.
- **Geração BB:** sentaram-se sempre afastados do monitor.

Tabela 4.2: Relação entre o conteúdo saúde e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.

Probabilidade Condicionada (%)	Faixas Etárias				Gênero	
	Z	Y	X	BB	Feminino	Masculino
P(Saúde $d = A$)	51,9	40,6	6,6	0,9	47,1	52,9
P(Saúde $d = B$)	37,3	38,6	22,9	1,2	54,5	45,5
P(Saúde $d = C$)	38,3	27,6	28,6	5,6	46,1	53,9
P(Saúde $d = D$)	42,1	25,1	29,4	3,4	54,2	45,8

Fonte: Próprio autor.

Tabela 4.3: Relação entre o conteúdo política e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.

Probabilidade Condicionada (%)	Faixas Etárias				Gênero	
	Z	Y	X	BB	Feminino	Masculino
P(Política $d = A$)	45,4	17,7	34,6	2,3	68,0	32,0
P(Política $d = B$)	31,4	25,2	40,5	2,8	51,0	49,0
P(Política $d = C$)	36,6	29,6	30,2	3,6	48,8	51,2
P(Política $d = D$)	38,8	26,0	32,7	2,5	48,9	51,1

Fonte: Próprio autor.

Tabela 4.4: Relação entre o conteúdo tecnologia e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.

Probabilidade Condicionada (%)	Faixas Etárias				Gênero	
	Z	Y	X	BB	Feminino	Masculino
P(Tecnologia $d = A$)	8,49	61,0	30,2	0,3	91,1	8,9
P(Tecnologia $d = B$)	15,9	53,3	29,5	1,3	83,1	16,9
P(Tecnologia $d = C$)	30,4	33,9	32,3	3,4	57,8	42,2
P(Tecnologia $d = D$)	37,3	25,4	34,6	2,7	59,3	40,7

Fonte: Próprio autor.

Tabela 4.5: Relação entre o conteúdo entretenimento e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.

Probabilidade Condicionada (%)	Faixas Etárias				Gênero	
	Z	Y	X	BB	Feminino	Masculino
P(Entretenimento $d = A$)	9,2	60,4	30,3	0,1	83,9	16,1
P(Entretenimento $d = B$)	12,7	54,8	30,9	1,6	83,3	16,7
P(Entretenimento $d = C$)	42,0	29,2	25,1	3,7	52,9	47,1
P(Entretenimento $d = D$)	38,5	30,1	28,5	2,9	56,7	43,3

Fonte: Próprio autor.

Tabela 4.6: Relação entre o conteúdo economia e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.

Probabilidade Condicionada (%)	Faixas Etárias				Gênero	
	Z	Y	X	BB	Feminino	Masculino
P(Economia $d = A$)	8,6	61,5	29,8	0,1	93,0	7,0
P(Economia $d = B$)	10,8	58,5	29,5	1,1	86,6	13,4
P(Economia $d = C$)	39,4	25,8	32,9	1,9	57,1	42,9
P(Economia $d = D$)	40,7	24,6	33,1	1,5	55,6	44,4

Fonte: Próprio autor.

Tabela 4.7: Relação entre o conteúdo policial e a proximidade que os espectadores estavam em relação à tela, considerando o gênero e a idade.

Probabilidade Condicionada (%)	Faixas Etárias				Gênero	
	Z	Y	X	BB	Feminino	Masculino
P(Policial $d = A$)	7,0	62,9	30,1	0,1	92,5	7,5
P(Policial $d = B$)	9,2	60,2	29,6	1,1	88,1	11,9
P(Policial $d = C$)	29,1	51,8	16,7	2,4	58,2	41,8
P(Policial $d = D$)	29,9	41,8	27,0	1,4	63,9	36,1

Fonte: Próprio autor.

4.5 Comparação com Outras Pesquisas

Esta subseção visa a apresentar a relação dos resultados obtidos nesta dissertação com os resultados de outras pesquisas, de forma a comparar e validar as análises apresentadas.

4.5.1 Pesquisa de Mercado Realizada pela Nielsen Company

A pesquisa global realizada pela Nielsen sobre Estilos de Vida, aconteceu entre 23 de fevereiro e 13 de março de 2015, contando com a participação de mais de 30 mil consumidores

entrevistados em 60 países em todo o globo. É considerada representativa para consumidores da Internet, por ter sido feita online e sua amostra possuir quotas fundamentadas na idade e sexo para cada país, com base em seus usuários de Internet (NIELSEN, 2015).

Foi perguntado aos entrevistados quais são as suas maiores aspirações pessoais, podendo optar por: Ficar em forma e saudável, Ganhar dinheiro e outros. A Figura 4.9, apresenta o resultado dessa pesquisa. Ao se considerar a idade dos entrevistados, observa-se que existe uma tendência crescente quanto a ficar em forma e saudável, isto é, quanto mais elevada é a faixa etária, maior essa aspiração. O inverso ocorre quanto à necessidade de ganhar dinheiro, que caiu com o avanço da faixa etária.

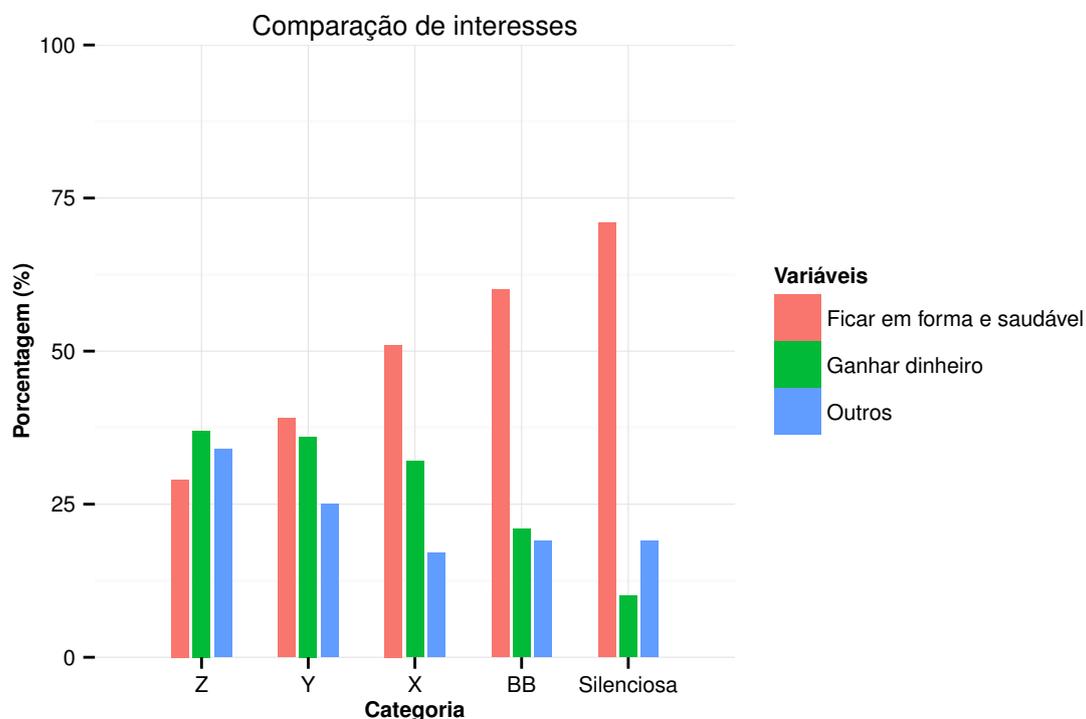


Figura 4.9: Comparação de interesses para diferentes categorias de idade.

Fonte: Inspirado em Nielsen (2015).

Os resultados da pesquisa apresentados na Figura 4.9 estão diretamente relacionados com os resultados obtidos nesta dissertação (apresentados na Figura 4.6). Tratando isoladamente estes dois conjuntos de vídeos, saúde e economia, observamos uma tendência crescente no interesse dos espectadores ao se considerar os vídeos sobre saúde. Isto é notado ao se ordenar os vídeos preferidos por cada uma das gerações, conforme apresentado na Tabela 4.1. Assim,

homens e mulheres, prestaram mais atenção aos vídeos sobre saúde à medida que as faixas etárias aumentam. Além disso, percebeu-se que os espectadores de gênero feminino tiveram uma tendência de interesse decrescente com o avanço da idade. Por fim, percebeu-se que os espectadores mais jovens, da geração Z, caracterizaram-se como *outliers* nessas conclusões, visto que nem sempre obedeceram às tendências das demais faixas etárias. Talvez por que o experimento realizado nesta dissertação represente a realidade específica da região que ocorreu o experimento.

4.5.2 Pesquisa de Mercado Realizada pela Ibope Inteligência

Outro estudo que reforça os resultados obtidos na pesquisa de mercado realizada pela Nielsen, é a pesquisa realizada pela Ibope Inteligência desenvolvida em parceria com Multivitamínico Centrum. Contando com 613 entrevistas com homens e mulheres com mais de 50 anos, de todas as classes econômicas, em todo país, com amostra proporcional à população brasileira, por região demográfica. A margem de erro foi de 4 pontos percentuais para mais ou para menos (INTELIGÊNCIA, 2015).

O levantamento mostrou que 77% dos brasileiros com mais de 50 anos consideram a saúde a principal preocupação na vida, correspondendo ao principal tema de interesse observado nesta dissertação para os indivíduos da geração BB. Seguido por estabilidade financeira, que corresponde a apenas 11% dos entrevistados. Ou seja, as evidências dessa pesquisa corroboram com as observações feitas nesta dissertação.

4.6 Considerações Finais

Nesse capítulo foram apresentadas e discutidas as avaliações de desempenho dos classificadores de gênero e de idade. Observou-se que o classificador de gênero obteve uma taxa de classificação alta (maior que 90% de *F1-score*), enquanto que o classificador de idade obteve taxa de 72% e a sua validação na base de dados MORPH, apresentou desempenho superior às técnicas estado-da-arte. Além disso, foram analisadas as relações existentes entre a faixa etária, o gênero e a distância do espectador em relação ao monitor para permitir análises demográficas dos espectadores. Foi observado que é possível categorizá-los utilizando redes bayesianas para relacionar o perfil demográfico ao conteúdo exibido. Ou seja, com a abor-

dagem proposta é possível selecionar automaticamente o conteúdo de maior interesse dos espectadores sabendo-se o seu perfil, ou, estimar qual público que mais se interessará pelo vídeo à ser exibido com base no seu conteúdo.

Capítulo 5

Considerações Finais

Este capítulo apresenta as considerações finais desta pesquisa. Como pôde ser observado ao longo desta pesquisa, foi detalhado um amplo experimento envolvendo a análise dos espectadores na sinalização digital. A pesquisa apresentou uma abordagem *bottom-up*, desde o estudo sobre efeitos culturais explorados pelos sociólogos na definição das faixas etárias, que foram aplicadas para categorizar os espectadores, passando por uma detalhada validação dos classificadores de gênero e de idade, até a realização de um experimento prático de sinalização digital, para o qual foi necessário criar uma base de vídeos para exibição, rotulá-los manualmente, e gerar estatísticas para então, analisar os espectadores demograficamente. Dessa maneira, observa-se que os objetivos da pesquisa foram alcançados, os quais são devidamente elencados adiante.

Este capítulo está organizado como se segue. Na Seção 5.1 são apresentadas as conclusões desta pesquisa. Enquanto a Seção 5.2, apresentada uma série de proposições para investigações em pesquisas futuras.

5.1 Conclusões

Ao fim deste estudo, observa-se que os objetivos de pesquisa foram alcançados. O Quadro 5.1 sintetiza as principais contribuições desta dissertação, assim como a sua respectiva localização no texto.

Nesta seção, cada objetivo específico é destacado em negrito para facilitar a associação dos objetivos alcançados aos resultados obtidos.

Quadro 5.1: Levantamento das principais contribuições desta dissertação.

Contribuição	Localização nesta Dissertação
Estudo para definições das faixas etárias dos espectadores.	Seção 3.3.1.
Uma metodologia para categorização de espectadores que frequentam determinado ambiente.	Seção 3.2.
Proposição de uma metodologia para aumento de dados aplicado a classificadores de idade.	Seção 4.1.
Organização de uma base de vídeos pública e com rotulagem manual do tipo de conteúdo exibido.	Capítulos B e C.
Publicações.	Capítulo E.

Fonte: Próprio autor.

Para a realização desse experimento, **foram exploradas diferentes formas de realizar a análise dos espectadores**. Entre as diversas opções existentes, observam-se que as principais são análise via dados fornecidos pelos aparelhos celulares, questionários e a utilização de sistemas inspirados em visão computacional. Entretanto, observa-se que as abordagens inspiradas em visão computacional apresentam métodos propícios para análises mais escaláveis, menos intrusivas e de baixo custo de operação. Porém, essas abordagens não fornecem resultados tão precisos para classificação de gênero e de idade quando comparadas com os questionários (por exemplo), assim como não fornecem resultados personalizados para indivíduos isoladamente e não considerando o aspecto demográfico no qual estão inseridas.

Assim, optou-se pela realização de experimentos considerando a abordagem inspirada em visão computacional. Ao explorar os materiais plausíveis de serem utilizados nesta pesquisa, observou-se que os classificadores de idade apresentavam desempenhos pouco satisfatórios, motivando assim, a necessidade de se realizar experimentos detalhados para aprimorar os resultados encontrados na literatura e integrar aos demais módulos do sistema. De tal forma, **todos os materiais utilizados puderam ser integrados entre si para a realização dos experimentos**.

Em seguida, foi realizado um estudo sobre **como definir as faixas etárias para agrupar os espectadores**. Foi observado que alguns pesquisadores utilizaram as faixas etárias de acordo com o que a base de dados fornece para treinar seus classificadores. Assim, motivou-se a realização de um estudo sobre os aspectos socioculturais que influenciam nas definições

das faixas etárias. Uma vez realizada essa análise, o classificador pôde ser devidamente treinado e conseqüentemente, **a estrutura de sinalização digital foi montada para a realização de uma experimentação prática**. Entretanto, para a realização do experimento, foi necessário criar uma **base de dados de vídeos de exibição**. O que levou à criação de uma base de mais de cem vídeos, publicamente disponíveis na internet, que foram manualmente rotulados para serem utilizados na **categorização dos espectadores de forma automática**.

