UNIVERSIDADE FEDERAL DO PARÁ

INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ROBERTO YURI DA SILVA FRANCO

Um Modelo para Classificação e Visualização Automática de Dados de Sentimentos

Belém

2019

ROBERTO YURI DA SILVA FRANCO

Um Modelo para Classificação e Visualização Automática de Dados de Sentimentos

Tese de Doutorado apresentado ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas e Naturais como requisito para obtenção do título de doutor em Ciência da Computação.

Universidade Federal do Pará

Orientador: BIANCHI SERIQUE MEIGUINS

Belém

2019

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

D111m da Silva Franco, Roberto Yuri
Um Modelo para Classificação e Visualização Automática de Dados de Sentimentos / Roberto Yuri da Silva Franco. — 2019.
139 f. : il. color.

Orientador(a): Prof. Dr. Bianchi Serique Meiguins
Tese (Doutorado) - , , Universidade Federal do Pará, Belém, 2019.

1. Análise de Sentimentos Multimodal. 2. Visualização da Informação. 3. Avaliação. 4. Inteligência Artificial. I. Título.

CDD 004

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ROBERTO YURI DA SILVA FRANCO

UM MODELO PARA CLASSIFICAÇÃO E VISUALIZAÇÃO AUTOMÁTICA DE DADOS DE SENTIMENTOS

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal do Pará como requisito para obtenção do título de Doutor em Ciência da Computação, defendida e aprovada em 17/12/2019, pela banca examinadora constituída pelos seguintes membros:

Prof. Dr. Bianchi Serique Meiguins
Orientador – PPGCC/UFPA

Prof. Dr. Jefferson Magalhães de Morais
Membro Interno – PPGCC/UFPA

Prof. Dr. Carlos Gustavo Resque dos Santos
Membro Interno – PPGCC/UFPA

Profa. Dra. Fabíola Pantoja Oliveira Araújo
Membro Externo – FACOMP/UFPA

Prof. Dr. Marcelo de Paiva Guimarães
Membro Externo – UNIFESP

Visto:
Prof. Dr. Nelson Cruz Sampaio Neto
Coordenador do PPGCC/UFPA

Prof. Dr. Nelson Cruz Sampaio Neto
Coordenador do PPGCC/UFPA
SIAPE: 2650210

Agradecimentos

Uma das tarefas mais difíceis de realizar nesta tese foi escrever os agradecimentos. Foram tantas pessoas que contribuíram para a execução desse trabalho e que forneceram ideias, comentários e considerações que esta seção não seria suficiente, porém, tentarei fazer essa lista de contribuintes.

Primeiramente, agradeço a Deus pela vida, pela capacidade de pensamento e por todas as coisas que Ele tem permitido. Agradeço, também, à minha mãe, Margarete Franco, por todos os seus esforços, dedicação, incentivo, ensinamentos e carinho, principalmente no aprendizado de como superar dificuldades. Agradeço ao meu pai, Tocantins Franco, por tudo que fez em vida ou pós vida para me ajudar. Agradeço ao meu irmão Roberto Allen e os meus sobrinhos Samia e Salim pelas diversões e distrações em momentos familiares.

Agradeço também à minha namorada Carolina Menezes por todas suas contribuições, revisões, apoio e por tudo que tem contemplado para a realização desse trabalho. Agradeço também à família da Carol, em especial, seu pai Max Reis, por todo o incentivos.

Agradeço, principalmente, ao meu orientador Bianchi Serique Meiguins por todo apoio, dedicação e, principalmente, pela abdicação dos seus dias de folga para que esse trabalho fosse feito tão breve quanto possível.

Agradeço também aos amigos do LabVis que contribuíram incessantemente nesta tese, em especial ao, Tiago Araújo, Rafael Paixão, Rodrigo Divino, Alexandre Freitas, Anderson Soares, Elvis Thermo, Gustavo Resque, Nelson, Yvan, Diego, Brunelli, Jairo e Luan por dicas, comentários, códigos e pizza.

Agradeço ainda aos amigos que me receberam tão bem em momentos de aprendizado e pesquisa: Wanderson Cunha, Fabrício Garcia, Paulo Malcher, Mariana Oliveira, Manuelle Ribeiro, Eulália da Mata e Koite Yasojima.

Por fim, agradeço a todos que não foram diretamente citados, mas que forneceram contribuições valiosas.

"É a cerração... Vida de bordo numa sala que palpita de emoção [...] Mas na hora em que na adriça, cessada a tempestade, a bandeira subia garbosa..."

(Bruno de Menezes, 1939)

Resumo

Análise de sentimentos é o processo de associar um ou mais sentimentos a uma pessoa durante uma determinada atividade, por exemplo, utilizando um produto físico ou interagindo com um software. Técnicas de aprendizado de máquina têm sido aplicadas para a classificação de sentimentos considerando os mais variados tipos de dados, tais como: vídeo, áudio, texto, dados fisiológicos, logs de interação, entre outros. A complexidade das técnicas de aprendizado de máquina e da interpretação dos resultados da análise de sentimentos torna, muitas vezes, a utilização dessas tecnologias inacessíveis a muitos profissionais que querem associar esses dados em suas análises. Além disso, apresentar os resultados obtidos e extrair alguma informação útil também é uma tarefa complexa, principalmente em função da grande quantidade de dados geradas, como o fluxo do tempo, fala, atividades e os sentimentos associados, tornando-se difícil buscar padrões e apresentá-los de modo a ajudar na tomada de decisões. A partir deste contexto, esta tese de doutorado apresenta um modelo que integra e abstrai a utilização das técnicas de aprendizado de máquina aplicadas à análise de sentimentos multimodal, incluindo o pré-processamento necessário, e propõe os resultados através de técnicas de visualização. Como cenário de utilização, o modelo foi aplicado em dados reais relativos a um processo de avaliação de usabilidade e experiência do usuário na interação com um software de visualização. Esta implementação do modelo gerou a ferramenta UXmood, que realiza a análise de sentimentos combinando dados de vídeo, áudio, texto em abordagem multimodal, e associando-os aos logs de interação do mouse, rastreador de olhos e a fala do usuário durante as tarefas que foram realizadas pelos participantes do teste, apresentando-os em um *dashboard* de visualizações coordenadas. Além disso, foram analisadas a viabilidade de utilização de arquiteturas de redes neurais convolucionais para as tarefas de análise de sentimento em diversos tipos de mídia. Para a validação do modelo e sua implementação foram realizados as seguintes etapas: 1 - foram testadas e comparadas arquiteturas de redes neurais convolucionais para análise de sentimentos em vídeo, áudio e texto, de forma individual, e, posteriormente, aplicadas a uma abordagem multimodal baseada em pesos (fusão por nível de decisão); 2- foram desenvolvidos cenários de visualização de dados e avaliadas por meio de questionários on-line com o objetivo de identificar os principais problemas de usabilidade; 3 – após realizar o redesign da ferramenta com os resultados da etapa 2, a ferramenta foi testada junto a usuários especialistas no contexto do teste. Por fim, os resultados da aplicação das técnicas de aprendizado de máquina e abordagem multimodal se mostraram compatíveis com a literatura e a utilização por especialistas apresentou boa aceitação da ferramenta e ótima avaliação na facilidade de uso para geração e visualização dos dados de sentimento.

Palavras-chave: Análise de Sentimentos Multimodal. Visualização da Informação. Avaliação. Inteligência Artificial.

Abstract

Sentiments analysis is the process of associating one or more Sentiments with a person during a particular activity, for example, using a physical product or interacting with software. Machine learning techniques have been applied for this type of classification considering the most varied types of data, such as video, audio, text, physiological data, interaction logs, among others. The complexity of machine learning techniques and the interpretation of the results about Sentiments analysis often makes the use of these technologies inaccessible to many professionals who want to associate this data in their analysis. In addition, presenting the results obtained and extracting useful information is also a complex task, especially due to the large amount of data generated, such as the flow of time, speech, activities, associated Sentiments, so it becomes difficult to search for patterns and present them to help in decision making. From this context, this doctoral thesis presents a model that integrates and abstracts the use of machine learning techniques applied to multimodal sentiment analysis, including the necessary preprocessing, and proposes the presentation of results through data visualization techniques, coordinates for analysis of these data together with others. As a scenario of use, the model was applied to real data related to a usability and user experience evaluation process in the interaction with a multidimensional data visualization software. This model implementation generated the UXmood tool, which performs sentiment analysis by combining data from video, audio, text in a multimodal approach, and associating it with mouse interaction logs, eye tracker and user speech during tasks that were performed by the test participants, presenting them in a configurable panel of coordinated data visualizations for the Web platform. In addition, the feasibility of using convolutional neural network architectures for sentiment analysis tasks in various types of media was analyzed. For the validation of the model and its implementation, the following steps were performed: 1 - Convolutional neural network architectures were tested and compared to analyze feelings in video, audio and text individually, and subsequently applied to a multimodal approach based on in weights (fusion by decision level); 2- Data visualization scenarios were developed and evaluated through online questionnaires in order to identify the main usability problems; 3 - After redesigning the tool with the results from step 2, the tool was tested with expert users in the context of the test. Finally, the results of the application of machine learning techniques and multimodal approach were compatible with the literature and the use by experts showed good acceptance of the tool and excellent evaluation of ease of use for generation and visualization of sentiment data.

Keywords: Multimodal Sentiment Analysis. Information Visualization. Evaluation. Artificial Intelligence.

Lista de ilustrações

Figura 1. Representação do método de fusão por nível de características 25
Figura 2. Representação do método de fusão por nível de decisão 26
Figura 3. Representação do método híbrido de fusão 26
Figura 4. Processo genérico para aprendizado de um modelo inteligente de aprendizado de máquina. As etapas podem ser adaptadas dependendo do problema e dos algoritmos utilizados. 28
Figura 5. Classificação de imagens seguindo duas abordagens, (a) a abordagem clássica, com extração de características e classificadores tradicionais. (b) A abordagem de aprendizado de representações, com as características sendo aprendidas durante o treino. 29
Figura 6. Rede Neural e seus componentes. No lado esquerdo, o neurônio e sua descrição formal. No lado direito, a rede completa, onde cada neurônio (círculo) segue a forma do lado esquerdo. 31
Figura 7. Arquitetura da LeNet, primeira CNN publicada. As camadas de convolução são apresentadas, assim como camadas de pooling e ao final as camadas totalmente conectadas. Os planos são os mapas de ativação entre as camadas. 32
Figura 8. Exemplo de aplicação do método do gradiente. 32
Figura 9. Processo genérico de geração de visualizações de dados. 34
Figura 10. Processo de geração de visualizações de dados. 35
Figura 11. Representação visuais das percepções da marca nas mídias sociais. . . . 37
Figura 12. Visualizações mais modernas com base em múltiplas visões coordenadas. 38
Figura 13. Distribuição anual dos artigos selecionados. 40
Figura 14. Percentual da área de aplicação dos artigos selecionados. 40
Figura 15. Visualização em layout de matrix do uso de impressoras por usuário. . 41
Figura 16. As técnicas de visualização de dados mais utilizadas para a representação visual de dados de sentimentos. 41
Figura 17. Mapa com a concentração de sentimentos positivos. 42
Figura 18. Trabalhos com até três visualizações para representar os dados de sentimentos. 42
Figura 19. Dados de sentimentos geoespaciais representado por cores através de mapa. 43
Figura 20. Aplicações que utilizaram mais de uma visão de dados. 43
Figura 21. Dados de sentimentos combinado com dados do domínio da aplicação. . 44
Figura 22. Trabalhos que usam polaridade, valência, emoções e polaridade e emoções. 44
Figura 23. Treemap com informações de polaridade de dados de redes sociais. . . . 45

Figura 24. Aplicação para análise de emoções de dados de redes sociais. 45
Figura 25. Classes de análise do survey. 46
Figura 26. Visão Geral de métodos comuns de entrada para análise de sentimentos, com respectivos métodos para classificação. 48
Figura 27. Taxonomia de reconhecimento de faces, aplicável para análise de sentimentos. 49
Figura 28. Classificadores de análise de sentimentos apresentando: (a) a fusão por nível de característica, (b) a fusão por nível de decisão e (c) abordagem intermediária, com fusão em diferentes níveis. 50
Figura 29. Visão macro do modelo. 51
Figura 30. Visão geral detalhada do modelo e módulos. 52
Figura 31. Módulo de gerenciamento de entrada de dados. 53
Figura 32. Fluxograma do ajuste temporal dos arquivos. 53
Figura 33. Fluxograma da localização e identificar ausência de dados dos canais. . 54
Figura 34. Funcionamento do módulo de construção do modelo inteligente. 55
Figura 35. Abordagem de fusão por nível de decisão. 56
Figura 36. Abordagem de fusão por nível de características. 57
Figura 37. Representação do módulo de análise de sentimentos multimodal. 59
Figura 38. Representação do módulo de visualização de dados. 61
Figura 39. Uma visão geral da interface principal do UXmood. 62
Figura 40. Visualizações configuradas para serem exibidas em dois monitores. . . . 63
Figura 41. Arquitetura da ferramenta UXmood. 67
Figura 42. Diagrama de classe do protótipo UXmood. 70
Figura 43. Os menus (a) para registrar testes ou tarefas em um projeto, que são exibidos como uma lista (b). 71
Figura 44. Formulário de cadastro do teste. 72
Figura 45. Exemplo de arquivo de log com dados de *Mouse / Eye Tracker*. 72
Figura 46. Figura 16. Formulário de cadastro da tarefa. 73
Figura 47. Legendas de Emoção, polaridade e resposta; as legendas atuam como filtros nas visualização. 74
Figura 48. O controle que manipula de forma síncrona as imagens do teste. 74
Figura 49. Detalhes sob demanda que fornecem informações adicionais sobre a tarefa (a) o vídeo (b) e a fala transcrita (c). 75
Figura 50. Tela de configuração do UXmood. 75
Figura 51. O gráfico de Gantt mostra o sentimento classificado de segmentos de tempo para cada tipo de entrada; o tempo e a duração das tarefas são mostrados na parte superior do gráfico. 77
Figura 52. O gráfico de Gantt configurado para exibir informações emocionais de todo o teste (a) ou apenas dos segmentos de tarefas (b). 77

Figura 53. Gráfico de Gantt configurado para representar tarefas como segmentos: a visualização mostra a duração, resultado e sentimento predominante das tarefas. 77
Figura 54. Gráfico de Gantt configurado para comparar diferentes testes. 78
Figura 55. Um gráfico de barras empilhadas de emoções classificadas (esquerda) e polaridade (direita). 78
Figura 56. Wordcloud de palavras faladas pelo usuário. 79
Figura 57. Um gráfico de caminho de verificação mostrando a fixação do usuário com o sentimento inferido associado. 79
Figura 58. Um gráfico de dispersão com os pontos interagidos pelo usuário (rastreamento de mouse ou olho), colorido pela emoção associada no momento da interação. 80
Figura 59. Vídeo da face do usuário com ícone de sentimento. 80
Figura 60. EmojiText, uma técnica de visualização para representar sentimentos em textos. 81
Figura 61. Visualização de cenário de uso. 86
Figura 62. Visualização de cenário de uso. 87
Figura 63. Modelo de avaliação da ferramenta 88
Figura 64. Modelo de avaliação da ferramenta 90
Figura 65. A precisão das respostas nas tarefas do questionário. 93
Figura 66. A precisão das respostas nas tarefas do questionário. 94
Figura 67. As pontuações do SUS de cada visualização. 95
Figura 68. Redesign para o gráfico de Gantt. A principal faixa é a multimodal e as outras aparecem como detalhes sob demanda. 96
Figura 69. Uma reformulação proposta para os gráficos Wordcloud e Scanpath / Scatter. Múltiplos pequenos podem reduzir a confusão visual e facilitar a interpretação das visualizações. 96
Figura 70. Reformulação do painel interativo de visualizações. 97
Figura 71. Tempo médio de realização das tarefas por tarefas. 100
Figura 72. Média de tempo gasto para a realização das tarefas junto com a emoção predominante. 101
Figura 73. Recursos de interação mais utilizados para solução das tarefas. 101
Figura 74. Visualizações mais utilizadas para solução das tarefas. 102
Figura 75. Maiores regiões de fixação visual pelos usuário. 102

Lista de quadros

Quadro 1. Tarefas de teste do protótipo UXmood 99

Lista de tabelas

Tabela 1. Classificação e catalogação dos trabalhos relacionados. 65
Tabela 2. Melhores resultados das CNNs. Xception foi retreinada com mais 150 épocas. Cada arquitetura gerou 12 modelos, totalizando 49 modelos treinados (48 + 1 final), com melhor resultado 68%. 84
Tabela 3. Interpretação da Escala SUS 91
Tabela 4. Questões mais citadas na pergunta aberta do questionário. 93
Tabela 5. Dados qualitativos dos recursos de interação e visualizações stackedbar e gantt chart. 103
Tabela 6. Dados qualitativos das visualizações wordcloud, scanpath, scatterplot e EmojiText. 103

Lista de abreviaturas e siglas

API	*Application Programming Interface* Interface de Programação de Aplicativos
d3js	*Data-Driven Documents JavaScript* Documentos orientados a dados JavaScript
infoVis	*Information Visualization* (Visualização da Informação)
MVC	*Model View Controller* Modelo Visão Controle
SUS	*System Usability Scale* Escala de Usabilidade do Sistema
TCLE	Termo de Consentimento Livre e Esclarecido
UX	*User Experience* (Experiência do Usuário)

Sumário

1 INTRODUÇÃO 16
1.1 Justificativa e Motivação 19
1.2 Objetivos 19
1.2.1 Objetivo Geral 19
1.2.2 Objetivos Específicos 20
1.3 Contribuições 20
1.3.1 Publicações Relacionadas com a Tese 20
1.4 Organização do Trabalho 21

2 FUNDAMENTAÇÃO TEÓRICA 22
2.1 Análise de Sentimentos 22
2.1.1 Sentimentos, Emoções e Polaridade 22
2.1.2 Computação Afetiva 23
2.1.3 Técnicas de Análise de Sentimentos e Reconhecimento de Emoções 24
2.1.4 Análise de Sentimentos Multimodal 25
2.2 Aprendizado de Máquina 26
2.3 Visualização de Dados 33
2.3.1 Pipeline de visualização de dados 33
2.3.2 Visualização de sentimentos 35

3 TRABALHOS RELACIONADOS 39
3.1 Visualização de Dados 39
3.2 Aprendizado de Máquina 47

4 MODELO 51
4.1 Módulo Entrada de dados: 52
4.2 Módulo Construção do Modelo Inteligente: 55
4.3 Escolher o tipo de classificador 56
4.3.1 Definir regime de treinamento 57
4.3.2 Obter base de dados 58
4.3.3 Realizar treino e validação 58
4.3.4 Criar modelo inteligente para produção 58
4.4 Módulo Análise de sentimentos Multimodal: 59
4.5 Módulo Visualização de dados: 60

5 IMPLEMENTAÇÃO DO PROTÓTIPO DO MODELO 62

5.1 Justificativa da Área de Aplicação . . . 63
5.2 Arquitetura . . . 66
5.3 Tecnologias . . . 68
5.4 Aspectos da Implementação . . . 69
5.5 Funcionalidades . . . 70
5.5.1 Projetos, teste e organização de tarefas . . . 70
5.5.2 Legenda, Filtros, Controle, Detalhes sob demanda e Configurações . . . 73
5.5.3 *Dashboard* de Visualizações . . . 76
5.6 Classificação de sentimentos . . . 81
5.6.1 Bases de dados . . . 82
5.6.2 Arquiteturas de CNN . . . 83
5.6.3 Treinamento e Classificação por CNN . . . 83
5.6.4 Cenários de Uso . . . 86

6 AVALIAÇÕES DO PROTÓTIPO . . . 88
6.1 Etapa 1 - Questionário On-line . . . 89
6.2 Escala de Usabilidade do Sistema (SUS) . . . 90
6.3 Resultados do Questionário On-line . . . 92
6.4 Redesenho das Técnicas de Visualização . . . 95
6.4.1 Avaliação com especialistas . . . 98

7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS . . . 105

REFERÊNCIAS . . . 109

APÊNDICE A – PARECER DO COMITÉ DE ÉTICA . . . 121

APÊNDICE B – QUESTIONÁRIO ON-LINE . . . 122

APÊNDICE C – TCLE - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO . . . 137

1 Introdução

A expressão das emoções é um importante canal de comunicação entre pessoas. Essa característica dos seres humanos tem sido explorada pela Computação Afetiva também no entendimento da comunicação de pessoas e máquinas durante a interação com as mesmas (ROBINSON; KALIOUBY, 2009). A Análise de Sentimentos é área que aborda a classificação das emoções de pessoas através de expressões faciais, fala, escrita, entre outros. (TAO; TAN, 2009). Esta área têm se beneficiado com a evolução das técnicas de Inteligência Artificial (IA), como as redes neurais convolucionais (PORIA; HUSSAIN; CAMBRIA, 2018) e estão sendo aplicadas em diversos campos, tais como: gerenciamento do relacionamento de clientes, sistemas de recomendações, direcionamento em investimentos de estratégias de marketing, predição do comportamento do público em geral sobre produtos e marcas, etc. (ZADEH et al., 2017)

Apesar das técnicas de IA permitirem a utilização mais eficiente da análise de sentimentos, essa nem sempre é uma tarefa ou processo fácil de incluir em ferramentas ou análises de dados (PORIA; HUSSAIN; CAMBRIA, 2018), ficando restrita a especialistas da área. A utilização dessas técnicas requer conhecimento especializado e, muitas vezes, a configuração de grande quantidade de parâmetros, pré-processamentos de mídias, e, não menos importante, a leitura e interpretação desses resultados (KWIATKOWSKA; SZÓSTEK; LAMAS, 2014).

A Análise de Sentimentos Multimodal é uma abordagem que tem sido utilizada para melhorar os resultados de classificação da análise de sentimentos pelas técnicas de IA, que inclui, além da tradicional análise de sentimentos baseado em texto, a análise em conjunto de outros tipos de dados, tais como: áudio, vídeo, dados fisiológicos, entre outros, envolvendo modelos mais complexos de análise de sentimentos. Contudo, a maioria das aplicações de análise de sentimentos tem sido voltada para o texto, considerando o contexto envolvido (Ebrahimi; Yazdavar; Sheth, 2017; Fong et al., 2013; Khozyainov; Pyshkin; Klyuev, 2013a; Patil; Atique, 2015).

A utilização de técnicas de IA para imagens e áudios tem obtido resultados excelentes em várias tarefas (PORIA; HUSSAIN; CAMBRIA, 2018), porém, essas soluções são mais complexas do que as de texto, o que torna a sua utilização ainda mais difícil. Alguns trabalhos aplicam essas técnicas (KAUR; KAUTISH, 2019; SOLEYMANI et al., 2017a) na classe de problemas de análise de sentimentos com bons resultados, mas sem aplicação prática dos resultados em cenários de utilização (HUSSEIN, 2018; MEDHAT; HASSAN; KORASHY, 2014).

Uma abordagem para mitigar a análise da complexidade e quantidade dos resultados

gerados pelas técnicas inteligentes é a utilização de técnicas de Visualização da Informação (VI) que permitem ao usuário visualizar dados abstratos em uma organização visual que pode facilitar o entendimento sobre o conjunto de dados e seus relacionamentos, padrões, outliers, tendências etc. Além disso, o aspecto interativo das ferramentas de VI permite criar diversos cenários de análise utilizando filtros de dados, relacionamento entre as visualizações e possibilita configurações dinâmicas que possam evidenciar mais facilmente os dados de análise (Ebrahimi; Yazdavar; Sheth, 2017; Khozyainov; Pyshkin; Klyuev, 2013b).

De acordo com o *survey* apresentado por Kucher, Paradis e Kerren (2018a) sobre Visualização de Sentimentos, dados resultantes do processamento de análise de sentimentos podem e devem ser combinados a outros dados. Esses conjuntos de dados gerados são caracterizados por serem multidimensionais, complexos (por exemplo, por apresentarem dados não estruturados), bem como apresentam aspectos temporais, relacionais e geo-espaciais. A grande maioria das técnicas ou ferramentas de VI apresentadas combina o uso de várias variáveis visuais, tais como: tamanhos/área, linhas ou gráfico de pizza ou barras ou apresentam múltiplas visões de dados. Apesar disso, as técnicas de VI ainda são pouco exploradas para dados de sentimentos, pois frequentemente são utilizadas de maneira simples representando dados sumarizados.

Com o objetivo de diminuir a complexidade no processamento e na análise de dados de sentimento em cenários exploratórios e de análise, esta tese de doutorado apresenta um modelo computacional que facilita a utilização das técnicas de aprendizado de máquina aplicadas à análise de sentimentos multimodal, incluindo o pré-processamento necessário, e propõe a apresentação dos resultados através de múltiplas técnicas de visualização de dados coordenadas.

A parte do modelo que abstrai a análise de sentimentos multimodal integra um processo que combina os resultados de análise de vídeo, áudio, texto, com expansão para novos tipos de dados, ou combinação entre eles, gerando uma única classificação multimodal. Já a parte do modelo para visualização de dados considera o tipo de dado disponível para análise para determinar a técnica de visualização a ser utilizada. Uma parte importante da abstração do modelo são as etapas de pré-processamento tanto para as técnicas de IA quanto visualização, desta forma o usuário não especialista nessas áreas apenas precisa informar os dados e analisar as visualizações geradas.

A implementação do modelo proposto gerou uma ferramenta computacional para criação e análise de dados de sentimentos, a UXmood foram consideradas para as tarefas de análise de sentimento o uso de arquitetura de redes neurais convolucionais unimodais para vídeo, áudio e texto, por já apresentarem bons resultados iniciais nesta área, mas ainda pouco explorada, e uma abordagem de fusão por nível de decisão para integrar o resultado dos vários classificadores. Para análise dos dados gerados foram consideradas

técnicas de visualização de dados multidimensionais, podendo também representar relações hierárquicas e temporais de acordo com os dados do domínio do problema. A representação visual principal para dados de sentimentos foram cores e símbolos (emojis), e foi considerada a utilização de múltiplas visões de dados coordenadas entre si para diminuir a complexidade das visualizações.

Como cenário de utilização, foram utilizados dados reais relativos a um processo de avaliação de usabilidade e experiência do usuário na interação com um software de visualização multidimensional de dados. A escolha deste cenário de uso é motivada pelas pesquisas recentes de Dove et al. (2017) e Yang et al. (2018), que apontam que pesquisadores e praticantes de avaliações de Experiência do Usuário (UX) possuem pouca ou nenhuma experiência na utilização de técnicas de IA ou entendem as capacidades e limitações de dessas técnicas para os seus cenários de análise, o que dificultaria o uso de dados sobre análise de sentimentos, ou demandaria a necessidade de usuários especialista em IA e visualização. Pesquisas recentes em avaliações de UX apontam o uso de dados de análise de sentimentos como fundamental para um melhor entendimento do comportamento do usuário durante uma atividade (DOVE et al., 2017; YANG et al., 2018).

A ferramenta UXmood desenvolvida para este cenário de uso foi denominada UXmood, que realiza a análise de sentimentos combinando dados de vídeo, áudio, texto em abordagem multimodal, além de associar esses dados gerados aos dados de logs de interação do mouse, rastreador de olhos e a fala do usuário durante as tarefas que foram realizadas pelos participantes do teste, apresentando-os em um painel configurável de visualizações de dados coordenadas entre si para plataforma Web.

Para validar a acurácia das informações geradas, os cenários de análise e a facilidade de uso da ferramenta foram realizadas as seguintes etapas: 1 - foram testadas e comparadas técnicas do estado da arte de aprendizado máquina para análise de sentimentos para vídeo, áudio e texto, de forma individual e, posteriormente, aplicadas a uma abordagem multimodal baseada em pesos (fusão por nível de decisão); 2- foram desenvolvidos cenários de visualização de dados e avaliadas através de questionários on-line, com o objetivo de identificar os principais problemas de usabilidade; 3 – após realizar o redesign da ferramenta com os resultados da etapa 2, a ferramenta foi testada junto a usuários que já haviam realizado, no mínimo, um processo de avaliação de usabilidade ou experiência do usuário. Por fim, os resultados da aplicação de técnicas de aprendizado de máquina e abordagem multimodal se mostraram compatíveis com a literatura, a utilização por especialistas apresentou boa aceitação da ferramenta e obteve ótima avaliação na facilidade de uso para geração e visualização dos dados de sentimentos.

1.1 Justificativa e Motivação

A utilização de dados de análise de sentimentos tem sido aplicada em vários contextos, tais como: sistemas de recomendações, análise de opiniões do público, predição do comportamento de pessoas, etc. (ZADEH et al., 2017). Para relacionar dados de sentimentos ao conjunto de dados de um domínio de problema são utilizadas técnicas inteligentes.

Além disso, a área de análise de sentimentos apresentou um forte crescimento nos últimos 10 anos (PORIA; HUSSAIN; CAMBRIA, 2018; KAUR; KAUTISH, 2019; KUCHER; PARADIS; KERREN, 2018a), principalmente em um função do desenvolvimento tecnológico e de novas técnicas de IA, como as redes neurais de aprendizado profundo, ainda com grande potencial de investigação nesta área. Apesar de já apresentar alguns bons resultados, a utilização e a interpretação dos resultados da aplicação dessas técnicas de IA nem sempre é um processo fácil (PORIA; HUSSAIN; CAMBRIA, 2018; HUSSEIN, 2018).

Para análise e interpretação dos dados, visualização de dados dados devem ser utilizadas considerando dados integrados do domínio do problema e dados de classificação de sentimentos, apresentando desafios de alta dimensionalidade e grande quantidade de dados, representação de aspectos temporais e hierárquicos, múltiplas visões de dados, variedade de cenários, entre outras. De acordo com Kucher, Paradis e Kerren (2018a), de maneira geral esses requisitos não são atendidos em função da utilização de visualizações de dados simples como gráficos de barras e linhas.

Assim, a apresentação de um modelo, e posteriormente uma aplicação que implementa esse modelo, possibilitará a classificação de dados de sentimento, integração ao conjunto de dados do domínio do problema e a análise desses dados integrados através de visualizações de dados, o que proporcionará a utilização de dados de classificação de sentimento para vários tipos de dados e em vários contextos, aumentando as dimensões de análise do comportamento humano em diversas atividades.

1.2 Objetivos

A seguir serão apresentados os objetivos geral e específicos desta tese.

1.2.1 Objetivo Geral

Apresentar e implementar um modelo para integrar a classificação e visualização de dados de sentimentos.

1.2.2 Objetivos Específicos

- Criar um modelo para gerar classificações de análise de sentimentos multimodal e visualizar esses dados juntamente com outros dados do domínio do problema;
- Desenvolver uma aplicação que suporte o modelo criado, contemplando a fases de pré-processamento, análise de sentimento e visualização de dados interativa;
- Avaliar a viabilidade da utilização de arquiteturas de redes neurais convolucionais para as tarefas de análise de sentimento multimodal, considerando vídeo, áudio e texto;
- Avaliar técnicas de visualização da informação multidimensionais para apresentação de dados de classificação de sentimentos e outros dados;
- Avaliar a integração e usabilidade da aplicação desenvolvida.

1.3 Contribuições

As principais contribuições deste trabalho são:

- Um modelo computacional para classificação e visualização de dados de sentimentos a outros dados.
- Aplicação para auxiliar e abstrair a complexidade de análise de sentimento dados e visualização desses dados.
- Uma nova técnica de visualização de dados baseada em sentimentos para representar e analisar frases.

1.3.1 Publicações Relacionadas com a Tese

DA SILVA FRANCO, Roberto Yuri et al. UXmood — A Sentiment Analysis and Information Visualization Tool to Support the Evaluation of Usability and User Experience. **Information**, v. 10, n. 12, p. 366, 2019. **Qualis: B2**

FRANCO, Roberto Yuri Da Silva et al. UXmood — A tool to investigate the user experience (UX) based on multimodal Sentiment analysis and information visualization (InfoVis). In: **2019 23rd International Conference Information Visualisation (IV)**. IEEE, 2019. p. 175-180. **Qualis: B1**

1.4 Organização do Trabalho

Esta tese está organizada nos seguintes capítulos: introdução, fundamentação teórica, trabalhos relacionados, modelo, implementação do protótipo do modelo, avaliação do protótipo, considerações finais e trabalhos futuros, e referências.

O capítulo de introdução apresentou a contextualização da tese, visão geral, a área a qual este trabalho abrange e apresenta, ainda, as subseções justificativa e motivação, objetivos geral e específicos, contribuições e publicações relacionada com a tese.

O segundo capítulo é a fundamentação teórica que aborda os conceitos fundamentais e relacionados a este trabalho. São eles: sentimentos, emoção e polaridade, computação afetiva, técnicas de análise de sentimentos, análise de sentimentos multimodal, aprendizado de máquina, e visualização de dados.

No terceiro capítulo são apresentados os trabalhos relacionados a esta pesquisa, sendo selecionados artigos que discorram acerca das ferramentas e métodos relacionados a essas diversas áreas e *surveys* nas áreas de visualização de dados, análise de sentimentos e análise de sentimentos multimodal e aprendizado de máquina.

No quarto capítulo evidenciou-se o novo modelo proposto para abstrair a complexidade do uso de análise de sentimentos multimodais e visualização de dados em análise de produtos.

No quinto capítulo é apresentada a implementação do modelo proposto intitulado de ferramenta UXmood e todas as suas características e funcionalidades, como análise de sentimentos em texto, áudio, vídeo e multimodal, além de um *dashboard* de visualizações de dados coordenadas e relacionadas temporalmente.

No sexto capítulo são apresentadas as avaliações que foram efetuadas na implementação do modelo para a validação do modelo. Como avaliações das representações visuais das técnicas de visualização de dados escolhidas e uma avaliação de usabilidade e experiência do usuário na ferramenta UXmood.

Por fim, no sétimo capítulo são apresentadas as considerações finais e trabalhos futuros, sendo, portanto, elementos estruturais deste trabalho, bem como as referências utilizadas para realizar esta pesquisa.

2 Fundamentação Teórica

Neste capítulo serão apresentados os principais conceitos utilizados nesta tese, relacionados, principalmente, às áreas de análise de sentimentos, técnicas de aprendizado de máquina aplicadas à análise de sentimentos e visualização de dados.

2.1 Análise de Sentimentos

Um conceito fundamental para esta tese é a análise de sentimentos multimodal. De maneira geral, a análise de sentimentos multimodal é uma área que estuda a classificação de sentimentos a partir de diversos tipos de dados como vídeo, áudio, texto, imagem, dados fisiológicos, entre outros. Para abordar esse conceito é importante apresentar outros conceitos como: sentimentos, emoções, polaridade, computação afetiva, técnicas de análise de sentimentos e reconhecimento de emoções. Por fim, uma definição mais detalhada de análise de sentimentos multimodal é apresentada.

2.1.1 Sentimentos, Emoções e Polaridade

Os conceitos de sentimentos e emoções são herdados da psicologia, porém, para a área da Computação, estes possuem abordagens diferentes. Munezero et al. (2014) De acordo com (Munezero et al., 2014), as definições de emoções são muitas e variadas na literatura que podem gerar experiências afetivas, gerar processo cognitivo, ajustes fisiológicos ou comportamento expressivo. Scherer et al. (2000) afirma que o excesso de definições de emoções é causado pelo dificuldade em distinguir claramente entre as diferentes classes de fenômenos diferentes.

Gordon (2017) define sentimentos como: ”padrões de sensações socialmente construídos, gestos expressivos e significados culturais organizados em torno de um relacionamento com um objeto social, outra pessoa, grupo, etc”. Em concordância, Murray e Morgan (1945) apresentam sentimento como ”uma disposição mais ou menos duradoura (predileção ou prontidão) em uma personalidade para responder com um efeito positivo ou negativo a uma entidade especificada.” Broad (1954) explicou que um sentimento é formado quando um determinado objeto é constantemente percebido ou pensado por uma pessoa e, com o tempo, ela cria uma ideia disposicional em relação ao objeto. De acordo com Mohammad e Turney (2013), o sentimento pode ser definido como uma resposta orgânica de duração e intensidade variadas, que pode ser analisado para gerar resultados como emoções e polaridade.

Destacam-se os conceitos de sentimentos, emoções e polaridade definidos nos

trabalhos de Cattell (1940), Mohammad e Turney (2013), Gordon (2017), em que, de maneira geral, sentimentos representam "uma disposição neuropsíquica principal, adquirida e relativamente permanente de reagir de modo emocional, cognitiva e intencional em relação a um determinado objeto (ou situação)." As reações neuropsíquicas de uma pessoa em relação aos sentimentos pode ser expressa a partir de alterações fisiológicas como expressão facial, tom, timbre e volume da fala ou som, etc Friedenberg e Silverman (2011).

Em concordância com essa definição, Ekman (1999) conceitua a existência de emoções universais denominada de emoções básicas, e define que estas podem ser reconhecidas a partir das micro-expressões faciais, além de classificar as emoções básicas em alegria, tristeza, raiva, medo, surpresa, nojo e desprezo.

Outro conceito importante nessa tese é o de polaridade definido por Benevenuto, Ribeiro e Araújo (2015), em que polaridade pode ser usado como saída para classificadores de análise de sentimentos e representa o grau de positividade ou negatividade de um dado. Alguns autores tratam a polaridade como um conjunto binário (positivo e negativo) ou ternário (positivo, negativo e neutro). (PANG; LEE, 2004)

2.1.2 Computação Afetiva

Picard (2000) define Computação Afetiva (CA) como a área que congrega as pesquisas da aplicabilidade da afetividade em sistemas não-biológicos, ou seja, o uso dos sentimentos em diferentes aspectos nos sistemas computacionais (personalização de interfaces, simulação de sistemas com emoções, pesquisas que envolvem sentimentos nas interações Homem Computador).