As análises sobre as preferências dos espectadores geraram diversas evidências sobre o seu comportamento demográfico. Por exemplo, foi observado que ao se considerar apenas o gênero e a faixa etária do espectador associado ao conteúdo exibido, o gênero é uma variável influenciada apenas pelo conteúdo exibido. Enquanto que a faixa etária depende tanto do gênero quanto do conteúdo exibido. Além disso, ao se considerar a distância que o espectador está localizado em relação ao monitor, observa-se que o gênero e a idade dependem da distância e a idade não mais depende do gênero. Essas relações permitem gerar diferentes tipos de inferências sobre os espectadores. Por exemplo, que no geral, as mulheres se posicionaram mais próximas do monitor do que os homens. Enquanto que os espectadores da categoria BB se posicionaram sempre mais afastados do monitor. Outras evidências interessantes de serem destacadas, foram que, as mulheres da geração Z ficaram mais afastadas do monitor nas exibições dos anúncios sobre saúde e política, os quais foram os temas de menor interesse dos espectadores dessa categoria. Por fim, as **publicações** oriundas dessa pesquisa são apresentadas no Apêndice E.

5.2 Pesquisas Futuras

O sistema proposto nesta dissertação se caracteriza de Análise Automática e Anônima dos Espectadores por não realizar reconhecimento facial ou armazenar gravações que revelem a identidade dos indivíduos envolvidos nas capturas, trazendo mais conforto para os usuários desse sistema devido a sua privacidade, conforme defendido por Kinebuchi et al. (2009). O diagrama apresentado na Figura 5.1 ilustra o sistema proposto nesta dissertação como uma sequência de blocos que representam módulos de processamento, cada um deles podendo ser modificados de acordo com o contexto, como exemplo, o detector de faces e o classificador de gênero e idade poderão ser atualizados para atender uma determinada necessidade do

usuário do sistema, que precise utilizar um outro código de menor custo computacional ou que necessite ser atualizado para *softwares* mais robustos. Além disso, o item "Outros dados" foi incluso, pois, em trabalhos futuros outras variáveis poderão ser acrescentadas para a segmentação automática dos espectadores, exemplos:

- **Horário e data:** Essa variável pode ser coletada da máquina que realiza as capturas dos vídeos e ser incluída para segmentá-los de acordo com o horário do dia, o dia da semana ou a data. Por exemplo, as seguintes hipóteses poderiam ser testadas: (i) As mulheres assistem mais aos anúncios no período da manhã do que os homens; (ii) As mulheres assistem mais aos anúncios no começo da semana do que os homens; e (iii) Nos primeiros dias do mês, as mulheres assistem mais aos anúncios do que os homens. Semelhante ao que foi abordado na pesquisa de Maeda et al. (2010), que associou a quantidade de trafegantes pelo local ao horário do dia para estimar os horários do dia de maior movimentação. Dessa maneira, sabendo-se que um determinado grupo demográfico prefere o tipo de conteúdo "A" sobre o tipo de conteúdo "B", então, sabendo-se também que em um determinado horário do dia há mais indivíduos desse grupo demográfico, é de interesse do consumidor, que o tipo de conteúdo "A" seja mais exibido naquele horário do dia. Essa investigação pode ser de interesse de companhias que invistam na satisfação do cliente, como é o caso das companhias de trens da Holanda que investigaram a satisfação de seus usuários na pesquisa de Moritz (2016).
- **Associação com a emoção:** Essa variável pode ser obtida por um classificador que utilize a face detectada para estimar a emoção esboçada pelo espectador, e assim associar a emoção despertada ao tipo de conteúdo exibido, de modo a verificar quais anúncios provocam emoções positivas/negativas. Nesse sentido, a pesquisa de Kramer, Guillory e Hancock (2014) pode ser utilizada como referência, pois foi analisado o impacto que um anúncio positivo/negativo têm sobre as postagens dos usuários nas redes sociais. Dessa maneira, pode ser de interesse da empresa de exibição que, preferencialmente, sejam exibidos mais anúncios positivos do que negativos, por exemplo, uma campanha de combate ao tabagismo pode avaliar a expressão de felicidade nas faces dos espectadores, logo após ser exibida uma mensagem para estimular o abandono do vício. Além disso, outras associações podem ser realizadas, como em Feng et al. (2011), que obser-

vou que há diferenças na atenção dos homens e das mulheres, assim, alguns anúncios podem despertar emoções diferentes nos espectadores, permitindo segmentá-los em um novo nível ao considerar a emoção, como feito na pesquisa de Santos, Gomes e Campos (2011).

- **Informação do vídeo exibido:** Os vídeos exibidos também podem ser processados e utilizados para produzirem-se inferências com respeito à atenção dos espectadores. Podem-se extrair informações como: (i) presença/ausência de face no vídeo exibido e relacionar com a atenção dos espectadores (vídeos com face chamam mais atenção do que vídeos sem face?); (ii) medição do brilho médio ou da distribuição de cores (vídeos com cores mais vibrantes chamam mais atenção do que vídeos com cores opacas?); (iii) estimação da quantidade de movimento (vídeos mais dinâmicos chamam mais atenção do que vídeos estáticos?).
- **Agrupamento das faces:** Semelhante à técnica de reconhecimento facial aplicada na pesquisa de (RAVNIK; SOLINA, 2014), em pesquisas futuras pode-se investigar a utilização de técnicas de agrupamento de faces, permitindo analisar a preferência dos espectadores individualmente. Isto é, por meio desse agrupamento é possível associar as preferências dos espectadores ao grupo demográfico que está inserido, observando se há correlações ou não com o grupo que está inserido. Além disso, poderá ser possível gerar análises considerando o tempo de visualização dedicado por cada espectador.

5.2.1 Análise em Tempo Real

A abordagem proposta de análise de atenção dos espectadores inspirada na construção de redes bayesianas permite que o usuário do sistema treine e gere a sua própria rede adaptada ao seu contexto, assim, pode-se utilizar dados da rede para gerar estatísticas sobre o perfil de consumidores e que podem ser úteis para a realização de contratos entre as empresas transmissoras e anunciantes. De tal forma, essa ferramenta poderia auxiliar a pesquisa de Roggeveen, Nordfalt e Grewal (2016), que investigaram os impactos nas vendas dos supermercados após a inserção da sinalização digital. As diferenças demográficas dos espectadores poderiam ser acrescentadas em cada estabelecimento comercial para análises mais detalhadas.

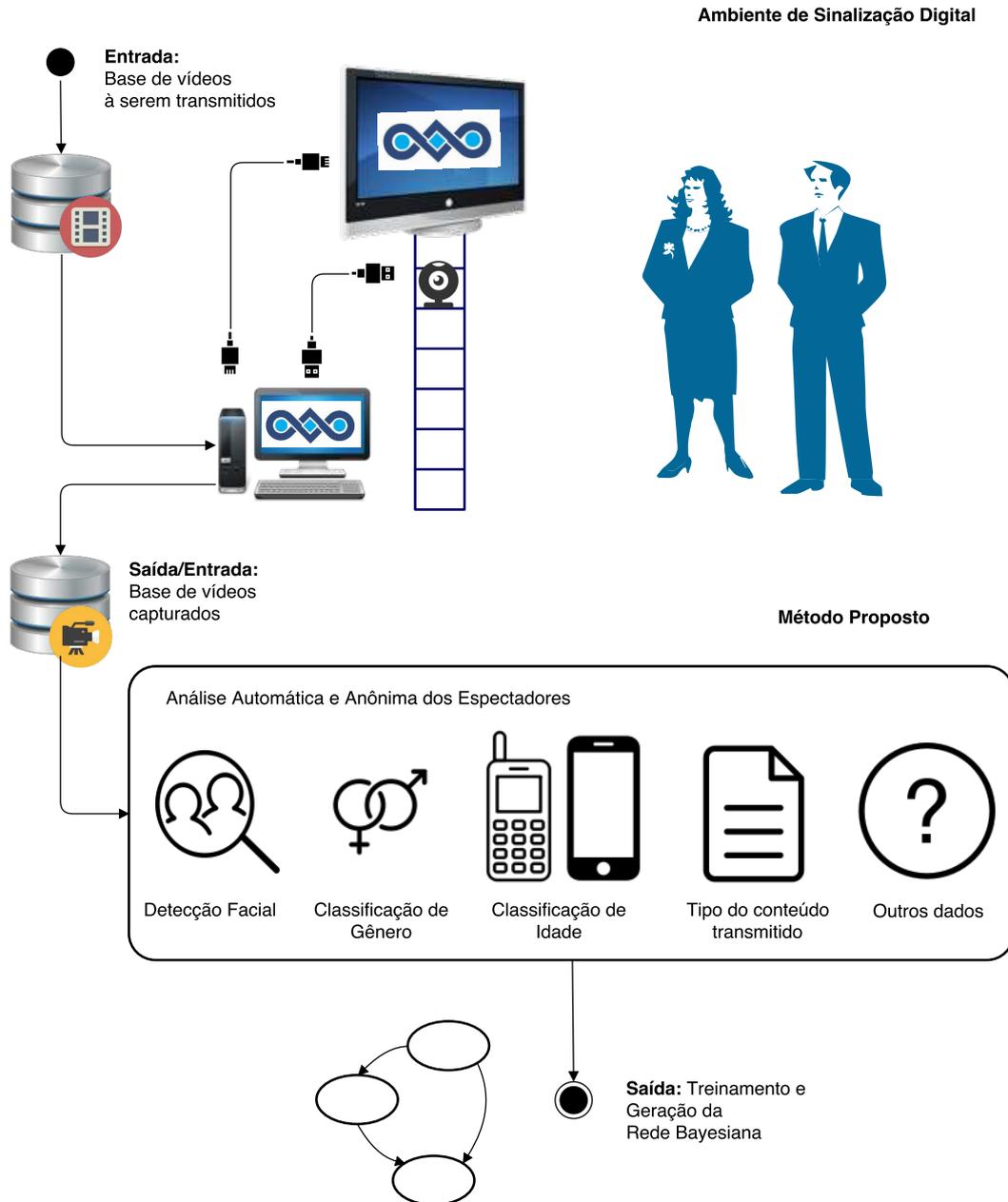


Figura 5.1: Diagrama de alto nível, representando o processamento em tempo de execução.
Fonte: Próprio autor.

Na pesquisa de Taneja e Wu (2013), foi observado que a diferença geográfica é um dos principais fatores na estimação demográfica dos espectadores de anúncios *on-line*, mas quando os experimentos ocorrem na mesma localização (e.g., lugares diferentes da mesma cidade), a rede bayesiana pode apresentar resultados satisfatórios dada as investigações realizadas por Bublitz et al. (2014), que conseguiu categorizar os espectadores demograficamente utilizando informações de sua localização (lugares frequentados). Desse modo, o dono de um

estabelecimento comercial que deseje inaugurar uma filial de sua companhia noutra lugar, poderia importar os dados treinados de uma outra loja sua que já possua uma rede treinada.

Por outro lado, a ferramenta desenvolvida poderia ser utilizada para adaptar a exibição dos vídeos automaticamente. Nesse caso, ao capturar as filmagens dos espectadores, o sistema poderia verificar as matrizes de probabilidade daquele grupo demográfico e selecionar o conteúdo de sua maior preferência. Por exemplo, ao empregar o sistema de análise em uma fila do cinema, a rede poderia aprender que jovens de sexo masculino preferem filmes de ficção científica a comédias românticas, sendo assim, ao notar a presença de muitos homens jovens, o sistema poderia exibir automaticamente o trailer do próximo filme de ficção científica a ser lançado. Essa análise em tempo real, pode ser de muito interesse para o usuário do sistema (e.g. uma empresa) e para o consumidor, podendo evitar que ocorra o paradoxo destacado por Bublitz et al. (2014), que ressaltou que se os melhores anúncios são entregues aos consumidores, os objetivos dos anunciantes podem ser, eventualmente, negligenciados. Por outro lado, se considerarmos apenas os interesses dos anunciantes, a relevância dos anúncios para os consumidores pode diminuir consideravelmente.

5.2.2 Pré-análise do Aspecto Tempo

No processamento dos dados, o instante em que a exibição ocorreu também pode ser considerado nas análises. Como muitos resultados já foram explorados, permitindo gerar conclusões úteis, foi considerado investigar o aspecto tempo mais profundamente em trabalhos futuros. Como exemplo de análise, foi construída uma rede bayesiana utilizando o algoritmo K2 (COOPER; HERSKOVITS, 1992), implementada na ferramenta Weka (HALL et al., 2009), incluindo a variável tempo, relacionando as partes dos vídeos que foram exibidos com os picos de atenção de determinados grupos demográficos.

Assim, a rede bayesiana gerada é apresentada na Figura 5.2. O vídeo processado foi dividido em intervalos de tempo de 30 segundos sendo as análises realizadas isoladamente por trecho, o que gerou complexas matrizes de confusão. Com isso, pode-se perceber que a faixa etária do espectador passou a depender apenas do instante de tempo e do conteúdo exibido. Já o gênero depende da distância em relação à tela, o que depende do instante de tempo. Isto é, os espectadores de diferentes faixas etárias prestam atenção de maneira diferenciada ao longo da exibição do vídeo. Além disso, homens e mulheres também são

influenciados pela distância da tela.

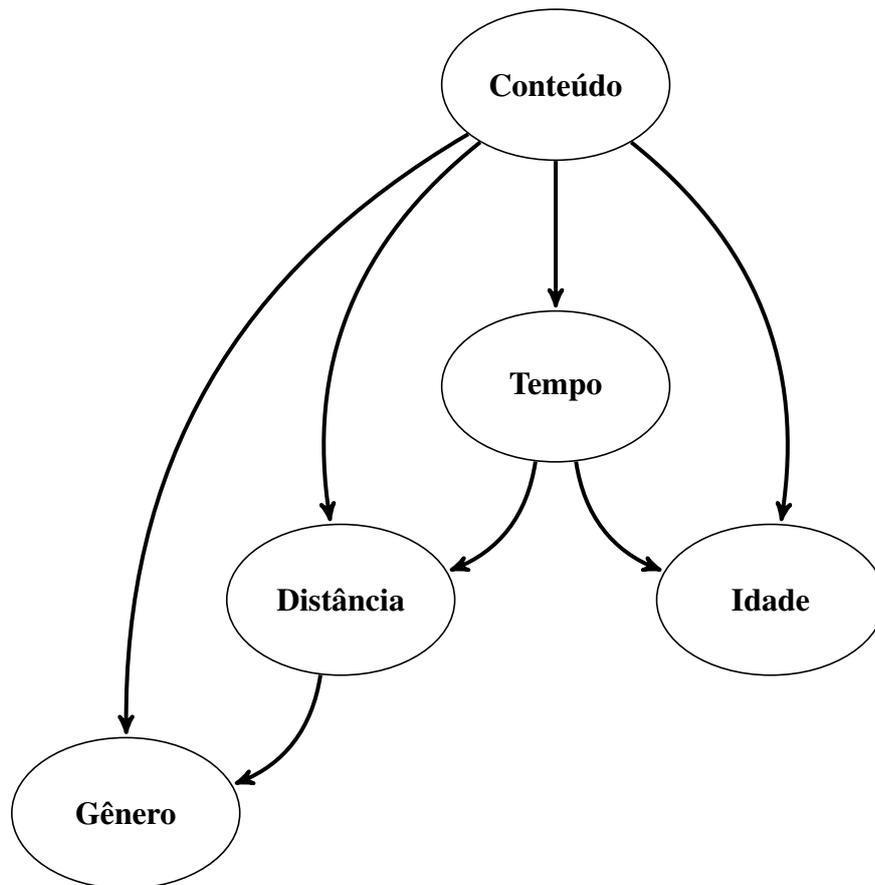


Figura 5.2: Estrutura da rede bayesiana considerando as variáveis tempo e distância.

Fonte: Próprio autor.

A Tabela 5.1 mostra a tabela de probabilidade para os vídeos de noticiários, as três maiores probabilidades foram destacadas em negrito. Assim, percebe-se uma tendência de que os espectadores de gerações mais avançadas, X e BB, prestem mais atenção nos minutos iniciais dos vídeos, enquanto que, os mais jovens, geração Z, tiveram pico de atenção nos segundos finais dos vídeos. Tais fatos evidenciam que existe certa relação entre o instante de tempo da exibição de um vídeo e a faixa etária.

Tabela 5.1: Tabela de probabilidades condicionadas incluindo a variável tempo nas análises.

Probabilidade Condicionada (%)	Faixas Etárias			
	Z	Y	X	BB
P(Noticiário $t = 1-30s$)	35,1	26,3	35,7	2,9
P(Noticiário $t = 31-60s$)	33,9	26,7	35,3	4,1
P(Noticiário $t = 61-90s$)	31,7	25,9	39,5	2,9
P(Noticiário $t = 91-120s$)	30,1	36,7	31,5	1,8
P(Noticiário $t = 121-150s$)	31,5	36,6	30,2	1,6
P(Noticiário $t = 151-180s$)	31,7	29,4	37,4	1,5
P(Noticiário $t = 181-210s$)	38,8	30,1	28,7	2,4
P(Noticiário $t = 211-240s$)	43,4	27,2	26,7	2,7
P(Noticiário $t = 241-270s$)	45,1	33	20	1,9
P(Noticiário $t = 271-300s$)	52,4	33,2	13,3	1,2
P(Noticiário $t = 301-330s$)	32,2	52,5	15	0,2
P(Noticiário $t = 331-360s$)	31,2	44,2	23,3	1,3

Fonte: Próprio autor.

Referências Bibliográficas

ANS, A. N. d. S. **Resolução Normativa, N° 63**. 2003. Acessado em 20 de junho de 2016. Disponível em: <<http://www.ans.gov.br/component/legislacao/?view=legislacao&task=TextoLei&format=raw&id=NzQ4>>.

BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Speeded-Up Robust Features (SURF). **Computer Vision and Image Understanding**, v. 110, p. 346–359, 2008. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1077314207001555>>.

BERGSTRA, J.; BREULEUX, O.; BASTIEN, F.; LAMBLIN, P.; PASCANU, R.; DESJARDINS, G.; TURIAN, J.; WARDE-FARLEY, D.; BENGIO, Y. Theano: A CPU and GPU math compiler in Python. In: **9th Python in Science Conference**. [s.n.], 2010. abs/1605.02688. Acessado em 20 de Junho de 2016. Disponível em: <<http://arxiv.org/abs/1605.02688>>.

BEYER, G.; ALT, F.; MÜLLER, J.; SCHMIDT, A.; ISAKOVIC, K.; KLOSE, S.; SCHIEWE, M.; HAULSEN, I. Audience behavior around large interactive cylindrical screens. In: **Conference on Human Factors in Computing Systems**. ACM, 2011. p. 1021–1030. Acessado em 20 de Junho de 2016. Disponível em: <<http://doi.acm.org/10.1145/1978942.1979095>>.

BROKAW, T. **The Greatest Generation**. [S.l.]: Random House, 2004. 464 p.

BUBLITZ, F. M.; ROSNER, M.; ALMEIDA, H. O.; LIMA, A. M. de; BATISTA, T. S. Using statistical data for context sensitive pervasive advertising. In: **International Conference on Consumer Electronics**. [S.l.: s.n.], 2014.

BUBLITZ, F. M.; SILVA, L. C.; OLIVEIRA, E. A. S.; LUIZ, S. O. D.; ALMEIDA, H. O. de; PERKUSICH, A. A petri net model specification for delivering adaptable ads through digital signage in pervasive environments. In: **International Conference on Software Engineering and Knowledge Engineering**. [S.l.: s.n.], 2013.

BUBLITZ, F. M.; SILVA, L. S. E.; ALMEIDA, H. O.; LUIZ, S. O. D.; PERKUSICH, A. Using location to infer user contextual information: A case study on pervasive advertising. In: **International Conference on Consumer Electronics**. [S.l.: s.n.], 2014.

BURKE, R. R. Behavioral effects of digital signage. **Journal of Advertising Research**, Journal of Advertising Research, v. 49, n. 2, p. 180–185, 2009.

- CARDOSO, F. F. **Crise Financeira Global: Impactos na Economia Brasileira, Política, econômica e resultados**. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2013.
- CARLSON, E. **The Lucky Few: Between the Greatest Generation and the Baby Boom**. 1. ed. [S.l.]: Springer, 2008. 216 p.
- CASTELLUCCIO, M.; POGGI, G.; SANSONE, C.; VERDOLIVA, L. Land use classification in remote sensing images by convolutional neural networks. **Computing Research Repository**, abs/1508.00092, 2015. Acessado em 20 de Junho de 2016. Disponível em: <<http://arxiv.org/abs/1508.00092>>.
- COES. **Informe Epidemiológico N° 02/2015 - Semana Epidemiológica 47 (22 A 28/11/2015)**. [S.l.], 2016. Acessado em 20 de junho de 2016. Disponível em: <http://combateaes.saude.gov.br/images/sala-de-situacao/informe_microcefalia_epidemiologico25.pdf>.
- COLLOBERT, R.; KAVUKCUOGLU, K.; FARABET, C. Torch7: A matlab-like environment for machine learning. In: **BigLearn, NIPS Workshop**. [S.l.: s.n.], 2011.
- COOPER, G. F.; HERSKOVITS, E. A bayesian method for constructing bayesian belief networks from databases. In: **Conference on Uncertainty in Artificial Intelligence**. Morgan Kaufmann Publishers, 1991. p. 86–94. Acessado em 20 de Junho de 2016. Disponível em: <<http://dl.acm.org/citation.cfm?id=2100662.2100674>>.
- COOPER, G. F.; HERSKOVITS, E. A bayesian method for the induction of probabilistic networks from data. **Machine Learning**, v. 9, p. 309–347, 1992. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1007/BF00994110>>.
- CROCOMO, M. K. **Algoritmo de otimização bayesiano com detecção de comunidades**. Tese (Doutorado) — Universidade de São Paulo, 2012.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: **Conference on Computer Vision and Pattern Recognition**. IEEE Computer Society, 2005. p. 886–893. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1109/CVPR.2005.177>>.
- DALY, R.; SHEN, Q.; AITKEN, S. Review: Learning bayesian networks: Approaches and issues. **Knowledge Engineering Review**, Cambridge University Press, v. 26, p. 99–157, 2011. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1017/S0269888910000251>>.
- DONAHUE, J.; JIA, Y.; VINYALS, O.; HOFFMAN, J.; ZHANG, N.; TZENG, E.; DARRELL, T. Decaf: A deep convolutional activation feature for generic visual recognition. **Computing Research Repository**, abs/1310.1531, 2013. Acessado em 20 de Junho de 2016. Disponível em: <<http://arxiv.org/abs/1310.1531>>.
- EIDINGER, E.; ENBAR, R.; HASSNER, T. Age and gender estimation of unfiltered faces. **IEEE Transactions on Information Forensics and Security**, v. 9, p. 2170–2179, 2014.

FAWCETT, T. An introduction to roc analysis. **Pattern Recognition Letters**, Elsevier Science, p. 861–874, 2006. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2005.10.010>>.

FENG, Q.; ZHENG, Y.; ZHANG, X.; SONG, Y.; LUO, Y. jia; LI, Y.; TALHELM, T. Gender differences in visual reflexive attention shifting: Evidence from an {ERP} study. **Brain Research**, p. 59–65, 2011. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0006899311009735>>.

FENTON, N.; NEIL, M. **Risk Assessment and Decision Analysis with Bayesian Networks**. [S.l.]: CRC Press, 2012.

FERNÁNDEZ, C.; HUERTA, I.; PRATI, A. A comparative evaluation of regression learning algorithms for facial age estimation. In: _____. Springer, 2015. p. 133–144. Acessado em 20 de Junho de 2016. Disponível em: <http://dx.doi.org/10.1007/978-3-319-13737-7_12>.

FISHER, R. A. The use of multiple measurments in taxonomic problems. **Annals of Eugenics**, Blackwell Publishing Ltd, v. 7, p. 179–188, 1936. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x>>.

FITCH, N. **Sylvia Beach and the Lost Generation: A History of Literary Paris in the Twenties and Thirties**. Norton, 1985. 447 p. Acessado em 20 de Junho de 2016. Disponível em: <https://books.google.com.br/books?id=Zpu5_c2o-2IC>.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. **Computer and System Sciences**, v. 55, p. 119–139, 1995.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). **Annals of Statistics**, v. 28, p. 337–407, 2000. Acessado em 20 de Junho de 2016. Disponível em: <<http://projecteuclid.org/euclid.aos/1016218223>>.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. Bayesian network classifiers. **Machine Learning**, v. 29, p. 131–163, 1997. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1023/A:1007465528199>>.

GROBELNY, J.; MICHALSKI, R. The role of background color, interletter spacing, and font size on preferences in the digital presentation of a product. **Computers in Human Behavior**, v. 43, p. 85–100, 2015. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0747563214005597>>.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: An update. **ACM SIGKDD Explorations Newsletter**, ACM, v. 11, p. 10–18, 2009.

HELANDER, F. **Meaduring Digital Signage ROI: A combination of Digital Signage and Mobile Advertising as a method for measuring Digital Signage ROI**. [S.l.], 2010.

HOPE, J. Get your campus ready for generation z. **Dean and Provost**, v. 19, p. 1–7, 2016. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1002/say.30253>>.

HOWE, N.; STRAUSS, W. **Millennials Rising: The Next Great Generation**. [S.l.]: Knopf Doubleday Publishing Group, 2009. 432 p.

HUERTA, I.; FERNÁNDEZ, C.; SEGURA, C.; HERNANDO, J.; PRATI, A. A deep analysis on age estimation. **Pattern Recognition Letters**, v. 68, parte 2, p. 239–249, 2015. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865515001683>>.

HYUN, W.; HUH, M. Y.; KIM, S. H.; KANG, S. G. Study on design and implementation of audience measurement functionalities for digital signage service using kinect camera. In: **International Conference on Advanced Communication Technology**. [S.l.: s.n.], 2014.

INTELIGÊNCIA, I. **Sinais da Nutrição**. 2015. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.revistapublicitta.com.br/atitude/pesquisas/sinaisdanutricaoodepoisdos50>>.

JAIN, V.; LEARNED-MILLER, E. **Fddb: A Benchmark for Face Detection in Unconstrained Settings**. [S.l.], 2010.

JAIN, V.; LEARNED-MILLER, E. Online domain adaptation of a pre-trained cascade of classifiers. In: **Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2011. p. 577–584.

JIA, Y.; SHELHAMER, E.; DONAHUE, J.; KARAYEV, S.; LONG, J.; GIRSHICK, R. B.; GUADARRAMA, S.; DARRELL, T. Caffe: Convolutional architecture for fast feature embedding. **Computing Research Repository**, ACM, p. 675–678, 2014. Acessado em 20 de Junho de 2016. Disponível em: <<http://doi.acm.org/10.1145/2647868.2654889>>.

JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. In: **Conference on Uncertainty in Artificial Intelligence**. Morgan Kaufmann Publishers, 1995. p. 338–345. Acessado em 20 de Junho de 2016. Disponível em: <<http://dl.acm.org/citation.cfm?id=2074158.2074196>>.

JUN, B.; KIM, D. Robust face detection using local gradient patterns and evidence accumulation. **Pattern Recognition**, v. 45, p. 3304–3316, 2012. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2012.02.031>>.

KINEBUCHI, T.; ARAI, H.; MIYAGAWA, I.; ANDO, S.; KATAOKA, K.; KOIKE, H. Image processing techniques for measuring advertising effectiveness of digital signage. In: . [S.l.: s.n.], 2009.

KRAMER, A. D. I.; GUILLORY, J. E.; HANCOCK, J. T. Experimental evidence of massive-scale emotional contagion through social networks. In: . [S.l.: s.n.], 2014. v. 24, p. 8788–8790.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in Neural Information Processing Systems**. Curran Associates, 2012. p. 1097–1105. Acessado em 20 de Junho de 2016. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>>.