Assim tornando o uso dos sistemas mais personalizado para cada usuário e permitindo a extração de informações subjetivas como as emoções humanas para diferentes usos nos setores comerciais ou nas pesquisas (PICARD, 2003).

Sistemas afetivos contam com tecnologias de Inteligência Artificial com elementos de Aprendizado de Máquina (AM) e Processamento de Linguagem Natural (NLP) para detectar, reconhecer, interpretar e processar os sentimentos, além de gerar respostas adequadas às emoções de seus usuários, sintetizando comportamentos humanos com empatia. Tudo isso podendo ser viabilizado por meio de várias mídias em vídeo, áudio e texto. (FORTIN-CÔTÉ et al., 2019)

Computação afetiva é a humanização da tecnologia, combinando várias áreas multidisciplinares e interdisciplinares como a ciência computacional, a engenharia, a ciência cognitiva dentre outras. Assim, construindo-se modelos computacionais, ou um robô, é possível ser capaz de entender o estado emocional do ser humano ao qual está interagindo e se adaptar, respondendo de maneira apropriada de acordo com as diferentes circunstâncias identificadas.

Segundo Poria, Hussain e Cambria (2018), a identificação das pistas emocionais a partir de dados possui diferentes modalidades e granularidade. Por exemplo, a classificação de polaridade têm uma alta granularidade, pois é um conjunto ternário e o reconhecimento de emoções possui uma baixa granularidade, pois possui um conjunto octal.

2.1.3 Técnicas de Análise de Sentimentos e Reconhecimento de Emoções

A Análise de Sentimentos é uma área que estuda a classificação do sentimento das pessoas por meio de dados para identificar o estado emocional de uma pessoa em um determinado momento (PANG; LEE, 2004). Existem várias abordagens para a análise de sentimentos. As primeiras que surgiram realizavam a classificação do texto por meio do uso de dicionários que mediam diretamente a valência, a atenção e a dominância associados à reação afetiva (BRADLEY; LANG, 1999). No entanto, as abordagens mais modernas de análise de sentimentos se concentram no uso de técnicas automatizadas para extrair informações subjetivas (LIU et al., 2010) de dados de diferentes tipos, como texto, áudio e vídeo. Quando a abordagem usa mais de um tipo de dados é possível inferir o sentimento utilizando um único classificador ou um classificador para cada tipo de dado, sendo chamada de abordagem multimodal.

Ao inferir sentimentos a partir de dados de vídeo, a principal tarefa é o reconhecimento de expressões faciais, em que os classificadores tentam identificar expressões predefinidas (RUDOVIC; PAVLOVIC; PANTIC, 2014). Segundo Poria, Hussain e Cambria (2018), as expressões desempenham um papel significativo na identificação de emoções em um fluxo multimodal. Um reconhecedor de emoções em vídeo identifica automaticamente pistas emocionais associadas a expressões faciais e classifica essas expressões em categorias de sentimentos. Os reconhecedores usam técnicas de aprendizado de máquina das mais variadas que vão de extrair pontos característicos da face ao aprendizado das características.

No que se refere ao áudio, a análise de sentimentos das declarações faladas pode ser feita de duas maneiras: (1) pesquisando pistas emocionais usando estratégias de aprendizado de máquina para classificar sentimentos analisando propriedades sonoras como tom, volume e timbre, ou (2) analisando o conteúdo do discurso (SCHULLER et al., 2011). Na primeira, assim como no vídeo, pode ser feito o aprendizado das características do áudio ou a extração dessas características como pitch, volume, tom, timbre, MFCC, Spectral Centroid, Spectral Flux ou etc.

Técnicas de reconhecimento de emoções em texto podem ser usadas com dicionários lexicais (LIU, 2012) ou algoritmos de aprendizado de máquina treinados em análise semântica e estatística de palavras (MEDHAT; HASSAN; KORASHY, 2014). Segundo Benevenuto, Ribeiro e Araújo (2015), a análise de sentimentos pode ser por nível de sentença, onde o texto é subdividido em frases ou sentenças que pode conter pistas

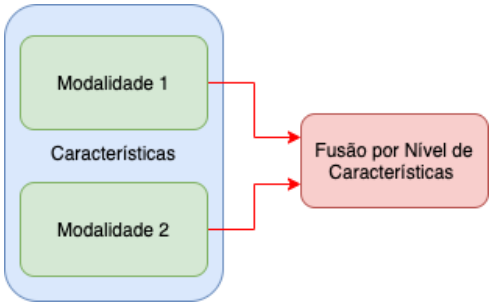
emocionais do autor ou nível de palavra, onde, de maneira geral, são utilizados dicionários para se identificar o sentimento de cada palavra.

2.1.4 Análise de Sentimentos Multimodal

Segundo Wagner et al. (2011), as principais abordagem de análise de sentimentos multimodais, onde são usados vários tipos de mídias (vídeo, áudio, texto, etc), são por fusão em nível de características, onde o classificador analisa em conjunto parâmetros dos múltiplos arquivos, ou por fusão em nível de decisão, onde é aproveitado a redundância dos classificadores unimodais para alcançar maior robustez através da combinação de seus resultados (PLANET; IRIONDO, 2012).

A fusão por nível de características utiliza a estratégia de extração das características de duas ou mais modalidades, obtendo o vetor de características por modalidade, e fazendo a fusão dos vetores para então efetuar a classificação. Segundo Poria, Hussain e Cambria (2018), a principal motivação para o desenvolvimento dessas estruturas é executar o processo de fusão em menos tempo. A Figura 1 apresenta a representação da fusão por nível de características.

Figura 1. Representação do método de fusão por nível de características

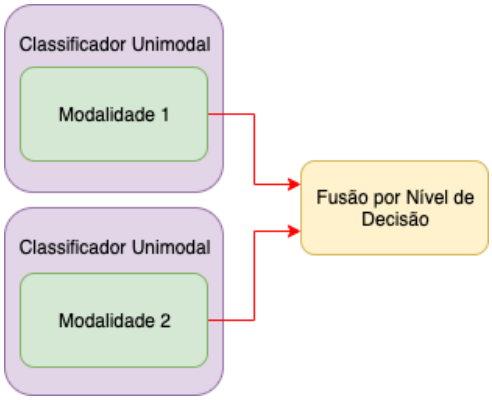


FONTE: O autor (2019)

Na fusão no nível de decisão, os vetores de características da modalidade utilizam um classificador de análise de sentimentos exclusivo para a classificação unimodal. Para então, efetuar a classificação por aplicação de pesos, modelos estatísticos ou um novo classificador. A classificação por pesos usa classificadores separados para cada modalidade em vez de concatenar vetores de características como na fusão no nível de características. A saída de cada classificador pode ser uma pontuação de classificação que pode ser probabilidade para cada classe de sentimento, duração, etc (Wagner et al., 2011). A Figura 2 apresenta a representação da fusão por nível de de decisão.

Ambas abordagens possuem vantagens e desvantagens como ganho de tempo de treinamento de modelos (fusão por nível de características), maior robustez (fusão por nível de decisão), etc. A literatura não descreve a melhor solução, as duas são frequentemente

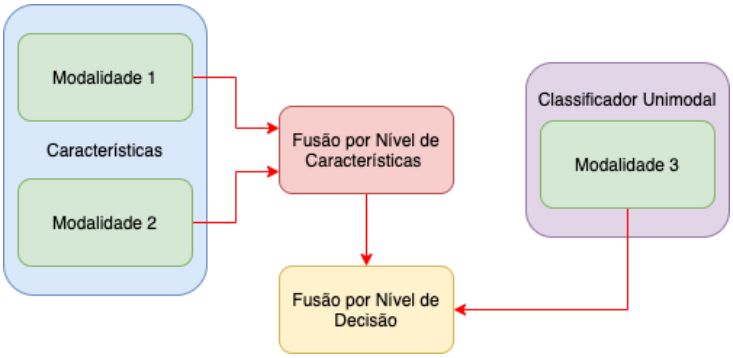
Figura 2. Representação do método de fusão por nível de decisão



FONTE: O autor (2019)

usadas e apresentam bons resultados independente dos tipos de dados e arquiteturas utilizadas. (ZADEH et al., 2017)

Figura 3. Representação do método híbrido de fusão



FONTE: Adaptado de Poria, Hussain e Cambria (2018)

Além disso, ainda há as abordagens híbridas que combinam tanto a fusão no nível de decisão quanto por características. A Figura 3 uma proposta híbrida de fusão, as partes verdes na Figura 3 representam tipos de dados, por exemplo, vídeo, áudio, texto, etc. A parte azul representa a extração das características e os vetores de características para ambas modalidades e a parte vermelha representa o classificador de fusão por nível de características. A parte amarela recebe a saída dos classificadores de cada modalidade e efetua a fusão por nível de decisão.

2.2 Aprendizado de Máquina

Uma forma comum de resolver o problema de análise de sentimentos é utilizar formas naturais de expressão do usuário, seja em texto, áudio ou imagem, e buscar padrões de forma automática para identificar um sentimento (PORIA et al., 2017). Esse cenário,

em que existe uma entrada x, representando as expressões de sentimento do usuário, que pode ser utilizada de maneira $y(x)$ para identificar um sentimento y é um típico problema de aprendizado de máquina, e é utilizado como principal abordagem para problemas de análise de sentimentos.

Aprendizado de máquina é o conjunto de técnicas utilizado para realizar tarefas de modo automático sem instruções explícitas, aprendendo através da experiência (GOODFELLOW; BENGIO; COURVILLE, 2016). É uma área da Inteligência Artificial (IA) em que um mapeamento $y(x)$ é realizado de forma implícita, através de experiência, avaliado por uma métrica, que é o que o algoritmo de aprendizado deve utilizar para ajustar a função. A experiência comumente é adquirida em forma de um conjunto de dados X, dada uma certa tarefa. Essa etapa em que o algoritmo ajusta a função em relação a entrada de acordo com uma métrica é chamado de treinamento (BISHOP, 2006). Durante o treino, um algoritmo pode ver o conjunto de dados dezenas ou centenas de vezes para ajustar os parâmetros da função. O mapeamento $y(x)$, também expressado como $y(x, w)$, onde w são os parâmetros da função aprendidos no treinamento, é chamado de modelo, que neste trabalho será denominado modelo inteligente.

Um dos objetivos principais em aprendizado de máquina é treinar modelos inteligentes que consigam categorizar corretamente novas entradas, que não estavam presentes no treinamento, como um ser humano que consegue identificar várias categorias a partir de pouca informação. Essa característica é chamada generalização, e para várias tarefas cognitivas os modelos inteligentes do estado da arte já ultrapassam o desempenho humano (KARPATHY, 2014).

Para alguns problemas e algoritmos, antes do treino é necessário realizar manipulação dos dados de entrada, seja para extrair ou gerar características, seja para normalização para entrada na fase de treino. Essa fase é chamada de pré-processamento, e é crucial para reduzir a variabilidade das entradas, tornando a etapa de treino mais fácil, impedindo que características indesejadas sejam generalizadas. Essa etapa também pode ser adicionada como forma de acelerar a computação, uma vez que alguns aspectos dos dados podem ser selecionados para o processamento, e outros com pouca importância sejam descartados. Adicionalmente, transformações podem ser feitas nos dados a fim de extrair características relevantes para o problema, ao invés de processar a entrada completa. A Figura 4 apresenta um processo genérico para uma solução de aprendizado de máquina, compreendendo todas as fases descritas até o momento.

Para aprendizado supervisionado, a métrica associada aos algoritmos utilizados são expressas através de uma função objetivo, $L(y', y)$, também conhecida como função *loss*, que orienta a função $y(x, w)$ ao aprendizado dos parâmetros w, através da computação da distância ou diferença das classes y' em relação a predição y. A função objetivo é essencial para o desenvolvimento de algoritmos e sua minimização em relação a $y(x, w)$ é parte

Figura 4. Processo genérico para aprendizado de um modelo inteligente de aprendizado de máquina. As etapas podem ser adaptadas dependendo do problema e dos algoritmos utilizados.

FONTE: O autor

central do estudo de aprendizado de máquina (BISHOP, 2006).

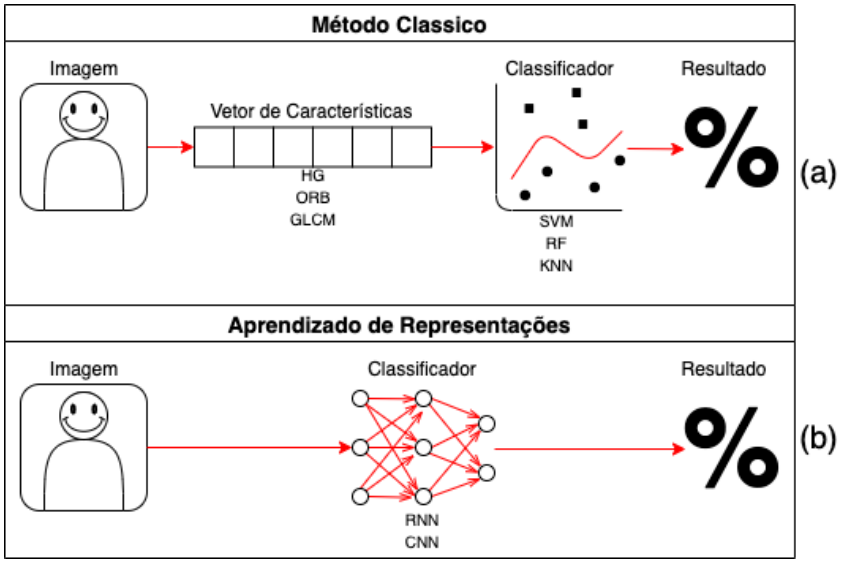
Algoritmos de aprendizado supervisionado são o tipo de algoritmos utilizados em análise de sentimentos, em que a entrada (X, Y) contém imagens, texto ou áudio como entrada x e sentimento y associado a essa entrada. A quantidade de dados e métodos disponíveis tornam essa abordagem eficiente para o problema (GOODFELLOW; BENGIO; COURVILLE, 2016), apesar da complexidade. Com métodos de pré-processamento para as entradas, seja extraindo características ou normalizando valores, vários trabalhos conseguem utilizar modelos inteligentes em várias aplicações de análise de sentimentos (MUKESHIMANA et al., 2017).

Para problemas de aprendizado supervisionado baseado em automação cognitiva, como identificar categorias em imagens, áudio e texto, a quantidade de dimensões e variações melhores resultados são obtidos através de redes neurais, para classificação e outras tarefas associadas (GOODFELLOW; BENGIO; COURVILLE, 2016), e a seguir apresenta o estado da arte para o problema.

Tradicionalmente as abordagens para aprendizado supervisionado dependem de uma seleção cuidadosa de características dado um problema. Um vetor de características é gerado através de um extrator de características, e esses atributos são utilizados por modelos inteligentes, tirando a dificuldade de processar os dados brutos. Por outro lado, construir esses extratores de características exige muito conhecimento do domínio do problema, normalmente adquirido por meio de um especialista, além do trabalho de criar a computação que traduz esse conhecimento para meios digitais. Para evitar que essa representação seja explicitamente declarada, o aprendizado de representações é um conjunto de métodos que permite que um algoritmo aprenda as características de uma certa entrada e automaticamente descubra características para classificação (BENGIO; COURVILLE; VINCENT, 2013).

Com um grande volume de dados, modelos inteligentes podem aprender além da intrínseca relação entre as características de um conjunto de dados, também como

Figura 5. Classificação de imagens seguindo duas abordagens, (a) a abordagem clássica, com extração de características e classificadores tradicionais. (b) A abordagem de aprendizado de representações, com as características sendo aprendidas durante o treino.



FONTE: O autor

extrair essas características. A partir desse cenário, aprendizado profundo (*Deep Learning*) através de Redes Neurais Profundas (*Deep Neural Networks* - DNN) apresenta os maiores avanços já vistos em aprendizado de máquina, vencendo antigos desafios da comunidade de IA (LECUN; BENGIO; HINTON, 2015). Métodos de aprendizado profundo são métodos de aprendizado de representações, obtidas compondo módulos simples não lineares que transformam uma representação em um nível (começando pela entrada) em outra representação em um nível maior, mais abstrato. A composição dessas transformações permite que uma função $y(x)$ aprenda representações complexas (BENGIO; COURVILLE; VINCENT, 2013). Tarefas de Visão Computacional mostram a eficiência do aprendizado de representações através de aprendizado profundo em relação aos métodos clássicos. A Figura 5 apresenta a comparação dessas abordagens para reconhecimento de imagens.

Para classificação de imagens, os métodos clássicos utilizam extratores de características em conjunto com um classificador. O classificador é treinado utilizando um vetor de características da imagem, e o classificador não tem acesso a entrada original, apenas a representação da mesma através da extração de características. Mesmo com a robustez desses extratores de características, informações intrínsecas sobre a imagem e domínio espacial são perdidas se não forem explicitamente extraídas. Uma CNN (Rede Neural Convolucional - *Convolutional Neural Network*) é uma abordagem de aprendizado profundo que utiliza filtros em operações de convolução com parâmetros aprendidos por meio de um conjunto de dados (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Esses filtros são organizados em camadas, em que o modelo inteligente aprende representações

mais abstratas conforme avança na rede.

Para o problema de análise de sentimentos essa abordagem de aprendizado de representações é necessária, já que a expressão de sentimentos por humanos é complexa, e criar um extrator de características para essas expressões pode não ser possível para todas as entradas (LECUN; BENGIO; HINTON, 2015), e também é a abordagem com melhores resultados para esse problema (PORIA et al., 2017; MUKESHIMANA et al., 2017).

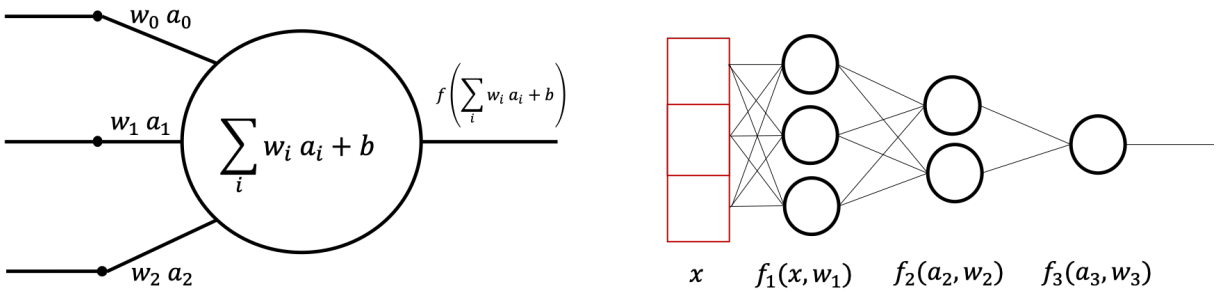
O bloco básico de construção de um modelo de aprendizado profundo é uma rede neural. Uma rede neural é um tipo de modelo inteligente vagamente baseado no cérebro animal, com unidades conectadas (como neurônios) que realizam processamento não linear através de uma função de ativação (como sinapses), com conexões baseadas em parâmetros ponderados, chamados de pesos. Esses neurônios formam uma rede organizada em camadas, que processa uma entrada e a partir dela produz uma inferência sobre classe ou valor. Essas camadas intermediárias de neurônios são chamadas camadas escondidas. O resultado da otimização da função $L(y', y)$ é expressado na função $y(x, w)$, que compreende todos os pesos por camadas. Essas camadas são responsáveis por distorcer o espaço de busca da função e facilitar a separação nos hiperplanos.

O neurônio de uma rede neural pode ser descrito pela função $f(\sum_i w_i \times a_i + b)$ onde a é a resultado do processamento da camada anterior, chamado ativação (na primeira camada $a = x$), i indica o índice da camada (i_0 é a entrada), b é o bias, um fator aditivo e f é uma função de ativação, que aplica não linearidade, permitindo a construção das distorções do espaço de busca. A função $y(x, w)$ pode ser descrita em função das camadas da rede, como por exemplo, em uma rede de três camadas $y(x, w) = f_3(f_2(f_1(x, w_1), w_2), w_3)$, com a divisão dos pesos em suas respectivas camadas. Experimentos modernos indicam ReLU como uma função de ativação com bons resultados (LI; YUAN, 2017), e normalmente é utilizada nas camadas escondidas, enquanto que a função final depende do problema. A ligação dessas camadas, juntamente com o número de neurônios e suas ligações é chamado de arquitetura. A Figura 6 apresenta a arquitetura dessa rede. Esse tipo de rede tem boa generalização para problemas práticos que podem ser expressados através de um vetor de características unidimensional sem relação espacial.

Para imagens e áudio, onde além dos valores a relação espacial também é importante no problema, o uso de CNNs é comum, aproveitando a operação de convolução para criar representações robustas. Esse tipo de rede neural tem como característica a presença da operação de convolução no lugar dos pesos conectados, onde a operação é aplicada e os pesos aprendidos são para aplicação na operação, não diretamente nas representações apreendidas. CNNs foram o primeiro tipo de DNN que alcançou resultados além do estado da arte em um antigo problema de Visão Computacional, o ILSVRC (RUSSAKOVSKY et al., 2015).

A convolução é uma operação em duas funções $f(x)$ e $g(x)$ gerando uma terceira

Figura 6. Rede Neural e seus componentes. No lado esquerdo, o neurônio e sua descrição formal. No lado direito, a rede completa, onde cada neurônio (círculo) segue a forma do lado esquerdo.



FONTE: Adaptado de Karpathy (2017)

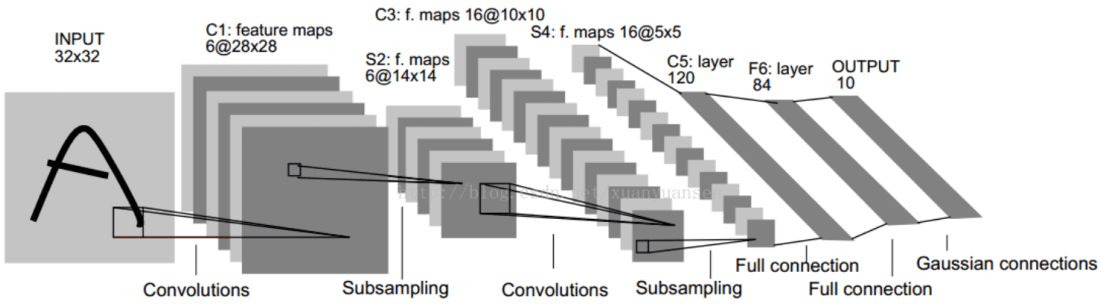
função $h(x)$ que expressa a forma de $f(x)$ em $g(x)$, e é definida em 1D ou 2D no contexto de DNNs do mesmo jeito que em processamento de sinal. Essas funções são o sinal de áudio ou imagem ($f(x)$), e os filtros aprendidos ($g(x)$). Essa operação tem três grandes vantagens, duas em relação a representação, mantendo a espacialidade das características e torna-as invariante à translação, e relação a compartilhamento de parâmetros, já que a conectividade dos filtros em relação às imagens e características intermediárias é menor que a conectividade completa dos pesos entre uma camada e outra (GOODFELLOW; BENGIO; COURVILLE, 2016).

Além da convolução, a operação de *pooling* ajusta a escala entre uma camada e outra para melhorar o desempenho da computação. Ao fim da rede, camadas totalmente conectadas são utilizadas como forma de ajustar os parâmetros para a função objetivo. Uma arquitetura popular, a primeira CNN, presente no trabalho de (LECUN et al., 1998) é apresentada na Figura 7 como forma de ilustração dos conceitos.

Para aprender representações complexas e suficientes para um problema, uma rede neural deve utilizar um grande volume de parâmetros, e arquiteturas do estado da arte chegam a ter milhões deles (SIMONYAN; ZISSERMAN, 2014) (SZEGEDY et al., 2017) (VINYALS et al., 2019). Essas redes ajustam esses parâmetros com um método eficiente de minimização de erro, o algoritmo de retro-propagação (*backpropagation*) (LECUN; BENGIO; HINTON, 2015).

A aplicação mais comum do retro-propagação é modificando o regime de treinamento, o conjunto de dados é dividido em lotes e cada lote tem seu gradiente computado de forma separada, sendo adicionado iterativamente. Desse modo nenhuma entrada é sobreposta por entradas que ativam mais a rede, impedindo o aprendizado por todos os itens do conjunto de dados, ao mesmo tempo que garante que itens que a rede não consiga

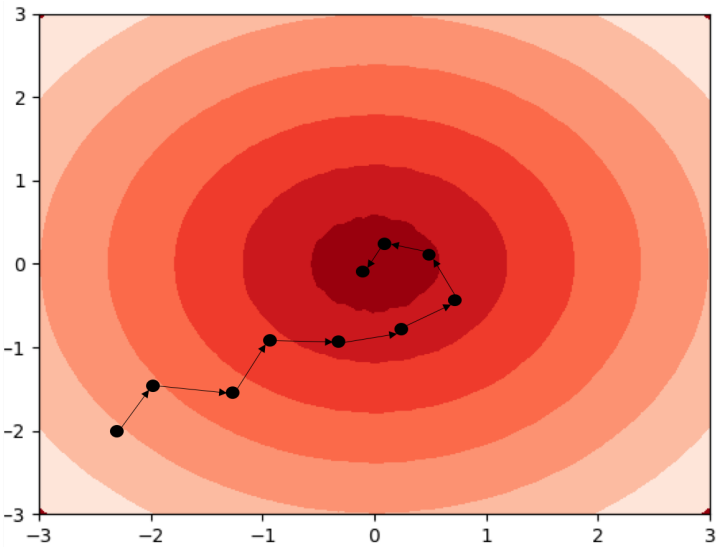
Figura 7. Arquitetura da LeNet, primeira CNN publicada. As camadas de convolução são apresentadas, assim como camadas de pooling e ao final as camadas totalmente conectadas. Os planos são os mapas de ativação entre as camadas.



FONTE: LeCun et al. (1998)

generalizar atrapalhe a generalização do resto dos dados. Cada vez que o conjunto de dados for processado pela rede e os gradientes fluírem de volta se alcança uma época. Esse processo é repetido por várias épocas até a generalização desejada. Esse método é chamado de método estocástico do gradiente (BOTTOU, 2010) e por várias vezes é só chamado de método do gradiente. A Figura 8 apresenta o espaço de busca em uma função e como o método do gradiente realiza a busca nesse espaço.

Figura 8. Exemplo de aplicação do método do gradiente.



FONTE: O autor

O objetivo do método é a partir do ponto inicial chegar no mínimo global da função. Através desses pequenos passos, indicado pelas setas e pontos intermediários, garante-se que a cada época se esteja mais perto do objetivo.

Por fim, como os parâmetros são atualizados automaticamente, resta escolher os hiper-parâmetros, indicando como a rede deve se comportar durante o treino. Na Figura 8,

o tamanho do passo entre um ponto e outro é indicado por um hiper-parâmetro chamado taxa de aprendizado, que indica o quanto deve ser aprendido em cada lote. Regularização do modelo, aceleração e velocidade também podem ser ajustados. O ajuste e seleção desses hiper-parâmetros é uma área ativa de pesquisa, com várias recomendações e análises que serão abordadas conforme se avança nesse trabalho.

2.3 Visualização de Dados

A visualização de dados é uma área que estuda a representações de dados abstratos em representações gráficas e interfaces de software que permitam ao usuário a manipulação de itens de dados visuais (PLAISANT, 2004).

Essas representações permite ao usuário fazer descobertas, tomar decisões e construir apresentações, sobre padrões (tendências, agrupamentos, intervalos, exceções, etc), grupos de itens e itens individuais. O uso de suporte computacional para visualizar e interagir com dados abstratos amplifica ou reforça a cognição humana, habilitando o usuário a ganhar conhecimento sobre os dados e o relacionamento entre eles, (CARD; MACKINLAY; SHNEIDERMAN, 2009) facilitando tarefas de pesquisa, análise, comunicação, comparação e exploração para ampliar a descoberta de padrões, outliers, e tendências (CARD; MACKINLAY; SHNEIDERMAN, 2009)

De acordo com Shneiderman e Plaisant (2006), uma ferramenta de visualização de informação deve fornecer ao usuário, pelo menos, as seguintes funcionalidades: visão geral dos dados, filtro, zoom, detalhes sob demanda, histórico de ações e extração de um subconjunto dos dados. Outros pesquisadores (LEE et al., 2006; SCHULZ et al., 2013; KERRACHER; KENNEDY; CHALMERS, 2015) apontam outras tarefas essenciais, como a configuração do layout de visualização, mapeamentos de dados visuais, classificação visual de itens, localização de valores, recuperação de valores (por exemplo, máximo e mínimo) e correlações entre itens de dados.

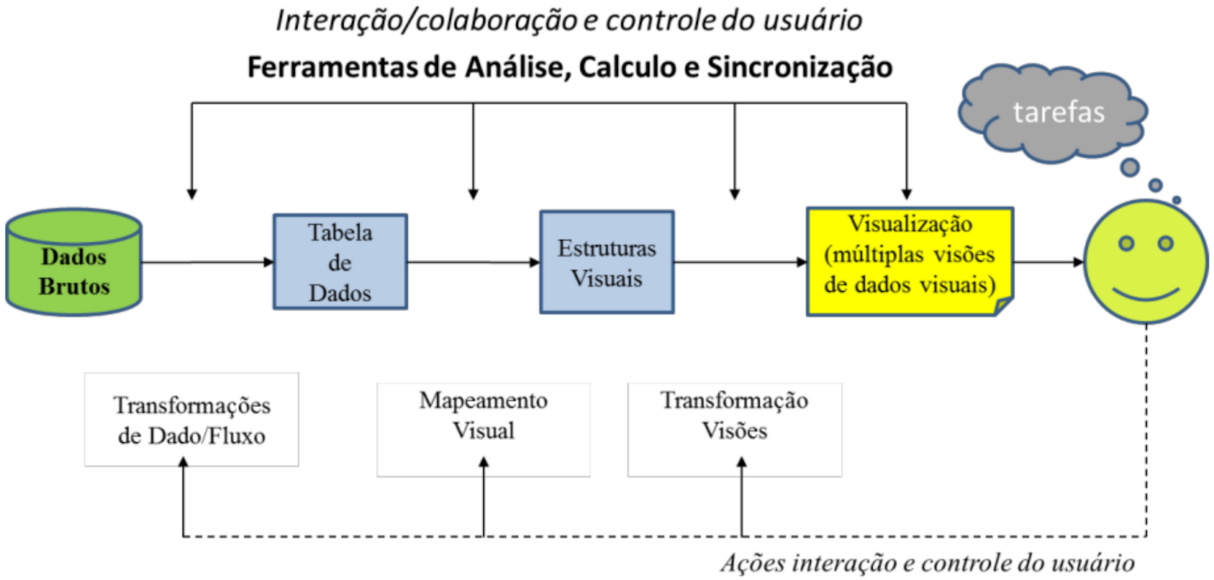
Um tipo de ferramenta de visualização de dados é o *dashboard*. De acordo com Few (2006), o *dashboard* é um painel de visualizações dos dados mais críticos necessários para alcançar um ou mais objetivos, geralmente organizados em uma única tela, e que podem ser analisados rapidamente sem muitas interações ou configurações.

2.3.1 Pipeline de visualização de dados

De maneira geral, o processo de visualização de dados faz parte de um processo de análise exploratória, descoberta de conhecimento ou análise visual. É importante que uma ferramenta de visualização de dados seja flexível o suficiente para se adaptar aos diversos perfis de usuário: experientes, iniciantes, com mais ou menos habilidades no domínio do

problema, até que os mesmos se sintam confortáveis para atingir seus objetos em relação a visualização proposta.

Figura 9. Processo genérico de geração de visualizações de dados.



FONTE: Adaptado de Ward, Grinstein e Keim (2015)

Segundo Ward, Grinstein e Keim (2015), a interação com uma ferramenta de visualização ocorre em diversos níveis e gera diversas visualizações intermediárias. A Figura 9 apresenta as principais etapas do processo / pipeline de visualização de dados e são descritas a seguir:

Modelagem de dados: necessidade de estruturar os dados que serão visualizados, em um padrão a ser definido, para permitir o uso de diversas fontes com layouts diferentes.

Seleção de Dados: identificar um subconjunto dos dados relevantes a serem visualizados, podendo ser realizado manualmente pelo usuário ou de forma automatizada através de métodos algorítmicos;

Mapeamento Visual dos dados: é a etapa em que os dados abstratos são vinculados a uma representação visual. Por exemplo, podem ser mapeados para características visuais, tais como: tamanho, posição, cor, forma, etc.;

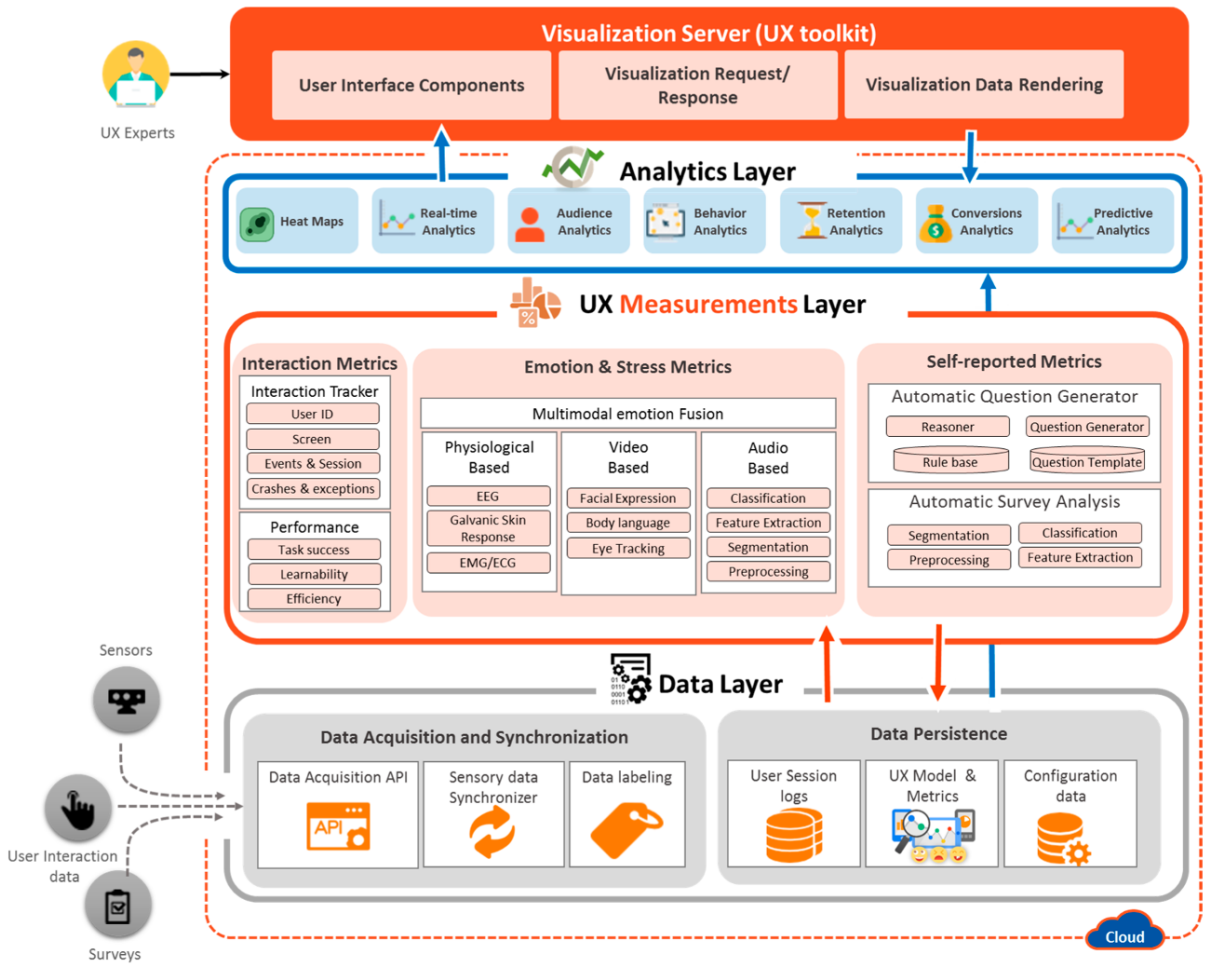
Configuração de parâmetros da cena (Transformação da Visão): a ferramenta deve permitir que o usuário possa especificar vários atributos da visualização que são relativamente independentes dos dados. Por exemplo: cor, som, luz, etc.;

Renderização ou geração da visualização: renderização da visualização é a representação da imagem no dispositivo de saída, como um monitor LCD, e que varia de acordo com o mapeamento visual que está sendo utilizado para dados. Além disso, a maioria das visualizações incluem informações complementares para facilitar a interpretação dos

dados, tais como: eixos, anotações, imagens, etc.;

Um exemplo prático do pipeline de visualização está presente no trabalho do Hussain et al. (2018) e apresentado no modelo descrito na Figura 10.

Figura 10. Processo de geração de visualizações de dados.



FONTE: Hussain et al. (2018)

No módulo de *Data Acquisition and Synchronization* este recebe os dados brutos para ser convertido em tabelas de dados no módulo de *Data Persistence*, além de realizar outros processamentos que não fazem parte do pipeline de visualização. O módulo de *Visualization Server* é responsável pelas etapas de mapeamento visual e renderização. Apesar do modelo apresentar as características de um pipeline de visualização, um ponto não tratado neste trabalho é a coordenação de dados entre as visualizações e a composição de um painel interativo, pontos que foram tratados nesta tese.

2.3.2 Visualização de sentimentos

Dentro da área de visualização de informação (ou dados) uma subárea importante para essa tese é a visualização de sentimentos, uma vez que esta tese utiliza métodos

combinados de dados de análises de sentimentos multimodais e técnicas de visualização de informação (ou dados).