LEE, J.; YOON, K. The application of digital signage system using smart device. In: **International Conference on Advanced Communication Technology**. [S.l.: s.n.], 2014.

LI, J.; WANG, T.; ZHANG, Y. Face detection using surf cascade. In: **International Conference on Computer Vision Workshops**. [S.l.: s.n.], 2011. p. 2183–2190.

LUCHT, T. **Sylvia Porter: America's Original Personal Finance Columnist**. Syracuse University Press, 2014. v. 31. 289-290 p. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1080/08821127.2014.905376>>.

MACEDO, R. G. A história da propaganda digital no brasil: sua estrutura e perspectivas de crescimento. In: **Publicidade e Propaganda: 200 anos de história no Brasil**. ACM, 2009. p. 533–540. Acessado em 20 de Junho de 2016. Disponível em: <<http://doi.acm.org/10.1145/1141911.1141920>>.

MAEDA, K.; NISHI, M.; YOSHIDA, T.; SUZUKI, K.; INOUE, H. Digital signage with audience detection using tv broadcasting waves. In: **International Symposium on Applications and the Internet**. [s.n.], 2010. p. 225–228. Acessado em 20 de Junho de 2016. Disponível em: <<http://doi.ieeecomputersociety.org/10.1109/SAINT.2010.99>>.

MARKUS, N.; FRLJAK, M.; PANDZIC, I. S.; AHLBERG, J.; FORCHHEIMER, R. A method for object detection based on pixel intensity comparisons. **Computing Research Repository**, abs/1305.4537, 2013. Acessado em 20 de Junho de 2016. Disponível em: <<http://arxiv.org/abs/1305.4537>>.

MELO, A. S. O.; MALINGER, G.; XIMENES, R.; SZEJNFELD, P. O.; SAMPAIO, S. A.; FILIPPIS, A. M. Bispo de. Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: tip of the iceberg? **Ultrasound in Obstetrics & Gynecology**, John Wiley & Sons, v. 47, 2016. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1002/uog.15831>>.

MIKAWA, M.; IZUMI, S.; TANAKA, K. Book recommendation signage system using silhouette-based gait classification. In: **International Conference on Machine Learning and Applications and Workshops**. IEEE Computer Society, 2011. p. 416–419. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1109/ICMLA.2011.43>>.

MILBORROW, S.; MORKEL, J.; NICOLLS, F. The MUCT Landmarked Face Database. **Pattern Recognition Association of South Africa**, 2010. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.milbo.org/muct>>.

MILLER, J. D. The generation x report: Active, balanced, and happy: These young americans are not bowling alone. **Longitudinal study of American youth**, v. 1, 2011.

Moritz, B. **Content in the train: Taking a closer look at in-train screens: A study on what types of content, on in-train digital signage, lead to a more positive customer experience**. 2016. Acessado em 20 de Junho de 2016. Disponível em: <<http://essay.utwente.nl/69391/>>.

MU, G.; GUO, G.; FU, Y.; HUANG, T. S. Human age estimation using bio-inspired features. In: **Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2009. p. 112–119. ISSN 1063-6919.

NATIONS, U. **World Economic Situation and Prospects**. 2013. Acessado em 20 de Junho de 2016. Disponível em: <http://www.un.org/en/development/desa/policy/wesp/wesp_current/2016wesp_ch1_en.pdf>.

NELSON, J. L.; WEBSTER, J. G. Audience currencies in the age of big data. **International Journal on Media Management**, v. 18, p. 9–24, 2016.

NIELSEN. **Geração Global Estilos de Vida**. [S.l.], 2015. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.nielsen.com/br/pt/insights/reports/2015/Geracao-global-estilos-de-vida.html>>.

OLIVEIRA, I. de P.; MEDEIROS, J. ao L. P.; SOUSA, V. F. de; JÚNIOR, A. G. T.; PEREIRA, E. T.; GOMES, H. M. A data augmentation methodology to improve age estimation using convolutional neural networks. In: **SIBGRAPI Conference on Graphics, Patterns and Images**. IEEE Computer Society, 2016. Acessado em 20 de Junho de 2016. Disponível em: <<http://sibgrapi.sid.inpe.br/col/sid.inpe.br/sibgrapi/2016/07.23.00.14/doc/PID4374341.pdf>>.

PALESTINI, C. Digital heritage and earthquake emergency. In: **Digital Heritage International Congress**. [S.l.: s.n.], 2013. v. 1, p. 769–769.

PHILLIPS, P. J.; MOON, H.; RAUSS, P.; RIZVI, S. A. The feret evaluation methodology for face-recognition algorithms. **Transactions on Pattern Analysis and Machine Intelligence**, p. 137–143, 1997.

POPAI, P. D. S. **Aplicações do Digital Signage**. 2014. Acessado em 3 de Agosto de 2014. Disponível em: <<http://www.portaldigitalsignage.com.br>>.

RAVNIK, R.; SOLINA, F. Audience measurement of digital signage: Quantitative study in real-world environment using computer vision. **Interacting with Computers**, p. 597–600, 2013.

RAVNIK, R.; SOLINA, F. Interactive and audience adaptive digital signage using real-time computer vision. **International Journal of Advanced Robotic Systems**, v. 13, p. 29–30, 2013.

RAVNIK, R.; SOLINA, F. Interactive and audience-adaptive information interfaces. **Electronic Letters on Computer Vision and Image Analysis**, 2014. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.5565/rev/elcvia.616>>.

RICANEK, K.; TESAFAYE, T. Morph: a longitudinal image database of normal adult age-progression. In: **International Conference on Automatic Face and Gesture Recognition**. IEEE Computer Society, 2006. p. 341–345. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1109/FGR.2006.78>>.

ROGGEVEEN, A. L.; NORDFALT, J.; GREWAL, D. Do Digital Displays Enhance Sales? Role of Retail Format and Message Content. **Journal of Retailing**, v. 92, p. 122–131, 2016. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022435915000755>>.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M. S.; BERG, A. C.; LI, F. ImageNet Large Scale Visual Recognition Challenge. **Computing Research Repository**, v. 115, p. 211–252, 2014.

SANTOS, B. P. d.; GOMES, C. G.; CAMPOS, G. M. P. **Sensus: Reconhecedor Automático de Expressões Faciais para Publicidade e Propaganda**. [S.l.], 2011.

SCHAEFER, S.; MCPHAIL, T.; WARREN, J. Image deformation using moving least squares. **ACM Transactions On Graphics**, ACM, v. 25, p. 533–540, 2006. Acessado em 20 de Junho de 2016. Disponível em: <<http://doi.acm.org/10.1145/1141911.1141920>>.

SERMANET, P.; EIGEN, D.; ZHANG, X.; MATHIEU, M.; FERGUS, R.; LECUN, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. **Computing Research Repository**, abs/1312.6229, 2013. Acessado em 20 de Junho de 2016. Disponível em: <<http://arxiv.org/abs/1312.6229>>.

SOKOLOVA, M.; LAPALME, G. **A systematic analysis of performance measures for classification tasks**. [s.n.], 2009. v. 45. 427-437 p. Acessado em 20 de Junho de 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0306457309000259>>.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S. E.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. **Computing Research Repository**, abs/1409.4842, 2014. Acessado em 20 de Junho de 2016. Disponível em: <<http://arxiv.org/abs/1409.4842>>.

TANEJA, H.; WU, A. X. Does the great firewall really isolate the chinese? integrating access blockage with cultural factors to explain web user behavior. **Computing Research Repository**, 2013.

TOPI, M.; TIMO, O.; MATTI, P.; MARICOR, S. Robust texture classification by subsets of local binary patterns. In: **International Conference on Pattern Recognition**. [S.l.: s.n.], 2000. v. 3, p. 935–938.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. **Cognitive Neuroscience**, v. 1, p. 71–86, 1991. Acessado em 20 de Junho de 2016. Disponível em: <<http://dx.doi.org/10.1162/jocn.1991.3.1.71>>.

VALENTIM, F. L. **Estudo e Implementação de Algoritmos de Inferência e Aprendizado em Redes Bayesianas**. 2007. Acessado em 7 de junho de 2016. Disponível em: <http://repositorio.ufla.br/bitstream/1/5398/1/MONOGRAFIA_Estudo_e_implemen-ta%C3%A7%C3%A3o_de_algoritmos_de_infer%C3%Aancia_e_aprendizado_em-redes_bayesianas.pdf>.

VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: **Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2001. v. 1, p. 511–518.

WANG, R.; XU, Z. A pedestrian and vehicle rapid identification model based on convolutional neural network. In: **7th International Conference on Internet Multimedia**

Computing and Service. ACM, 2015. Acessado em 20 de Junho de 2016. Disponível em: <<http://doi.acm.org/10.1145/2808492.2808524>>.

WENG, R.; LU, J.; YANG, G.; TAN, Y.-P. Multi-feature ordinal ranking for facial age estimation. In: **Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on.** [S.l.: s.n.], 2013. p. 1–6.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques.** 3. ed. [S.l.]: Morgan Kaufmann Publishers, 2005.

YIN, K.-C.; WANG, H.-C.; YANG, D.-L.; WU, J. A study on the effectiveness of digital signage advertisement. In: **International Symposium on Computer, Consumer and Control.** [S.l.: s.n.], 2012. p. 169–172.

ZHANG, E.; ZHANG, Y. F-measure. In: _____. **Encyclopedia of Database Systems.** Boston, MA: Springer US, 2009. p. 1147–1147. ISBN 978-0-387-39940-9.

Apêndice A

Fundamentação Teórica

A análise estatística empregada na análise da atenção dos espectadores, utilizou redes bayesianas para relacionar a dependência entre as variáveis coletadas no experimento, que são elas: o tipo de conteúdo exibido, o gênero, a idade e a distância do espectador em relação ao monitor. Foi utilizado o Teorema de Bayes para gerar as redes bayesianas e a definição da arquitetura da rede foi realizada utilizando o algoritmo de aprendizagem K2 (COOPER; HERSKOVITS, 1992), os quais, são discutidos neste capítulo. O desenvolvimento das redes bayesianas é apresentado na Seção A.1. Na Seção A.2, são apresentados os conceitos sobre o Teorema de Bayes. Enquanto, a descrição do método utilizado na construção da rede é abordado na Seção A.3.

A.1 Teorema de Bayes

Esta subseção se destina a apresentar as terminologias e notações matemáticas relacionadas ao Teorema de Bayes, que foi amplamente utilizado na análise dos dados desta dissertação. Assim, entende-se que o espaço amostral Ω é definido como um conjunto de resultados possíveis de um experimento aleatório, isto é, $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, em que, ω são os elementos deste conjunto com tamanho n . Um evento A em Ω é um subconjunto de Ω , isto é, $A \in \Omega$. Esses resultados possíveis são eventos unitários, em que a distribuição de probabilidade P é uma função que calcula a probabilidade de um evento ocorrer em meio ao espaço amostral, fornecendo um número real entre 0 e 1, $P : P(\Omega) \rightarrow [0, 1]$ (DALY; SHEN; AITKEN, 2011). Para fins de ilustração, o espaço amostral para o tipo de conteúdo do vídeo

exibido neste estudo, pode ser definido como:

$$\Omega = \{saúde, política, tecnologia, entretenimento, economia, policial\}.$$

A probabilidade de que um quadro capturado seja proveniente de um vídeo do tipo *saúde* é de 20,2%, logo, $P(saúde) = 20,2\%$ ¹. Considerando que a probabilidade de dois eventos ocorrerem, A e B por exemplo, é especificada como $P(A \cap B)$, então a *probabilidade condicionada* do evento A ocorrer, dado que o evento B já ocorreu, é definida por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (A.1)$$

Obviamente, para que a Equação A.1 seja válida, $P(B)$ deve ser estritamente positivo. Alternativamente, a probabilidade condicionada é denotada como:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \quad (A.2)$$

O que pode ser reescrito como:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (A.3)$$

Assim, utilizando os conceitos supracitados e com base nos vídeos capturados, pode-se calcular as probabilidades condicionadas para realizar a análise da atenção dos espectadores. Por sua vez, foi criada uma rede bayesiana, que representasse a estrutura dos dados coletados neste estudo de caso para realizar as inferências, o que será discutido na subseção seguinte.

A.2 Redes Bayesianas

As redes bayesianas são diagramas que representam a dependência entre as variáveis e organizam os dados em um grafo orientado entre *causa* e *efeito*, em que, a variável que se conhece as probabilidades *a priori* se chama *causa* e a variável *efeito* é aquela que se calcula a probabilidade condicionada. De acordo com Fenton e Neil (2012), a descrição de uma rede bayesiana consiste em:

¹A Tabela 3.1 apresenta as probabilidades *a priori*.

- Grafos Acíclicos Dirigidos (*Directed Acyclic Graph* - DAG): é o tipo de topologia (também chamado de estrutura) da rede bayesiana, consiste em um conjunto de nós e arcos, em que os nós correspondem às variáveis e os arcos representam os vínculos entre os nós. Um arco que liga A para B indica que existe uma dependência de A em B ; Então, o nó A é pai de B . Por ser um grafo acíclico, se existe um arco de A para B e um outro arco de B para C , então, não pode existir um arco de C para A .
- Tabela de probabilidade: representa a distribuição de probabilidades de um determinado conjunto de pais de A . Para um nó raiz, isto é, um nó sem pai, a distribuição de probabilidades deste nó é a própria probabilidade *a priori*.

Na Figura A.1, é ilustrada a estrutura de uma rede bayesiana, em que, os nós representam as variáveis e os arcos indicam as dependências. Observe que, o sentido da seta é sempre da variável *causa* para a variável *efeito*, ou da variável pai para a filha. Com isso, é possível estruturar os fatores envolvidos neste experimento, como o conteúdo do vídeo exibido, o gênero e a idade do espectador, para realizar uma análise objetiva da atenção dada e relacionar as dependências destas variáveis.

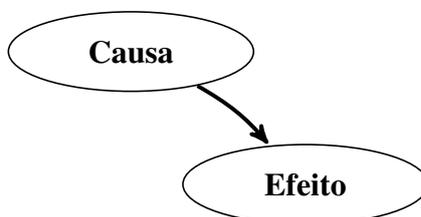


Figura A.1: Representação gráfica de uma rede bayesiana.