A visualização de sentimentos é entendida como um desafio de pesquisa em visualização de informações (InfoVis) para analisar sentimentos descobertos em dados de texto, áudio, imagens, vídeo, entre outros. (KUCHER; KERREN, 2015) As aplicações e tarefas de visualização de sentimentos incluem, por exemplo, monitoramento da opinião pública nas mídias sociais, análise de literatura, apoio à pesquisa de sentimentos, postura em linguística, NLP, etc.

Alguns dos primeiros trabalhos que mencionam a visualização de sentimentos se originam na mineração de dados (DM) ou NLP e usam representações visuais básicas na maioria dos casos, como gráficos de barras, linhas ou pizza (KUCHER; PARADIS; KERREN, 2018b). Técnicas de visualização mais recentes geralmente refletem os avanços na área de visualizaão de informação e incorporam sentimentos em configurações complexas que envolvem dados heterogêneos com características temporais, relacionais e geoespaciais.

Kucher, Paradis e Kerren (2018b) lista as propriedades gerais e os desafios da visualização de sentimentos da seguinte maneira:

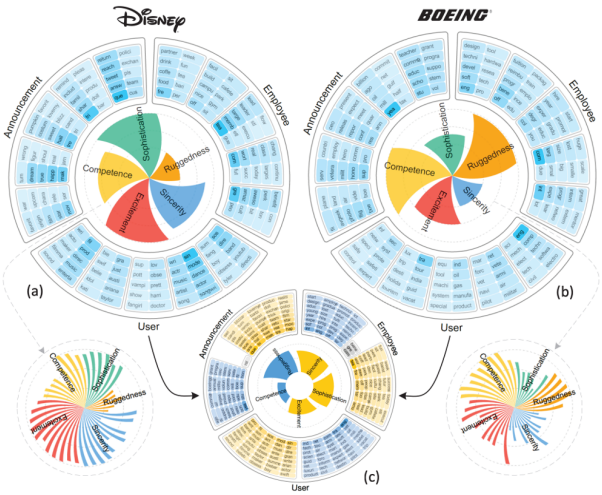
- a visualização de sentimentos abrange uma variedade de tarefas de análise de sentimentos, desde detecção de subjetividade até reconhecimento de emoções e análise de postura;
- as técnicas de visualização de sentimentos podem ter que usar dados específicos para o modelo de análise de sentimentos (por exemplo, baseado em léxico ou aprendizado de máquina) e escopo (nível de palavra, nível de expressão, etc.);
- A visualização de sentimentos reflete uma variedade de domínios de dados e tarefas do usuário existentes em pesquisas e aplicações de análise de sentimentos, que vão desde a pesquisa teórica em linguística e NLP até a mídia social e o monitoramento de notícias, implicando o uso de vários canais e representações visuais.

Para a análise de sentimentos, *dashboards* com metáforas de velocímetro, gráficos de linha e histogramas são frequentemente usados para representar informações quantitativas dos sentimentos, enquanto que *wordclouds* são usadas para representar a frequência e a qualidade da emoção ou polaridade. Outras técnicas menos usadas são radares para representar e comparar informações quantitativas, gráficos de dispersão para executar correlações, técnicas hierárquicas para agrupar e correlacionar dados e representações temporais.

Em geral, os conjuntos de dados de sentimentos são compostos de dados categóricos de emoção (raiva, medo, tristeza, alegria, surpresa, nojo, desprezo e neutro) ou polaridade

(positivo, negativo e neutro), normalmente representados por cores (Figura 11), e outros dados numéricos como tempo ou geo-espaciais referentes ao domínio do problema. Segundo Holtz e Healy (2018), técnicas de visualização como boxplot, pie chart, wordcloud, treemap e gantt chart são mais recomendados para esse caso.

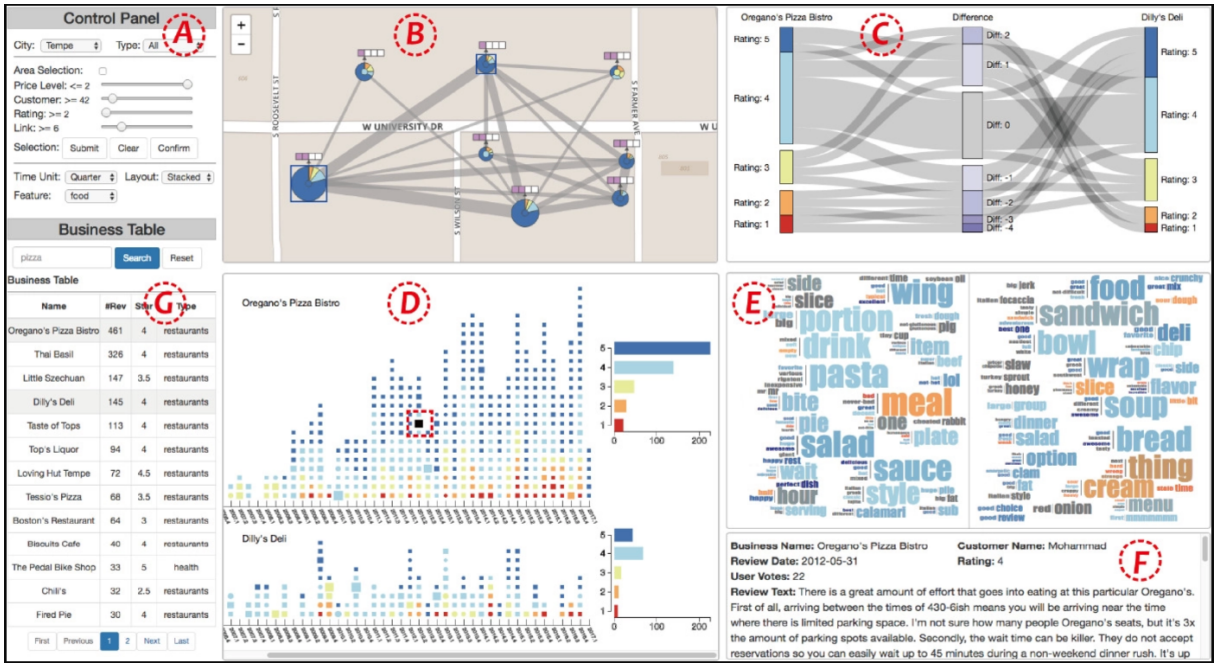
Figura 11. Representação visuais das percepções da marca nas mídias sociais.



FONTE: Liu et al. (2016)

Em casos, em que não há ordenamento do conjunto de dados como os *logs* numéricos de posicionamento, como rastreamento de olhos e *mouse*, as técnicas que relacionam dois valores numéricos são mais recomendadas, tais como: dispersão de dados, *violin*, *boxplot*, etc. Além disso, os trabalhos em análise de sentimento tem apresentado *dashboards* com múltiplas visões coordenadas, integrando dados de sentimento e dados do domínio do problema. A Figura 12 apresenta um exemplo de *dashboard* com múltiplas visões de dados coordenadas, incluindo representação de sentimento por cores, que permeia as visualizações, e utilizam as técnicas de grafo em B, matriz em D, conjuntos paralelos em C e *wordcloud* (nuvem de palavras) em E.

Figura 12. Visualizações mais modernas com base em múltiplas visões coordenadas.



FONTE: Wang et al. (2018)

3 Trabalhos Relacionados

A revisão dos trabalhos relacionados apresentou duas partes principais. A primeira relacionada visualização de dados de sentimento, e a segunda relacionada à análise de sentimentos multimodal.

3.1 Visualização de Dados

Em relação a visualização de dados de sentimento foram realizadas buscas nos sites do IEEE e ACM, com os termos "*visualization*" ou "*information visualization*" ou "*data visualization*" ou "*visual analytics*" e "*sentiment analysis*" ou "*emotion*" no resumo. Para o site do IEEE foram retornados 181 resultados, considerando o período de 2009-2019, e para site da ACM foram retornados 277 resultados, considerando o mesmo período. Foram lidos os *abstracts*, não foram considerados *posters*, artigos duplicados nas duas bases de dados, e o artigo deveria ter pelo menos uma visualização e a mesma deveria conter representação visual de sentimentos, emoções ou polarização. Como resultado final foram selecionados 51 artigos, serão apresentadas as análises de algumas categorizações, tais como: distribuição por ano, se apresenta múltiplas visões de dados coordenadas, tipos de técnicas de visualização utilizada, domínio da aplicação, tipo de sentimento, e se a visualização é 2D ou 3D. Além disso, foram adicionados 3 artigos selecionados do *survey* "*SentimentVis Browser* - A *Visual Survey of Sentiment Visualization Techniques*", bem como considerações do próprio *survey* citado, totalizando 55 artigos. Por fim, são apresentados de forma breve alguns trabalhos considerados representativo para cada uma das categorias.

A Figura 13 apresenta a distribuição por ano dos artigos selecionados. Os últimos 5 anos apresentou aproximadamente 75% das publicações envolvendo visualização de sentimentos, com destaque para os anos de 2018 e 2019, com aproximadamente 35% dos artigos selecionados, indicando que é um tema atual de pesquisa.

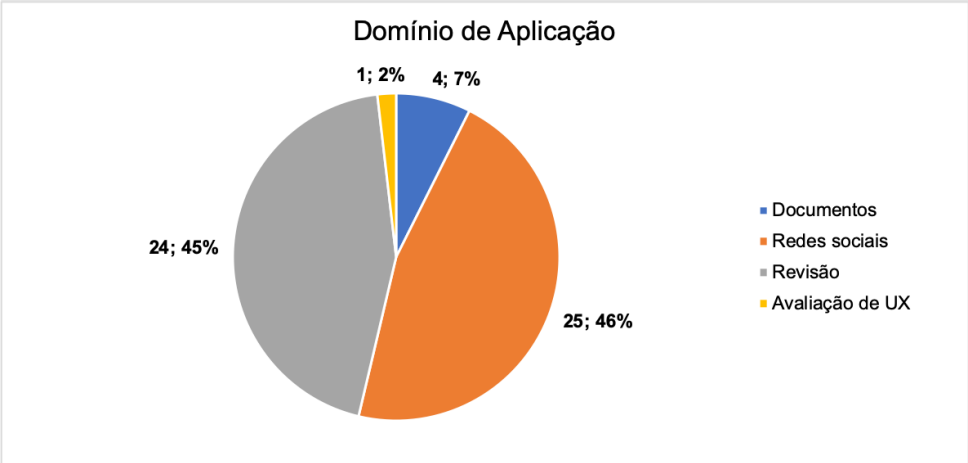
A Figura 14 apresenta os percentuais de área de aplicação dos artigos selecionados, com aproximadamente 46% voltados para revisão de produto, serviço ou marca, o que inclui a opinião de pessoas, 45% voltados para conteúdo de redes sociais, como texto de do Twitter, 7% análise de Documentos, e 1% voltado a avaliação de processo de Experiência do usuário, artigo originado desta tese.

Um exemplo revisão de produtos pode ser visto na Figura 15, que apresenta uma visualização em layout de matrix do uso de impressoras por usuários, avaliando determinadas características das mesmas. A cor azul representa comparativamente opiniões

Figura 13. Distribuição anual dos artigos selecionados.

FONTE: O autor

Figura 14. Percentual da área de aplicação dos artigos selecionados.

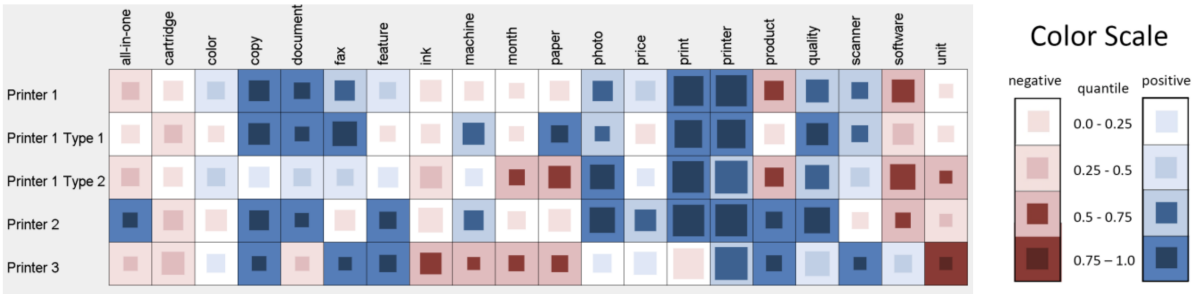


FONTE: O autor

positivas dos usuários e cores vermelhas comparativamente negativas (escala de cores), o tamanho do retângulo interno indica a quantidade de usuário que comentaram sobre um atributo. Quanto maior o retângulo, mais comentários são fornecidos pelos clientes.

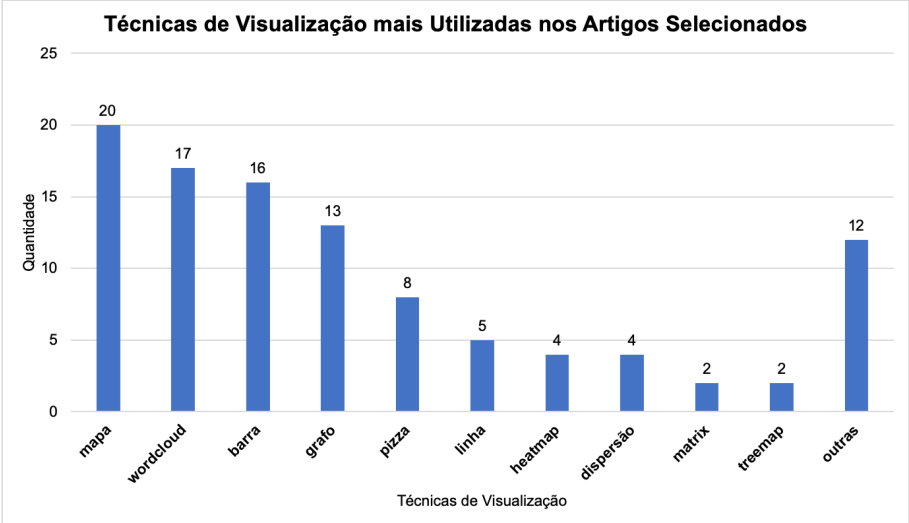
A Figura 16 apresenta as técnicas de visualização de dados mais utilizadas para apresentar dados de sentimentos nos artigos selecionados, mais de uma técnica pode ocorrer no mesmo artigo. A representação visual mais utilizada foi mapas, principalmente para representar dados de redes sociais com dados geo-espaciais, e podem até ser combinadas com outras técnicas, como *wordcloud*, gráfico de barras, grafo e pizza. Outras técnicas são utilizadas, mas com baixa ocorrência de uso. De maneira geral, as técnicas mais utilizadas são simples, com análise unidimensional ou bidimensional dos dados.

Figura 15. Visualização em layout de matrix do uso de impressoras por usuário.



FONTE: Oelke et al. (2009)

Figura 16. As técnicas de visualização de dados mais utilizadas para a representação visual de dados de sentimentos.



FONTE: O autor

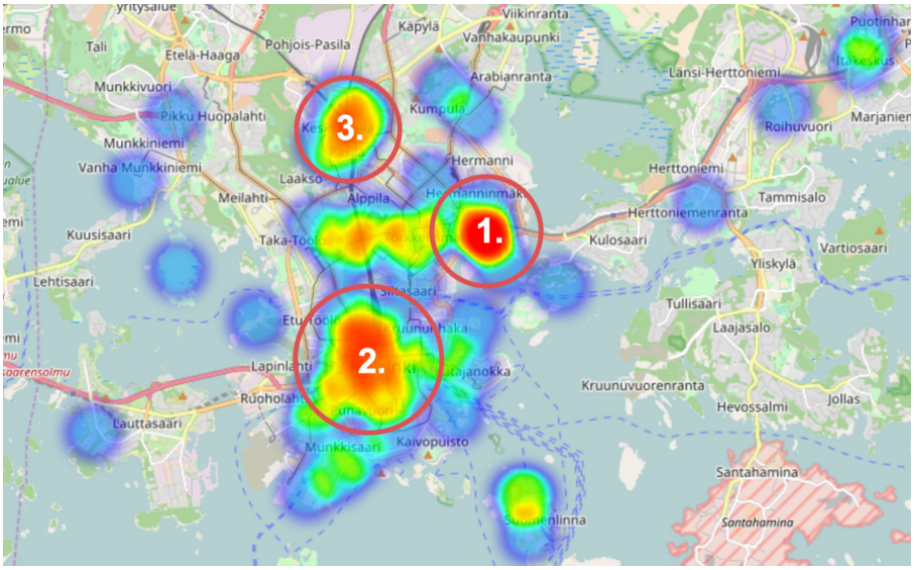
A Figura 17 apresenta a concentração de pontos de sentimento positivo em mapas, quanto mais vermelho maior a quantidade de mensagens de conteúdo positivo, dado uma determinada área geográfica para um período de tempo específico.

A Figura 18 apresenta que quase a totalidade dos trabalhos possui até três visualizações para representar os dados de sentimento e dados do domínio do problema, geralmente em cenários, tais como: mapa, barra e *wordcloud*; ou barra, linha e *wordcloud*; ou grafo, barra e *wordcloud*, entre outros. Contudo, destaca-se que aproximadamente 37% dos trabalhos apresenta apenas uma técnica de visualização de dados.

A Figura 19 apresenta uma única visualização integrando a os dados de sentimento representado por cores, os dados geo-espaciais através do uso do mapa, e informações de relações entre locais por uma representação de grafo. Quanto mais técnicas são representadas em conjunto, mas difícil será a análise por parte do usuário.

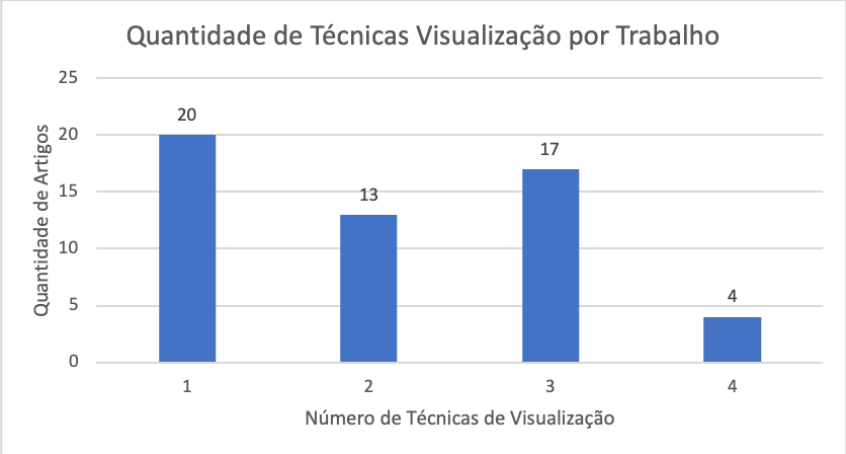
Figura 20 apresenta o percentual de aplicações, relacionadas aos artigos selecionados,

Figura 17. Mapa com a concentração de sentimentos positivos.



FONTE: Strengell e Sigg (2018)

Figura 18. Trabalhos com até três visualizações para representar os dados de sentimentos.



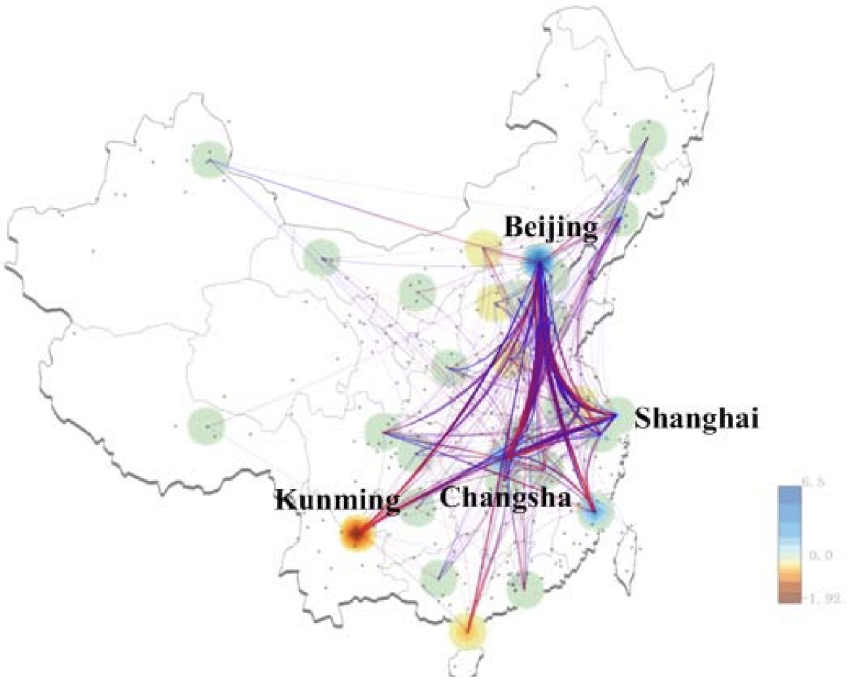
FONTE: O autor

que utilizam mais de uma visão de dados.

A Figura 21 mostra uma aplicação que apresenta dados de sentimentos junto com dados de domínio da aplicação, utilizando gráficos de barras (A), wordcloud (C), uma técnica para visualizar relacionamento conhecida como non-ribbon chord (B), e representação de conjuntos (D), e também estão relacionadas com o texto de redes sociais que deu origem a classificação (E).

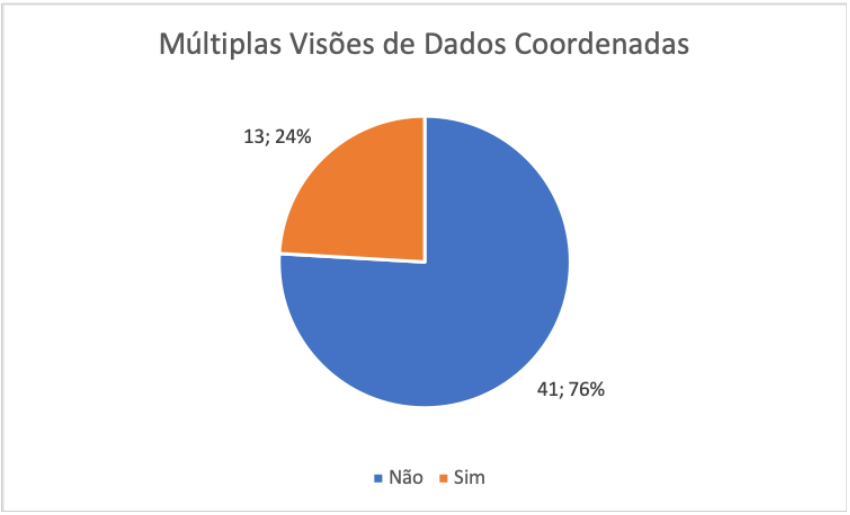
Por fim, a Figura 22 apresenta que a maioria dos trabalhos representam os dados de sentimento por polaridade categórica (positivo, negativo e neutro), em torno de 60%. A valência da polaridade, que utiliza uma escala numérica para especificar a polaridade, variando do negativo ao positivo, totaliza 20%, assim quase 80% dos trabalhos utiliza

Figura 19. Dados de sentimentos geoespaciais representado por cores através de mapa.



FONTE: Zhang, Liu e Wang (2013)

Figura 20. Aplicações que utilizaram mais de uma visão de dados.

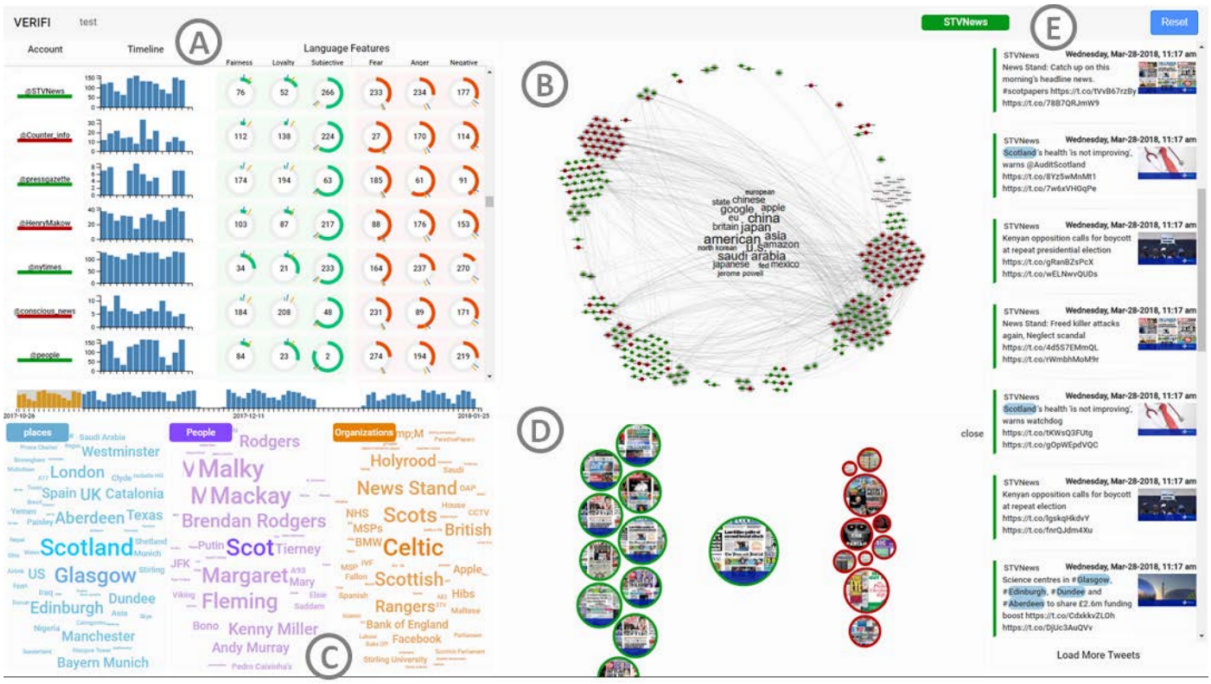


FONTE: O autor

apenas polarização como aplicação de análise de sentimentos, de maneira geral porque é menos complexo a classificação de polarização do que emoções, e são mais simples de representar visualmente. E apenas, 2% utiliza ambas.

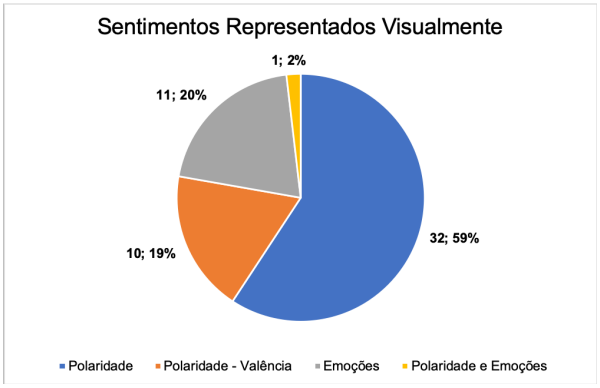
A Figura 23 apresenta um *treemap* que apresenta informações de polaridade em relação a mensagens em redes sociais sobre o nascimento de um bebê na realeza britânica. cada retângulo representa uma localidade, a área representa quantidade de mensagens

Figura 21. Dados de sentimentos combinado com dados do domínio da aplicação.



FONTE: Karduni et al. (2019)

Figura 22. Trabalhos que usam polaridade, valência, emoções e polaridade e emoções.



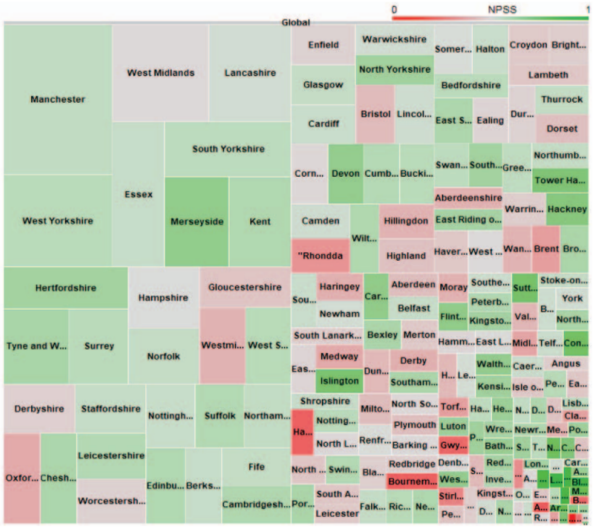
FONTE: O autor

analisadas, e a cor a polaridade média associada aquela localidade.

A Figura 24 apresenta uma aplicação que permite a análise de emoções relacionadas a mensagens de redes sociais para um evento esportivo através de uma escala temporal. O anel mais externo apresenta a escala de cores das emoções, e os círculos mais internos representam um intervalo de tempo, e a frequência de ocorrência de emoções naquele intervalo de tempo.

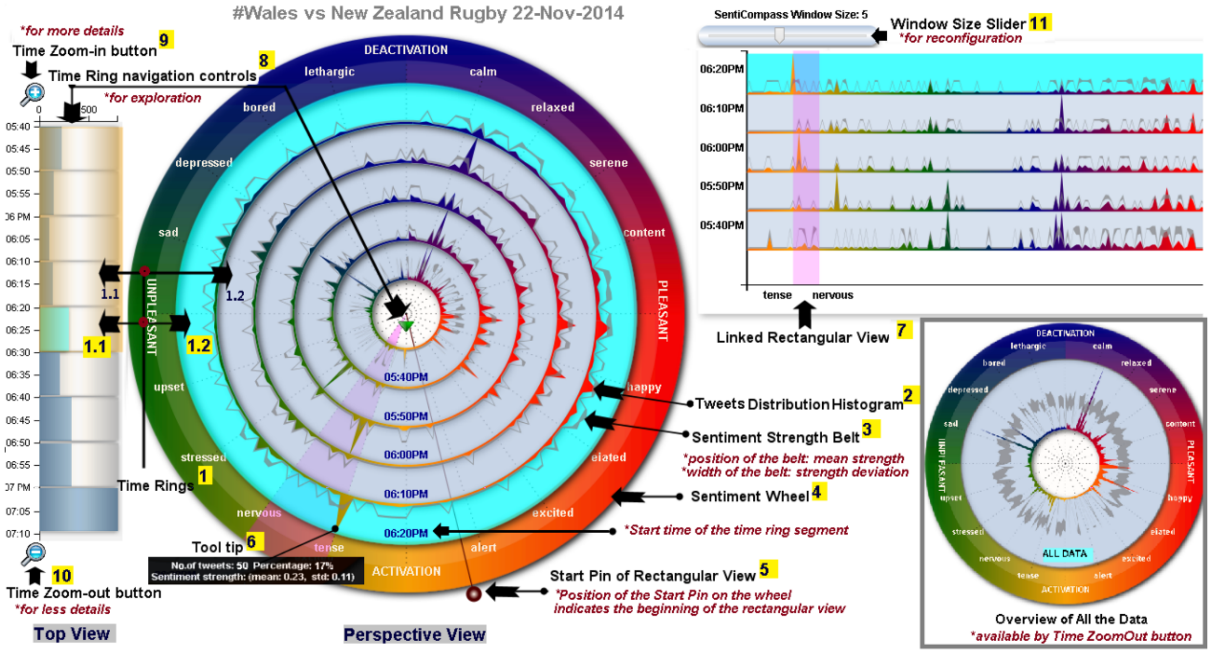
O *survey "The State of the Art in Sentiment Visualization"* (KUCHER; PARADIS; KERREN, 2018b) também merece destaque, porque apresenta novas perspectivas de análise considerando a utilização de técnicas de visualização de dados para representar

Figura 23. Treemap com informações de polaridade de dados de redes sociais.



FONTE: Nguyen, Varghese e Barker (2013)

Figura 24. Aplicação para análise de emoções de dados de redes sociais.

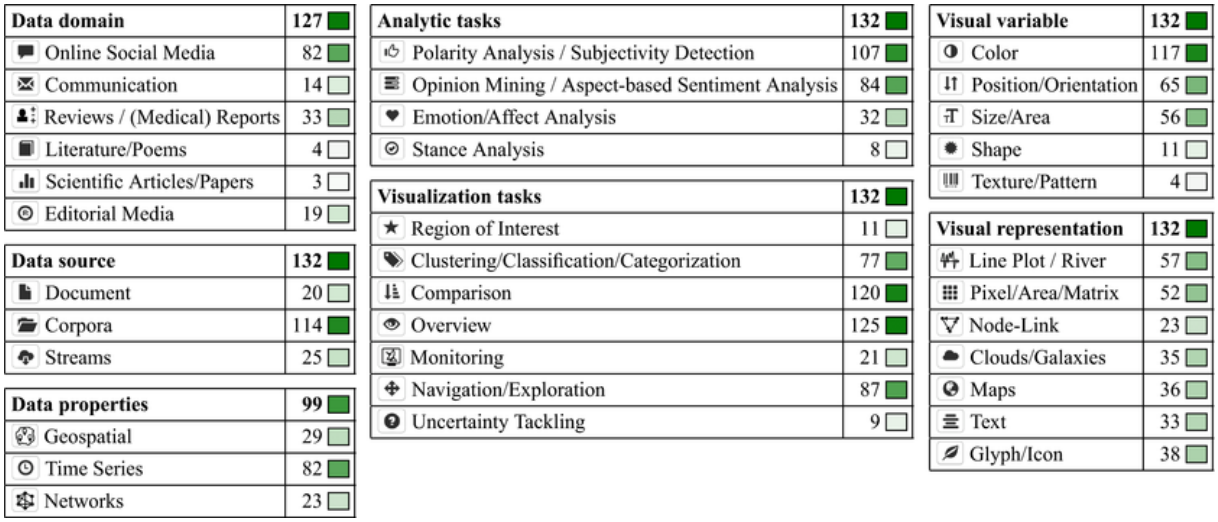


FONTE: Wang et al. (2015)

dados de sentimento, dentre elas destaca-se: como principal nicho de aplicação as redes sociais; o dado de tempo como suporte a análise dos dados de sentimento; representação dos dados de sentimento por polaridade como base das tarefas analíticas; as principais tarefas de visualização que as ferramentas devem suportar são visão geral, exploração, comparação, o que demanda algum grau de configuração e interatividades com as aplicações na criação de cenários de análise; as principais variáveis visuais utilizadas para representar os dados de sentimento são predominantemente cores, seguidos de posição e área; por

fim, os principais layouts de visualização utilizados são gráficos de linhas, matriz, nuvens, mapas, sem grande destaque de uso para algum em específico. A Figura 25 apresenta as classes de análise do *survey*.

Figura 25. Classes de análise do survey.



FONTE: Wang et al. (2015)

De forma resumida, algumas diretrizes adotadas para essa tese em relação às visualizações são: utilizar múltiplas visões de dados coordenadas, utilização de cores para representar dados de sentimento e devem permear todas as visualizações, suporte a dados temporais, geo-espaciais ou de *logs* que demandem sistemas de coordenadas, suporte à representação de dados por polarização ou emoções, suporte interativo as tarefas de visualizações como filtro, detalhes sob demanda, configurações, entre outros.

Em relação a análise de sentimentos multimodal foram realizadas buscas nos sites do IEEE e ACM, com os termos *"multimodal sentiment analysis"* contidos no resumo dos artigos. Para o site do IEEE foram retornados 86 resultados, considerando o período de 2009-2019. Foram lidos os *abstracts*, não foram considerados *posters*, artigos duplicados nas duas bases de dados, e o artigo deveria considerar no mínimo a classificação em conjunto de sentimento em dados de texto, áudio e vídeo.

Como resultado final foram selecionados 51 artigos, serão apresentadas as análises de algumas categorizações, tais como: distribuição por ano, se apresenta múltiplas visões de dados coordenadas, tipos de técnicas de visualização utilizada, domínio da aplicação, tipo de sentimento, e se a visualização é 2D ou 3D. Além disso, foram adicionados 3 artigos selecionados do *survey "SentimentVis Browser - A Visual Survey of Sentiment Visualization Techniques"*, bem como considerações do próprio *survey* citado, totalizando 55 artigos. Por fim, são apresentados de forma breve alguns trabalhos considerados representativo para cada uma das categorias.

3.2 Aprendizado de Máquina

Foram encontrados 216 surveys na pesquisa pelo indexador ACM, utilizando a *string "multimodal sentiment analysis survey"* e 14 continham título e resumo condizentes com a temática do trabalho, mas apenas 10 foram selecionados após a leitura completa.

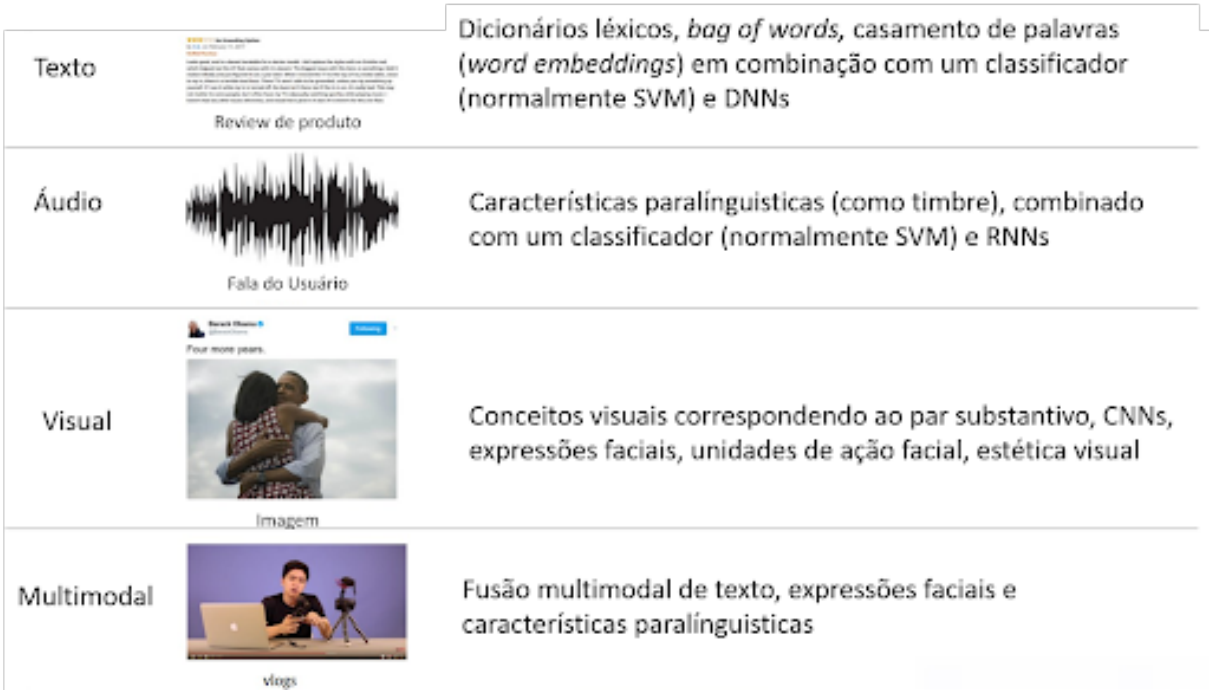
O trabalho de (SOLEYMANI et al., 2017b) já destaca o potencial inexplorado da análise de sentimentos multimodal, com atenção especial ao fato de ser uma área orientada a dados (*Data driven*) o que implica que a quantidade e qualidade dos dados é diretamente responsável pela qualidade da solução, indicando que as soluções atuais dependem dos dados envolvidos, não das regras pré-definidas. O uso de entradas visuais, aural (com áudio) e textuais são normalmente unificadas em nível de decisão, enquanto abordagens a nível de características tem ganhado destaque recentemente. O survey (Hakak et al., 2017) trata de multimodalidade somente em texto, e um outro trata somente das bases de dados para análise de sentimentos (Dudzik et al., 2019), sem apresentar soluções ou técnicas.

As entradas costumam ser categorizadas no sentido visual (com imagens e vídeos), aural e textual (com análise léxica e extração de texto a partir de áudio) (PORIA et al., 2017) (MUKESHIMANA et al., 2017) (CORNEANU et al., 2016). Há muito foco no uso das faces dos usuários nesse tipo de problema (CORNEANU et al., 2016) (MUKESHIMANA et al., 2017), já que a face carrega muitas informações de sentimentos, e variações do uso dessas três entradas são vistos e comparadas frequentemente. Uma visão geral dos tipos de mídia, exemplos e abordagens podem ser vistos na Figura 26.

Vários trabalhos citam o processo de análise de sentimentos em duas abordagens relacionadas à extração de características (MUKESHIMANA et al., 2017) (PORIA et al., 2017) (RAMACHANDRAM; TAYLOR, 2017) (CORNEANU et al., 2016) (FULSE; SUGANDHI; MAHAJAN, 2014), com extração manual e por meio de aprendizado. Mesmo nos problemas ditos não estruturados (com conjuntos de dados não anotados na forma $(entrada, saída)$, essas abordagens podem ser encontradas (SENG; ANG, 2019). Além disso, esses mesmos trabalhos também tratam dos algoritmos em nível de classificação com fusão a nível de decisão, em que cada resultado dos classificadores unimodais tem um peso para a decisão final, e fusão a nível de característica, onde todas as características das entradas são treinadas no mesmo classificador.

Outra decisão importante abordada por esses *surveys* é a seleção do tipo de classificação, se será polarizada ou com as emoções. Muitos trabalhos optam somente pela polarização (PORIA et al., 2017), principalmente os baseados em texto, mas o uso de entradas multimodais também pode ser visto diretamente com a classificação de emoções (FULSE; SUGANDHI; MAHAJAN, 2014). Todas essas decisões levam diretamente a decisão dos classificadores, que podem ser dos mais clássicos, convergindo para o SVM, e ao mais modernos, convergindo para aprendizado profundo, sendo citado como abordagem futura

Figura 26. Visão Geral de métodos comuns de entrada para análise de sentimentos, com respectivos métodos para classificação.



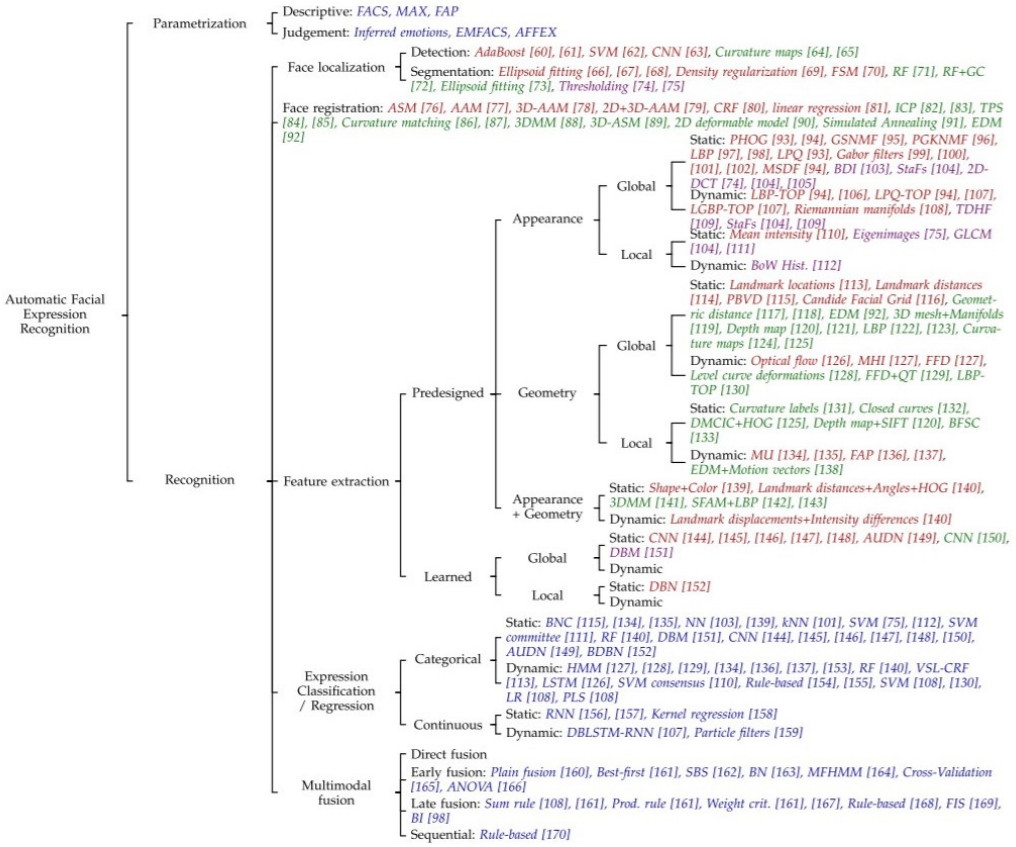
FONTE: Adaptado de (SOLEYMANI et al., 2017b)

em alguns trabalhos (SOLEYMANI et al., 2017b) (FULSE; SUGANDHI; MAHAJAN, 2014) (SENG; ANG, 2019). Exemplificando com a detecção de faces, a taxonomia na Figura 27 apresenta os aspectos que podem ser utilizados nesse problema, e alguns já abordados aqui podem ser generalizados para análise de sentimentos.

Para análise de sentimentos, de acordo com a taxonomia de (CORNEANU et al., 2016), o ramo de extração de características (*feature extraction*) pode ser em questão de aprendizado ou pré-definido. Em termos de fusão das entradas e saídas, podem ser diretas, em nível de entrada única e único classificador, mais cedo (*early fusion*), que são em nível de característica, mais tarde (*late fusion*), em nível de decisão e sequencial, baseado em regras. Por fim, os métodos automáticos podem ser categóricos, levando em consideração um estado das entradas ou contínuos, que levam em consideração a sequência das entradas.

Trabalhos recentes já utilizam aprendizado profundo, alcançando bons resultados práticos (PORIA; HUSSAIN; CAMBRIA, 2018) (RAMACHANDRAM; TAYLOR, 2017) (CORNEANU et al., 2016) (ZHANG; WANG; LIU, 2018), inclusive com técnicas modernas de visão computacional, como modelos generativos para criação de novos itens nas bases de dados. Uma mudança importante no paradigma dos classificadores tradicionais para aprendizado profundo é a possibilidade de criar redes neurais multimodais com a entrada direta sem pré-processamento. Trabalhos que tem propósito de descrever imagens a partir delas mesmas, usam entradas de imagem e texto diretamente na rede, o que permite

Figura 27. Taxonomia de reconhecimento de faces, aplicável para análise de sentimentos.



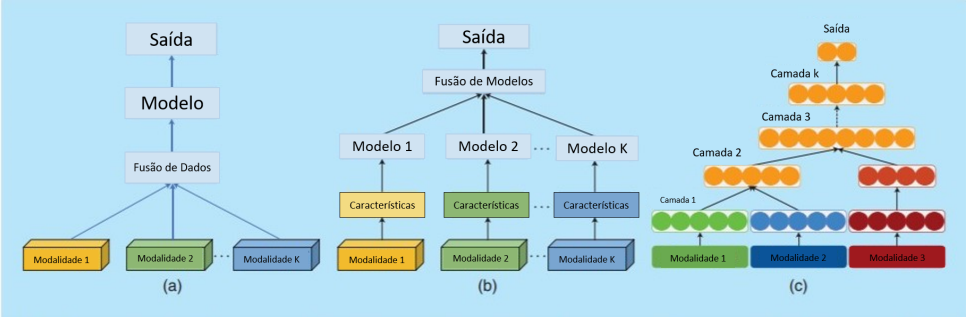
FONTE: Corneanu et al. (2016)

que a generalização aconteça de forma multimodal e intermediária na rede. A Figura 28 apresenta níveis de fusão, na Figura 28 (a) por fusão de características com um modelo inteligente único, (b) com fusão por nível de decisão, baseado em classificadores unimodais e (c) com uma abordagem intermediária, que mostra decisões e características sendo unidas em vários níveis. Não apresentado na figura, uma quarta abordagem mistura fusão por nível de decisão com a fusão por nível de característica, criando uma fusão híbrida.

Os surveys também apontam para trabalhos que considerem o tempo como uma dimensão importante, para mapeamento das microexpressões através de CNNs 3D e RNNs, sendo um desafio a ser vencido na nova era de análise de sentimentos (PORIA et al., 2017). Os melhores resultados indicam 80% de acurácia (PORIA; HUSSAIN; CAMBRIA, 2018), com uma variação multimodal entre 52% e 80% de acurácia (ZHANG; WANG; LIU, 2018) (PORIA; HUSSAIN; CAMBRIA, 2018) (CAMBRIA et al., 2017; CORNEANU et al., 2016; PORIA; HUSSAIN; CAMBRIA, 2018) entre várias combinações de texto, áudio e imagem.

Por fim, existe um grande interesse pela comunidade de IHC (Interação Humano Computador), expressada através dos trabalhos descritos em (MUKESHIMANA et al., 2017), que trata principalmente do uso da face dos usuários para análise de sentimentos.

Figura 28. Classificadores de análise de sentimentos apresentando: (a) a fusão por nível de característica, (b) a fusão por nível de decisão e (c) abordagem intermediária, com fusão em diferentes níveis.



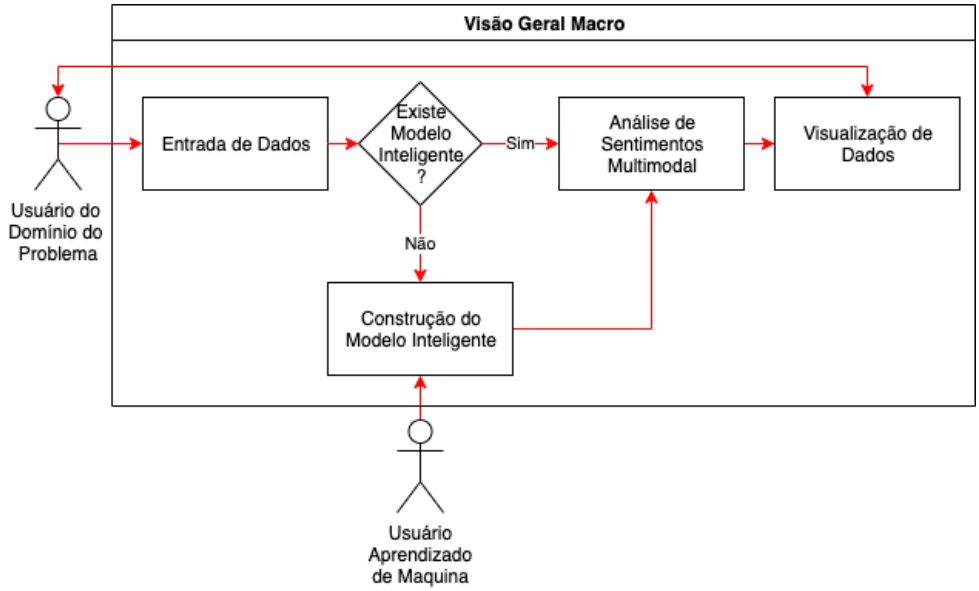
FONTE: Ramachandram e Taylor (2017)

Para IHC, o uso dessas técnicas é especialmente importante para avaliações e testes com usuários, já que pode ser feita uma análise além dos tradicionais questionários pós-teste.

4 Modelo

Este capítulo descreve o modelo desenvolvido por esta tese para abstrair a complexidade da classificação e análise de dados de sentimento multimodal. Além disso, integra os dados de sentimento gerados ao conjunto de dados pertinentes ao domínio do problema para visualização e análise pelo usuário. A figura 29 apresenta a visão mais macro do modelo.

Figura 29. Visão macro do modelo.



FONTE: O autor

A Figura 30 apresenta uma visão geral de todo o modelo com os principais sub-módulos expandidos e interações de usuários. O modelo proposto é composto de quatro módulos principais:

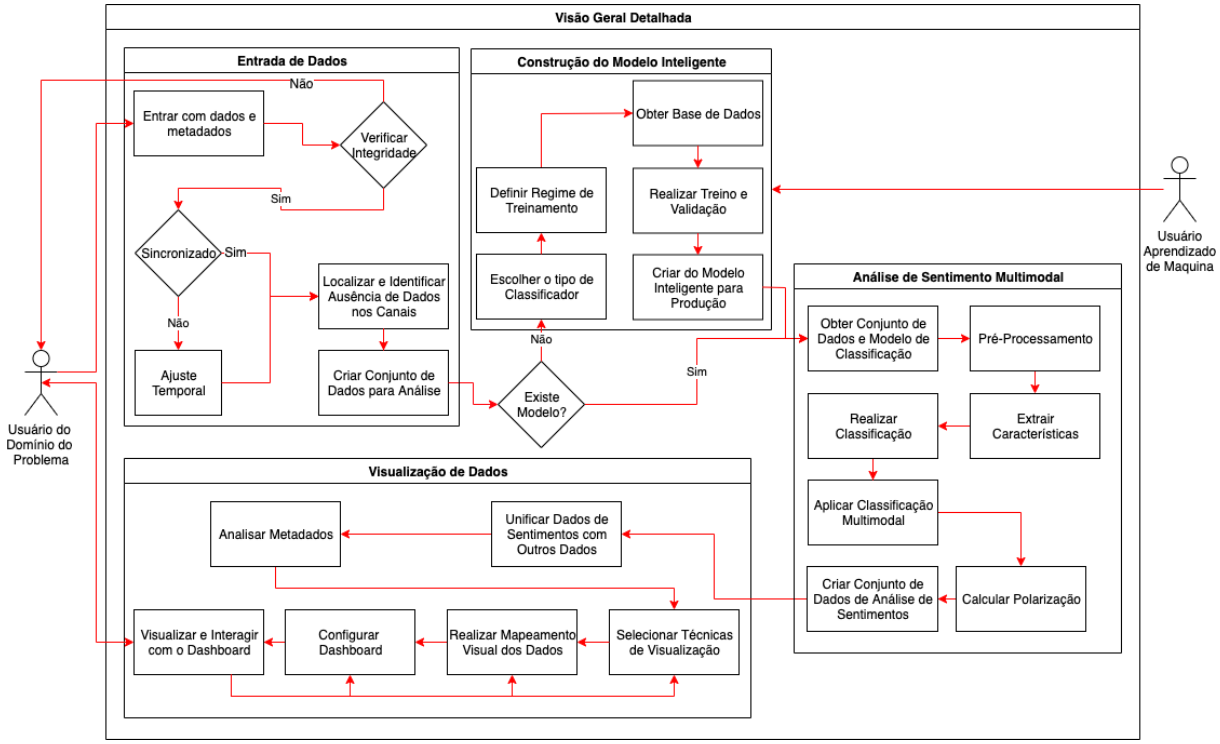
Entrada de Dados - é o módulo responsável por controlar a entrada de dados e garantir a integridade, confiabilidade e sincronismo dos dados, e tem ação direta do usuário do Domínio do Problema;

Construção de Modelo Inteligente - é responsável por criar os modelos de aprendizado de máquina. Caso o modelo já exista este é utilizado diretamente para classificação dos dados. Se um novo modelo precisa ser criado para um conjunto de dados, este pode receber ação do usuário especialista em aprendizado de máquina;

Análise de Sentimentos Multimodal - é responsável por extrair características, classificar e gerar dados de sentimentos a partir dos modelos treinados no módulo de Construção de Modelos Inteligentes e dos dados inseridos no módulo de Entrada de Dados;

Visualização de dados - é responsável por pré-processar, integrar, e representar visualmente os dados em componentes de visualização para que possam ser melhor interpretados e manipulados pelos usuários. A seguir será apresentado o detalhamento de cada módulo:

Figura 30. Visão geral detalhada do modelo e módulos.



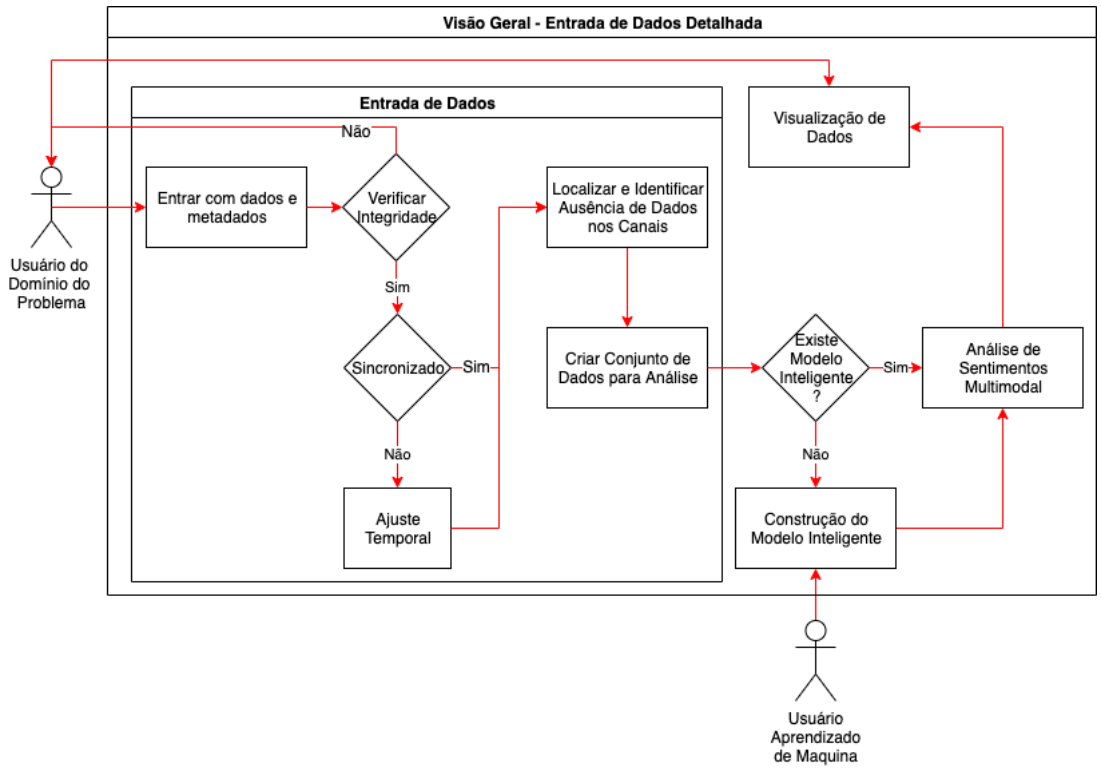
FONTE: O autor

4.1 Módulo Entrada de dados:

O módulo de entrada de dados é responsável por gerenciar, controlar e sincronizar os arquivos de mídia e *logs* que são utilizados para o reconhecimento de emoções. Este módulo recebe interação do usuário do Domínio do Problema e este começa com a submissão dos dados e metadados. Os dados podem ter várias fontes diferentes e formatos: vídeo da face do usuário, vídeo da interação do usuário, áudio do usuário e *logs* de sensores como *Eye Tracker*, Eletroencefalograma (EEG) , eletrocardiograma (ECG), etc. Os metadados principais são informações de tempo relacionados aos arquivos de dados para análise de sentimentos e outros dados do domínio do problema que serão sincronizados e integrados posteriormente. A Figura 31 apresenta o módulo de entrada de dados.

De maneira geral, para este módulo o usuário faz o envio de arquivos, *logs* e metadados. A partir desta etapa é feito uma verificação de integridade nos arquivos, caso não esteja íntegro o usuário recebe um retorno, e caso esteja íntegro é passada para fase de verificação da sincronização temporal dos arquivos submetidos. Caso os arquivos não

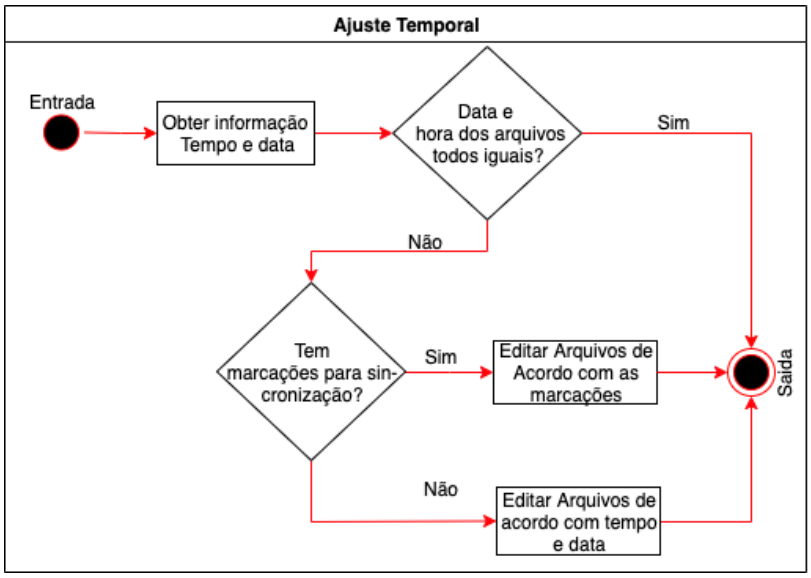
Figura 31. Módulo de gerenciamento de entrada de dados.



FONTE: O autor

estejam com valores de tempo e data sincronizados, um ajuste temporal automático é feito baseado nos metadados (Figura 32).

Figura 32. Fluxograma do ajuste temporal dos arquivos.

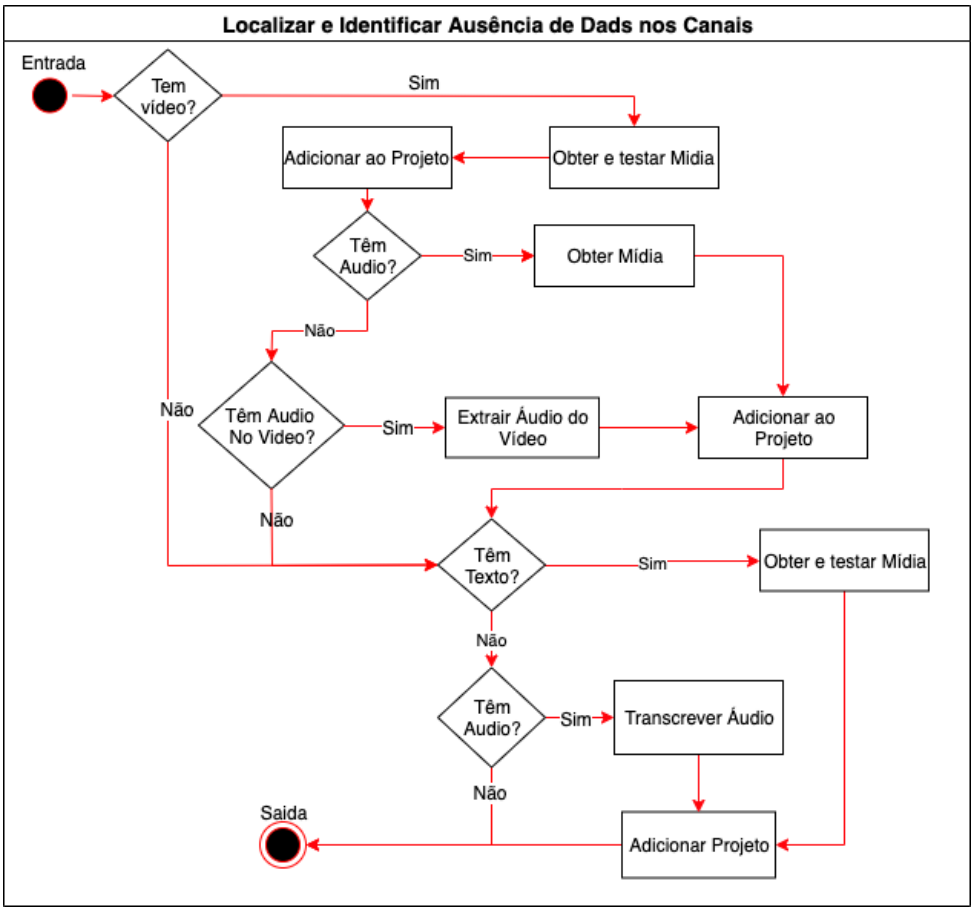


FONTE: O autor

Caso esteja sincronizado ou após o ajuste temporal inicia a fase de localizar e

identificar ausência de dados nos canais (Figura 33). Esta etapa identifica se estão ausentes os tipos de dados em que será realizada a análise de sentimentos, e tentará realiza extração automática de alguns deles. Por exemplo, considerando que a análise de sentimentos será realizada utilizando vídeo, áudio e texto, e foi identificado que há os dados de vídeo, mas não há dados de áudio e texto, é possível extrair o áudio do vídeo, e transcrever o texto do áudio, completando todos os canais de dados definidos pelo usuário. Após essa etapa são criados os conjuntos de dados para armazenamento e manipulação nos próximos módulos.

Figura 33. Fluxograma da localização e identificar ausência de dados dos canais.



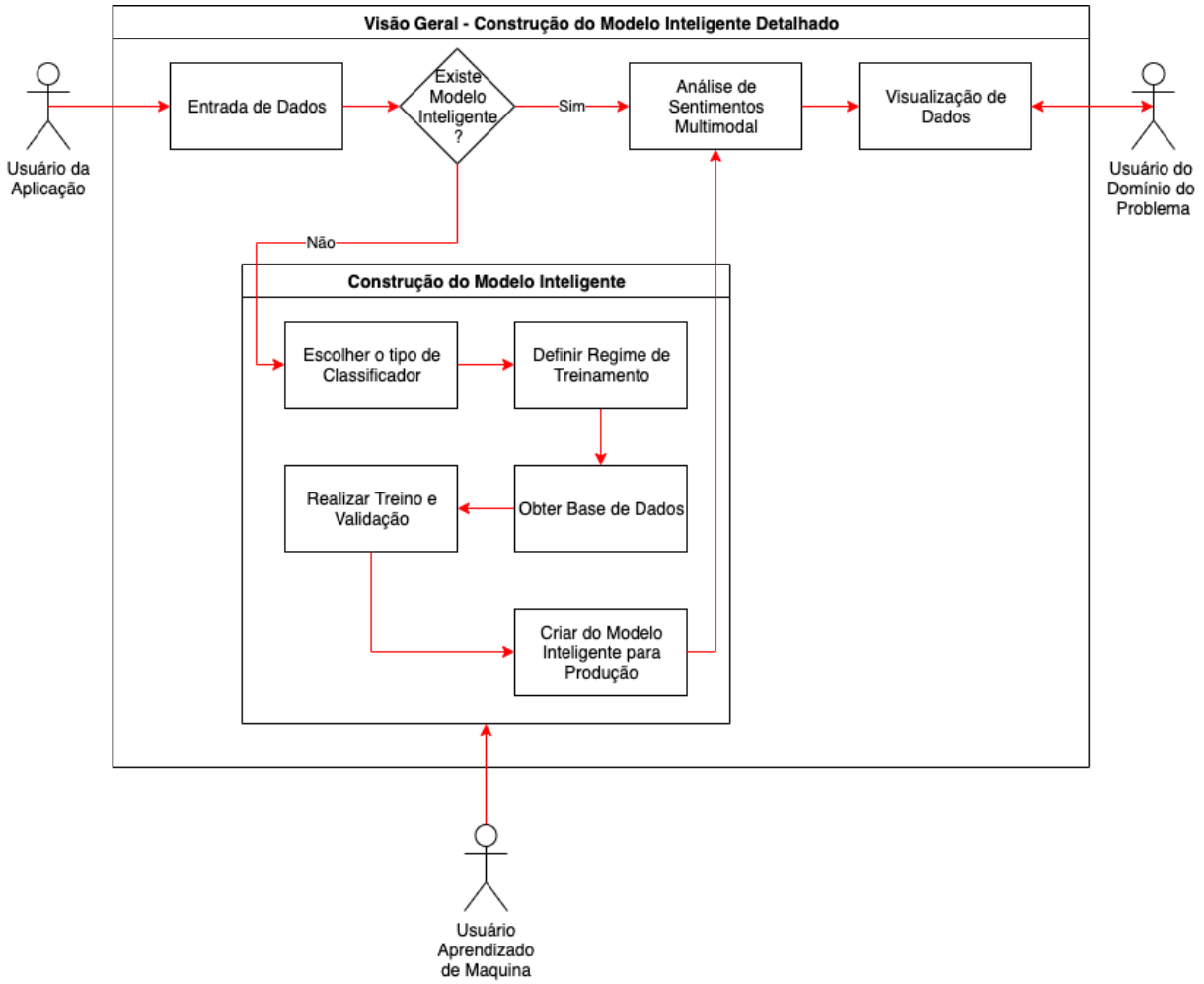
FONTE: O autor

As principais características deste módulo são: Interação do usuário do Domínio do Problema para envio dos arquivos e *logs*, além de verificação da integridade e sincronismo dos arquivos. Etapas de localização e identificação de ausência de dados nos canais para localizar e identificar os dados de áudio e texto, além de criar os conjuntos de dados para posterior utilização.

4.2 Módulo Construção do Modelo Inteligente:

De todos os módulos do modelo, esse é o único sem o envolvimento do Usuário do domínio do problema de aplicação, o ator envolvido nesse módulo é o Usuário Especialista em técnicas de aprendizado de máquina. Esse ator é responsável por criar métodos inteligentes para análise de sentimentos automática, e esse módulo engloba o passo a passo tradicionalmente utilizado por esses especialistas para esse problema. Esse ator não precisa ser uma pessoa envolvida diretamente com o domínio do problema, a comunidade de IA vem amadurecendo questões em torno de reprodutibilidade de experimentos e testes (PORIA; HUSSAIN; CAMBRIA, 2018), e esse comportamento permite a reutilização de soluções já construídas por especialistas. Além disso, serviços de nuvem e processamento de tarefas de IA podem economizar esforço de implementação voltados a parte mais técnica, como configuração de placas para processamento configuração de rede. O modelo suporta esses cenários através do ator, que mesmo sendo indicado como um usuário especialista, não está necessariamente envolvido em tempo presente na aplicação do modelo.

Figura 34. Funcionamento do módulo de construção do modelo inteligente.



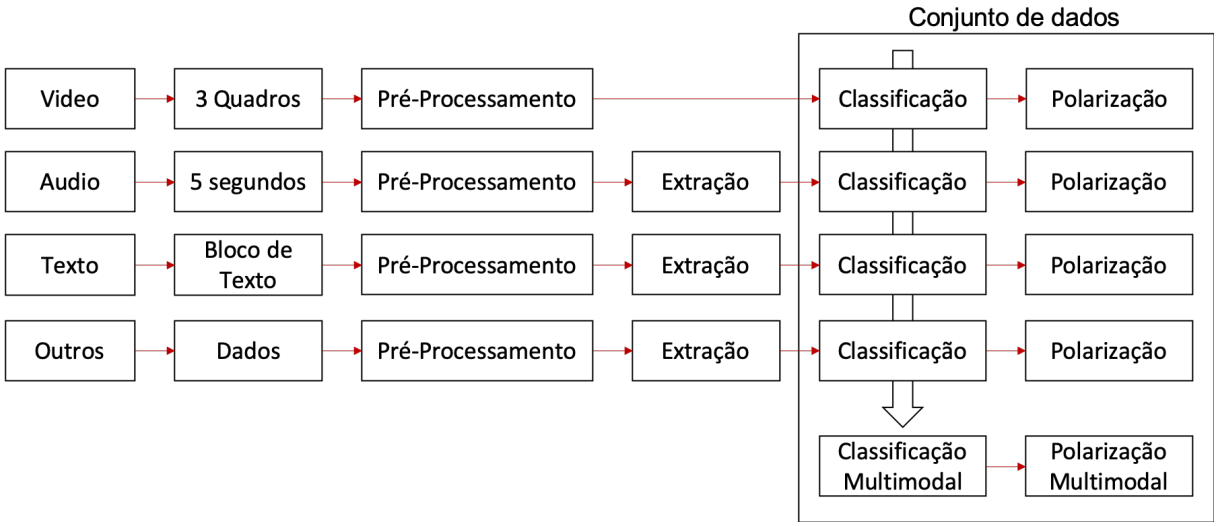
FONTE: O autor

Tradicionalmente, a análise de sentimentos é um problema de classificação à dados, tais como: como *frames* de vídeo, trechos de áudio, textos, etc. (KAUR; KAUTISH, 2019) Para isso, um modelo inteligente classifica esses dados em um dos tipos de sentimentos esperados pelo problema, por exemplo emoções ou polarização. Da definição do problema até a implementação do modelo em produção, esse passo a passo é bem definido pela literatura (PORIA; HUSSAIN; CAMBRIA, 2018; KAUR; KAUTISH, 2019; HUSSEIN, 2018; MEDHAT; HASSAN; KORASHY, 2014) e será descrito nessa seção com destaques necessários para o problema de análise de sentimentos. A Figura 34 apresenta os sub-módulos em detalhe, bem como sua descrição a seguir.

4.3 Escolher o tipo de classificador

A primeira decisão do Usuário especialista em aprendizado de máquina é na escolha do tipo de classificador utilizado. Dado a natureza multimodal do problema de classificação de sentimentos, a escolha a ser feita define todos os outros passos diretamente. A decisão entre os classificadores de fusão por nível de decisão e fusão por nível de características depende de como o modelo será utilizado e a entrada dos dados.

Figura 35. Abordagem de fusão por nível de decisão.



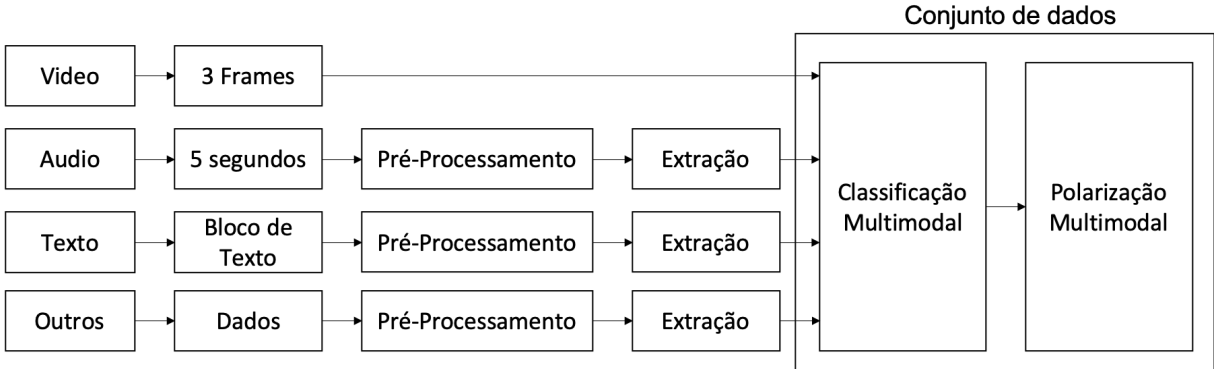
FONTE: O autor

A abordagem de fusão por nível de decisão (Figura 35) utiliza um classificador para cada tipo de entrada, de modo que cada classificador é treinado de maneira independente e a decisão final é feita por votação utilizando pesos (PORIA; HUSSAIN; CAMBRIA, 2018). Esse tipo de abordagem é apropriada para problemas em que a entrada é multimodal mas nem sempre é uniforme, como uma entrada de vídeo que não tem todas as transcrições de áudio associadas ou uma entrada de vídeo que tem muitos períodos silenciosos. Treinar

modelos com essa abordagem também é mais simples que treinar somente um modelo que engloba tudo (HUSSEIN, 2018).

A abordagem de fusão por nível de características (Figura 36) utiliza somente um classificador para todas as entradas, utilizando as características extraídas de cada entrada em somente um classificador. O treino é mais complexo com essa abordagem (HUSSEIN, 2018), e não é simples representar a falta de uma entrada, já que os pesos devem ser definidos diretamente na função objetivo. A vantagem dessa abordagem é que a representação apreendida pelo modelo leva em consideração todas as entradas ao mesmo tempo e nesse tipo de problema uma solução com um treino apropriado generaliza melhor que uma contraparte utilizando um classificador para cada entrada, além de compartilhar a computação entre várias entradas, reduzindo tempo de inferência (PORIA; HUSSAIN; CAMBRIA, 2018).