Fonte: Próprio autor.

A.3 Algoritmo de Aprendizagem K2

O algoritmo de aprendizagem K2 foi utilizado para gerar a estrutura da rede bayesiana. Este algoritmo gera diferentes possibilidades de topologia da rede bayesiana e calcula a probabilidade dela gerar o conjunto de dados de entrada (COOPER; HERSKOVITS, 1992). O algoritmo K2 é formalmente definido a partir de:

- CD : Conjunto de dados de entrada com m observações (ou casos).
- ER : Estrutura da rede.
- $P(ER|CD)$: Probabilidade da estrutura da rede condicionada do conjunto de dados. Busca-se a melhor estrutura da rede com base no maior valor obtido.

Na construção da rede bayesiana, o algoritmo K2 inicialmente admite que os nós não possuem pais, então as variáveis são adicionadas como pais da variável *a priori*. A cada passo o valor de $P(ER|CD)$ é calculado e verificado o aumento quando uma nova variável é adicionada como pai, buscando maximizar este valor. Quando não é possível aumentar mais o valor de $P(ER|CD)$, então a próxima variável é selecionada e o ciclo é repetido até que todas as variáveis estejam interconectadas na rede (CROCOMO, 2012). Sabendo-se que o problema possui n variáveis, assim, existem $2^{\frac{n(n-1)}{2}}$ estruturas possíveis (VALENTIM, 2007), que são determinadas com base no valor obtido pela equação A.4, definida em (COOPER; HERSKOVITS, 1991).

$$Q_{K2}(ER, CD) = P(S) \prod_{i=0}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(r_i - 1 + N_{ij})!} \prod_{k=1}^{r_i} N_{ijk}! \quad (\text{A.4})$$

em que:

- $P(S)$: Probabilidade *a priori* da variável.
- X_i : Estados possíveis das variáveis contidas no CD .
- N_{ijk} : Quantidade total de casos em CD .
- N_{ij} : Quantidade de ocorrências que possuem o conjunto de pais de X_i .
- n : Número de variáveis em X_i .
- q_i : Número de combinações possíveis para os pais de X_i .
- r_i : Quantidade total de valores possíveis que a variável $X_i = (i = 1, \dots, n)$ pode assumir.

Nesta dissertação, outras métricas foram testadas e comparadas, tais como: *Tree Augmented Naive Bayes* (TAN) (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997), *Naive*

Bayes (JOHN; LANGLEY, 1995), K2 (COOPER; HERSKOVITS, 1992), *Tabu Search* e *Simulated annealing* (COOPER; HERSKOVITS, 1991). Foi realizada uma validação cruzada nos dados, com 10 *folds* e foi observado que o algoritmo K2 obteve o melhor desempenho, com Área Sob a Curva *Receiver Operating Characteristic* (ROC) de 0,82. Mais detalhes sobre essas medidas estatísticas podem ser vistos em Fawcett (2006) e Sokolova e Lapalme (2009).

Apêndice B

Base de Vídeos Própria: Exibição

O Quadro B.1 apresenta a relação dos vídeos exibidos na avaliação experimental conduzida nesta pesquisa, as suas respectivas categorias e as URL (*Uniform Resource Locator*) para *download*. Dessa maneira, outros autores poderão replicar o experimento desta dissertação ou, fazer adaptações ao seu contexto. Por exemplo, adicionando nova(s) categoria(s) de vídeos à presente lista, removendo alguma categoria de vídeo indesejada ou atualizando a relação com vídeos mais atuais.

Quadro B.1: Relação completa dos vídeos rotulados manualmente e que foram exibidos aos espectadores.

URL	Título	Conteúdo
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-acidente-na-br-230-deixa-tres-mortos-proximo-a-soledade-na-paraiba/4629958/	Acidente na BR-230 deixa três mortos próximo a Soledade, na Paraíba	Polícia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/adolescente-e-apreendida-com-10kg-de-maconha-em-cabedelo/4635182/	Adolescente é apreendida com 10kg de maconha em Cabedelo	Polícia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-bandidos-roubam-motos-para-colocar-pecas-a-venda-em-joao-pessoa/4629961/	Bandidos roubam motos para colocar peças à venda, em João Pessoa	Polícia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-corpo-de-tecnica-em-enfermagem-assassinada-em-joao-pessoa-e-enterrado-hoje/4663764/	Corpo de técnica em enfermagem assassinada em João Pessoa é enterrado hoje	Polícia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-gerente-e-outros-funcionarios-sao-acusados-de-aplicar-golpe-em-churrascaria-em-joao-pessoa/4658753/	Gerente e outros funcionários são acusados de aplicar golpe em churrascaria em João Pessoa	Polícia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/mulher-e-atropelada-em-frente-a-faculdade-particular-na-br-230-em-cabedelo/4629957/	Mulher é atropelada em frente à faculdade particular na BR-230, em Cabedelo	Polícia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/policial-e-dois-suspeitos-sao-baleados-em-tiroteio-em-joao-pessoa/4637738/	Policial e dois suspeitos são baleados em tiroteio em João Pessoa	Polícia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/preso-do-regime-semi-aberto-e-assassinado-em-padaria-em-campina-grande/4629972/	Preso do regime semi-aberto é assassinado em padaria em Campina Grande	Polícia

(continuação)

http://g1.globo.com/videos/t/g1-em-1-minuto/v/g1-em-1-minuto-britanica-descobre-que-estava-sendo-traida-e-se-vinga-do-marido/4630188/	G1 em 1 Minuto: britânica descobre que estava sendo traída e se vinga do marido	Polícia
https://www.youtube.com/watch?v=e9CAZknD3Hc	Passageiros do Uber são agredidos por taxistas em SP e acabam no hospital	Polícia
https://www.youtube.com/watch?v=MzQh9Oru61M	Polícia Federal solta casal dono do Mega Filmes HD	Polícia
http://globoplay.globo.com/v/4645395/	Polícia procura Papai Noel que roubou helicóptero em São Paulo	Polícia
http://globoplay.globo.com/v/4645428/	Conheça quem eram os jovens metralhados por PMs no Rio	Polícia
http://globoplay.globo.com/v/4668741/	Rodovias de São Paulo têm dezenas de radares desligados	Polícia
http://g1.globo.com/videos/t/g1-em-1-minuto/v/g1-em-1-minuto-pf-prende-jose-carlos-bumlai/4630068/	G1 em 1 Minuto: PF prende José Carlos Bumlai	Política
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-senado-federal-discute-crise-hidrica-no-nordeste/4653477/	Senado Federal discute crise hídrica no nordeste	Política
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-feira-contracorrupcao-acontece-em-joao-pessoa/4663702/	Feira contra corrupção acontece em João Pessoa	Política
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-moradores-reclamam-da-frequencia-no-corte-no-abastecimento-de-agua-em-joao-pessoa/4666400/	Moradores reclamam da frequência no corte no abastecimento de água em João Pessoa	Política
http://g1.globo.com/videos/t/g1-em-1-minuto/v/g1-em-1-minuto-pf-prende-lider-do-governo-no-senado-e-banqueiro-do-btg/4632542/	G1 em 1 Minuto: PF prende líder do governo no Senado e banqueiro do BTG	Política

(continuação)

http://g1.globo.com/videos/t/g1-em-1-minuto/v/g1-em-1-minuto-votacao-sobre-cunha-no-conselho-de-etica-e-adiada/-4631164/	G1 em 1 Minuto: votação sobre Cunha no Conselho de Ética é adiada	Política
http://g1.globo.com/videos/t/todos-os-videos/v/camaroes-liberta-900-refens-do-boko-haram/4653444/	Camarões liberta 900 reféns do Boko Haram	Política
http://g1.globo.com/videos/t/todos-os-videos/v/publico-grita-nao-vai-ter-golpe-durante-cerimonia-da-conferencia-nacional-de-saude/4653653/	Público grita 'não vai ter golpe' durante cerimônia da Conferência Nacional de Saúde	Política
http://globoplay.globo.com/v/4634866/	STF nega qualquer influência de Delcídio para tentar livrar Nestor Cerveró	Política
http://globoplay.globo.com/v/4634875/	Senado decide manter Delcídio do Amaral preso	Política
http://globoplay.globo.com/v/4634935/	Gravação mostra discussão de um plano de fuga para Nestor Cerveró	Política
http://globoplay.globo.com/v/4642818/	STF decide que André Esteves e chefe de gabinete de Delcídio vão continuar presos	Política
http://globoplay.globo.com/v/4645538/	Conselho de Ética decide nesta terça (1º) o andamento do processo contra Eduardo Cunha	Política
http://globoplay.globo.com/v/4653147/	Governo quer apressar votação do pedido de processo de <i>impeachment</i>	Política
http://globoplay.globo.com/v/4663493/	Supremo Tribunal Federal suspende instalação da comissão especial do <i>impeachment</i>	Política
http://globoplay.globo.com/v/4663517/	Dilma e Temer têm primeiro encontro nesta quarta (9) após carta-desabafo do vice	Política
http://globoplay.globo.com/v/4663527/	Conselho de Ética adia mais uma vez votação do processo contra Eduardo Cunha	Política
http://globoplay.globo.com/v/4666125/	Dilma Rousseff e Michel Temer se reúnem e prometem relação fértil e profícua	Política

(continuação)

http://globoplay.globo.com/v/4666161/	Excesso de postes em cidade do Paraná leva Ministério Público a investigar prefeitura	Política
http://globoplay.globo.com/v/4668802/	STF decide regras para processo de <i>impeachment</i> na próxima quarta-feira (16)	Política
http://globoplay.globo.com/v/4668848/	Mulheres participam da primeira eleição na Arábia Saudita	Política
http://globoplay.globo.com/v/4668851/	Conselho de Ética adia decisão sobre processo contra Cunha após deputados trocarem tapas	Política
https://www.youtube.com/watch?v=A9DRyzc-ftk	Por que Alckmin quer fechar escolas em São Paulo?	Política
https://www.youtube.com/watch?v=h06j3gtALlw	Como funcionam as eleições nos EUA - Debates	Política
https://www.youtube.com/watch?v=iROOVkdqahk	6 estratégias do Estado Islâmico para espalhar o terrorismo pela <i>internet</i>	Política
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-circo-dos-7-anoes-chega-a-joao-pessoa/4669004/	Circo dos 7 Anões chega a João Pessoa	Entretenimento
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/-v/encontro-da-musica-nacional-de-raiz-acontece-em-campina-grande/4635164/	Encontro da Música Nacional de Raiz acontece em Campina Grande	Entretenimento
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/-v/encontro-da-musica-nacional-de-raiz-acontece-em-campina-grande/4635164/	Encontro da Música Nacional de Raiz acontece em Campina Grande	Entretenimento
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-programe-se-com-a-agenda-cultural-do-bdpb/4669015/	Programe-se com a agenda cultural do BDPB	Entretenimento
https://www.youtube.com/watch?v=0EansuPy3bo	10 heróis da vida real	Entretenimento
https://www.youtube.com/watch?v=7riLTnAoi5U	X-MEN <i>theme on piano</i> (x-men 90's cartoon opening music)	Entretenimento
https://www.youtube.com/watch?v=aDDeyhftpc	10 cães que não gostam nem um pouco de banho	Entretenimento

(continuação)

https://www.youtube.com/watch?v=AiC7lkqgvFo	Reginaldo Rossi no Piano - Música Brega - Garçon	Entretenimento
https://www.youtube.com/watch?v=aZJQ5ITWB0o	10 pessoas fazendo coisas que você jamais faria	Entretenimento
https://www.youtube.com/watch?v=bGT-BPhR0x8	Musica dos Simpsons na Pianola - Simpsons <i>theme in a Piano Roll</i>	Entretenimento
https://www.youtube.com/watch?v=bqq1UQGZcv0	Os 10 tombos mais engraçados em piscinas	Entretenimento
https://www.youtube.com/watch?v=EpIga48tzdw	Desafio Qual é a Música?	Entretenimento
https://www.youtube.com/watch?v=fpIDOVnWn2A	Super Mario <i>World - Overworld theme on piano</i>	Entretenimento
https://www.youtube.com/watch?v=HfnXPqdV8rU	<i>Iron Man 3 Music on Piano - Main Theme Soundtrack</i>	Entretenimento
https://www.youtube.com/watch?v=i-JIvpOsWz0	<i>Mortal Kombat Theme Song on Piano</i>	Entretenimento
https://www.youtube.com/watch?v=It0cowaiU80	Morre Chaves hoje dia 28/11/2014 e esta é a minha homenagem ao Roberto Bolaños	Entretenimento
https://www.youtube.com/watch?v=jM8hu-XTT3Q	10 grandes trapalhadas do futebol	Entretenimento
https://www.youtube.com/watch?v=K76DD__q-hY	Os 10 maiores motoqueiros em fazer vergonha	Entretenimento
https://www.youtube.com/watch?v=keWN9eKyy3A	Stormtrooper <i>plays Cantina Band music on piano - Starwars soundtrack</i>	Entretenimento
https://www.youtube.com/watch?v=kGz59kj62og	Vinheta Futebol Globo - Música de Futebol da Globo	Entretenimento
https://www.youtube.com/watch?v=krQMnfmFsKc	<i>Attack on Titan theme on piano - Shingeki no Kyojin music - Anime Theme Songs</i>	Entretenimento
https://www.youtube.com/watch?v=LyCINRLF-3U	Pokémon versão Sertanejo com Fábio Lima	Entretenimento
https://www.youtube.com/watch?v=MEJXWRjjyek	10 coisas que devem ter sido bem dolorosas	Entretenimento
https://www.youtube.com/watch?v=OxSW5_ZIZmA	Os 10 cães mais engraçados da <i>internet</i>	Entretenimento
https://www.youtube.com/watch?v=p21TNeudd6M	<i>Walking Dead Music on Piano - Main Title Opening Theme</i>	Entretenimento
https://www.youtube.com/watch?v=q9dWbqWPU5U	Piradinha em 5 ritmos diferentes - Música da Novela Amor a Vida - piri piri pira pira	Entretenimento

(continuação)

https://www.youtube.com/watch?v=qzMo-nkHsSM	10 truques de magia que deram errado	Entretenimento
https://www.youtube.com/watch?v=rAAMA5hDDqo	<i>Star Wars - Duel of Fates on Piano - Epic lightsaber music</i>	Entretenimento
https://www.youtube.com/watch?v=S0Mz5ZpZhvY	Os 10 cachorros mais sonhadores da <i>internet</i>	Entretenimento
https://www.youtube.com/watch?v=tCb76mCouZI	<i>Naruto theme on piano - Blue Bird - Naruto Ship-puden music - Anime Theme Songs</i>	Entretenimento
https://www.youtube.com/watch?v=v0FtUMxrmc0	10 vídeos engraçados de pessoas tomando choque	Entretenimento
https://www.youtube.com/watch?v=yMMycSccobc	<i>Brazilian Music - Tico Tico no fuba on piano</i>	Entretenimento
https://www.youtube.com/watch?v=ysWcDAGyyF0	<i>Darude - Sandstorm (Live Keyboard Playing Re-mix)</i>	Entretenimento
https://www.youtube.com/watch?v=zGS7EKC-O-U	10 pessoas que correram mais que Usain Bolt para escapar da morte	Entretenimento
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-acoes-de-conscientizacao-tem-ajudado-a-dimimuir-os-casos-de-dengue-em-jp/4635147/	Ações de conscientização têm ajudado a diminuir os casos de dengue em JP	Saúde
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-casas-fechadas-oferecem-risco-de-proliferacao-do-mosquito-da-dengue-em-campina-grande/4669013/	Casas fechadas oferecem risco de proliferação do mosquito da dengue em Campina Grande	Saúde
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-incluir-sementes-de-gerimum-na-dieta-ajuda-na-perda-de-peso/-4632413/	Incluir sementes de gerimum na dieta ajuda na perda de peso	Saúde
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-kako-marques-continua-maratona-de-exercicios-no-projeto-la-vem-o-verao/4637758/	Kako Marques continua maratona de exercícios no projeto "Lá Vem o Verão"	Saúde
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/-medico-explica-diferencas-entre-doencas-causadas-pelo-mosquito-aedes-aegypti/4666389/	Médico explica diferenças entre doenças causadas pelo mosquito <i>Aedes aegypti</i>	Saúde

(continuação)

http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/-v/numeros-de-criancas-com-microcefalia-dobrou-na-paraiba/-4645638/	Números de crianças com microcefalia dobrou na Paraíba	Saúde
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/-v/telespectadores-mostram-fotos-de-possiveis-foco-do-aedes-aegypt-em-joao-pessoa/4658754/	Telespectadores mostram fotos de possíveis focos do <i>Aedes aegypti</i> em João	Saúde
http://globoplay.globo.com/v/4635209/	Gestante faz exame do líquido amniótico e descobre que foi contaminada pelo Zika Vírus	Saúde
http://globoplay.globo.com/v/4635226/	Entenda de onde veio o Zika Vírus	Saúde
http://globoplay.globo.com/v/4635228/	Com um mês de vida, Helena faz reabilitação para tratar a microcefalia	Saúde
http://globoplay.globo.com/v/4645458/	Ministério da Saúde comprova ligação do mosquito <i>Aedes aegypti</i> com casos de microcefalia	Saúde
http://globoplay.globo.com/v/4645590/	Brasil tem má distribuição de médicos e atendimento no interior é precário	Saúde
http://globoplay.globo.com/v/4658314/	Greve de agentes que controlam o <i>Aedes aegypti</i> agrava combate a doenças	Saúde
http://globoplay.globo.com/v/4658330/	Brasil gasta mais no tratamento da dengue do que no combate ao mosquito	Saúde
http://globoplay.globo.com/v/4663604/	Casos de microcefalia sobem quase cinco vezes em um mês	Saúde
http://globoplay.globo.com/v/4666190/	Pesquisadores brasileiros começam a pensar em vacina contra o zika vírus	Saúde
http://globoplay.globo.com/v/4668697/	Ameaça do zika vírus faz aumentar procura por repelente nas farmácias	Saúde
http://globoplay.globo.com/v/4668828/	Fiscalização nas lojas do país verifica segurança dos brinquedos para o Natal	Saúde

(continuação)

http://globoplay.globo.com/v/4668847/	Técnica promete reverter a queda de cabelos causada pela quimioterapia	Saúde
https://www.youtube.com/watch?v=pm3do0nEuuM	Vírus Zika Nerdologia 112	Saúde
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/-v/saiba-o-que-fazer-quando-problemas-nos-fios-da-rua-prejudicam-servicos-de-internet/4632420/	Saiba o que fazer quando problemas nos fios da rua prejudicam serviços de <i>Internet</i>	Tecnologia
https://www.youtube.com/watch?v=f_ozBUARwuc	Hoje no TecMundo (09/12/2015): <i>Pixel C</i> , Marshmallow no Moto E 2015, computação quântica e mais	Tecnologia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/-v/veja-as-opcoes-de-lazer-para-o-fim-de-semana-em-campina-grande-e-joao-pessoa/4637732/	Veja as opções de lazer para o fim de semana em Campina Grande e João Pessoa	Tecnologia
http://g1.globo.com/videos/t/oficina-g1/v/guia-pratico-134-seu-carro-fala-aprenda-a-identificar-sintomas-de-problemas/-4389124/	Guia Prático #134: Seu carro "fala". Aprenda a identificar sintomas de problemas	Tecnologia
http://globoplay.globo.com/v/4657543/	Um acessório para ter luz diurna com LED no carro	Tecnologia
http://globoplay.globo.com/v/4657552/	Conheça novos motores V8 e entenda por que são tão apaixonantes	Tecnologia
https://www.youtube.com/watch?v=4q2JGULC0pE	Motorola atualiza lista de aparelhos que vão receber o Android Marshmallow	Tecnologia
https://www.youtube.com/watch?v=4UeMshwSU6k	Dicas para Fuvest: biologia	Tecnologia
https://www.youtube.com/watch?v=7DvrzwSJ4zQ	7 Mitos desmentidos sobre a bateria do seu celular	Tecnologia
https://www.youtube.com/watch?v=7GW0HBo55A0	Novos rumores sugerem iPhone 7 com 3 GB de RAM e <i>design</i> à prova d'água	Tecnologia
https://www.youtube.com/watch?v=ACi8wTAuvRI	Hoje No TecMundo (30/11/2015) - Casal do Mega Filmes HD solto, novo fone do iPhone 7 e AMD Crimson	Tecnologia

(continuação)

https://www.youtube.com/watch?v=aLo_iDNkv7U	Google finalmente lança o <i>Pixel C</i> , o seu primeiro <i>tablet</i> Android	Tecnologia
https://www.youtube.com/watch?v=_Asjgd0a3ZQ	Cases vazados mostram o que podemos esperar dos Galaxy S7 e S7+ [rumor]	Tecnologia
https://www.youtube.com/watch?v=baD2uP9tI0w	"30 anos do Windows Saga Mega Filmes HD e Ex-namorados no Facebook - Hoje no TecMundo (20/11/2k15)"	Tecnologia
https://www.youtube.com/watch?v=C2fxf_02m8c	Olhar Digital: Por que os smartphones esquentam tanto?	Tecnologia
https://www.youtube.com/watch?v=cAzU7qYAXYc	Este vídeo mostra por que você não deve acreditar em todo viral no YouTube	Tecnologia
https://www.youtube.com/watch?v=CKfFeM6zSa0	Por que o Netflix faz tanto sucesso?	Tecnologia
https://www.youtube.com/watch?v=DWWDCpr-L_Q	Hoje no TecMundo (23/11/2015)? Mega "Fake"HD, usuários do Popcorn Time rastreados e mais	Tecnologia
https://www.youtube.com/watch?v=-DzBF4qt5sk	Dicas para Fuvest: português e literatura	Tecnologia
https://www.youtube.com/watch?v=ehBPCu71nJs	"Um dia depois Mega Filmes HD volta ao ar em novo domínio"	Tecnologia
https://www.youtube.com/watch?v=Fa1L6QAwGco	5 coisas que podem ser hackeadas e você não sabia - TecMundo	Tecnologia
https://www.youtube.com/watch?v=h4TimFltbkg	WhatsApp e Telegram fazem cópias de suas imagens sem que você fique sabendo	Tecnologia
https://www.youtube.com/watch?v=hL1Z6kdJ1ds	10 probabilidades de algumas coisas acontecerem na sua vida	Tecnologia
https://www.youtube.com/watch?v=iK2Q6j7XDS8	Olhar Digital - Mitos e verdades sobre baterias de celular	Tecnologia

(continuação)

https://www.youtube.com/watch?v=IXAb1Kn7MEs	Hoje no TecMundo (02/12/2015)? herdeira do Facebook, caso Mega Filmes HD, iPad Pro no Brasil e mais	Tecnologia
https://www.youtube.com/watch?v=JurEsndsp6s	Engenheiro cria carro esportivo feito quase que inteiramente de madeira	Tecnologia
https://www.youtube.com/watch?v=KOmDZ-hsc2s	Olhar Digital - <i>Smartphones</i> à prova d'água realmente funcionam debaixo d'água?	Tecnologia
https://www.youtube.com/watch?v=LI5B2iWNmbc	<i>Streaming</i> : um jeito novo de ouvir música	Tecnologia
https://www.youtube.com/watch?v=LQ5UZ5O2DnE	Absurdo: empresa de cartão de crédito troca sobrenome de cliente por xingão	Tecnologia
https://www.youtube.com/watch?v=moWsmphMQFE	7 tecnologias de Star Wars que existem na vida real? TecMundo	Tecnologia
https://www.youtube.com/watch?v=OU9NjLV9-tg	De onde vem o Halloween?	Tecnologia
https://www.youtube.com/watch?v=pe91VGN5qT8	Em parceria com a NASA, Google demonstra avanços na computação quântica	Tecnologia
https://www.youtube.com/watch?v=PHWh8kTkRT8	Melhores do ano 2015: <i>Smartphone</i> super top de linha (até R\$ 5 mil)? TecMundo	Tecnologia
https://www.youtube.com/watch?v=PvGmZEprK08	10 pessoas fazendo coisas incrivelmente rápidas	Tecnologia
https://www.youtube.com/watch?v=qcahi843z6o	Xperia M5: Android com maior resolução de câmera no mundo chega ao Brasil	Tecnologia
https://www.youtube.com/watch?v=WCQKWbbIuS4	Sony pode anunciar um novo <i>smartphone</i> poderoso em janeiro	Tecnologia
https://www.youtube.com/watch?v=Wrlj1BzV_5h0	Conheça todos os celulares da Motorola e descubra qual é o melhor para você? TecMundo	Tecnologia
https://www.youtube.com/watch?v=yb82JQVWYS0	ASUS e LG fazem piada sobre novo case com bateria extra para o iPhone 6s	Tecnologia
https://www.youtube.com/watch?v=YWsOmV8Ix4U	É necessário ter cabos HDMI especiais para ver conteúdo em 4K? - TecMundo	Tecnologia

(continuação)

https://www.youtube.com/watch?v=zwla3k37i8k	Parece que o futuro não chegou: <i>smartphone</i> da Pepsi pode não sair do papel	Tecnologia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/economia/v/entenda-como-o-aumento-do-dolar-influencia-na-economia-brasileira/4386973/	Entenda como o aumento do dólar influencia na economia brasileira	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/beneficiarios-do-bolsa-familia-recebem-abono-natalino-em-campina-grande/4663699/	Beneficiários do Bolsa Família recebem abono natalino em Campina	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/cerca-de-300-mil-paraibanos-estao-com-contas-de-energia-atrasadas/4635183/	Cerca de 300 mil paraibanos estão com contas de energia atrasadas	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/combustivel-e-mais-carro-em-campina-grande-do-que-na-capital-segundo-pesquisa/4632424/	Combustível é mais caro em Campina Grande do que na Capital segundo pesquisa	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/profissionais-paraibanos-apostam-na-criatividade-para-superar-a-criese-economica/4658556/	Profissionais paraibanos apostam na criatividade para superar a crise econômica	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/saiba-os-direitos-do-consumidor-nas-compras-do-black-friday/-4637798/	Saiba os direitos do consumidor nas compras do <i>Black Friday</i>	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/tire-suas-duvidas-sobre-o-pagamento-do-13-salario/4629956/	Tire suas dúvidas sobre o pagamento do 13º Salário	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/veja-dicas-de-como-aproveitar-os-descontos-da-black-friday-sem-riscos/4629963/	Veja dicas de como aproveitar os descontos da <i>Black Friday</i> sem riscos	Economia
http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/v/veja-dicas-para-economizar-no-periodo-de-criese/4632406/	Veja dicas para economizar no período de crise	Economia

(continuação)

http://g1.globo.com/pb/paraiba/bom-dia-pb/videos/t/edicoes/-v/veja-dicas-para-fazer-compras-vantajosas-durante-a-black-friday/4635184/	Veja dicas para fazer compras vantajosas durante a <i>Black Friday</i>	Economia
http://globoplay.globo.com/v/4663621/	Consumo de combustíveis cai pela primeira vez em mais de dez anos	Economia
http://globoplay.globo.com/v/4666154/	Alta de preços também atinge alimentos vendidos nas feiras livres	Economia
http://globoplay.globo.com/v/4666162/	Conta de luz sobe até 300% de uma hora para outra no Amapá	Economia
http://globoplay.globo.com/v/4666179/	Preço do material escolar deve subir novamente em 2016	Economia
http://globoplay.globo.com/v/4668728/	Crise na economia diminui o movimento nas estradas brasileiras	Economia
https://www.youtube.com/watch?v=_dmWn8-Uu9A	Economia de Água e Energia Elétrica	Economia
https://www.youtube.com/watch?v=HcpDuTLIyWg	Hoje no TecMundo (26/11/2015) - Black Friday, Raspberry Pi Zero e abandono da AMD	Economia
https://www.youtube.com/watch?v=NtPFpa-a8zM	5 Dicas para Economizar Água	Economia
https://www.youtube.com/watch?v=rBu1tjuCy_s	Brasileiro abandona linhas de celular por causa de crise e do WhatsApp	Economia

Fonte: Próprio autor.