Figura 36. Abordagem de fusão por nível de características.



FONTE: O autor

A escolha de uma dessas abordagens, ou até uma híbrida (SOLEYMANI et al., 2017a), deve levar principalmente em consideração os tipos de dados utilizados e suas características para o problema de classificação de sentimento, bem como a interpretabilidade de seus resultados.

4.3.1 Definir regime de treinamento

Definido o tipo de classificador, é necessário definir o regime de treinamento dos modelos inteligentes. A utilização de métodos do estado da arte para classificação converge para área de aprendizado profundo (LECUN; BENGIO; HINTON, 2015), que é orientado à utilização de grandes bases de dados, e isso deve ser levado em consideração. Para esses métodos, uma etapa de treinamento é utilizada para que o modelo aprenda uma representação baseada nesse conjunto de dados.

O regime de treinamento deve levar em consideração o tempo disponível para a treinamento dos modelos inteligentes, pré-processamento, custos de treinamento (energia,

servidores, placas de vídeo, anotação de dados), utilização de dados e licenças, métodos para avaliação e seleção de parâmetros. O usuário especialista de aprendizado de máquina deve ponderar sobre essas questões e definir no fim como disponibilizar o modelo inteligente treinado para a aplicação que implementa o modelo.

4.3.2 Obter base de dados

Os métodos de classificação com melhores resultados utilizam bases de dados com uma grande quantidade de itens. Uma base de dados deve conter a melhor aproximação possível dos dados reais que serão utilizados pelo usuário do domínio do problema. A obtenção da base de dados deve levar em consideração o tipo de abordagem que será utilizada, seja de extração de características ou aprendizado de representações. Para extração de características, é necessário conhecimento do domínio e um método para traduzir esse conhecimento para características computáveis, já modelos treinados de aprendizado profundo aprendem representações complexas, tem capacidade de generalização altamente dependente do conjunto de dados (BENGIO; COURVILLE; VINCENT, 2013) e não precisam de um especialista do domínio para definir as características que devem ser utilizadas no problema, já que as aprende sozinho. Independente do formato da entrada e a abordagem utilizada, o usuário especialista deve levar em consideração a qualidade da base quando for realizar o treinamento.

4.3.3 Realizar treino e validação

O treino dos modelos deve ser realizado de acordo com o regime de treinamento, levando em consideração a seleção de parâmetros, mas com atenção do usuário especialista para qualquer mudança que possa ser feita baseada na validação do modelo. A manipulação da base de dados é a decisão mais importante nessa etapa, com decisões envolvendo sua divisão em conjuntos de treino, validação e teste, que dá mais certeza sobre a generalização dos modelos, aumentar a base de dados, realizar (ou não) extração de características e outros pré-processamentos. Regularizar o treinamento e seleção de arquiteturas são outras escolhas que podem ser feitas nessa etapa, baseado nos resultados do treinamento dos modelos.

4.3.4 Criar modelo inteligente para produção

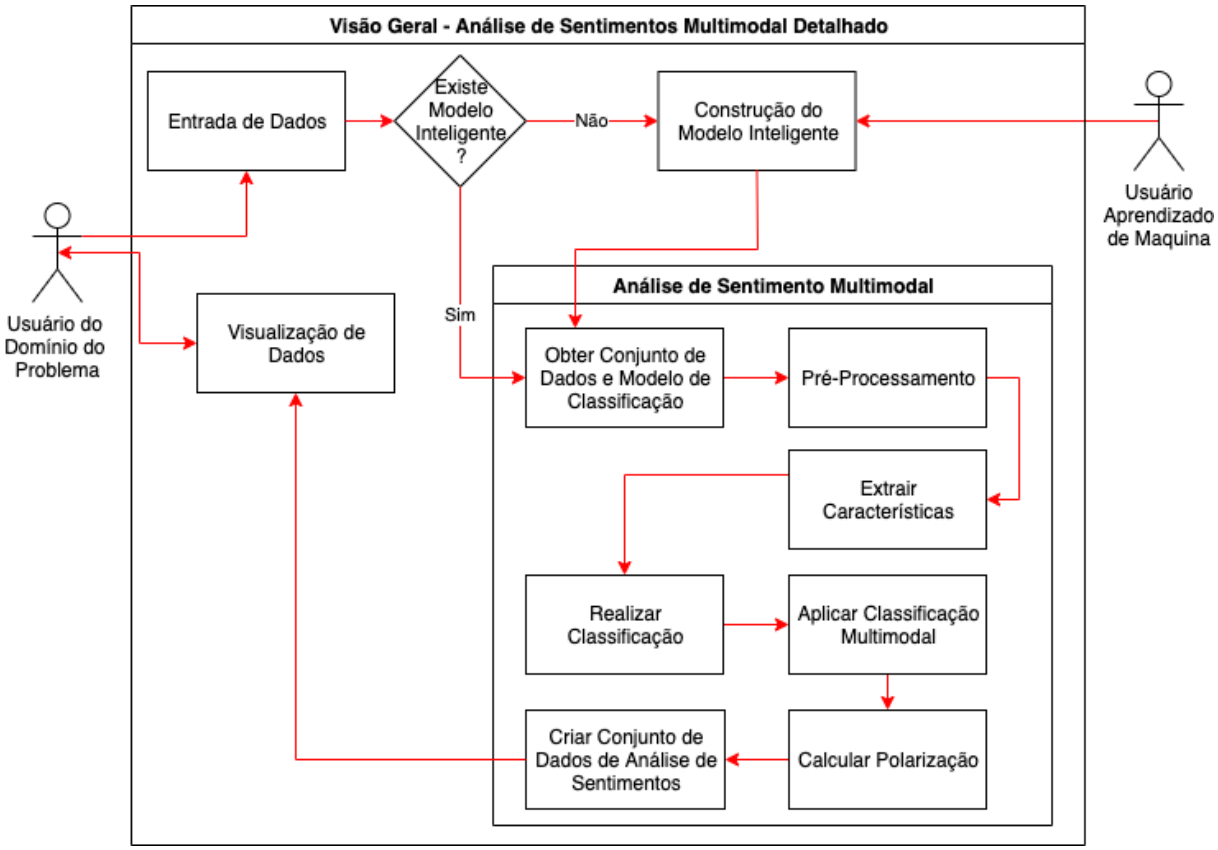
Como última etapa deste módulo, um modelo inteligente que seja capaz de fazer reconhecimento automático de sentimentos deverá ser colocado em produção, de acordo com as necessidades do usuário do Domínio do Problema. Para tal, deve ser levado em consideração o tipo de aplicação do modelo, equipamento selecionado para realizar classificação e o tempo disponível do usuário para que seja feita a análise de sentimentos.

Essas decisões são de caráter técnico e devem ser tratadas de forma objetiva, com testes e avaliações quando possível.

4.4 Módulo Análise de sentimentos Multimodal:

O módulo de análise de sentimentos multimodal (Figura 37) é responsável por classificar os dados de sentimentos. Esse módulo precisa de um modelo inteligente para ser usado no classificadores de análise de sentimentos unimodal e multimodal. Esse modelo é fornecido pelo módulo de Construção do Modelo Inteligente. A classificação dos sentimentos funciona de maneira modular e de acordo como os modelos inteligentes existentes, ou seja, podem haver muitos classificadores de sentimentos unimodais e uma integração de seus resultados, bem como um único classificador multimodal e ainda um classificador híbrido, que tem entrada multimodal e vários resultados, para classificar os multi-parâmetros recebidos de diferentes arquivos de mídia e *logs*.

Figura 37. Representação do módulo de análise de sentimentos multimodal.



FONTE: O autor

De posse dos dados e modelo inteligente, o módulo inicia o pré-processamento onde normaliza e ajusta os dados de acordo com o modelo inteligente para então extrair as características dos dados, se necessário, que podem variar de acordo com o tipo de classificador, modelo e dados.

De maneira geral, quando necessárias, as características extraídas do vídeo são as regiões do rosto, para o áudio podem ser MFCC ou parâmetros como *pitch*, volume, timbre, etc (PORIA; HUSSAIN; CAMBRIA, 2018). Para o texto podem ser usados blocos de texto, palavras-chaves ou utilização de dicionários. Além de características dos dados oriundos de sensores e *logs* que podem ser utilizados para compor um classificador unimodal ou multimodal.

A classificação multimodal pode ser realizada a partir de um novo classificador, aplicação de pesos ou outras técnicas no resultado das classificações unimodais. Outra abordagem pode ser a aplicação de um único classificador multi-parâmetro, sem a necessidade de especificar pesos. Por fim, abordagens híbridas das duas anteriores podem ser aplicadas, como pode ser vistas em (PORIA; HUSSAIN; CAMBRIA, 2018; KAUR; KAUTISH, 2019).

Após a classificação de sentimento multimodal a partir dos dados é possível realizar o cálculo da polaridade, este cálculo é baseado nos sentimentos encontrado pelos classificadores. Assim, tanto as classificações das emoções quanto polaridades são adicionadas a um conjunto de dados, relacionado ao seu dado de origem, para serem utilizadas no módulo de visualização de dados.

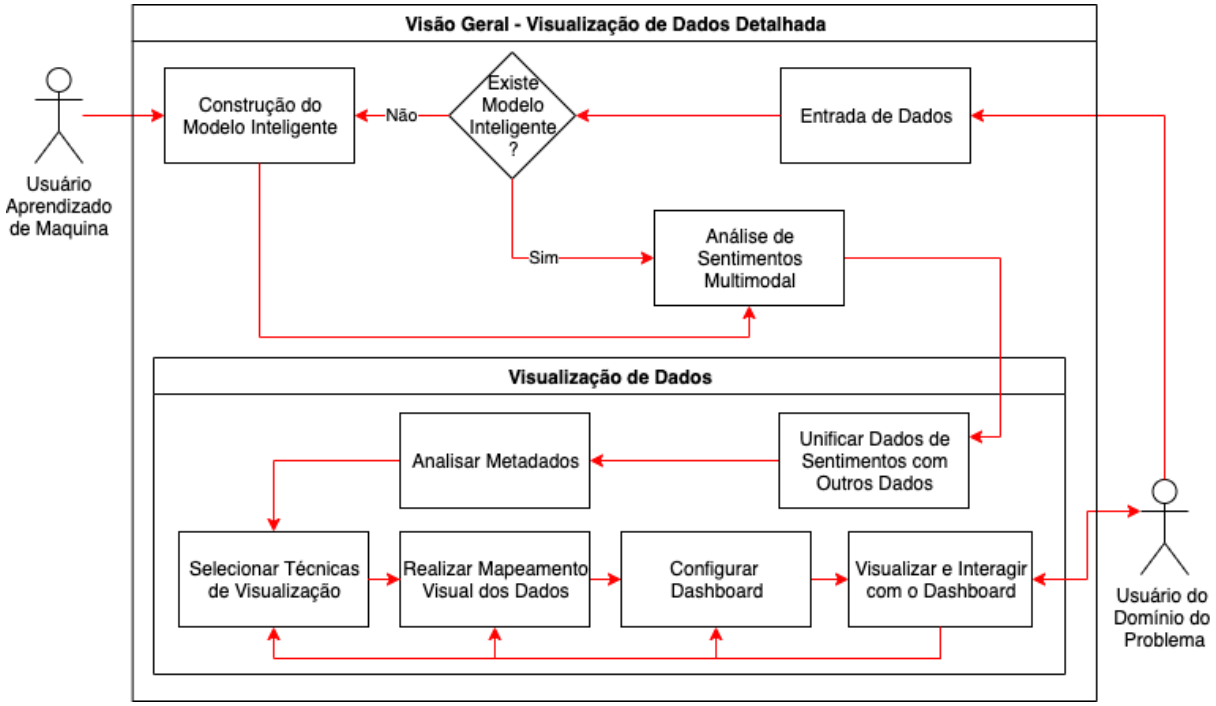
4.5 Módulo Visualização de dados:

Esse módulo é responsável por apresentar, através de técnicas de visualização, os dados de sentimentos gerados e integrado aos outros dados do domínio do problema para análise do usuário. A visão mais detalhado deste módulo é apresentado na Figura 38.

O funcionamento deste módulo inicia com a unificação dos dados de sentimentos e outros dados que foram disponibilizados pelo usuário no módulo de Entrada de Dados, ou seja, faz a integração de dados, por exemplo, de *logs* de rastreamento de olhos e mouse, de sensores de frequência de batimentos cardíacos, entre outros. Após essa unificação o módulo cria, configura e relaciona as diferentes visualizações de dados de análise baseado nos metadados temporais.

Técnicas de visualização são selecionadas para cada tipo de conjunto de dados disponível. Cada conjunto de dado é caracterizado através de metadados obtidos a partir de informações dos arquivos de dados ou informados pelo próprio usuário. Após, é realizado o mapeamento visual dos dados específicos para cada técnica de visualização, onde os dados são codificados em uma representação visual e organizado de acordo com um layout permitido pela técnica de visualização. As visualizações de dados criadas deverão ser organizadas e integradas em um *dashboard*, que nada mais é que um conjunto de visualização de dados organizados e relacionadas entre si em um painel. Além disso, o *dashboard* deve permitir interações básicas como configurações, filtros e detalhes sob

Figura 38. Representação do módulo de visualização de dados.



FONTE: O autor

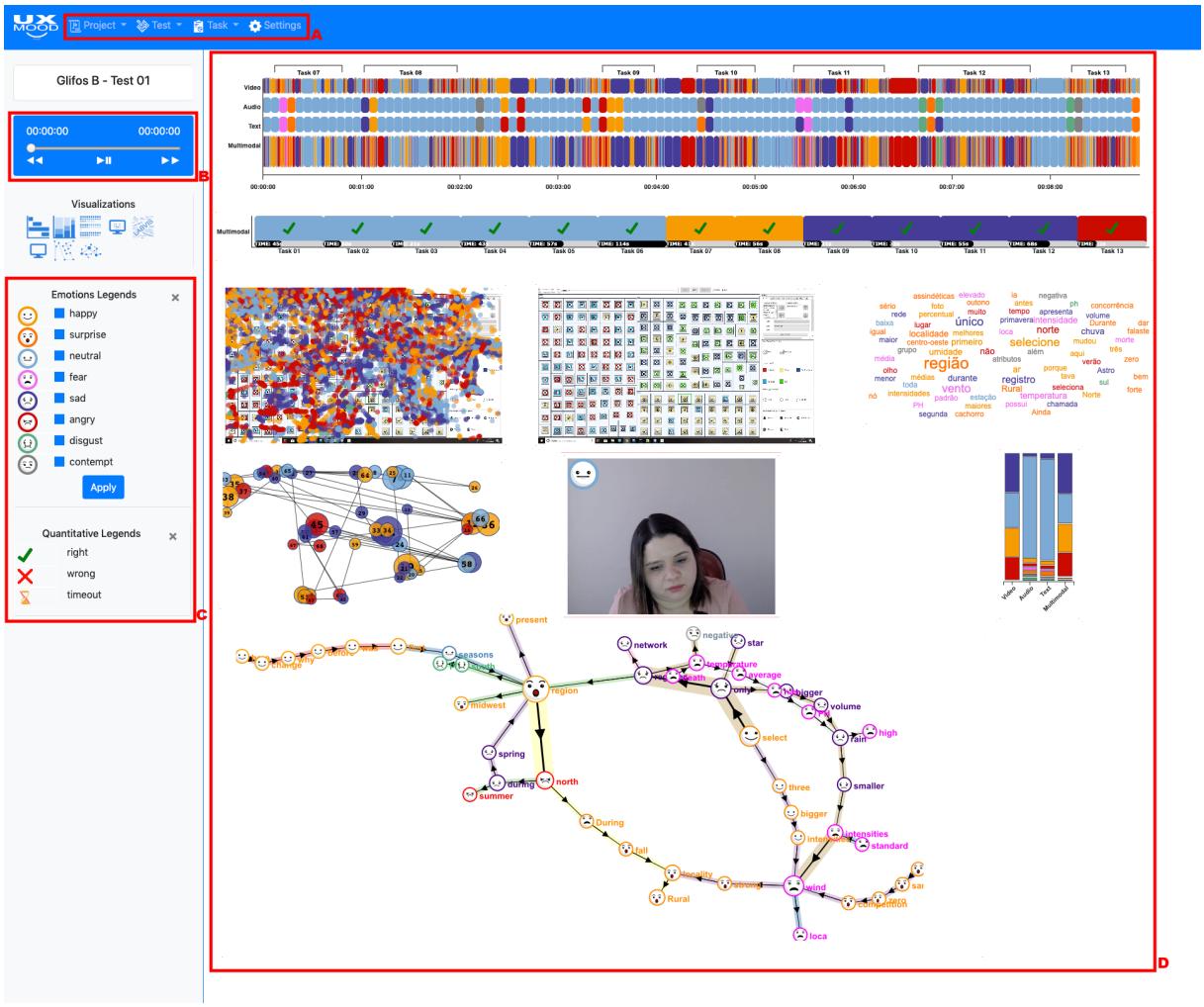
demanda.

Para o usuário do domínio da aplicação, é o módulo com maior relevância para as análises dos usuários, pois são nas visualizações que ele poderá ver as classificações de emoções ou polaridades associados aos outros de dados de análise, aumentando a possibilidade de análise mais precisas através de diversos cenários possibilitados pela interatividade que deverá ser proporcionada pelo *dashboard*.

5 Implementação do Protótipo do Modelo

Este capítulo apresenta uma implementação do modelo proposto nesta tese. Como contexto de uso a implementação, denominada UXmood, é aplicada para auxiliar o processo de avaliação de Testes de Usabilidade e Experiência do Usuário, adicionando dados de sentimento aos dados coletados de um processo de avaliação e apresentando-os por meio de múltiplas visualizações de dados, todo processo é integrado e automático. De maneira geral, o usuário especialista do domínio do problemas apenas precisa inserir os dados do processo de avaliação para UXmood, e a aplicação abstrai a complexidade do uso de análise de sentimentos e a geração das visualizações de dados juntamente com outros dados.

Figura 39. Uma visão geral da interface principal do UXmood.



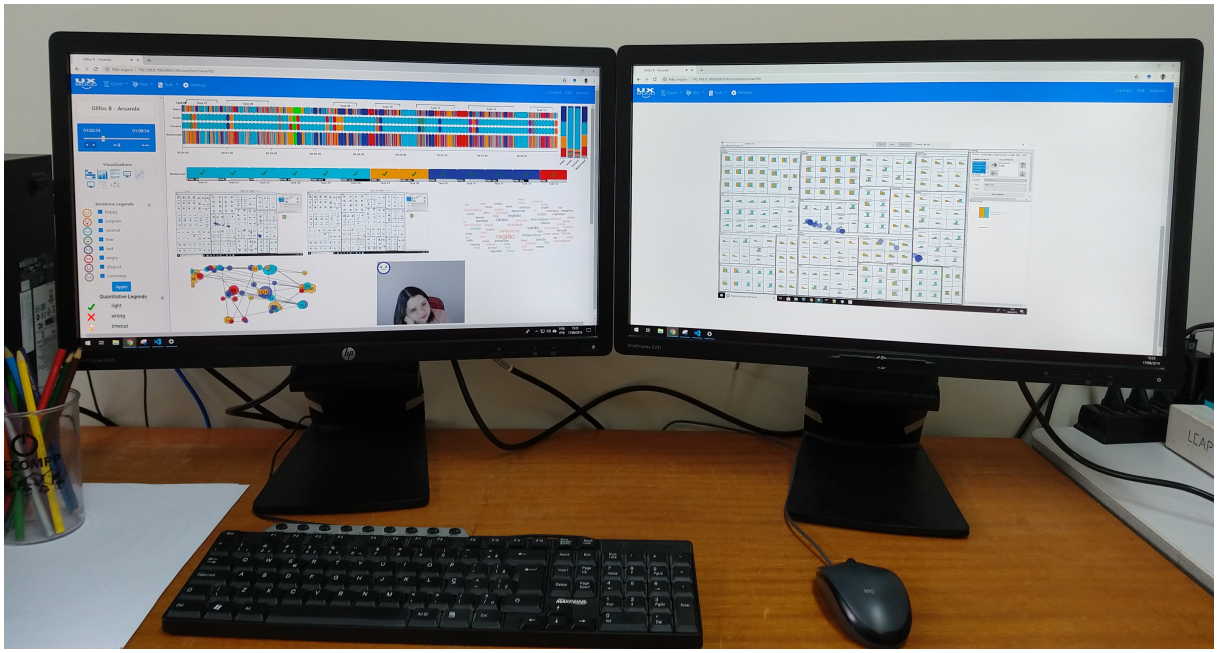
FONTE: O autor

A Figura 39 apresenta a interface principal do UXmood, consistindo em: menus de gerenciamento de projetos (A), uma área de interação que permite aos especialistas

reproduzir interações do usuário (B), legendas para apresentar a codificação visual de dados em visualizações (C) e uma área para múltiplas visualizações de de dados, que inclui os dados de sentimentos, dados de interações dos participantes da de avaliação, e outros dados como tempo de realização de tarefas, corretude de uma tarefa, entre outras (D).

A ferramenta é configurável na escolha de visualizações do painel e é ainda possível selecionar um conjunto de visualizações a serem organizadas em outros monitores (Figura 40). A ferramenta está disponível no site do grupo de pesquisa [1].

Figura 40. Visualizações configuradas para serem exibidas em dois monitores.



FONTE: O autor

Este capítulo, apresenta também os principais recursos presentes na ferramenta UXmood, tais como: arquitetura, dados de entrada e saída, módulos e funcionalidades, além da justificativa de aplicação na área de avaliação de usabilidade e experiência do usuário.

5.1 Justificativa da Área de Aplicação

O cenário de aplicação escolhido para direcionar a implementação do modelo apresentado nesta tese foi o de avaliações de usabilidade e experiência do usuário. Essa escolha se deve ao crescente uso das técnicas de análise de sentimentos multimodais para apoiar a análise da avaliação da experiência do usuário (HUSSAIN et al., 2018). Além disso, pesquisas recentes (KWIATKOWSKA; SZÓSTEK; LAMAS, 2014; MATTHEWS; JUDGE; WHITTAKER, 2012; HARTSON; PYLA, 2012) apontam que pesquisadores, profissionais

[1] http://labvis.ufpa.br/uxmood/

e praticantes da área de avaliação de experiência do usuário e usabilidade tem interesse em adicionar dados de sentimentos em seus contextos de análise, porém apresentam pouca habilidade ou grande dificuldade na utilização de técnicas de aprendizado de máquina para gerar esses dados, bem como entender os resultados gerados por essas técnicas. Assim, UXmood vem de encontro a essas expectativas, abstraindo a complexidade de análise de sentimento e facilitando a interpretação dos dados gerados com técnicas de visualização.

Para corroborar esse cenário, foi realizada uma pesquisa nas principais bibliotecas de artigos científicos (IEEExplore e ACM Digital Library) para verificar o uso combinado destas áreas, com os seguintes termos de busca: *"sentiment analysis", "data visualization", "user experience" or "usability"*, considerando o período de busca entre 2009 e 2019. IEEE retornou 360 resultados e ACM retornou 84 resultados, e destes foram selecionados 19 artigos mais relevantes e organizados de acordo com seis categorias: Dados Quantitativos, Dados Qualitativos, Análise de Sentimentos, Tipo de Avaliação, Técnica de Visualização e Outros (Tabela 1). Essa classificação auxilia na descoberta de padrões e lacunas na literatura, bem como na comparação das características do UXmood com os trabalhos selecionados.

Entre os trabalhos analisados, apenas alguns citam técnicas de análise de sentimentos multimodais para auxiliar no processo de análise de dados sobre avaliação de testes de usabilidade ou da experiência ou a usabilidade do usuário. O tipo mais comum de classificação de sentimento é para texto, seja na transcrição de áudio ou nas revisões de produtos na Internet (SETCHI; ASIKHIA, 2017; HEDEGAARD; SIMONSEN, 2013; PORAT; SCHCLAR; SHAPIRA, 2013; YADAV; BHUSHAN; GUPTA, 2015; VASA et al., 2012; VOIGT; KIESLINGER; SCHÄFER, 2017a; HUSSAIN et al., 2018). Os trabalhos (PORAT; SCHCLAR; SHAPIRA, 2013; YADAV; BHUSHAN; GUPTA, 2015; VASA et al., 2012; HUSSAIN et al., 2018; MUNIM et al., 2017) registram imagens do rosto do usuário para usar algoritmos de análise de sentimentos baseados em expressões faciais.

A captura de fala do usuário está presente em seis trabalhos, tais como (SETCHI; ASIKHIA, 2017; PORAT; SCHCLAR; SHAPIRA, 2013; YADAV; BHUSHAN; GUPTA, 2015; VASA et al., 2012; HUSSAIN et al., 2018; BAUMAN; LIU; TUZHILIN, 2016), mas apenas quatro trabalhos (YADAV; BHUSHAN; GUPTA, 2015; VASA et al., 2012; HUSSAIN et al., 2018; BAUMAN; LIU; TUZHILIN, 2016) usam características físicas do áudio para classificar os sentimentos.

Tabela 1. Classificação e catalogação dos trabalhos relacionados.

Referências	Similaridade	Tipo de Artigo	Tempo	Corretude	Rastreador de olhos	Rastreador do Mouse	Questionário	Protocolo Think Aloud	Vídeo da face do usuário	Vídeo da Interação	Áudio do Usuário	Transcrição do Áudio	Revisão de Produto	Esquema de Imagens	Vídeo	Áudio	Texto	Multimodal	Opinião do Especialista em UX	UX	Usabilidade	Heatmap	Gráfico de Pizza	Stackedbar	Radar	Gráfico de linhas	Scanpath	EmojiText	Wordcloud	Gráfico de Gantt	Scatterplot	Remote Access
✱ UXmood	22	Tool	✓	✓	✓	✓		✓	✓	✓	✓	✓			✓	✓	✓	✓		✓	✓			✓			✓	✓	✓	✓	✓	✓
Hussain et al. (2018)	16	Tool	✓	✓	✓		✓		✓	✓	✓				✓	✓	✓	✓		✓		✓	✓	✓								✓
Setchi e Asikhia (2017)	12	Method	✓	✓			✓	✓		✓	✓	✓		✓			✓			✓	✓								✓			
Fan et al. (2019)	12	Tool	✓	✓				✓		✓	✓	✓	✓				✓			✓	✓		✓	✓		✓						
Dodd, Athauda e Adam (2017)	9	Method	✓	✓			✓			✓									✓	✓	✓		✓			✓						
Malgaonkar et al. (2019)	8	Tool											✓				✓			✓	✓		✓	✓		✓						✓
Porat, Schclar e Shapira (2013)	8	Tool	✓	✓			✓		✓	✓	✓								✓		✓											
Dünser e Billinghurst (2011)	7	Tool	✓	✓			✓												✓	✓	✓	✓										
Figueiras (2018)	7	Method	✓	✓			✓	✓				✓								✓	✓											
Voigt, Kieslinger e Schäfer (2017b)	6	Tool					✓										✓			✓				✓	✓	✓						
Bakiu e Guzman (2017a)	6	Method											✓				✓			✓	✓		✓		✓							
Yadav, Bhushan e Gupta (2015)	6	Method							✓		✓				✓	✓		✓		✓												
Munim et al. (2017)	5	Method							✓	✓					✓					✓						✓						
Kharel e Fernstrom (2016)	5	Tool					✓						✓				✓				✓			✓								
Paramitha, Dantes e Indrawan (2018)	5	Method	✓	✓			✓												✓	✓												
Moizer et al. (2019)	5	Method		✓			✓												✓	✓	✓											
Sánchez-Adame et al. (2018)	5	Method	✓				✓												✓	✓	✓											
Hedegaard e Simonsen (2013)	5	Tool											✓	✓			✓			✓	✓											
Corrêa et al. (2017)	5	Method		✓			✓												✓	✓	✓											
Yu et al. (2019)	4	Tool	✓	✓			✓													✓												
Seguin, Scharff e Pedersen (2019)	4	Method					✓												✓	✓	✓											
Syahrir e Sfenrianto (2019)	4	Method	✓	✓			✓													✓												
Basri, Adnan e Baharin (2019)	3	Method					✓	✓												✓												
Souza et al. (2019)	3	Tool		✓		✓														✓												
Bakiu e Guzman (2017b)	3	Tool											✓							✓	✓											
Bauman, Liu e Tuzhilin (2016)	3	Method											✓				✓			✓												
Totals			12	14	2	2	16	5	5	7	6	4	6	2	4	3	10	3	8	24	16	2	5	6	2	5	1	1	2	1	1	3

Dados Quantitativos	Dados Qualitativo	Tipo de Avaliação	Análise de Sentimentos	Visualizações de Dados	Outros

Os trabalhos (DODD; ATHAUDA; ADAM, 2017; FIGUEIRAS, 2018; VASA et al., 2012; VOIGT; KIESLINGER; SCHÄFER, 2017a; HUSSAIN et al., 2018; MUNIM et al., 2017; BAKIU; GUZMAN, 2017a; KHAREL; FERNSTROM, 2016; BAUMAN; LIU; TUZHILIN, 2016) apresentam visualizações de dados para mostrar os resultados com as técnicas mais comuns sendo nuvem de palavras, mapas de calor, e histogramas, essa técnicas unidimensionais podem apresentar dificuldades para representar os dados multidimensionais destes cenários de análise. Apenas dois trabalhos (HUSSAIN et al., 2018; BAUMAN; LIU; TUZHILIN, 2016) usam sensores, tais como rastreadores oculares, para identificar a área que o usuário está observando.

Quase metade dos estudos (SETCHI; ASIKHIA, 2017; YU et al., 2019; MOIZER et al., 2019; SEGUIN; SCHARFF; PEDERSEN, 2019; BASRI; ADNAN; BAHARIN, 2019; SÁNCHEZ-ADAME et al., 2018; PORAT; SCHCLAR; SHAPIRA, 2013; VOIGT; KIESLINGER; SCHÄFER, 2017a; HUSSAIN et al., 2018; SYAHRIR; SFENRIANTO, 2019; LAW et al., 2009) usa alguma forma de questionário para obter informações qualitativas acerca da experiência do usuário, ou seja, não é um processo automatizado. A grande maioria realiza a avaliação *in loco*, com o especialista observando o teste no mesmo local do participante, exceto os artigos (VASA et al., 2012; HUSSAIN et al., 2018; BAUMAN; LIU; TUZHILIN, 2016), em que o especialista observa remotamente a interação do usuário.

De todos os trabalhos selecionados, o estudo mais se assemelha em características funcionais é (VASA et al., 2012) e (HUSSAIN et al., 2018). O trabalho Vasa et al. (2012) é um método para revisões de produtos e o trabalho (HUSSAIN et al., 2018) utiliza outros sensores, como BCI e eletrocardiograma, e apresenta uma análise de sentimentos com base em todos os dados de maneira multimodal. No entanto, o *dashboard* de sua aplicação apresenta apenas algumas visualizações desconectadas para apresentar dados do usuário, sem um controle refinado em relação ao aspecto temporal do teste. Ao comparar as ferramentas, o UXmood apresenta uma variedade de visualizações, e também adiciona uma metáfora do *media player* para controlar a revisão do processo de interação do usuário e é uma ferramenta para revisar tarefas do usuário.

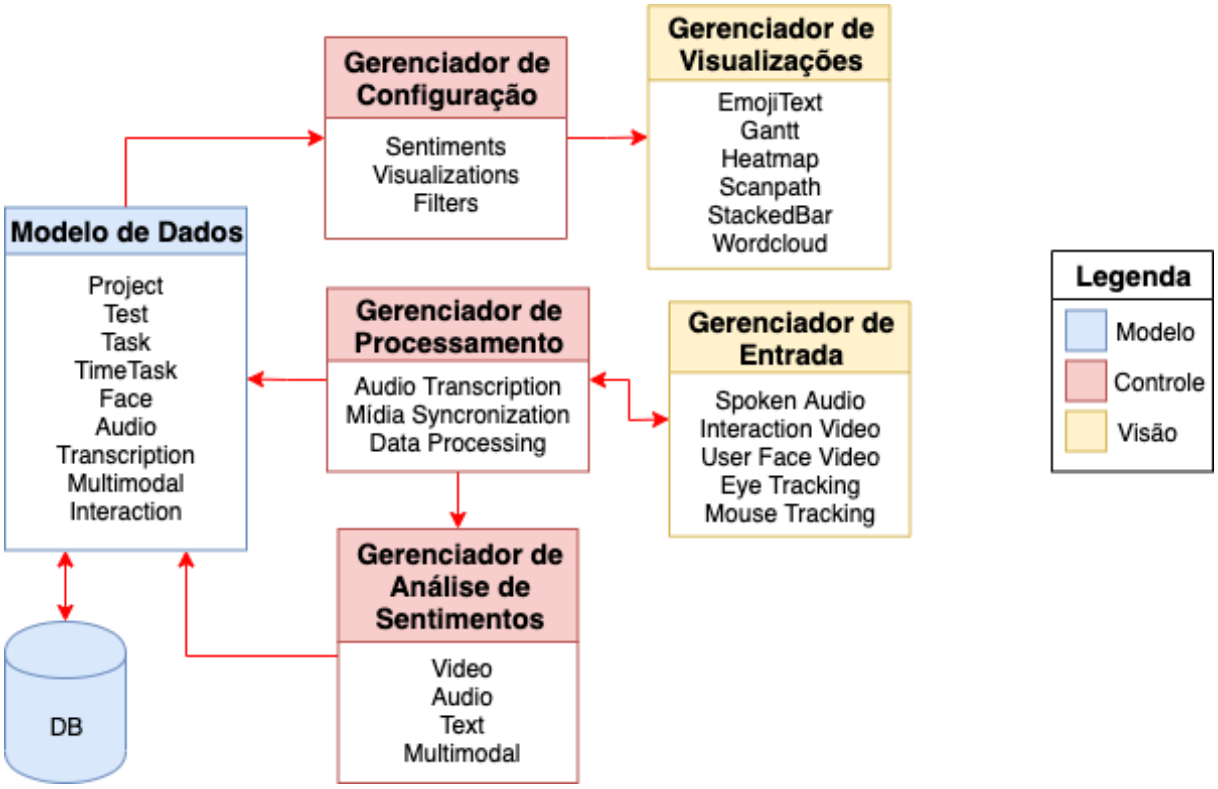
A Tabela 1 apresenta os trabalhos relacionados categorizados e agrupados. A partir dela, é evidente que poucos trabalhos utilizam técnicas de análise de sentimentos multimodais combinadas com técnicas de visualização para realizar uma avaliação qualitativa da experiência do usuário de forma integrada.

5.2 Arquitetura

A arquitetura da ferramenta UXmood é baseada no padrão de arquitetura de *software* MVC - *Model View Controller* (SOMMERVILLE, 2010) e no modelo descrito no capítulo anterior. A Figura 41 apresenta a arquitetura da ferramenta, onde os componentes

em azul pertencem ao modelo de dados, vermelho representam componentes do controle, e em amarelo componentes da visão, e subdivide-se em cinco gerenciadores de funções e recursos principais e o modelo de dados, cada um deles explicado a seguir.

Figura 41. Arquitetura da ferramenta UXmood.



FONTE: O autor

O **Modelo de Dados** é conjunto de classes que faz o encapsulamento dos dados para ser usado nos gerenciadores de Visualizações, Entrada, Configuração, Processamento e Análise de Sentimentos, além de ter as funções de acesso e persistência de dados no banco.

O **Gerenciador de Configuração** tem a responsabilidade de monitorar e controlar os detalhes sob demanda, configurações, coordenação e filtro dos dados para as visualizações. Os detalhes fazem parte dos dados mas não estão representados visualmente nas visualizações, por isso, este gerenciador fica encarregado de associar aos elementos visuais outras informações. As configurações das visualizações permitem que usuário modifique as representações de acordo com a necessidade, como definições duração de uma fixação, apresentação de polaridade ou emoção e o tipo de dados como vídeo, áudio, texto ou multimodal. Este gerenciador ainda controla a coordenação das visualizações que são os *highlights* nas visualizações. A utilização dos filtros cria subconjuntos dos dados para as visualizações. Outra função deste gerenciador é o controle da sincronização entre componentes do *dashboad* (vídeos e visualizações).

Gerenciador de Processamento é a implementação do módulo de Entrada de Dados (capítulo de Modelo 4.1), e é responsável por efetuar a transcrição do áudio, a sincronização dos dados e todo o processamento para carregamento inicial dos dados na ferramenta. Este recebe da camada de visão os dados que o usuário carregar na ferramenta, e faz a verificação da integridade, cortes, sincronização necessária, verificação dos dados ausentes, transcrição e salva os conjuntos de dados.

Gerenciador de Análise de Sentimentos é a implementação do módulo de Análise de Sentimentos Multimodal (capítulo 4.4) que tem a responsabilidade de efetuar as classificações de dados do tipo vídeo, áudio e texto, já processados pelo gerenciador de processamento. Este módulo é composto de três classificadores unimodais de análise de sentimentos, ambos redes neurais convolucionais e um método de fusão por nível de decisão que faz aplicação de pesos no resultado dos classificadores unimodais para gerar uma única classificação de sentimentos e posteriormente inferir a polarização do mesmo.

Gerenciador de Visualizações é a implementação do módulo de Visualização de Dados (capítulo 4.5) e faz parte da camada de visão do MVC. Este gerenciador é composto por um conjunto de classes que é responsável por controlar a criação das visualizações de dados e apresentar ao usuário da aplicação todas as representações visuais de acordo com as técnicas de visualizações de informações implementadas.

Gerenciador de Entrada é a parte do modelo responsável pela entrada de dados que está na camada de visão, por interagir com o usuário da aplicação para o upload dos arquivos e metadados que o usuário tem a intenção de enviar para a ferramenta, e efetuar o retorno de mensagens de sistema para os usuários.