Apêndice C

Amostra dos Vídeos Exibidos

Este apêndice apresenta uma amostra de todos os vídeos exibidos nesta dissertação. As Figuras C.1, C.2, C.3, C.4 e C.5 apresentam amostras de vídeos sobre Diversão, Economia, Noticiário, Policial e Saúde, respectivamente. A amostra de vídeo político não foi exibida neste apêndice pois foi apresentada na Seção 3.1.1.



Figura C.1: Amostra de um vídeo sobre diversão, em que são apresentados jogadores de futebol que se atrapalharam em suas jogadas.

Fonte: Proveniente do canal Lista10 no Youtube.



Figura C.4: Amostra de um vídeo de tema policial, em que é noticiado o furto de um helicóptero que ocorreu em Mairinque (SP).

Fonte: Proveniente do programa Bom Dia Brasil transmitido em rede aberta pela emissora Rede Globo.



Figura C.5: Amostra de um vídeo sobre saúde, em que são noticiados os cuidados no combate ao Zika vírus e a relação com o surto dos casos de microcefalia.
Fonte: Proveniente do programa Bom Dia Paraíba transmitido em rede aberta pela TV Paraíba afiliada da Rede Globo.

Apêndice D

Códigos Fonte

Este apêndice apresenta os principais códigos fonte desenvolvidos nesta pesquisa. Os códigos estão organizados em seções. A Seção D.1 apresenta o código-fonte utilizado para realizar a exibição dos anúncios aos espectadores sincronizada com a captura. O código-fonte utilizado para treinar a rede *bayesiana* com os resultados de análise dos espectadores é apresentado na Seção D.2. Na Seção D.3, é apresentado o código-fonte utilizado para converter a base de imagens para o formato LMDB, padrão de entrada para o Caffe, que é o *framework* utilizado para treinar os classificadores. Por fim, a Seção D.4, apresenta quais modificações devem ser feitas na definição da arquitetura de rede GoogLeNet, utilizada como classificador de gênero e de idade.

D.1 Sincronização da Exibição e Captura

Para realizar a exibição/captura sincronizada foram utilizados os seguintes pacotes no Ubuntu 14.04 *Long Term Support* (LTS):

- **mplayer.** Reprodutor de vídeo utilizado para exibir os vídeos aos espectadores, que é de código fonte aberto e com suporte à várias codificações de vídeo;
- **gststreamer-tools.** Programa utilizado para realizar gravações utilizando a *webcam* que foi serviu para gerar a base de vídeos de captura;
- **ffmpeg2theora.** Codificador/Decodificador (CODEC) de vídeo que dá suporte a diferentes tipos de formatos.

Os seguintes comandos são suficientes para instalar esses pacotes e as suas dependências pela linha de comando utilizando os comandos exibidos no Código Fonte D.5.

Código Fonte D.1: Instalação das depências utilizadas para realizar a transmissão e captura sincronizadas.

```
1 sudo apt-get install mplayer
2 sudo apt-get install gstreamer-tools
3 sudo apt-get install ffmpeg2theora
```

Uma vez instaladas as dependências, é possível iniciar a exibição/captura sincronizada executando o Código Fonte D.2. A seguir são apresentadas as variáveis que deverão ser atualizadas pelo pesquisador que deseje replicar o experimento.

- **diretorioEntrada.** Diretório de entrada que contém todos os vídeos a serem exibidos aos espectadores;
- **diretorioSaida.** Diretório de saída que conterà todos os vídeos capturados dos espectadores;
- **fps.** Taxa de quadros por segundo que o vídeo será capturado pela *webcam*;
- **largura.** Largura do vídeo capturado pela *webcam*;
- **altura.** Altura do vídeo capturado pela *webcam*.

Código Fonte D.2: Código utilizado para realizar a exibição e captura sincronizadas.

```
1 #!/bin/bash
2
3 diretorioEntrada=$1
4 diretorioSaida=$2
5
6 fps=30
7 largura=1920
8 altura=1080
9
10 for arquivo in $(ls -al $pathIn | shuf ); do
11
```

```

12 videoEntrada=${diretorioEntrada$arquivo
13 quantQuadros=$( ./ getQuantQuadros $videoEntrada $fps )
14
15 mplayer -fs "$videoEntrada" &
16 gst-launch v4l2src num-buffers=$quantQuadros ! video/x-raw-yuv,width=
    $largura,height=$altura,framerate=$fps/1 ! decodebin2 !
    ffmpegcolorspace ! x264enc ! mp4mux ! filesink location=
    $diretorioSaida$arquivo
17
18 done

```

Por fim, o *script* apresentado no Código Fonte D.2, faz chamada a outro código auxiliar, apresentado no Código Fonte D.3 que é utilizado para estimar a duração do vídeo de exibição e conseqüentemente, o tempo de captura necessário. Esse código utiliza a biblioteca *Open Source Computer Vision* (OpenCV), foi utilizada a versão 2.4.9. Os parâmetros recebidos por ele são apresentados na lista a seguir, entretanto, o usuário não precisa modificá-las diretamente visto que são recebidos pelo *script* que o instancia.

- **argv[1]**. Nome do vídeo a ser exibido;
- **argv[2]**. Taxa de quadros por segundo que deverá ser capturado pela *webcam*.

Ambos os parâmetros são fornecidos pelo *script shell* apresentado no pseudocódigo anterior.

Código Fonte D.3: Código utilizado para estimar o tempo do vídeo de captura.

```

1 #include <opencv2/opencv.hpp>
2 #include <string>
3 #include <stdio>
4 #include <stdlib>
5
6 using namespace cv;
7
8 void printfError();
9 int getRounded( float valueSrc );
10 long int getNumFrames( float fpsDst, float fpsSrc,
11     long int numFramesSrc );

```

```
12
13 int main(int argc , char* argv[]) {
14     if (argc != 3) {
15         printfError();
16         return EXIT_FAILURE;
17     }
18
19     string nameVideo = argv[1];
20     float fpsOut = atof(argv[2]);
21     VideoCapture cap(nameVideo);
22
23     if (!cap.isOpened()) {
24         return EXIT_FAILURE;
25     }
26
27     double intervalBetweenFramesIn;
28     float fpsIn , numFramesOut;
29     long int numFramesIn;
30
31     cap.set(CV_CAP_PROP_POS_FRAMES, 100);
32     Mat frame;
33     cap >> frame;
34     intervalBetweenFramesIn = cap.get(CV_CAP_PROP_POS_MSEC);
35     fpsIn = (100000/intervalBetweenFramesIn);
36     fpsIn = getRounded(fpsIn);
37     numFramesIn = cap.get(CV_CAP_PROP_FRAME_COUNT);
38     numFramesOut = getNumFrames(fpsOut , fpsIn , numFramesIn);
39
40     cout << numFramesOut << endl;
41     return EXIT_SUCCESS;
42 }
43
44 // Calcula a quantidade de quadros a serem capturados
45 // @param taxa de quadros por segundo do vídeo capturado
46 // @param taxa de quadros por segundo do vídeo transmitido
47 // @param número de quadros do vídeo transmitido
48 // @return número de quadros do vídeo capturado
```

```
49 long int getNumFrames( float fpsDst , float fpsSrc ,
50   long int numFramesSrc ) {
51   long int numFramesDst;
52   float numFrameDstAux;
53   numFrameDstAux = ( fpsDst / fpsSrc ) * numFramesSrc;
54   numFramesDst = ( int ) numFrameDstAux;
55   return numFramesDst;
56 }
57
58 // Arredonda um número real
59 // @param número real
60 // @return número inteiro arredondado
61 int getRounded( float valueSrc ) {
62   int valueDst;
63   valueDst = ( int ) valueSrc;
64   if ( ( valueSrc - valueDst ) > 0.5 ) {
65     valueDst++;
66     valueSrc = valueDst;
67   }
68   return valueDst;
69 }
70
71 // Imprime mensagem de erro em caso de chamada errada
72 void printfError() {
73   std::cout << "You must call: " << std::endl;
74   std::cout << "./run <input video name> <input video fps>";
75 }
```

D.2 Processamento dos Dados Utilizando o Weka

O Código Fonte D.4, apresenta um *script* que foi utilizado para estimar o melhor algoritmo de aprendizagem e gerar a rede *bayesiana*. Faz saber das seguintes variáveis:

- **weka-3-7-13/weka.jar**: Versão do Weka utilizado.
- **-Xmx6500m**: Quantidade de memória a ser alocada, no caso, 6500MB.

- **\$inputFile**: Nome do arquivo de entrada, o formato utilizado é o *Attribute-Relation File Format* (ARFF).
- **\$numCross**: Número de passos na validação cruzada, o valor utilizado no experimento apresentado nesta dissertação foi dez.
- **-P 3**: Número de filhos na rede *bayesiana*.
- **\$output.txt**: Nome do arquivo em que as informações de processamento serão salvas, tais como: ROC, Fscore e outros. Caso não seja fornecido o caminho para algum arquivo, as informações de processamento serão impressas na tela.

Código Fonte D.4: Código para construção da Rede *Bayesiana* utilizando a base de dados.

```
1 java -cp weka-3-7-13/weka.jar -Xmx6500m weka.classifiers.bayes.BayesNet -  
    t $inputFile -x $numCross -Q weka.classifiers.bayes.net.search.global.  
    K2 -- -P 3 -S BAYES -E weka.classifiers.bayes.net.estimate.  
    SimpleEstimator -- -A 0.5 > $output.txt
```

D.3 Conversão da Base de Dados para o Formato LMDB

Como discutido na Seção 3.3.2, no processo de treinamento de um classificador utilizando o *framework* Caffe, a base de dados é completamente convertida em um único arquivo no formato LMDB. O Código Fonte D.5 é responsável por realizar esta conversão. O pré-requisito para executá-lo é possuir o Caffe devidamente instalado na máquina.

- **caminho_lmdb**. Caminho em que as imagens serão salvas após serem convertidas para o formato de dado LMDB;
- **caminho_imagens**. Caminho para as imagens da base de dados;
- **caminho_caffe**. Caminho em que o Caffe foi instalado na máquina;
- **caminho_imagens_treinamento**. Localização das imagens de treinamento;
- **caminho_imagens_validacao**. Localização das imagens de validação (pode ser o mesmo caminho para o diretório das imagens de treinamento desde que no arquivo

que descreva a lista de imagens do treinamento/validação faça a separação da base adequadamente);

- **redimensionar.** Se a base de dados precisa ser redimensionada para a resolução padrão (256×256), então, esta variável deve ser definida como *true*. Caso todas as imagens da base de dados já estejam redimensionadas, então, basta atribuir o valor *false*;
- **dimensao_altura.** Número de *pixels* da altura da imagem de entrada (0 caso não seja necessário redimensionar);
- **dimensao_largura.** Número de *pixels* da largura da imagem de entrada.

Além disso, é preciso fornecer dois arquivos *treinamento.txt* e *validacao.txt*, que são responsáveis por conter um relação de imagens do conjunto de treinamento e validação, respectivamente. Esses arquivos possuem a seguinte sintaxe:

```
1 caminho_para_a_imagem_1 rotulo_da_imagem_1
2 caminho_para_a_imagem_2 rotulo_da_imagem_2
3 ( ... )
4 caminho_para_a_imagem_n rotulo_da_imagem_n
```

Ou seja, o nome da imagem (acompanhada de sua extensão), espaço e o seu rótulo. O rótulo é representado por um número inteiro que indica a classe que a imagem pertence. Normalmente, o rótulo varia de zero até o número de classes envolvidas no treinamento menos um.

Código Fonte D.5: Código para conversão da base de dados para o formato LMDB.

```
1 caminho_lmdb=lmdb
2 caminho_imagens=base_de_dados
3 caminho_caffe=caffe/build/tools
4
5 caminho_imagens_treinamento=/caminho/para/treinamento/
6 caminho_imagens_validacao=/caminho/para/validacao/
7
8 redimensionar=false
9 if $redimensionar; then
```

```
10  dimensao_altura=256
11  dimensao_largura=256
12  else
13  dimensao_altura=0
14  dimensao_largura=0
15  fi
16
17  if [ ! -d "$caminho_imagens_treinamento" ]; then
18  exit 1
19  fi
20
21  if [ ! -d "$caminho_imagens_validacao" ]; then
22  exit 1
23  fi
24
25  GLOG_logtostderr=1 $caminho_caffe/convert_imageset \
26  --resize_height=$dimensao_altura \
27  --resize_width=$dimensao_largura \
28  --shuffle \
29  $caminho_imagens_treinamento \
30  $caminho_imagens/treinamento.txt \
31  $caminho_lmdb/lmdb_treinamento
32
33  GLOG_logtostderr=1 $caminho_caffe/convert_imageset \
34  --resize_height=$dimensao_altura \
35  --resize_width=$dimensao_largura \
36  --shuffle \
37  $caminho_imagens_validacao \
38  $caminho_imagens/validacao.txt \
39  $caminho_lmdb/lmdb_validacao
```

D.4 Modificações na Definição da Arquitetura de Rede

A arquitetura da rede GoogLeNet que é distribuída com a ferramenta Caffe possui parâmetros com valores *default*. Faz-se necessário modificar a quantidade de saídas de camadas específicas da rede. Isto é, por padrão, o número de saídas da GoogLeNet é 1000, visto

que a arquitetura de rede foi elaborada para realizar o treinamento na base de dados ImageNet (RUSSAKOVSKY et al., 2014), que possui 1000 classes de objetos. Assim, para o classificador de gênero, é preciso alterar o número de saídas para 2 (visto que as classes possíveis são: homem ou mulher). Para o classificador de idade, utilizando a base de dados MORPH, é preciso alterar o número de saídas para 54 (visto que, o intervalo de 16 à 59 anos de idade, possuem 54 idades distintas).

Assim, as camadas da definição de rede a serem alteradas são as seguintes: "loss1/classifier", "loss2/classifier" e "loss3/classifier". O atributo "num_output: 1000" do campo "inner_product_param" deve ser modificado para "num_output: 2" e "num_output: 54", caso se deseje treinar o classificador de gênero e de idade, respectivamente.