5.3 Tecnologias

O UXmood utiliza várias tecnologias, incluindo as que executam análises de sentimentos multimodais. A análise de sentimentos subdivide-se em quatro etapas de análise - vídeo, áudio, fala transcrita e multimodal – que diferem no tipo de dados que utiliza e, consequentemente, na técnica apropriada para classificá-lo.

Na análise de vídeo, o tipo de dado é um vídeo da face do usuário, do qual três quadros (não sequenciais) por segundo são extraídos para classificação por uma CNN com a arquitetura Xception (CHOLLET, 2017). Na análise de áudio, o tipo de dado é um registro da fala do usuário durante as interações, que é pré-processado para cortar intervalos de silêncio e depois encaminhado para um algoritmo de reconhecimento de emoções baseado em CNN, embora ainda não esteja publicado, possui o código disponível em um repositório online (<https://github.com/Samarth-Tripathi/IEMOCAP-Emotion-Detection>). Na análise de texto, a ferramenta transcreve automaticamente o discurso da análise de áudio, direcionando os trechos de texto para um algoritmo de análise de sentimentos textuais

descrito no trabalho de Jain, Kumar e Fernandes (2017). Finalmente, na análise multimodal, o UXmood agrega os resultados dos classificadores anteriores por meio de uma média ponderada para realizar uma classificação multimodal, baseando-se nos pesos propostos pelo psicólogo Mehrabian e Friar (1969). O regime de treinamento será descrito a seguir (5.6).

Outras tecnologias auxiliares também estão presentes na ferramenta, como o *kernel* de visualização D3js (BOSTOCK; OGIEVETSKY; HEER, 2011), que auxilia no desenvolvimento de visualizações para a *web*, o *framework Django* (HOLOVATY; KAPLAN-MOSS, 2009) e facilita a construção da web interativa por meio de formulários. Para treinamento e utilização dos modelos inteligentes foi utilizada a biblioteca Keras com o *backend* Tensorflow 2. Além disso, as ferramentas FFMPEG (LEI; JIANG; WANG, 2013) e Pyaudio (PHAM, 2006) auxiliam na manipulação e edição de arquivos de vídeo e áudio, respectivamente, para sincronizar os vários arquivos enviados à ferramenta e executar operações de pré-processamento, como remoção de silêncio.

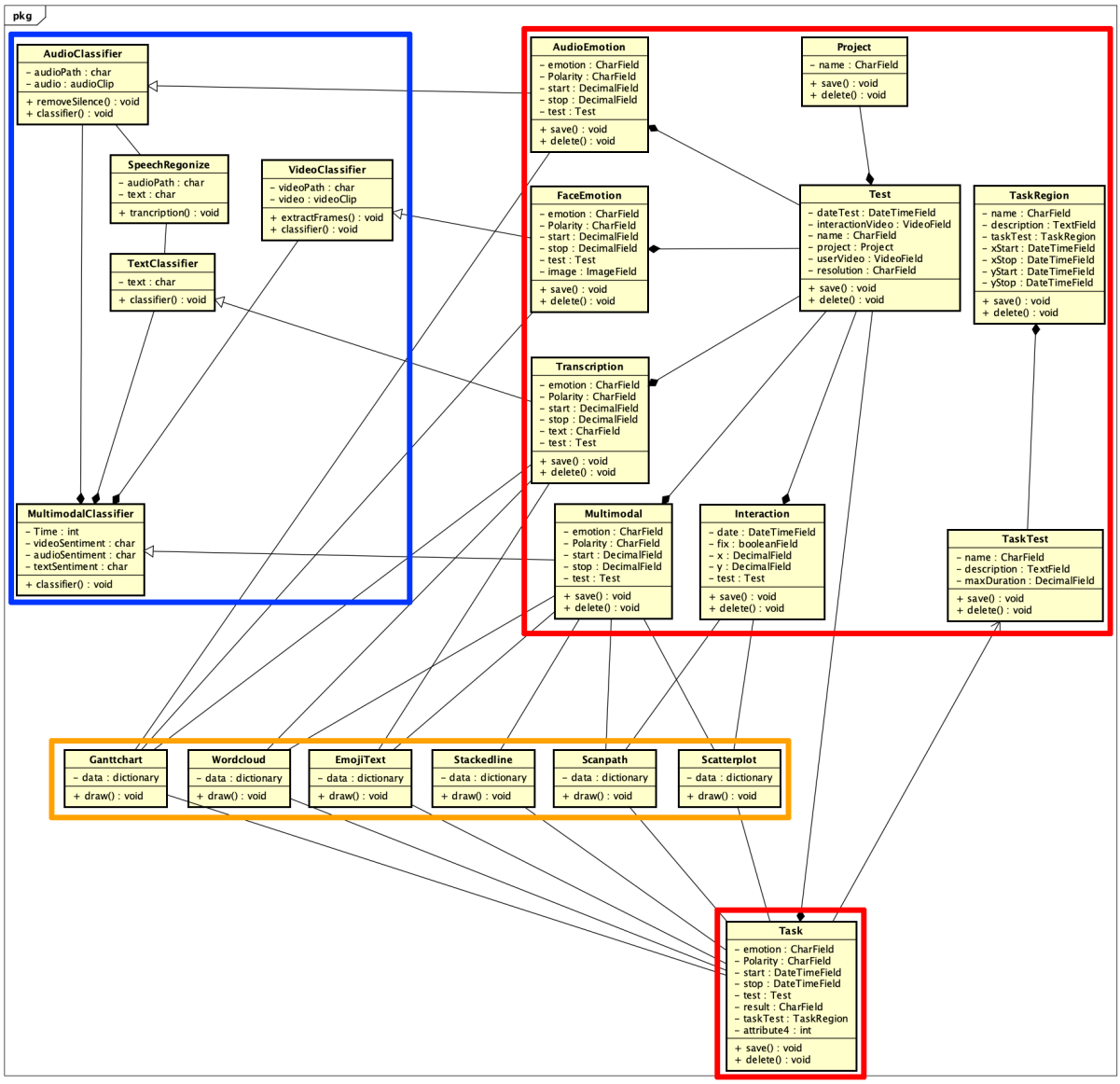
5.4 Aspectos da Implementação

A implementação do modelo se divide em três partes principais. A primeira é relacionada a dados de organização do projeto de avaliação, como: testes, tarefas, e dados para análise de sentimentos, como: áudio, transcrição, video, multimodal etc, a segunda está relacionada aos classificadores de análise de sentimentos e a terceira relacionada as visualizações de dados como: Gantt Chart, Wordcloud, EmojiText, StackedBar, Scanpath, Scatterplot e outros. A Figura 42 apresenta o diagrama de classes do protótipo.

A representação desse diagrama de classe foi simplificado para a melhor representação dos relacionamentos das classes. Seguindo a arquitetura MVC as classes de modelos apresentam características dos dados para ser instanciados, manipulado e armazenados pelas outras classes. Uma instância da classe *project* é um conjunto de instâncias de classes *tests* que por consequência é um conjunto de instâncias de classes *tasks*. Além de classes com AudioEmotions que é o modelo dos dados de classificação do áudio, FaceEmotions que é o modelo dos dados de classificação do vídeo, Transcriptions que é o modelo dos dados de transcrição e classificação dos textos transcritos e Multimodal que é o modelo dos dados de classificação multimodal.

As classes responsáveis por realizar as classificações de sentimentos são AudioClassifier, VideoClassifier, MultimodalClassifier e TextClassifier. A classe speechRegonize faz a transcrição do áudio. O conjunto de classes de borda laranja são usados para acesso aos dados e representação através de visualizações de informações.

Figura 42. Diagrama de classe do protótipo UXmood.



FONTE: O autor

5.5 Funcionalidades

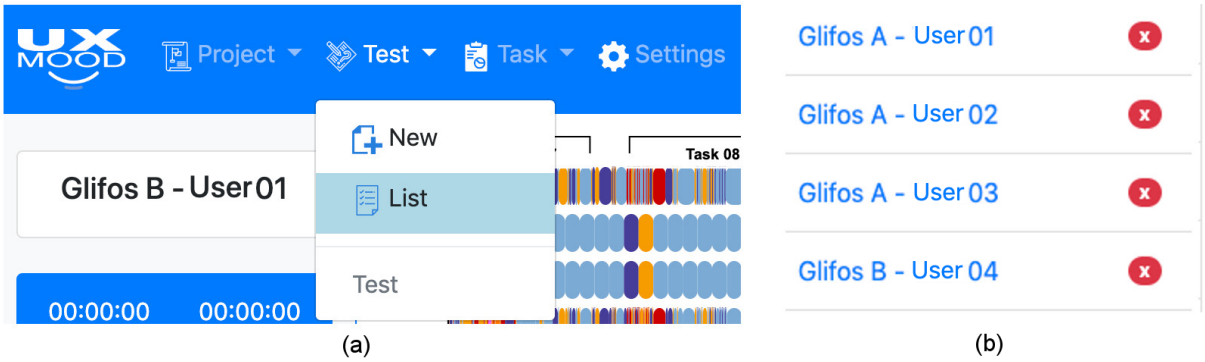
A seguir serão apresentados as principais funcionalidades do UXmood, dividido em três partes: a primeira relacionada às tarefas, testes e projetos, a segunda relacionado aos filtros, legenda, controle e detalhes sob demanda, e a terceira relacionada ao *dashboard* e as visualizações.

5.5.1 Projetos, teste e organização de tarefas

A organização de todos os dados inseridos e gerados pela ferramenta é baseada em uma hierarquia de projetos, testes e tarefas. Cada projeto pode conter vários conjuntos de testes, que, por sua vez, podem conter um conjunto de tarefas. A tarefa é a menor unidade

de uma atividade, por exemplo, em testes de usabilidade, a ação do usuário de abrir um site pode ser considerada uma tarefa. O teste passa a ser um conjunto de tarefas para um mesmo objetivos, por exemplo, ao fazer um teste de usabilidade em um navegador de internet, a ação anterior pode ser agrupada em teste. Assim, o projeto se torna um conjunto de vários testes. Essa organização possibilita executar uma análise visual por teste ou projeto. Outra forma é efetuar a organização do grupo de tarefas por participante (testes) e organização do grupo de testes por objetivo (projetos), por exemplo, todas as tarefas de um participante são agrupados em um testes e todos o testes do mesmo navegador são agrupados em um projeto. A Figura 43 exibe os menus projeto, teste e tarefa, e uma lista de testes registrados.

Figura 43. Os menus (a) para registrar testes ou tarefas em um projeto, que são exibidos como uma lista (b).

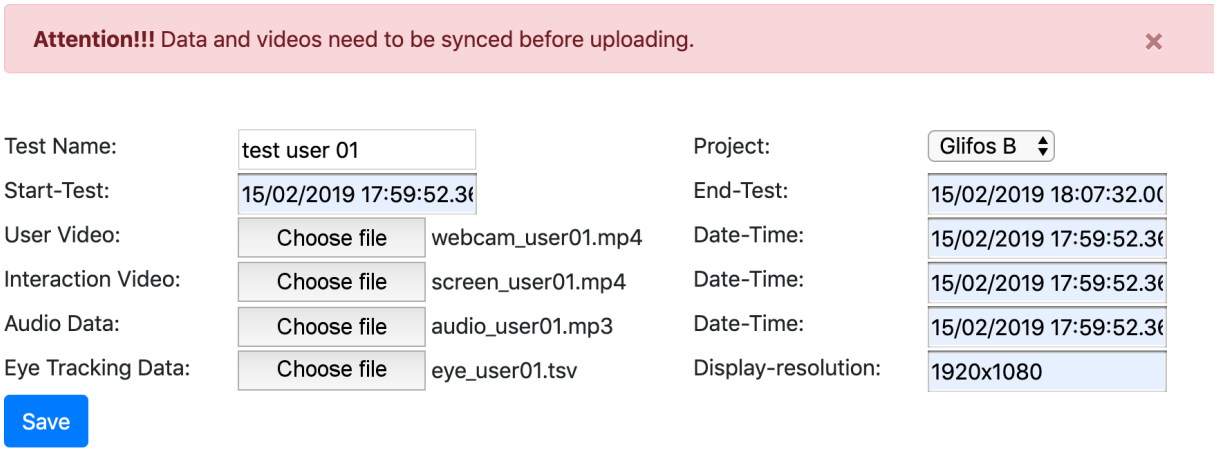


FONTE: O autor

Para registrar um novo teste na ferramenta, o usuário deve preencher o formulário mostrado na Figura 44 fornecendo o projeto ao qual o teste pertence, o nome do teste (geralmente um identificador do participante que testou o aplicativo), o horário de início e término do teste, os arquivos de mídia da interação do usuário durante o teste e os respectivos horários de início e término relativos do teste para sincronizar os arquivos. A ferramenta UXmood é capaz de processar os formatos mais comuns de arquivos de vídeo (MP4, MPG, MPEG, M4P, AVI, WMV e MOV), áudio (WAV, AIFF, BWF e MP3) e texto (CSV, TSV e TXT). Após o processamento, todos os arquivos são convertidos para MP4 vídeo, WAV áudio e CSV texto.

Os requisitos dos tipos de dados para arquivos de textos são: arquivos de rastreamento do ponteiro precisa conter coordenadas x e y e etiqueta de data e hora; rastreamento de olhos deve conter coordenadas x e y, fixação é uma duração, porém opcional, e etiqueta de data e hora, no arquivo de transcrição manual deve conter etiqueta de data e hora de início e término e o texto transcrito. A ferramenta faz a transcrição automática, mas o usuário pode optar por realizar a transcrição manualmente e realizar o carregamento do arquivo de dados na ferramenta (Figura 45).

Figura 44. Formulário de cadastro do teste.



FONTE: O autor

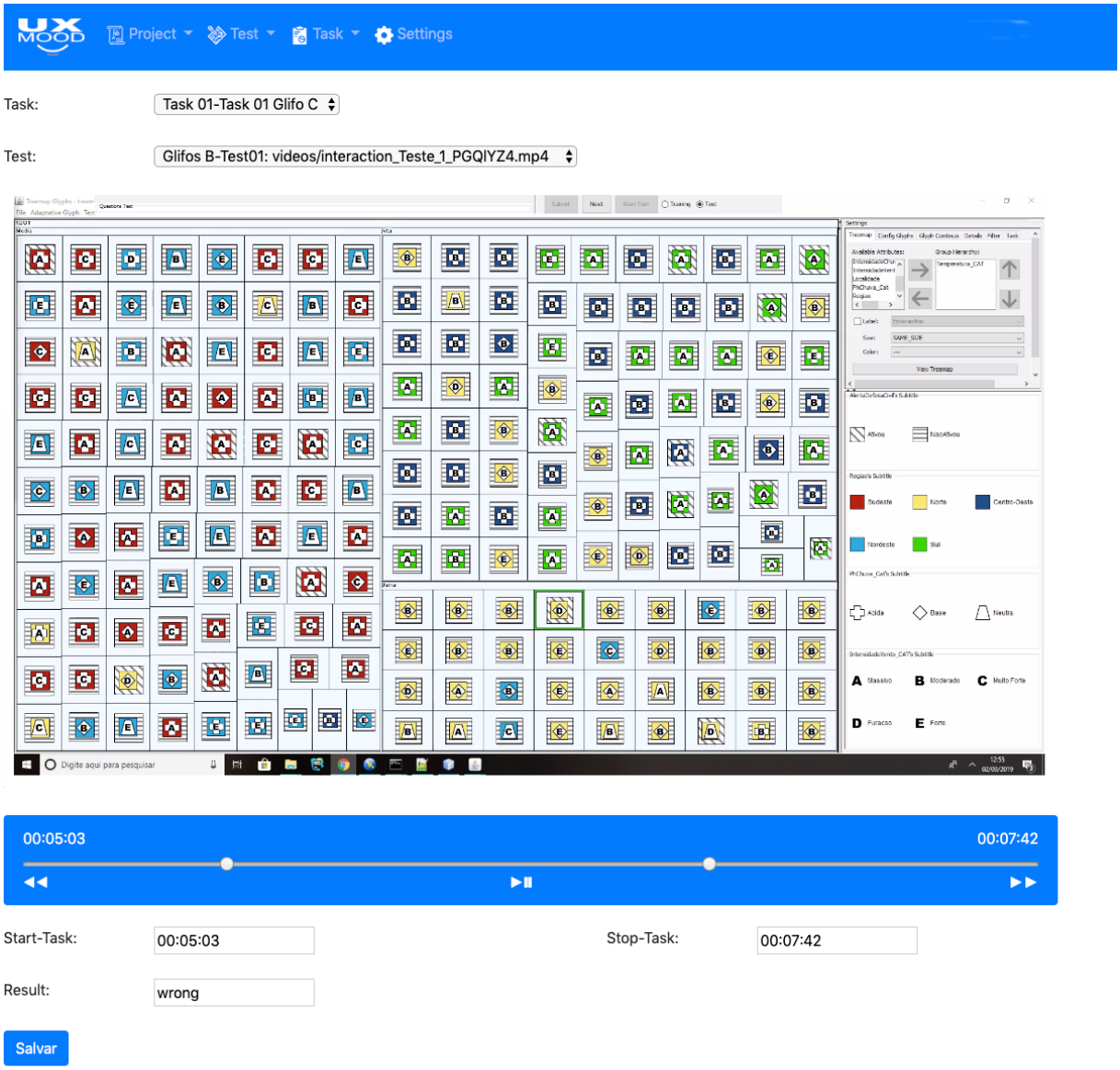
Figura 45. Exemplo de arquivo de log com dados de *Mouse / Eye Tracker*.

```
x          y          Fixo     Timestamp
1289.732155     479.389654      false   2019-12-03 18:47:28.727
1289.612709     480.874074      true    2019-12-03 18:47:28.728
1290.687059     485.798923      true    2019-12-03 18:47:28.737
1290.762727     485.918208      true    2019-12-03 18:47:28.748
1290.571223     484.974375      true    2019-12-03 18:47:28.760
1289.59411      481.949133      true    2019-12-03 18:47:28.771
1289.007055     482.938184      true    2019-12-03 18:47:28.782
1287.954406     483.131828      true    2019-12-03 18:47:28.793
1281.153294     489.952532      true    2019-12-03 18:47:28.803
1281.115347     488.77388       true    2019-12-03 18:47:28.815
1210.638208     482.197685      false   2019-12-03 18:47:28.826
1165.282242     479.047249      true    2019-12-03 18:47:28.836
1027.092221     445.004116      false   2019-12-03 18:47:28.848
769.377798      384.817976      false   2019-12-03 18:47:28.860
627.153459      356.39347       false   2019-12-03 18:47:28.870
591.300604      349.440779      true    2019-12-03 18:47:28.882
562.264473      343.782429      false   2019-12-03 18:47:28.893
536.172846      339.972134      true    2019-12-03 18:47:28.904
427.269047      317.859766      false   2019-12-03 18:47:28.915
372.16763       309.771159      false   2019-12-03 18:47:28.926
358.318234      309.37137       true    2019-12-03 18:47:28.937
```

FONTE: O autor

Para registrar as tarefas relacionadas a um teste, o especialista deve preencher o formulário mostrado na Figura 46, fornecendo: o nome da tarefa, a hora de início e término e o resultado da tarefa (por exemplo, sucesso, incorreto, tempo esgotado). O vídeo de teste está disponível nesta etapa para ajudar o especialista a definir o horário de início e término da tarefa.

Figura 46. Figura 16. Formulário de cadastro da tarefa.



FONTE: O autor

5.5.2 Legenda, Filtros, Controle, Detalhes sob demanda e Configurações

Os filtros de emoção, polaridade e corretude das tarefas estão presentes na ferramenta para que o analista possa diminuir o conjunto de dados representados nas visualizações do *dashboard*. Se apenas alguns tipos de emoções ou polarizações forem relevantes para o contexto do teste, o especialista poderá reduzir os tipos de emoções ou polarizações para restringir o foco aos dados relevantes (por exemplo, filtrar a polaridade neutra, ou a emoção raiva). Os filtros também atuam como uma legenda das cores e ícones usados nas visualizações. A Figura 47 mostra os componentes que apresentam os filtros e legendas disponíveis: emoções, polaridade e corretude das tarefas. Essas funcionalidades

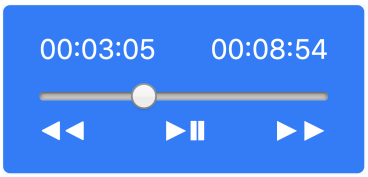
estão presente na literatura com boas práticas para ferramentas de visualização de dados (SHNEIDERMAN; PLAISANT, 2006).

Figura 47. Legendas de Emoção, polaridade e resposta; as legendas atuam como filtros nas visualização.

FONTE: O autor

Como a ferramenta apresenta vídeos e animações, a interface também contém um controle que segue a metáfora do *media player* (Figura 48), o qual o especialista pode pausar, retomar, avançar rapidamente e retroceder a gravação. Este controle permite o relacionamento explícito entre diversas visualizações dados através da sincronização e componentes visuais de destaque, as visualizações também incluem incluindo logs de interação, áudio e vídeos do teste.

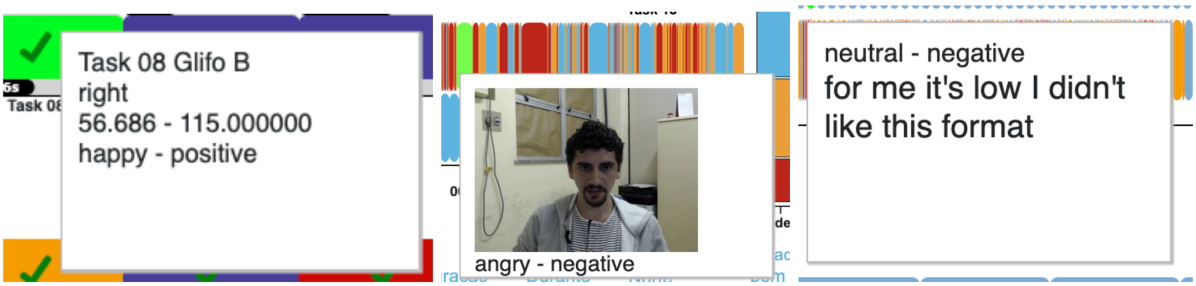
Figura 48. O controle que manipula de forma síncrona as imagens do teste..



FONTE: O autor

Outro tipo de interação são os detalhes sob demanda, que apresenta informações que não estão disponíveis visualmente nas visualizações. Os detalhes sob demanda aparecem como uma dica da ferramenta ao passar o mouse sobre um elemento visual, o que também aciona um destaque no mesmo elemento nas outras visualizações que é uma interação de *brushing*. Por exemplo, a Figura 49 mostra os detalhes ao passar o mouse sobre uma representação de sentimento, fornecendo detalhes da tarefa que estava sendo executada (Figura 49 a), o quadro de vídeo daquele momento (Figura 49 b) ou a fala transcrita do usuário (Figura 49 c) para que o especialista possa entender melhor o que causou o sentimento.

Figura 49. Detalhes sob demanda que fornecem informações adicionais sobre a tarefa (a) o vídeo (b) e a fala transcrita (c).

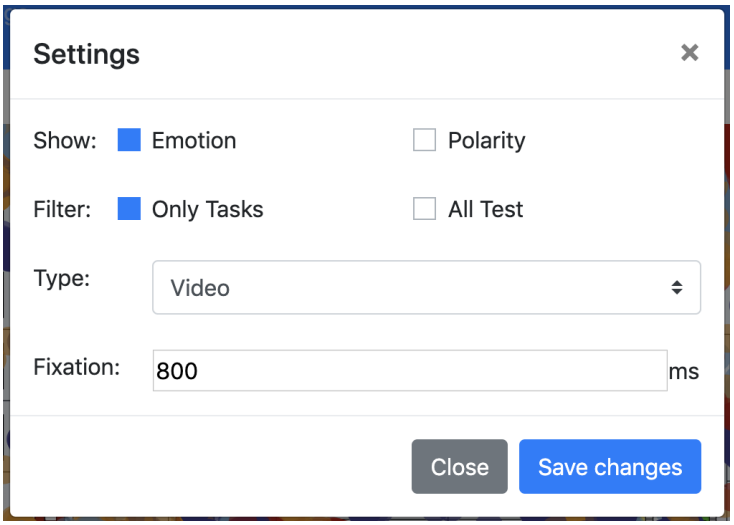


FONTE: O autor

Configuração do painel. Um menu de configuração (Figura 50) permite ao especialista alterar algumas configurações que afetam como a ferramenta apresenta os dados e quais visualizações são usadas. São elas:

- Escolhe a saída do reconhecimento do tipo de emoção (raiva, medo, feliz, triste, surpresa, nojo, desprezo e neutro) ou polaridade (positivo, negativo e neutro).
- Escolhe exibir apenas os intervalos relacionados às tarefas, desconsiderando o intervalo entre eles, se desejado.
- Define quais tipos de análise de sentimentos serão exibidos, escolhendo entre áudio, vídeo, texto e multimodal.
- Defina a duração do tempo que caracteriza uma fixação no rastreamento ocular.

Figura 50. Tela de configuração do UXmood.



FONTE: O autor

5.5.3 *Dashboard* de Visualizações

De maneira geral, o *dashboard* é uma interface gráfica composta por um conjunto de elementos que permite o especialista veja momentos específicos de um teste de avaliação por meio de diferentes visualizações, cada uma focada em um aspecto diferente dos dados. Os elementos que compõem o dashboard são técnicas de visualizações de dados, legenda e mecanismos de interação que foram selecionadas na literatura por melhor representar e manipular os conjuntos de dados que estão presentes na ferramenta (KUCHER; PARADIS; KERREN, 2018a; HOLTZ; HEALY, 2018; FUCHS et al., 2013; SHNEIDERMAN; PLAISANT, 2006).

As visualizações do UXmood apresentam as seguintes diretrizes: suportar a representação visual dos seguintes tipos de dados: rastreamento do olhar, rastreamento do ponteiro, de sentimentos para dados de vídeo, áudio, texto através de codificações de cor (KUCHER; PARADIS; KERREN, 2018a) e ícone; apresentar o contexto temporal da interação do usuário (KUCHER; PARADIS; KERREN, 2018a), auxiliar tarefas de visão geral do processo de interação do usuário, comparação entre participantes ou grupo de participantes (SHNEIDERMAN; PLAISANT, 2006), entre outras. Ao todo foram selecionadas cinco visualizações, sendo elas *Gantt Chart* (CLARK, 1922), scanpath (STARK et al., 1979), *wordcloud* (CUI et al., 2010), gráfico de barras empilhadas (ALSBURY; BECERRA, 2012), e gráfico de dispersão (FOREMAN-MACKEY, 2016). Todas devem correlacionar dados sobre sentimentos extraído dos arquivos de mídia e estarem relacionadas temporalmente entre si, de tal modo que possam ser complementares. Além disso, uma técnica foi proposta, o *emojiText* que tem a função de representar dados de texto transcrito com dados de sentimentos.

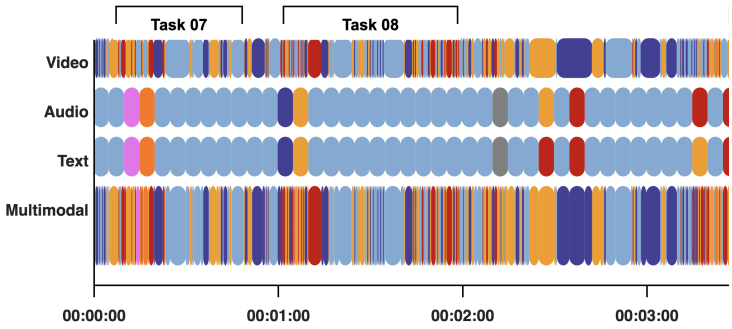
O *Gantt chart* é a visualização principal do *dashboard*, as demais visualizações estão relacionadas temporalmente e pela codificação visual do sentimento a. A visualização organiza temporalmente e por tarefa de testes a análise de sentimentos dos dados. Cada mídia tem sua visualização de análise de sentimento, bem como a análise multimodal, com alguns dados quantitativos associados: erro/acerto da tarefa e o tempo gasto na mesma.

Este atua como uma linha do tempo de sentimentos, apresentando a tempo e o sentimento de cada tipo de entrada (vídeo, áudio e texto) e uma análise combinada (multimodal). A Figura 51 apresenta os sentimentos classificados em emoções (representados pela cor dos segmentos). O tempo e a duração de cada tarefa também são mostrados na parte superior dos segmentos.

O *dashboard* tem a opção de apresentar os dados de todo o teste (Figura 52 a) ou filtrar apenas as partes que correspondem às tarefas (Figura 52 b).

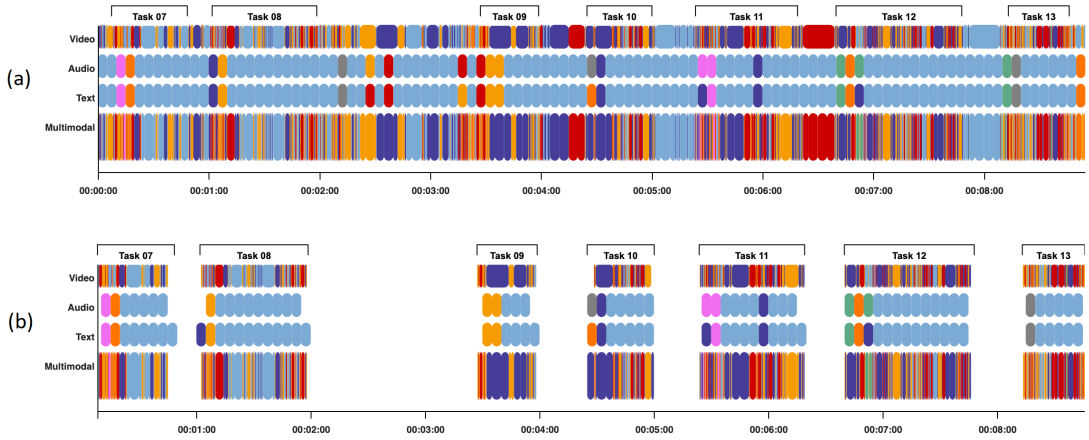
O *Gantt Chart* também pode ser configurado para agrupar os segmentos por tarefa e mostrar apenas o sentimento predominante, como mostra a Figura 53. Nesta configuração,

Figura 51. O gráfico de Gantt mostra o sentimento classificado de segmentos de tempo para cada tipo de entrada; o tempo e a duração das tarefas são mostrados na parte superior do gráfico.



FONTE: O autor

Figura 52. O gráfico de Gantt configurado para exibir informações emocionais de todo o teste (a) ou apenas dos segmentos de tarefas (b).



FONTE: O autor

os segmentos do *Gantt Chart* são as tarefas com o tempo de resolução e os resultados (correto, incorreto, tempo encerrado) de cada um, que também são apresentados por meio de uma barra e ícone de progresso, respectivamente.

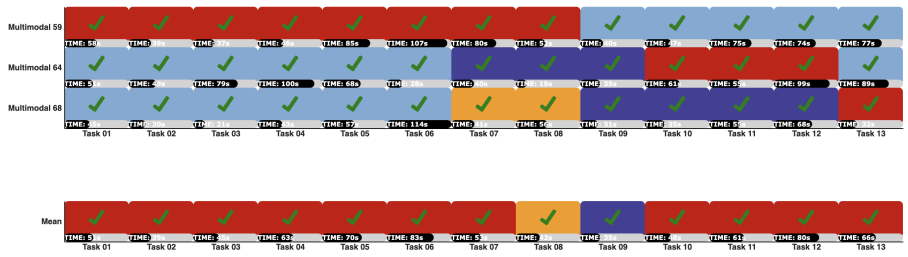
Figura 53. Gráfico de Gantt configurado para representar tarefas como segmentos: a visualização mostra a duração, resultado e sentimento predominante das tarefas.

FONTE: O autor

Outra configuração do *Gantt Chart* é a análise conjunta de participantes entre diferentes testes. Essa configuração permite uma comparação direta de testes em termos de duração, resultado e sentimento predominante. A Figura 54 ilustra essa configuração:

o especialista configurou o *Gantt Chart* para comparar o sentimento predominante na classificação multimodal de vários participantes. Essa configuração também apresenta um gráfico que representa o sentimento médio dos participantes em cada tarefa.

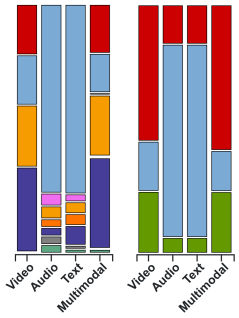
Figura 54. Gráfico de Gantt configurado para comparar diferentes testes.



FONTE: O autor

O gráfico de barras empilhadas apresenta a proporção total de cada sentimento classificado (em emoções ou polaridade), como mostra a Figura 55. Esta visualização mostra uma visão geral de como o participante se sentiu ao longo de todo o teste.

Figura 55. Um gráfico de barras empilhadas de emoções classificadas (esquerda) e polaridade (direita).

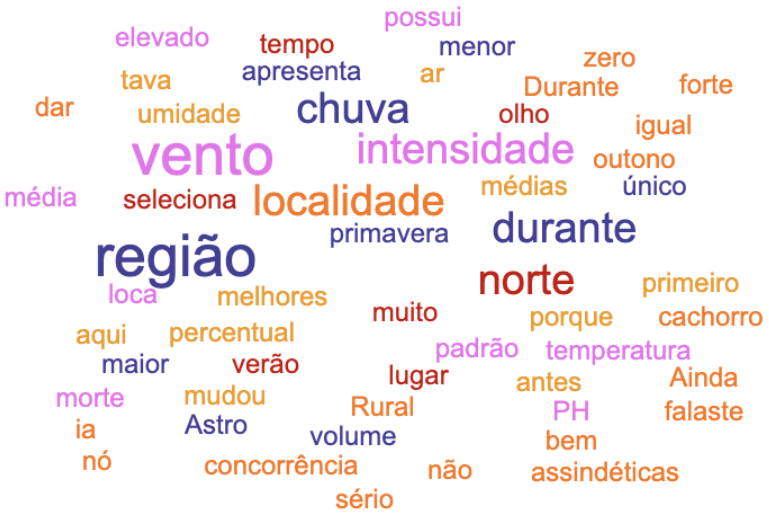


FONTE: O autor

A wordcloud apresenta as principais palavras faladas pelo participante durante a execução das tarefas, conforme mostrado na Figura 56. O tamanho da palavra representa sua frequência e a cor leva em consideração o sentimento do participante durante o contexto em que foi citado, que é determinado pelo classificador de texto. Por exemplo, na Figura 56, o participante citou a palavra ”região” em oito frases, a maioria delas classificadas como nojo, portanto esta palavra está associada a essa emoção.

Scanpath e gráficos de dispersão são usados para apresentar os dados dos registros de interação dos participantes, como rastreamento de mouse ou rastreamento ocular. A visualização do gráfico é um caminho de varredura (Figura 57) e é adequada para realizar o rastreamento ocular a fim de mostrar as fixações e pontos de vista do

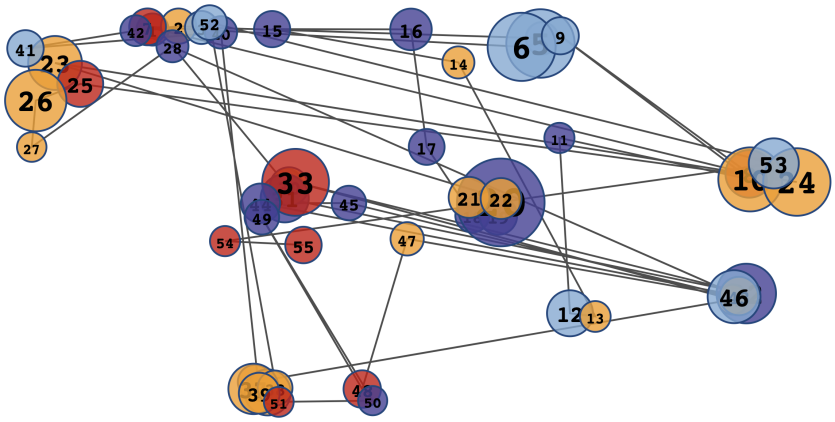
Figura 56. Wordcloud de palavras faladas pelo usuário.



FONTE: O autor

participante durante os testes, fixações são regiões onde foi fixado o olhar durante um tempo e pontos de vistas são o local que foi olhado.

Figura 57. Um gráfico de caminho de verificação mostrando a fixação do usuário com o sentimento inferido associado.

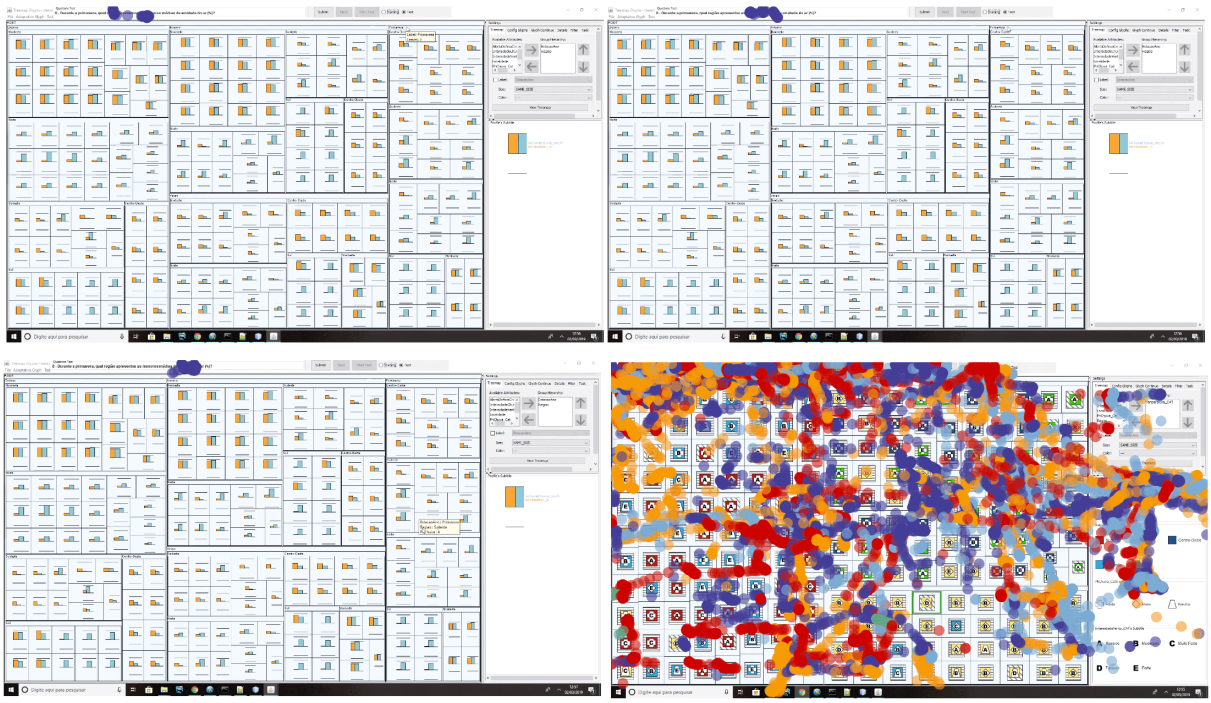


FONTE: O autor

O gráfico de dispersão (Figura 58) apresenta os pontos registrados no log pelo rastreamento ocular ou do ponteiro, por exemplo. Além disso, é desenhado em uma camada acima do vídeo da tela com as interações dos participantes, permitindo a correlação direta e sincronizada das coordenadas do log e o conteúdo na tela. Nas duas visualizações, as cores representam o sentimento associado durante esse momento, determinado pelo classificador de vídeo, para que o especialista saiba onde o participante estava interagindo e o sentimento associado à interação.

Ícones. Uma das visualizações no painel do UXmood é uma gravação de vídeo da

Figura 58. Um gráfico de dispersão com os pontos interagidos pelo usuário (rastreamento de mouse ou olho), colorido pela emoção associada no momento da interação.



FONTE: O autor

face do usuário sincronizada durante a interação. Além dela, um ícone de sentimento é exibido no canto superior esquerdo para indicar o sentimento que o classificador de vídeo deduziu naquele momento, conforme mostrado a figura 59.

Figura 59. Vídeo da face do usuário com ícone de sentimento.

FONTE: O autor

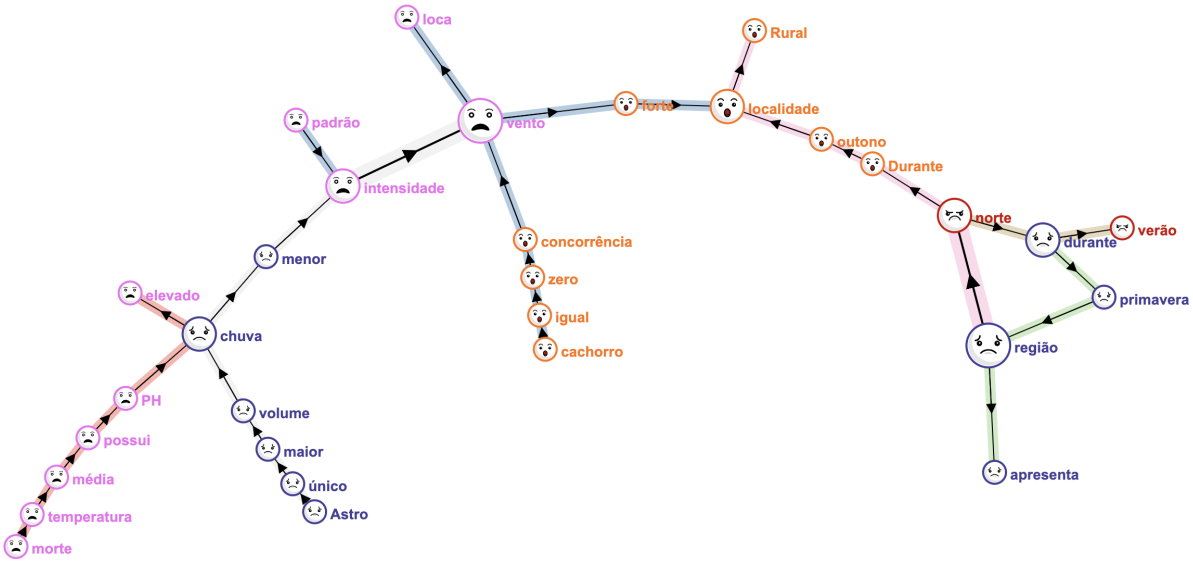
EmojiText é uma nova técnica de visualização de informações para representar dados transcritos de análise de fala e sentimento (FRANCO et al., 2019). Essa visualização (Figura 60) mostra as principais palavras de uma frase - ou seja, apenas os substantivos,

pronomes, verbos e adjetivos - representando a emoção ou polaridade associada a cada palavra enquanto preserva seu significado e contexto.

O EmojiText apresenta as palavras em um grafo direcional que indica a sequência em que foram faladas. Cada nó representa uma palavra e o emoji no centro do nó representa o sentimento predominante de todas as frases que a palavra apareceu. Dessa forma, se uma palavra aparecer em mais de uma frase, o sentimento associado será o sentimento médio nos diferentes contextos.

O tamanho dos nós e das arestas representa a frequência em que a sequência de palavras apareceu no texto; portanto, nós maiores representam palavras repetidas e, consequentemente, têm mais frases conectadas a ele. A cor das arestas representa frases completas, portanto, uma alteração na cor indica o final de uma frase e o início de outra.

Figura 60. EmojiText, uma técnica de visualização para representar sentimentos em textos.



FONTE: O autor

5.6 Classificação de sentimentos

A escala de emoções usada nesta tese foi o conjunto de emoções primárias proposto por Ekman (1999), composto por seis emoções básicas: raiva, medo, tristeza, felicidade, surpresa e neutro. Este trabalho também incluiu uma emoção neutra para quando o classificador não pode identificar uma das emoções primárias nas expressões do usuário, bem como quando houver condições que prejudicam a captura.

Para que o UXmood deduza o estado emocional do participante durante o teste em um determinado momento ou tarefa, os classificadores de análise de sentimentos são usados

para vídeo, áudio, texto e uma combinação deles (multimodal). A seguir, são apresentados as bases de dados, arquiteturas de CNN e descrição do treinamento.

5.6.1 Bases de dados

As bases de dados utilizadas abrangem os classificadores unimodais, de imagem, áudio e texto. A base FER2013 - Facial Expression Recognition 2013 foi criado por Pierre Luc Carrier e Aaron Courville utilizando a API de pesquisa de imagens do Google para procurar imagens de rostos que correspondam a um conjunto de 184 palavras-chave relacionadas à emoção, como ”feliz”, ”enfurecido” etc. Essas palavras-chave foram combinadas com outras palavras relacionadas a sexo, idade ou etnia, para obter quase 600 strings que foram usadas como consultas de pesquisa de imagens faciais. A partir disso, foram selecionadas as primeiras 1000 imagens retornadas por consulta e utilizadas no estágio de processamento. Nesse estágio o reconhecimento de faces foi utilizado para obter regiões delimitadoras em torno de cada face nas imagens coletadas. Além disso, rotuladores humanos rejeitaram imagens rotuladas incorretamente, corrigiram o corte, se necessário, e filtraram algumas imagens duplicadas. As imagens cortadas aprovadas foram redimensionadas para 48 x 48 pixels e convertidas para escala de cinza. Com isso, o conjunto de dados resultante contém 35887 imagens, com 4953 imagens de ”raiva”, 547 imagens de ”nojo” (não utilizadas no treinamento), 5121 imagens de ”medo”, 8989 imagens de ”felicidade”, 6089 imagens de ”tristeza”, 4002 imagens de ”tristeza”, 4002 imagens de ”surpresa” e 6198 imagens ”neutras”.

A base de dados *Voice Emotion Recognition dataBase in Portuguese Language* (VERBO) foi desenvolvida por pesquisadores da USP com 1167 gravações de voz de 12 atores 6 do sexo feminino e 6 do sexo masculino, todos entre 2 à 23 anos de experiência e são atores de grupo de teatro de diferente regiões do brasil. O material linguístico utilizado é composto de 14 frases que utilizam todos os fonemas da língua portuguesa, a criação das frases foi realizada por especialista em linguística. A gravação realizada expressa as sete emoções básicas: alegria, raiva, tristeza, surpresa, medo, nojo e desprezo. Todas as etapas de criação da base de dados foram baseadas no trabalho (NETO et al., 2018). A base de dados foi validada por juízes especialistas e obteve 76% de taxa de concordância.

O conjunto de dados denominado de *International Survey on Emotion Antecedents and Reactions* (ISEAR) coletado por Klaus R. Scherer e Harald Wallbott contém sete emoções principais: alegria, medo, raiva, tristeza, nojo, vergonha e culpa. O banco de dados contém 7666 frases classificadas e coletado de 1096 participantes com diferentes antecedentes culturais que responderam perguntas sobre sete emoções: raiva, nojo, medo, tristeza, vergonha, alegria e culpa.

5.6.2 Arquiteturas de CNN

Todo o treino baseado em imagens foi realizado utilizando arquiteturas de CNN vencedoras do ILSVRC (RUSSAKOVSKY et al., 2015), que apresentam resultados do estado da arte em problemas complexos em visão computacional. A VGG (SIMONYAN; ZISSERMAN, 2014) é uma arquitetura que popularizou o uso de filtros muito pequenos ($3x3$) nas camadas de convolução, permitindo menos parâmetros sem perder o campo receptivo, e foi utilizado para o treinamento de modelos mais profundos, com modelos de 16 e 19 camadas.

A ResNet (HE et al., 2016) é um tipo de arquitetura que utiliza conexões residuais, que além de ligar o resultado da computação de uma camada a seguinte, também pula algumas camadas e dispõe desse resultado novamente, criando saltos entre as ativações, o número de camadas de uma rede como essa varia entre 50 e 152.

MobileNets (SANDLER et al., 2018) são um conjunto de arquiteturas que utiliza convoluções separáveis, no sentido de separar as dimensões de profundidade ao invés de uni-las no processamento, e utilizam conexões residuais para garantir eficiência no treinamento.

A Inception (SZEGEDY et al., 2017) é o primeiro tipo de arquitetura que utiliza os módulos Inception. Esses módulos utilizam várias operações de convolução com tamanhos de filtros diferentes no mesmo nível da rede, e concatena o resultado de todas as computações. Além de impedir o problema do gradiente desaparecer, por não precisar que a rede seja tão profunda, também ajuda no campo receptivo da rede, já que filtros maiores podem aprender informações globais e filtros menores informações locais. Para ajudar o fluxo do gradiente durante o treino, a rede dispõe de saídas intermediárias, que computa o custo da função objetivo e adicionam o gradiente no fluxo que vem das camadas finais. Uma modificação dessa arquitetura é a Xception (CHOLLET, 2017), que utiliza convoluções separáveis ao invés das convoluções tradicionais.

5.6.3 Treinamento e Classificação por CNN

Para a classificação de sentimentos para o vídeo de rosto do participante, três quadros não sequências por segundo (3 frames em 24) são extraídos não sequencialmente de forma aleatória, já que Yan et al. (YAN et al., 2013) afirma que as microexpressões faciais rápidas duram de 260 ms a 500 ms. Esses quadros são classificados utilizando o melhor modelo Xception treinado, que será descrito abaixo. As emoções que esse classificador é capaz de classificar são: raiva, repulsa, medo, alegria, tristeza, surpresa, desprezo e neutralidade. Para avaliar este classificador, testou-se sua precisão usando a base FER2013 para treino e teste (LIVINGSTONE; RUSSO, 2018).

Neste trabalho avaliou-se o uso das arquiteturas VGG19 (indicando 19 camadas),

Tabela 2. Melhores resultados das CNNs. Xception foi retreinada com mais 150 épocas. Cada arquitetura gerou 12 modelos, totalizando 49 modelos treinados (48 + 1 final), com melhor resultado 68%.

Modelo	**Épocas**	**Taxa de aprendizado**	**Decaimento**	**Acurácia**	**Razão Treino / Validação**
Xception	50	10^-6	10^-6	0.54	0.001
Xception	200	10^-6	10^-6	0.68	-
ResNet	50	10^-6	10^-6	0.47	0.001
MobileNet	50	10^-6	10^-6	0.56	0.1
VGG	50	10^-6	10^-6	0.63	0.035

FONTE: O autor

Xception, ResNet152 (indicando 152 camadas) e MobileNet para treinamento em imagens da base de imagens FER2013, procurando classificar as imagens por sentimento. O pré-processamento das imagens foi em termos de escala por um fator de 2 ($48x48 * 2 = 96x96$), e durante o treino as imagens foram aumentadas em termos de inversão horizontal (metade das imagens), leve mudança de zoom e *shear* (fator de 0.2).

Todas as arquiteturas seguiram o mesmo regime de treinamento, gerando 12 modelos inteligentes para cada arquitetura, com seleção dos parâmetros taxa de aprendizado, com valores $10^-4, 10^-5, 10^-6, 10^-7$ e decaimento de peso com valores de $10^-4, 10^-5, 10^-6$. Além disso, todas as redes foram iniciadas com os pesos já treinados do desafio ImageNet (RUSSAKOVSKY et al., 2015). O treinamento foi realizado utilizando a biblioteca Keras (CHOLLET et al., 2015) com o *backend* do Tensorflow 2 (ABADI et al., 2015), em um computador com Intel I5, 16GB de RAM e placa de vídeo GTX 1060 6GB. Foram todas treinadas com 50 épocas, em uma divisão da base de dados 70/30 de treino e validação. A média da distância de Manhattan foi utilizada entre resultado do treino e validação foi utilizada para dar continuidade ao treino dos melhores modelos, com o critério de que a menor distância indica que a acurácia pode continuar o treinamento, já que está generalizando melhor que os outros. Os melhores resultados com seus respectivos parâmetros estão na Tabela 2.

O melhor modelo, Xception, continuou com o treinamento com mais 150 épocas e alcançou acurácia de 68%. Vale ressaltar que mesmo não sendo diretamente comparáveis, resultados para a mesma tarefa em outras bases não diferem tanto do resultado alcançado aqui, variando entre 52% e 80% (CAMBRIA et al., 2017; CORNEANU et al., 2016; PORIA; HUSSAIN; CAMBRIA, 2018). À medida que classificadores mais robustos são desenvolvidos, essa opção pode ser revisada e atualizada em versões posteriores do UXmood.

Depois que as emoções dos frames de vídeo são classificadas, o UXmood realiza um agrupamento linear de emoções para identificar sentimentos possivelmente não classificados. Se duas classificações únicas de um sentimento são cercadas por duas ou mais classificações

consecutivas de outro sentimento, então o sentimento da classificação única é substituído pelos outros, consentindo com a duração das micro-expressões. Por exemplo, se um intervalo de dois segundos inicialmente tem um segundo de emoção A, seguido por 330 ms de emoção B, e o restante sendo emoção A, então todo intervalo será considerado como emoção A, porque a emoção B foi classificada em um intervalo muito menor no meio de intervalos de emoção.

Para realizar a análise de áudio, este pode ser carregado em um arquivo separado ou extraído automaticamente do vídeo do usuário, se nenhum arquivo de áudio for fornecido. O UXmood divide esse áudio em blocos de cinco segundos e os envia para um classificador de análise de sentimentos. Dois classificadores de áudio foram treinados na base RAVDESS, um SVM (máquina de vetores de suporte), treinado por Giannakopoulos et al. (GIANNAKOPOULOS; PIKRAKIS; THEODORIDIS, 2009), com precisão de 66,34% e uma CNN, com precisão 88,64%, com modelo treinado online disponível em (<https://github.com/MITESHPUTHRANNEU/Speech-Emotion-Analyzer>).

Pedaços de áudio também são usados em um algoritmo de reconhecimento de fala a ser transcrito. Foi utilizada a API do Google (SCHUSTER, 2010) para transcrever segmentos de áudio de 5 segundos cada em texto. Os segmentos de texto são classificados por uma rede neural proposta por Tripathi e Beigi (TRIPATHI; BEIGI, 2018), que possui uma CNN treinada pronta para uso disponível em um repositório online (<https://github.com/Samarth-Tripathi/IEMOCAP-Emotion-Detection>), com acurácia de 62%. Além do classificador, o dicionário SentiwordNet (BACCIANELLA; ESULI; SEBASTIANI, 2010) é usado para identificar a polaridade, com classificadores Rocchio ternários, com acurácia de 75

Como o idioma falado pelos participantes deste teste era o português, houve um novo treinamento para os classificadores de áudio e texto antes do processamento dos dados. Para o classificador de áudio, utilizou-se o conjunto de dados VERBO (NETO et al., 2018) e obteve-se uma precisão de 78,64%. Utilizou-se a biblioteca LibROSA (MCFEE et al., 2015) para realizar a extração de características, nomeadamente o MFCC, que são coeficientes que representam o espectro de força do som, treinando uma CNN com os parâmetros sugeridos pelo autor do modelo. Para o classificador de texto, foi utilizada uma versão traduzida do conjunto de dados ISEAR (SCHERER; WALLBOTT, 1990) disponível on-line (https://www.kaggle.com/shrivastava/isears-dataset) e obtive-se a precisão de 70,4%. A API de transcrição detecta automaticamente o idioma, portanto, nenhuma alteração adicional é necessária.

A análise multimodal de sentimentos combina os resultados dos classificadores anteriores usando a regra 7-38-55 descrita por Mehrabian (MEHRABIAN; FRIAR, 1969), que sugere que apenas 7% da comunicação ocorre em palavras faladas (conteúdo da fala), 38% é na maneira como as palavras são faladas (recursos de áudio como tom e volume)

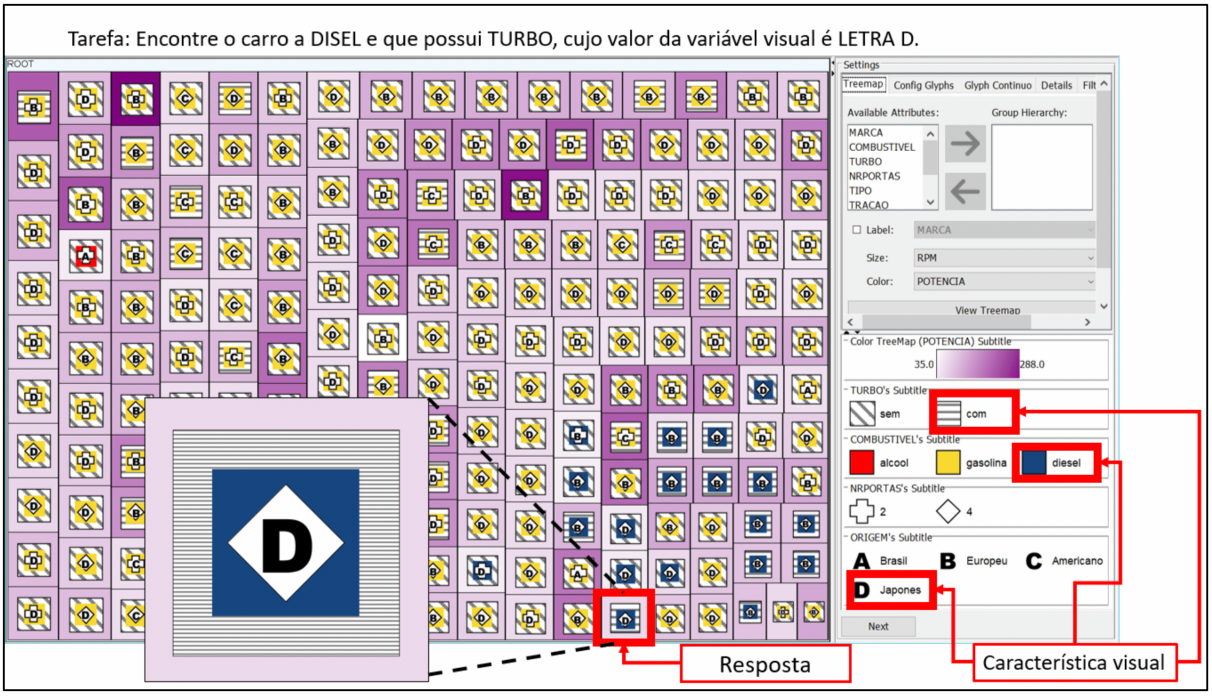
e 55% estão nas atitudes e expressões faciais. O UXmood usa essas porcentagens como pesos para compor as classificações de texto, áudio e vídeo em uma nova classificação multimodal.

O computador usado para testar os algoritmos foi um laptop com um processador Intel i7 de 7a geração, 16GB de RAM e 1,5 GB de memória de vídeo dedicada. O tempo de processamento foi aproximadamente duas vezes a duração do vídeo. Portanto, se os dados de entrada forem um vídeo de 3 minutos (com um áudio correspondente de 3 minutos), os dados estarão completamente disponíveis para análise após 6 minutos.

5.6.4 Cenários de Uso

Como cenário de uso para avaliação do UXmood, dados reais de testes com usuários para avaliação de um software de visualização da informação são utilizados, pertencentes ao trabalho (Soares et al., 2018). Trata-se da avaliação de uma técnica em árvore de visualização de dados com adição de glifos multidimensionais. A Figura 61 foi extraída deste trabalho para exemplificar um cenário de uso.

Figura 61. Visualização de cenário de uso.

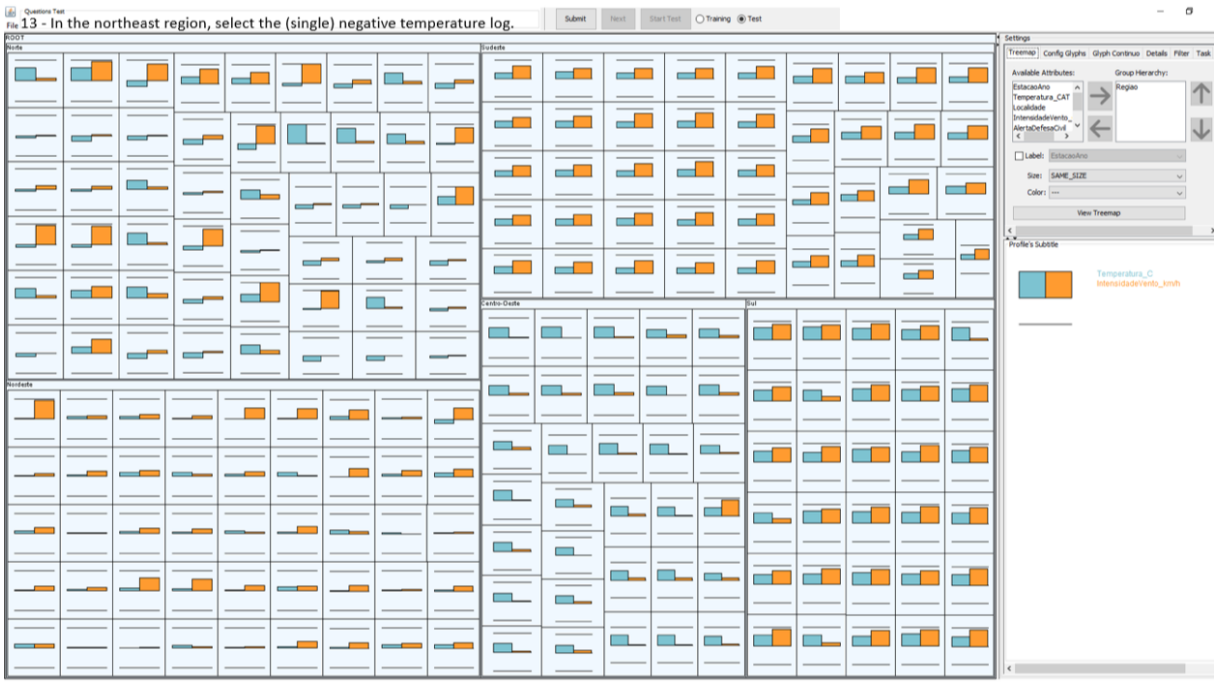


FONTE: O autor

A técnica de visualização da informação principal utilizada no testes conduzidos em (Soares et al., 2018) é a Treemap (SHNEIDERMAN; WATTENBERG, 2001), que é uma técnica que possui as seguintes características: se adaptar ao espaço visual disponível, representa itens de dados por retângulos, e é possível mudar o layout da visualização realizando agrupamento hierárquico dos itens visuais (Figura 62). A técnica treemap foi

adicionado um conjunto de dados, organizadas em um glifo através de textura, cor, forma e uma letra, aumentando a capacidade de representação da técnica Treemap. Um glifo pode ser observado com mais detalhes na Figura 61, no item destacado como resposta.

Figura 62. Visualização de cenário de uso.



FONTE: O autor

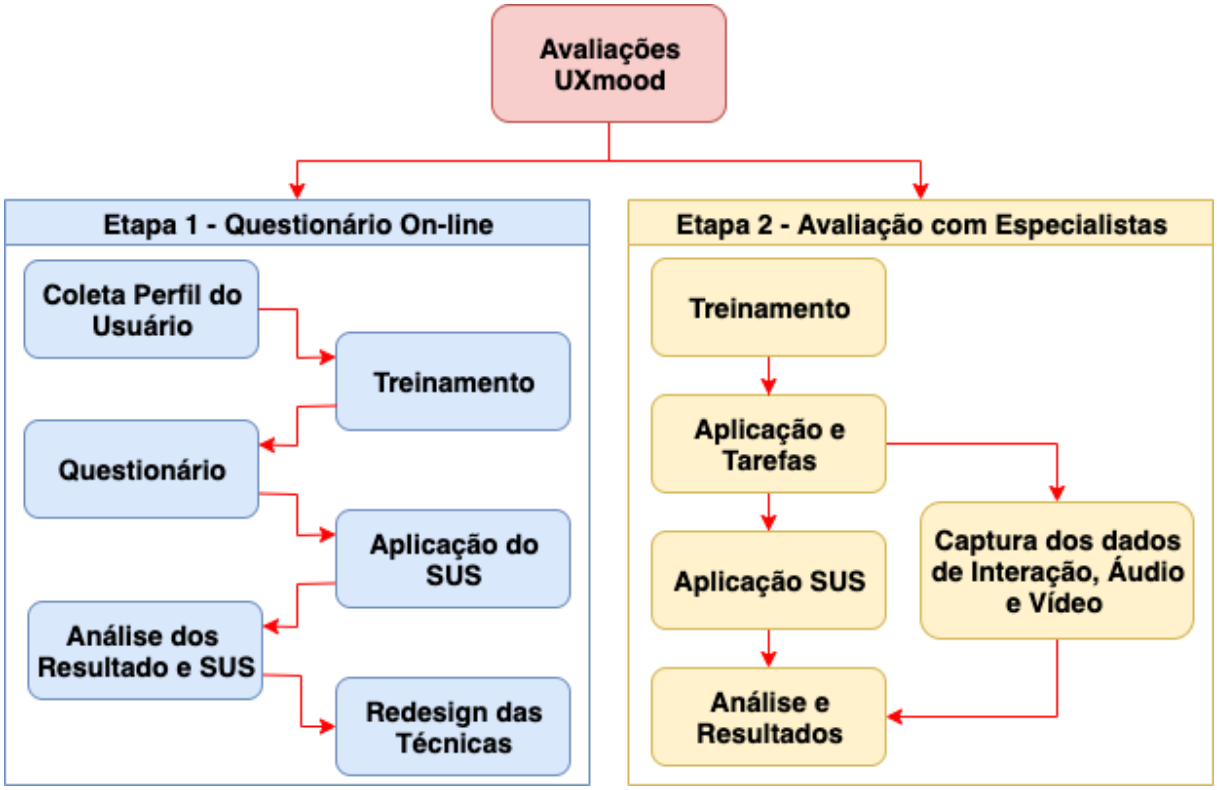
Os testes conduzidos por(Soares et al., 2018) apresentaram algumas variações de cenários de teste, três ao total, são eles: a técnica treemap com adição de glifos mas sem hierarquias, treemap com hierarquia e glifos, e treemap com hierarquias. As tarefas propostas durante o teste objetivaram avaliar tarefas de busca, combinando se o item de busca visualmente era conhecido ou não, e se o local de busca era restrito no cenário ou considerava todo cenário, um exemplo de tarefa pode ser vista no topo do cenário apresentada na Figura 62. Para esses testes foram capturados a dados de interação com mouse, dados de rastreamento do olhar do usuário, vídeo das expressões faciais do usuário e áudio de todo a processo de interação.

Esse conjunto de dados gerados pelos testes conduzido por 81gifos são utilizados para gerar dados de sentimento e visualização de dados, e avaliar o UXmood como ferramenta ao suporte ao processo de avaliação de testes de usabilidade e experiência do do usuário.

6 Avaliações do Protótipo

A metodologia de validação da ferramenta UXmood foi realizada em duas etapas, e é apresentada de maneira resumida na Figura 63. A primeira etapa tem como foco uma avaliação utilizando um questionário on-line, apresentando as visualizações que compõe o *dashboard*, objetivando identificar problemas de comunicação dos dados pelas visualizações que compõe o UXmood. Na segunda etapa é realizada uma avaliação de usabilidade mais tradicional com usuários possuem alguma experiência na condução de testes de usabilidade ou avaliação de experiência do usuário. Os procedimentos éticos da pesquisa e das avaliações envolvendo usuários foram aprovados pelo Comitê de Ética em Pesquisa do Instituto de Ciências da Saúde da Universidade Federal do Pará (CAAE: 15522319.2.0000.0018 sob o parecer nº: 3.618.227 Apêndice A).

Figura 63. Modelo de avaliação da ferramenta



FONTE: O autor

A Figura 63 apresenta as duas grandes etapas para a avaliação do protótipo UXmood, a primeira etapa é composta pela aplicação de um questionário online que apresenta e coleta os dados da realização de tarefas de análise de dados realizadas pelo participantes, além de coletar a opinião do mesmo sobre o uso da técnica de visualização, outras etapas são referentes análise dos dados do questionário, e a aplicação de seus

resultados para melhorar o design da visualização. A segunda grande etapa ocorre com UXmood após o redesenho da ferramenta, e objetiva avaliar a usabilidade da UXmood com base na realização de tarefas por usuários especialistas, também são coletados e analisados dados sobre opinião e interações dos usuário.

6.1 Etapa 1 - Questionário On-line

O dashboard do UXmood é essencialmente uma ferramenta de visualização, desta forma precisa comunicar as informações com eficiência ao usuário. Embora as visualizações escolhidas (exceto o EmojiText que foi proposto) estejam bem estabelecidas na comunidade de visualização, há poucos estudos que fazem uso dessas técnicas para representar dados de sentimento 1.

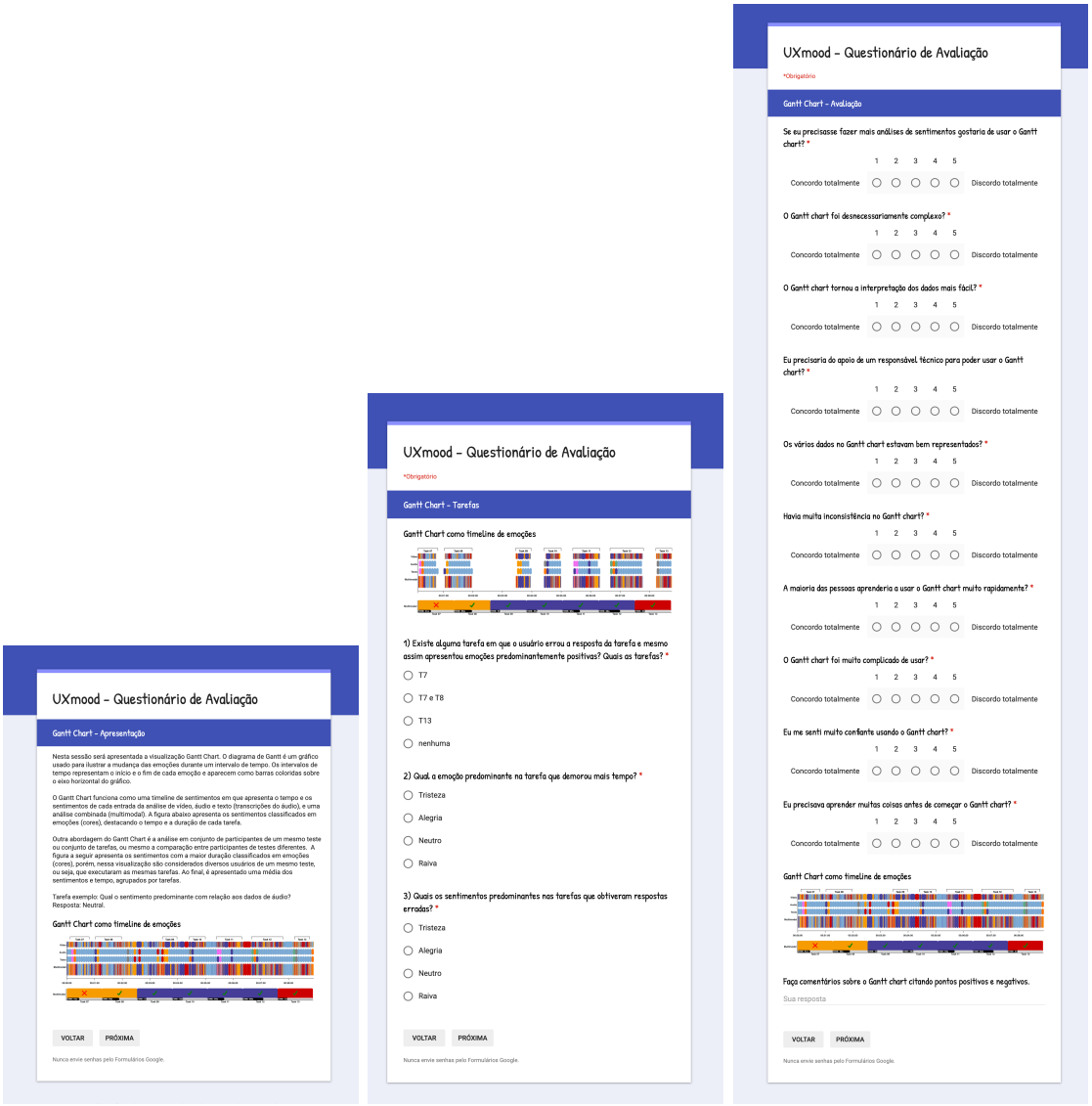
O questionário online tem como objetivo avaliar as principais visualizações de informações que compõem o *dashboard* da UXmood, sendo elas *Gantt Chart*, Wordcloud, Scanpath(grafo), dispersão de dados e Emojitext para saber se as visualizações utilizadas comunicam os dados que se propõem a representar, e destacar pontos positivos e negativos nas visualizações apresentadas borgo2018information.

O link para o questionário foi compartilhado com estudantes de graduação e pós-graduação em ciências da computação nas universidades locais. As imagens de visualização seguem uma explicação dos mapeamentos de dados visuais e conjuntos de dados. Um total de 150 pessoas participaram dos testes.

O questionário online consiste de quatro partes principais para os participantes: a primeira diz respeito a coleta de dados sobre o perfil do participante, alguns dados sobre experiência do mesmo em relação a leitura de gráficos de dados, frequência de uso, escolaridade, dispositivo utilizado, etc.; a segunda etapa é referente a uma página de conteúdo (Figura 64 a) representando um treinamento que descreve a técnica de visualização de informação e como ela codifica visualmente os dados; a terceira etapa apresentou três tarefas de análise de dados, onde cada tarefa possui quatro opções de respostas, com objetivo do participante do teste entender melhor a proposta da técnica de visualização como mostra a Figura 64 b; e, na última etapa, é apresentado um questionário com base no método SUS (Escala de Usabilidade do Sistema)(BROOKE et al., 1996a) e uma questão aberta de avaliação qualitativa da técnica de visualização de dados utilizada pelo participante, como mostra a Figura 64 c. Com exceção da primeira etapa, as outras etapas são similares para todas as técnicas de visualização avaliadas no questionário. O questionário completo está no apêndice B.

As perguntas no formato da Escala de Usabilidade do Sistema (SUS) (BROOKE et al., 1996a) são utilizadas para avaliar a facilidade de interpretação de cada visualização, e suas respostas utilizam a escala Likert de 1-5.

Figura 64. Modelo de avaliação da ferramenta



(a) (b) (c)

FONTE: O autor

Esse processo se repete para as outras técnicas de visualização: Wordcloud, Scanpath/Scatterplot e EmojiText. Os Scanpath e Scatterplots são testados juntos, pois mostram duas visualizações complementares dos mesmos dados e, portanto, podem ser usados juntos para responder às tarefas.

6.2 Escala de Usabilidade do Sistema (SUS)

Para a avaliação das técnicas de visualização foi utilizada e adaptada a Escala de Usabilidade do Sistema, do inglês *System Usability Scale* (SUS) (Figura 64 C) (BROOKE et al., 1996b) utilizada para quantificar o quão usável é um determinado produto, e, para

Tabela 3. Interpretação da Escala SUS

Intervalo SUS	**Nota GPA**	**Porcentagem**
84,1 - 100	A+	96 - 100
80,8 - 84,0	A	90 - 95
78,9 - 80,7	A-	85 - 89
77,2 - 78,8	B+	80 - 84
74,1 - 77,1	B	70 - 79
72,6 - 74,0	B-	65 - 69
71,1 - 72,5	C+	60 - 64
65,0 - 71,0	C	41 - 59
62,7 - 64,9	C-	35 - 40
51,7 - 62,6	D	15 - 34
0,0 - 51,6	F	0 - 14

Fonte: (SAURO; LEWIS, 2012)

isso, faz uso de dez perguntas seguindo *templates* que devem ser respondidas utilizando a escala Likert (ALBAUM, 1997) logo após a utilização da técnica de visualização.