Apêndice E

Publicações

Este capítulo se destina apresentar todas as publicações decorrentes desta pesquisa, as quais se encontram sumarizados nos Quadros E.1, E.2 e E.3. Em sequência, são apresentadas cópias das primeiras páginas dos artigos publicados.

O artigo referenciado no Quadro E.1, destaca uma análise inicial realizada pelo autor em um cenário real. Essa análise relacionou o gênero dos espectadores ao conteúdo exibido. Nesse estudo, foram considerados apenas dois tipos de conteúdos exibidos, os anúncios jornalísticos e os esportivos. Além disso, foram empregadas técnicas com desempenhos inferiores aos apresentados nesta dissertação. Por exemplo, o detector de faces proposto por Viola e Jones (2001) e o classificador de gênero *Fisherfaces* (FISHER, 1936). Tal fato motivou a pesquisa por técnicas mais robustas. Foi necessária a realização de uma nova captura dos espectadores, com a exibição de mais vídeos e com mais categorias, assim como a inclusão de outro aspecto demográfico, que foi a faixa etária do espectador.

O artigo apresentado no Quadro E.2, consiste em uma análise parcial dos resultados desta dissertação. Foram empregados os mesmos materiais utilizados nesta dissertação. Entretanto, com um escopo menor, pois foram considerados apenas dois tipos de conteúdos: político e economia.

Por fim, a última publicação é apresentada no Quadro E.3. Tratou do processamento da base de faces, gerando faces sintéticas com ligeiras modificações em características faciais. A geração de uma base com maior amostra de faces forneceu melhor desempenho no classificador testado, a arquitetura de rede GoogLeNet (SZEGEDY et al., 2014). Com esse aumento de dados aplicado no treinamento do classificador de idade, foi obtido desempenho

superior ao do estado-da-arte.

Quadro E.1: Publicação na Revista de Informática Teórica e Aplicada (RITA Vol. 23).

Título da revista	Revista de Informática Teórica e Aplicada (RITA Vol. 23)
Título do artigo	Estudo da Relação Entre o Gênero de Espectadores e o Conteúdo de Vídeo
Lista completa dos autores	Ítalo de Pontes Oliveira, Eanes Torres Pereira e Herman Martins Gomes
Keywords	Sinalização digital, Detecção de faces, Classificação de gênero
Situação	Publicado

Fonte: Próprio autor.

Quadro E.2: Publicação no Encontro Paraibano de Estatística (EPBEST).

Título da conferência	Encontro Paraibano de Estatística (EPBEST)
Título do artigo	Redes <i>Bayesianas</i> Aplicadas à Análise da Atenção de Espectadores
Lista completa dos autores	Ítalo de Pontes Oliveira, Carlos Daniel Oliveira Interaminense, Oeslle Alexander Soares de Lucena e Eanes Torres Pereira
Keywords	Sinalização Digital, Atenção dos Espectadores, Redes <i>Bayesianas</i> , Redes Neurais Artificiais
Situação	Apresentado

Fonte: Próprio autor.

Quadro E.3: Publicação no evento *Conference on Graphics, Patterns and Images* (29th SIBGRAPI 2016).

Título da conferência	<i>Conference on Graphics, Patterns and Images</i> (29 th SIBGRAPI 2016)
Título do artigo	<i>A Data Augmentation Methodology to Improve Age Estimation using Convolutional Neural Networks</i>
Lista completa dos autores	Ítalo de Pontes Oliveira, João Lucas Peixoto Medeiros, Vinícius Fernandes de Sousa, Adalberto Gomes Teixeira Júnior, Eanes Torres Pereira, Herman Martins Gomes
Keywords	<i>Data Augmentation; Age Estimation; Deep Learning; Fiducial Points; Face Detection</i>
Situação	Publicado

Fonte: Próprio autor.

Estudo da Relação Entre o Gênero de Espectadores e o Conteúdo de Vídeos

Study of the Relationship Between Spectators Gender and Video Content

Ítalo de Pontes Oliveira ^{1, 2, 3}

Eanes Torres Pereira ^{1, 4}

Herman Martins Gomes ^{1, 5}

Data de submissão: 15/09/2015, Data de aceite: 10/05/2016

Resumo: A sinalização digital pode ser vista como uma forma de publicidade exibida painéis posicionados em locais com circulação ou presença de pessoas, a qual pode ser incrementada com métodos de visão computacional para oferecer recursos de medição da audiência. Neste contexto, o presente trabalho apresenta a avaliação experimental de uma abordagem para estimar, a partir de vídeos, o perfil de espectadores relacionando o gênero e os conteúdos exibidos. A abordagem foi implementada utilizando o detector de faces proposto por Viola-Jones e um classificador de gênero com base em características *Fisherfaces*, validados utilizando a base de faces FERET. O foco central deste trabalho foi o levantamento e teste experimental de hipóteses relativas à associação entre o gênero e categorias de conteúdos exibidos em sinalização digital. A análise realizada a partir de dados obtidos de um ambiente de cantina, em que os espectadores, monitorados por uma câmera de vídeo, assistiam a conteúdos selecionados em uma TV enquanto se alimentavam, mostrou que existe diferença estatística na atenção dada por homens e mulheres em função do tipo de conteúdo. Mais especificamente, (i) homens assistiram mais aos anúncios do que as mulheres; e (ii), os homens e as mulheres se concentraram mais aos anúncios do tipo esportivo do que aos anúncios do tipo jornalístico.

Palavras-chave: sinalização digital, detecção de faces, classificação de gênero

¹Laboratório de Percepção Computacional (LPC), Departamento de Sistemas e Computação (DSC), Centro de Engenharia Elétrica e Informática (CEED), Universidade Federal de Campina Grande (UFCG). R. Aprígio Veloso, 882 - Bairro universitário, Campina Grande - PB, 58429-900.

²Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

³{italooliveira@copin.ufcg.edu.br}

⁴{eanes@computacao.ufcg.edu.br}

⁵{hmg@dsc.ufcg.edu.br}

Redes Bayesianas Aplicadas à Análise da Atenção de Espectadores

Ítalo de Pontes Oliveira^{1,ab}, Carlos Daniel Oliveira Interaminense^{1,a,b},
Oeslle Alexander Soares de Lucena^{1,b}, Eanes Torres Pereira^{1,a,b}

^aLaboratório de Percepção Computacional (LPC)

^bUniversidade Federal de Campina Grande (UFCG)

Resumo

Grandes empresas têm investido milhões de reais em propagandas publicitárias, muitas vezes sem dispor de um meio adequado para avaliar o interesse do público por tais propagandas. Segundo o IBOPE, entre as dez maiores marcas do mercado brasileiro, 80% de seus investimentos publicitários são feitos na TV, sendo que, em 2014, 56% dos investimentos publicitários foram dedicados a esse veículo de comunicação. Assim, com o intuito de maximizar o Retorno Sobre o Investimento (*Return On Investment - ROI*), faz-se necessária uma avaliação precisa do perfil dos espectadores. Além disso, é relevante conceber formas de analisar automaticamente o interesse do público pela propaganda. Neste contexto, o presente trabalho apresenta uma análise estatística fundamentada em uma rede *bayesiana*, a qual foi automaticamente construída em um estudo de caso sobre a atenção de espectadores num cenário real de sinalização digital. A rede *bayesiana* é uma representação gráfica gerada à partir das probabilidades condicionadas entre duas variáveis que são calculadas utilizando o Teorema de *Bayes*. Assim, a rede *bayesiana* foi gerada utilizando como dados de entrada, o tipo conteúdo do vídeo transmitido, o gênero e a idade dos espectadores. O experimento foi realizado em um refeitório da Universidade Federal de Campina Grande, por vários dias, no mesmo horário, alternando apenas o conteúdo dos anúncios transmitidos. Foi realizada uma análise demográfica dos espectadores, relacionando o gênero e a faixa etária dos espectadores com o conteúdo exibido. Utilizou-se de uma câmera para capturar os espectadores enquanto assistiam os anúncios. Os vídeos capturados foram processados utilizando o detector de faces proposto por Markuš *et al.* (2014), que foi utilizado para detectar as faces frontais em relação ao monitor e determinar, indiretamente, quando os indivíduos estavam prestando atenção ao vídeo. Também foram empregadas redes neurais artificiais treinadas para estimar o gênero e a idade a partir de imagens de faces. A base de dados utilizada para treinamento das redes neurais contém mais de 50 mil faces humanas, a qual foi incrementada sinteticamente para mais de um milhão de imagens, fato que caracteriza o problema de pesquisa como um problema de *Big Data*. Neste trabalho, observaram-se diferenças na atenção dos espectadores considerando os seus gêneros (feminino e masculino) e os conteúdos transmitidos (política e economia). Quando considerado a idade dos espectadores, foi observado que os homens mais jovens tiveram preferência sobre os vídeos econômicos, enquanto que as mulheres, tiveram preferência sobre política. Ao se considerar os espectadores de idade mais elevada, foi observado que as mulheres são indiferentes ao conteúdo transmitido, enquanto isso, os homens tiveram preferência aos vídeos políticos. Na maioria dos casos, foi observado que as mulheres deram maior atenção a vídeos sobre economia, enquanto que os homens preferiram vídeos políticos. Por fim, o presente trabalho contribui com uma detalhada análise estatística sobre a atenção dos espectadores, uma vez que, foram utilizadas redes neurais treinadas com grande volume de dados, permitindo assim, resultados de classificação de gênero e idade mais precisos.

Palavras Chaves: Sinalização Digital, Atenção dos Espectadores, Redes *Bayesianas*, Redes Neurais Artificiais.

¹ Autor Correspondente: Laboratório de Percepção Computacional (LPC), Universidade Federal de Campina Grande (UFCG), Av. Aprígio Veloso, 882, Bodocongó, Campina Grande - PB, 58429-970, Brasil. E-mail: italooliveira@copin.ufcg.edu.br, carlos.interaminense@ccc.ufcg.edu.br, oeslle.lucena@ee.ufcg.edu.br e eanes@computacao.ufcg.edu.br.

A Data Augmentation Methodology to Improve Age Estimation using Convolutional Neural Networks

Ítalo de Pontes Oliveira, João Lucas Peixoto Medeiros, Vinícius Fernandes de Sousa
Adalberto Gomes Teixeira Júnior, Eanes Torres Pereira, Herman Martins Gomes
Federal University of Campina Grande (UFCG)

Email: italooliveira@copin.ufcg.edu.br, joao.medeiros@ee.ufcg.edu.br, viniiciusfernandes@copin.ufcg.edu.br,
adalbertojunior8@gmail.com, eanes@dsc.ufcg.edu.br, hmg@dsc.ufcg.edu.br

Abstract—Recent advances in deep learning methodologies are enabling the construction of more accurate classifiers. However, existing labeled face datasets are limited in size, which prevents CNN models from reaching their full generalization capabilities. A variety of techniques to generate new training samples based on data augmentation have been proposed, but the great majority is limited to very simple transformations. The approach proposed in this paper takes into account intrinsic information about human faces in order to generate an augmented dataset that is used to train a CNN, by creating photo-realistic smooth face variations based on Active Appearance Models optimized for human faces. An experimental evaluation taking CNN models trained with original and augmented versions of the MORPH face dataset allowed an increase of 10% in the F-Score and yielded Receiver Operating Characteristic curves that outperformed state-of-the-art work in the literature.

Keywords—Data Augmentation; Age Estimation; Deep Learning; Fiducial Points; Face Detection

I. INTRODUCTION

In recent years, there has been great advances in image recognition and classification due to the progress on training large neural network models. Not surprisingly, deep learning approaches are among the solutions with best results found in the literature [1]. Several factors have contributed to that, being the increase in processing power of General Purpose Graphics Processing Units (GP-GPUs) an important one, which allows faster training of neural network models with large number of stacked layers (e.g., Deep Convolutional Neural Networks - CNN) and large training sets.

Models with many free parameters (such as CNN) require a large volume of data for training. More specifically, when such models are applied to the face analysis area, there is a crescent need for large face image datasets that are not currently available. Various approaches have been proposed for general data augmentation with the purpose of generating better classification models, including CNN [2]. However, they only consider very simple transformations on the input images, such as changing the view point, flipping, cropping or making color changes (e.g. [3]), and do not take advantage of specific knowledge regarding the problem at hand.

Within the above context, this paper proposes a novel methodology based on Active Appearance Models in order to perform data augmentation on human face datasets, with a special focus on training CNNs classifiers for face image

age estimation. Formally, the data augmentation term refers to methods for constructing iterative optimization or sampling algorithms via the introduction of unobserved data or latent variables [4], which are often utilized to generate additional training data without introducing extra labeling costs.

In this paper, we present a novel approach to perform data augmentation which was validated using a CNN model trained with original (about 50K images) and augmented (> 1M images) versions of the MORPH face dataset. Data augmentation allowed an increase of 10% in the Fscore and yielded Receiver Operating Characteristic curves that outperformed state-of-the-art work in the literature of age estimation from human faces.

This article is organized as follows: Section II describes the related work on data augmentation techniques. The CNN operation is briefly explained in section III. CNN was used to validate the proposed methodology for data augmentation. Section IV details how the proposed approach works. An experimental evaluation, implementation details and results are discussed in Section V. Finally, Section VI presents the conclusions of the work.

II. RELATED WORK

Dosovitskiy *et. al.* [5] investigated the use of data augmentation as the main component of an unsupervised feature learning architecture. Their approach obtains random patches of size 32×32 pixels from different images and applies random transformations combining four kinds variations that include: translation transformation, scale, color and contrast variation. A limitation of that work was the absence of a comparison of the accuracy scores with and without data augmentation. Another similar work is that of Gerke *et. al* [6], in which the data augmentation technique was applied to obtain random patches from the images to increase the size of the dataset in a ratio of 1 : 10. Classifier accuracy, in a problem of soccer jersey number recognition, has increased more than 10% when using the augmented dataset.

Data augmentation was used by Chen *et. al* [7] to reduce classifier overfitting. Translations and horizontal reflections have been applied to dataset images. On the training phase, the data augmentation process extracts random 224×224 patches from the 256×256 pixels images, including their horizontal reflections. The above discussed strategies do not consider semantic information of the data, only changes in the point of