Segundo Brooke (BROOKE et al., 1996b), um sistema, neste caso, uma técnica de visualização, pode ser considerado usável se apresentar 68 pontos ou mais na escala do SUS cujo intervalo é de 0-100, ou equivalente a uma nota C na Figura X. Com um resultado do SUS com valor menor que 51 pontos, ou nota F, a técnica de visualização precisará ser refeita. Já resultados no intervalo entre 51-68 a técnica de visualização precisará de melhorias antes de seu uso.

As perguntas utilizadas no questionário on-line com base nos *templates* do SUS para a visualização Gantt chart é apresentada a seguir:

1. Se eu precisasse fazer mais análises de sentimentos gostaria de usar o Gantt chart?
2. O Gantt chart foi desnecessariamente complexo?
3. O Gantt chart tornou a interpretação dos dados mais fácil?
4. Eu precisaria do apoio de uma pessoa técnica para poder usar o Gantt chart?
5. Os vários dados no Gantt chart estavam bem representados?
6. Havia muita inconsistência no Gantt chart?
7. A maioria das pessoas aprenderia a usar o Gantt chart muito rapidamente?
8. O Gantt chart foi muito complicado de usar?
9. Eu me senti muito confiante usando o Gantt chart?

10. Eu precisava aprender muitas coisas antes de começar o Gantt chart?

A resposta das perguntas do SUS são opções de uma escala Likert com valores entre 1 e 5 pontos (1 - Discordo fortemente, 5 - Concordo fortemente), categorizando a resposta do usuário de acordo com sua experiência durante a tarefa executada. Para as demais visualizações foram usadas as mesmas perguntas e respostas alterando apenas o nome da visualização.

O cálculo para o valor do SUS ocorre da seguinte forma:

- para cada pergunta de índice par, ou seja negativa, é atribuído à resposta um número resultado do seguinte cálculo: nr = 5 - resposta;
- todos os valores encontrados para perguntas par devem ser somados;
- para cada pergunta de índice ímpar, ou seja positiva, é atribuído à resposta um número através do seguinte cálculo: nr = resposta - 1;
- todos os valores encontrados para perguntas ímpar devem ser somados;
- por fim, soma-se os resultados obtidos e multiplica-se pelo fator 2,5.

Se um participante achar que não pode responder a um item específico, deverá marcar o ponto central da escala, ou seja, marcar o valor três na escala. A escolha da escala SUS se deu em função das seguintes características (78brooke1996sus):

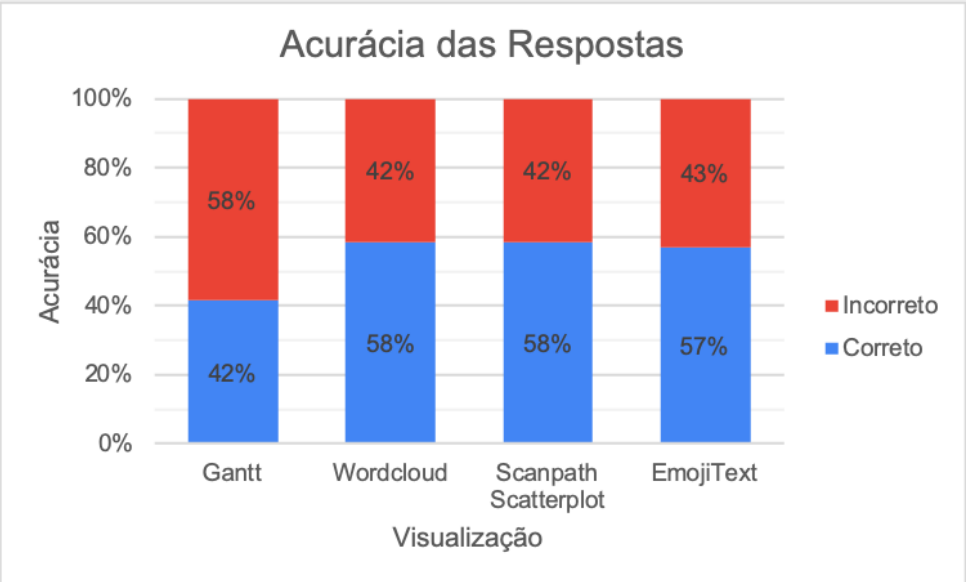
- Escalas fáceis de administrar aos participantes do teste;
- Pode ser utilizado em amostras pequenas com resultados confiáveis;
- Diferenciar efetivamente sistemas utilizáveis e inutilizáveis.

A seguir são apresentados os dados coletados através deste questionário, a análise da aplicação do SUS e discute como eles podem ser usados para melhorar o design das visualizações e da ferramenta UXmood.

6.3 Resultados do Questionário On-line

A Figura 65 mostra a corretude das respostas para as visualizações em cada tarefa e a tabela 4 apresenta os problemas mais citados na pergunta aberta do questionário. Juntos, eles podem gerar insights sobre os problemas relacionados a cada visualização.

Figura 65. A precisão das respostas nas tarefas do questionário.



FONTE: O autor

Tabela 4. Questões mais citadas na pergunta aberta do questionário.

Problemas mais citados nas visualizações			
Gráfico de Gantt	**Wordcloud**	**Scanpath / Scatterplot**	**EmojiText**
Muitos sentimentos e cores. O foco tende a estar nas palavras maiores.	Desordem visual.	Dificuldade de identificação de sentenças.	
Visualização pequena.	Dificuldade de comparação.	Muitas cores.	Confusão com cores de nós e arestas.
Dificuldade de interpretação.	Layout desorganizado.	Visualização complexa.	Sobreposição em alguns rótulos.

O *Gantt Chart* teve o pior desempenho, com a maioria dos participantes respondendo às perguntas incorretamente. A principal razão para essa baixa precisão, de acordo com a análise da questão em aberto, é que a visualização mostra muitos sentimentos, o que dificulta a distinção entre as cores. Os participantes também pensaram que as barras que compõem o *Gantt Chart* às vezes eram muito finas e, portanto, era difícil localizar as respostas.

O Wordcloud teve uma proporção maior de respostas corretas, mas ainda havia erros. Os participantes relataram que era difícil comparar o tamanho das palavras e que a visualização não possui nenhum componente temporal para ver a ordem das mesmas.

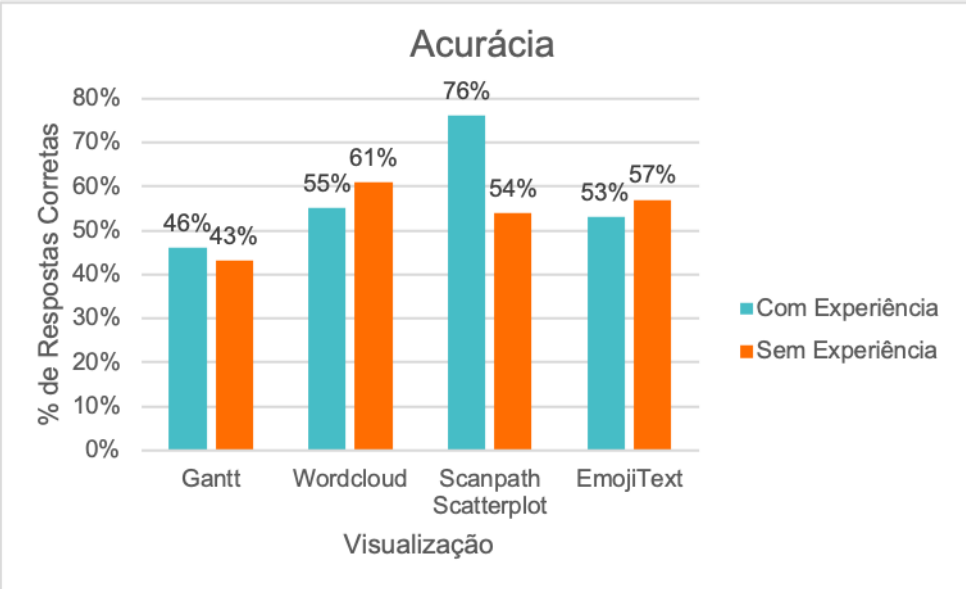
O Scanpath e o Scatterplot também apresentaram mais respostas corretas, similar ao *wordcloud*. A principal desvantagem apontada pelos participantes foi a oclusão que

ocorre nas duas visualizações quando muitos pontos ocorrem em pontos próximos. Alguns participantes apontaram que, se a visualização tivesse menos cores, poderia ser menos confusa.

O EmojiText também apresentou mais respostas corretas que erradas, mas similar em resultados às duas anteriores. Os participantes comentaram que as tarefas que perguntavam sobre a quantidade frases na visualização eram difíceis de responder em função da oclusão e cores. Além disso, muitos participantes relataram que a relação entre as frases não eram claras na visualização. Como comentário positivo, os participantes aprovaram a utilização de *emojis* como metáforas de sentimentos. A proposta de design inicial para esta visualização utilizou a cor das arestas para representar as frases, mas foi confundida com a cor dos nós.

A Figura 66 apresenta um cenário de análise que considera o quão experiente era o participante do teste , considerando os participantes que usam visualizações pelo menos uma vez por semana como usuários experientes em visualização de dados. A figura mostra que o Scanpath/Scatterplot tem mais precisão entre participantes experientes, sugerindo que essas visualizações podem, em um primeiro momento, não ser tão fáceis de interpretar. Por outro lado, o EmojiText e o Wordcloud tiveram um desempenho ainda melhor entre participantes inexperientes, sugerindo que as visualizações de texto escolhidas são mais intuitivas e fáceis de entender, independentemente da experiência anterior. O *Gantt Chart* também teve uma pequena vantagem entre os participantes experientes, mas a precisão é baixa nos dois grupos.

Figura 66. A precisão das respostas nas tarefas do questionário.

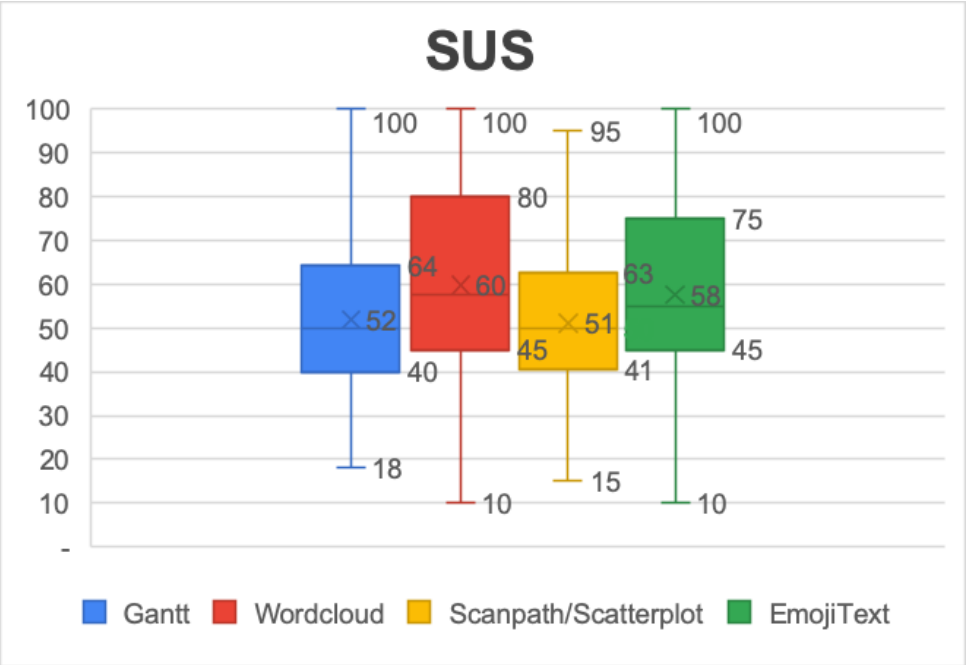


FONTE: O autor

A Figura 67 mostra um gráfico em caixa das pontuações do SUS para cada

visualização. A pontuação média do SUS de todas as visualizações cai no intervalo de 51 a 68, que é classificado como um nível regular de usabilidade que tem espaço para melhorias. O *Gantt Chart* e o Scanpath/Scatterplot apresentaram o pior nível de usabilidade e também foram os que apresentaram o menor nível de precisão. O Wordcloud e o EmojiText apresentaram melhores relatórios de usabilidade, em torno de 58 e 60 na pontuação.

Figura 67. As pontuações do SUS de cada visualização.



FONTE: O autor

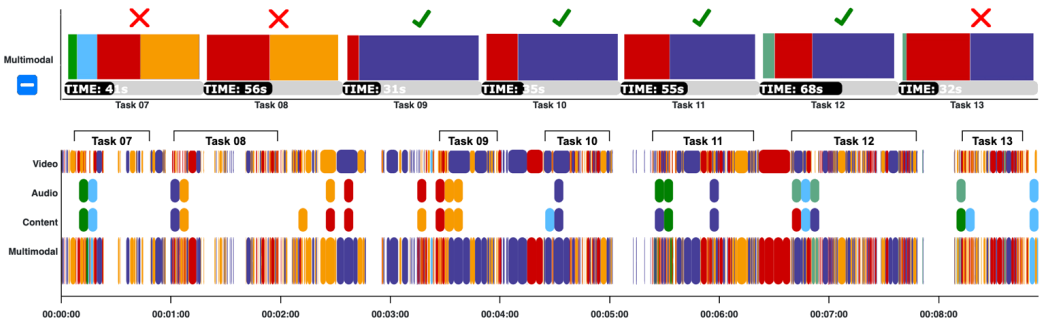
6.4 Redesenho das Técnicas de Visualização

Com a análise dos dados, é possível identificar pontos a serem melhorados e elaborar estratégias de redesenho para melhorar as visualizações. O *Gantt Chart* e o Scanpath/Scatterplot receberam muitos comentários sobre o excesso de cores diferentes na visualização. Como o número de cores causou confusão nas duas visualizações, uma abordagem para melhorar o entendimento da visualização seria reduzir a quantidade de emoções representadas no painel, de um total de 8 para um total de 6. Uma paleta de emoções mais reduzida pode levar a um painel mais limpo, o que pode auxiliar na compreensão das visualizações. Esse redesenho também permite a escolha de esquemas de cores mais distintos que podem melhorar ainda mais a diferenciação das emoções.

O *Gantt Chart* também recebeu comentários sobre a grande quantidade de informações que mostra, que inclui vídeo, áudio, texto e faixas multimodais de sentimentos. Outra opção para melhoria poderia ser exibir apenas a faixa multimodal do gráfico e adicionar um botão que exibe as outras visualizações sob demanda. Como a maioria dos

participantes relatou que prestou mais atenção à faixa multimodal, exceto quando a tarefa solicitou explicitamente informações sobre sentimentos de uma mídia específica, essa opção de design estaria mais alinhada às expectativas dos participantes.

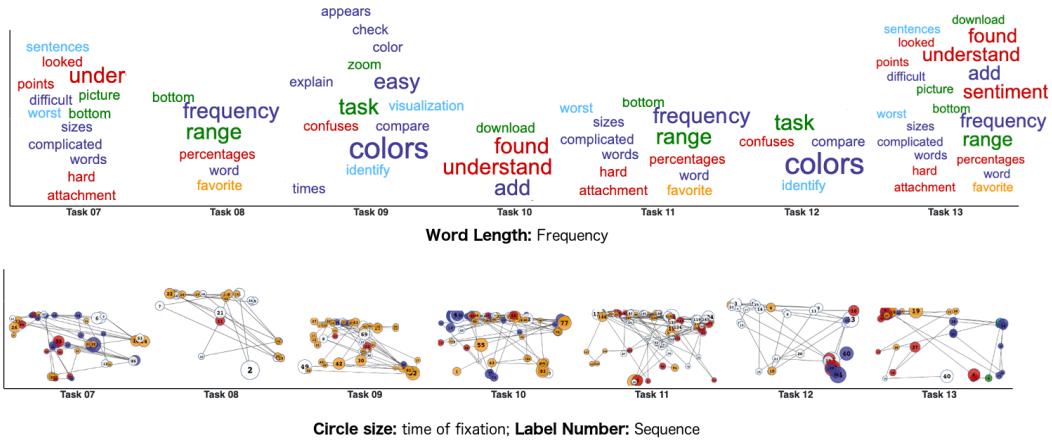
Figura 68. Redesign para o gráfico de Gantt. A principal faixa é a multimodal e as outras aparecem como detalhes sob demanda.



FONTE: O autor

Para o Wordcloud, uma abordagem de melhoria pode ser o uso de novos ou diferentes layouts, como exibir palavras com as mesmas emoções em grupos, posicionar as palavras para codificar um aspecto temporal ou dividir a visualização em pequenos múltiplos para representar tarefas diferentes. Um detalhe sob demanda para exibir a frequência da palavra também pode facilitar a quantificação e, consequentemente, a comparação. Um filtro interativo para eliminar palavras com baixa frequência de aparência pode ajudar a tornar a visualização mais limpa, se o analista quiser se concentrar apenas nas palavras mais citadas.

Figura 69. Uma reformulação proposta para os gráficos Wordcloud e Scanpath / Scatter. Múltiplos pequenos podem reduzir a confusão visual e facilitar a interpretação das visualizações.



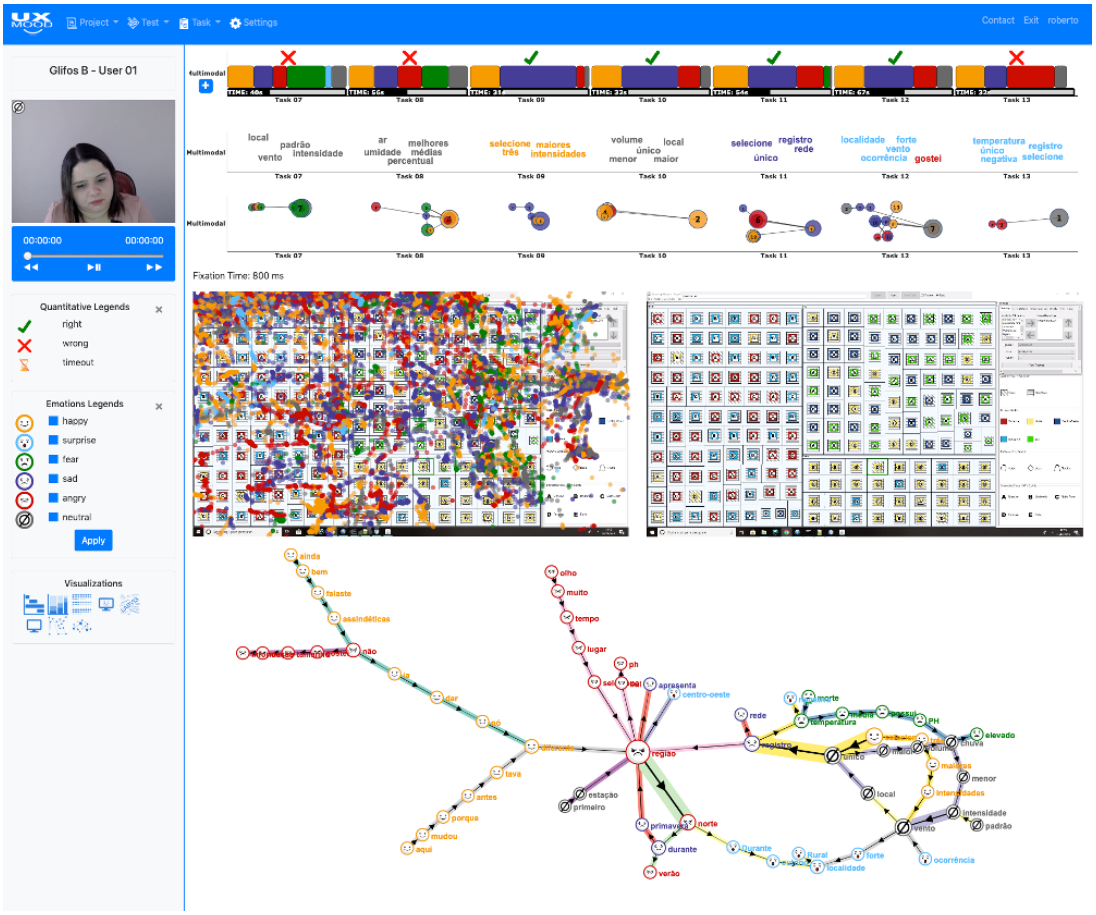
FONTE: O autor

Para o EmojiText, a maior desvantagem é a representação de sentenças. Uma

abordagem pode ser explorar layouts topológicos para permitir um posicionamento mais temporal dos nós. Interações também podem ser adicionadas para destacar nós que pertencem à mesma frase ou até reordenar o layout ao clicar em nós para tornar a frase mais explícita, posicionando-os da esquerda para a direita.

Uma possível reformulação do *Gantt Chart* é dar mais foco à faixa multimodal e exibir as outras faixas apenas ao clicar em um botão de expansão. A Figura 68 mostra como isso pode ser implementado. A faixa multimodal pode mostrar a proporção de sentimentos usando barras empilhadas horizontalmente. Esse novo design reduz a quantidade de cores e agrega informações para simplificar a visualização. Esses novos projetos também incluem uma breve descrição dos mapeamentos visuais para facilitar o entendimento das visualizações.

Figura 70. Reformulação do painel interativo de visualizações.



FONTE: O autor

Para o Wordcloud e o Scanpath, o uso de múltiplos pequenos (small multiples), que é série de gráficos comparáveis com quantidade de dados reduzida, pode mitigar os problemas de complexidade nas visualizações causada pela quantidade de elementos na tela. A Figura 69 apresenta essa reformulação na referida visualização. Nesta abordagem, cada tarefa tem sua própria versão da visualização. Durante a revisão da interação pelo

participante do teste, a tarefa em análise pode ser ampliada para melhor visualização.

O redesenho completo da ferramenta é apresentado na Figura 70 e foi baseado em todas os dados e comentários recebidos dos usuários a partir do questionário online. Modificações extras foram feitas para ajustar questões de design ou codificação dos elementos.

6.4.1 Avaliação com especialistas

A avaliação com especialista consiste em cinco partes principais: treinamento, tarefas de teste, aplicação da escala de usabilidade SUS e análise dos resultados. Todas essas etapas visam avaliar a ferramenta e o redesenho realizado para o *dashboard*.

Na fase de treinamento foi utilizado um roteiro explicando os objetivos e funcionalidades do protótipo UXmood, aspectos da análise de sentimento e seu uso na avaliação qualitativa do processo de interação do usuário; apresentação das técnicas de visualização sobre a codificação visual dos dados coletados nas referidas técnicas; e o aspecto de interação da ferramenta e das visualizações. Esse roteiro foi utilizado para garantir que todos os participantes recebessem o mesmo treinamento. Esta etapa ainda contempla tarefas de exemplo para os especialistas conhecerem mais sobre a ferramenta. Durante a realização destas tarefas eles poderiam tirar dúvidas e fazer perguntas sobre a base ou recursos do protótipo. Isto para que o participante do teste possa entender melhor a proposta e o funcionamento do UXmood.

As tarefas do teste servem para que os participantes utilizem o protótipo e consigam realizar avaliações referentes a utilização de funcionalidades da ferramenta, comunicação da interface e da visualização de dados. As tarefas são divididas em tarefas de busca, onde o usuário precisa encontrar um dado, tarefas comparação, onde o participante precisa comparar dados para resolver e tarefas de visão geral, onde o usuário precisa compreender os conjuntos de dados como um todo, para isso foram propostas quatro perguntas para análise de dados, apresentadas no quadro 1:

Além das tarefas de análise de dados, foi solicitado ao participante a criação de um projeto que simule uma nova avaliação de UX, com a seguinte hierarquia Projeto->Teste->Tarefas, após essa etapa deveria adicionar arquivos de vídeo, áudio, texto e *logs*, todos já sincronizados, fornecido pelo facilitador do teste, e visualizar o .

Diante das tarefas os usuários eram capazes de produzir um entendimento maior da base de dados e do teste de usuário que estava sendo analisado. Os participantes poderiam fazer comentários sobre a ferramenta, visualizações e recursos do protótipo ou do teste. Nesse conjunto de testes com especialistas também foi utilizado o SUS como escala de usabilidade (SAURO; LEWIS, 2012)

Por fim, perguntas abertas sobre aspectos positivos e negativos do UXmood foram

Quadro 1. Tarefas de teste do protótipo UXmood

1. Qual o sentimento predominante na tarefa de maior duração?
2. Encontre a frase de emoção alegre que tem a palavra de maior frequência?
3. Qual a região da tela que o usuário apresentou pouca surpresa?
4. Qual a tarefa que o usuário fixou o olhar por mais de 2200ms e teve emoção alegre?

Fonte: o autor

feitas aos participantes, essas perguntas tem como objetivo extrair informações qualitativas da ferramenta e encontrar possíveis funcionalidades que não estejam adequadas. São sete perguntas abertas com o foco em aspectos positivos e negativos para componentes de interação (*player*, filtro, detalhes sob demanda, etc), visualização de dados (*stackedbar*, *gantt chart*, *wordcloud*, *scanpath* e *scatterplot*, e *emojitext*)

A avaliação com especialistas foi conduzida, após o redesenho do protótipo, com sete participantes com experiência em avaliação de usabilidade ou experiência do usuário. (NIELSEN; LANDAUER, 1993) afirma que um estudo com cinco participantes encontra em torno 85% dos problemas de usabilidade, e faz recomendações para que se faça outro teste após a aplicação das melhorias do teste anterior.

Dos sete participantes do teste, os dados de um dos participantes não foi considerado nas análises, uma vez que este participante declarou possuir discromatopsia, essa doença afeta a capacidade de diferenciar algumas cores, e a ferramenta não apresenta características e codificação de cores para a inclusão de participantes com essa deficiência. Portanto somente seis voluntários tiveram seus dados utilizados.

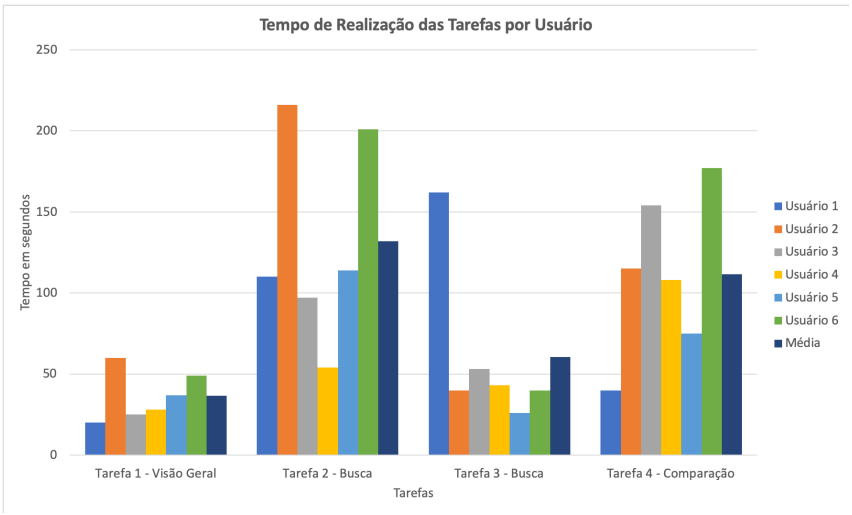
A acurácia dos participantes especialistas na realização das tarefas de análise de dados foi de aproximadamente 93%. O valor médio obtido da aplicação do SUS foi de 75 pontos ou conceito B, segundo Brooke et al. (1996b) e Sauro e Lewis (2012) a ferramenta apresenta boa usabilidade (Como pode ser visto comparando os resultados com a Tabela 3), mas ainda pode ser melhorada.

O tempo médio de realização total do teste para tarefas de análise de dados foi de seis (6) minutos e trinta e três (33) segundos considerando todos os participantes. A grande variação em relação ao tempo total gasto para realizar as tarefas está relacionada aos diferentes usos dos recursos de interação e das visualizações de dados pelos participantes. Participantes que levaram mais tempo para a solução utilizam mais recursos como filtros, detalhes sob demanda e configuração ou utilizaram mais de uma visualização para encontrar

a resposta.

As tarefas apresentam propósitos diferentes para que os participantes experimentassem todas as visualizações e recursos da ferramenta, portanto foram feitas tarefas de comparações, busca e visão geral. Como consequência o tempo de realização das tarefas foi diferente. A Figura 71 apresenta o tempo médio e individualizado das tarefas. Para tarefas de busca e visão geral os usuários necessitavam de menos tempo para a resolução. Tarefas de comparação, que precisavam da utilização de recursos como filtro ou configuração, os participantes precisaram de mais tempo para finalizar a tarefa.

Figura 71. Tempo médio de realização das tarefas por tarefas.



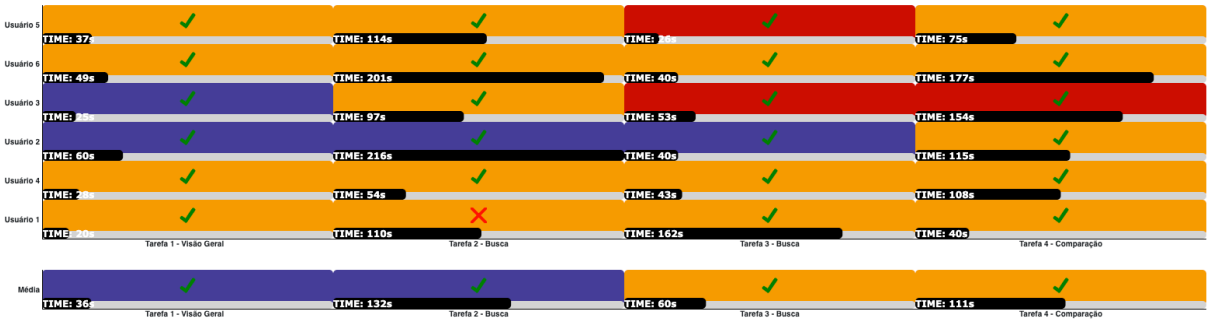
Fonte: O autor

A Figura 72 apresenta o média de tempo e emoção (cor) predominante da tarefa por usuário. De maneira geral os voluntários tiveram emoção alegre (amarelo) em toda a busca pela solução da tarefa. A Figura 72 apresenta o potencial da ferramenta UXmood na apresentação de dados sobre avaliação de experiência do usuário e dados testes de usabilidade, incluindo acertos/erros, sentimentos, tempo e tempo médio, em uma única visualização, quando comparado ao gráfico de barras da Figura 73.

Os recursos de interação mais utilizados pelos participantes para a resolver as tarefas foram filtro e detalhe sob demanda. O filtro diminui o conjunto de dados de análise e o detalhe sob demanda apresenta dados que não estão representados visualmente, mas que precisam ser analisados. De acordo com a Figura 73 estes recursos foram utilizados em três das quatro tarefas.

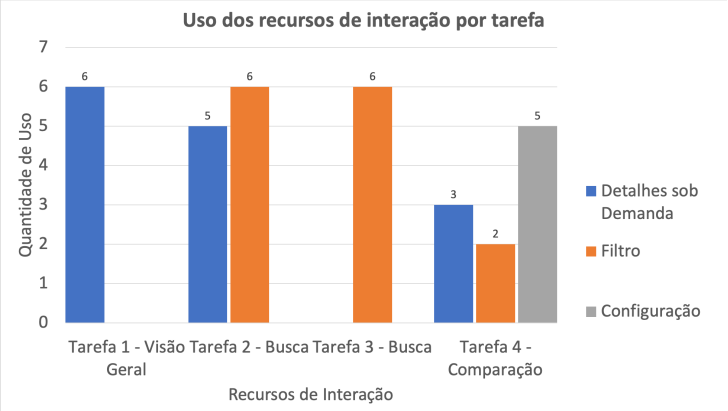
Um tipo de dado que foi coletado durante a realização das tarefas foram os dados de um dispositivo de rastreamento de olhos. O dispositivo utilizado nos testes foi o Tobii 4C desenvolvido pela Tobii Technology, com esse dispositivo não vestível é possível identificar para onde o usuário estava olhando ao interagir com o protótipo. É a partir deste que foi possível analisar as visualizações e recursos de interação mais utilizados.

Figura 72. Média de tempo gasto para a realização das tarefas junto com a emoção predominante.



Fonte: O autor

Figura 73. Recursos de interação mais utilizados para solução das tarefas.



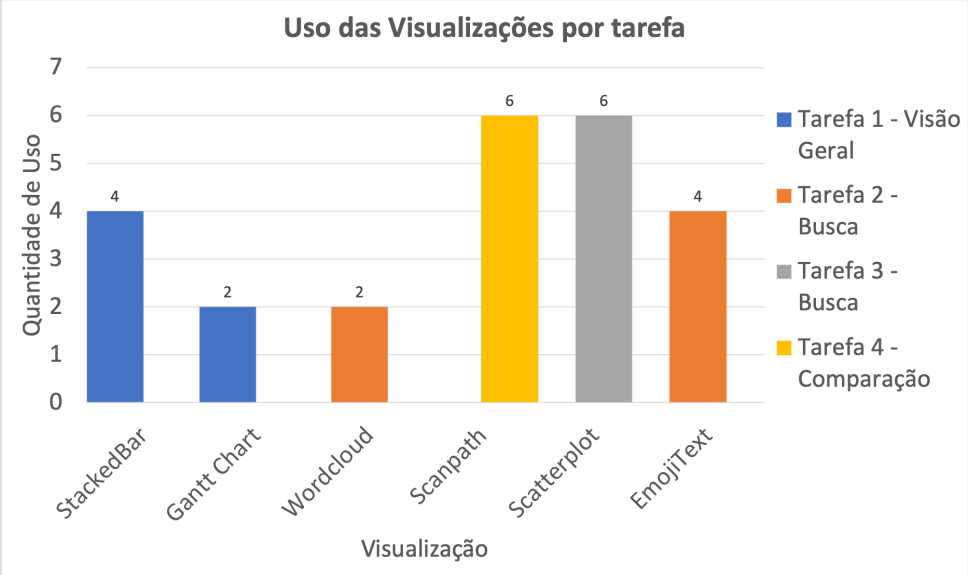
Fonte: O autor

A Figura 75 apresenta um heatmap produzido com os dados coletados por esse dispositivo para todos os usuários. Nesta existem três grandes regiões mais observadas pelos participantes, duas dessas dentro da área destinada as visualizações e uma na região do filtro, reforçando a afirmação de que o filtro foi o recurso de interação mais utilizado.

As técnicas de visualização mais utilizadas para realizar as tarefas são apresentadas na Figura 74. Destaca-se, a utilização das técnicas stacked bar e gantt chart para tarefas de busca mais simples ou visão geral, uma utilização maior do EmojiText em relação ao wordcloud para tarefas que envolviam texto e que a proposta de redesenho organizado por tarefas para scanpath, utilizando múltiplos pequenos, facilitou a comparação de informações.

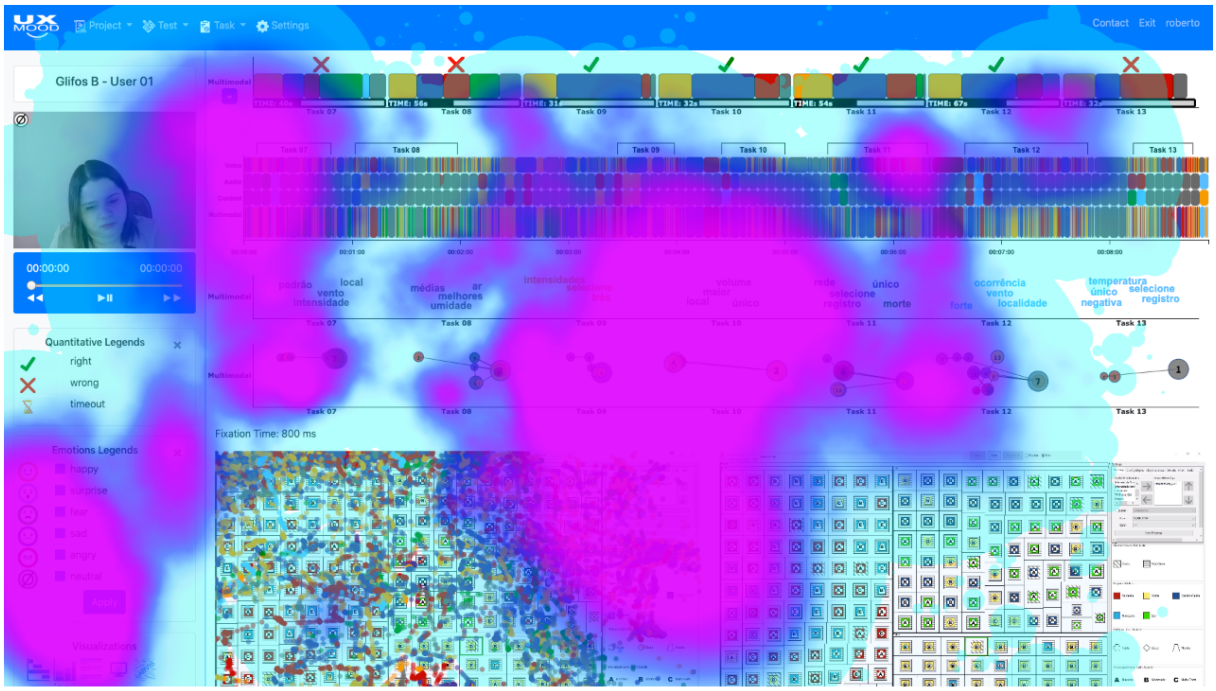
Utilizando o UXmood para analisar as interações dos participantes, através de técnicas como heatmap do rastreamento ocular e de reprise testes, por exemplo, é possível detalhar as etapas da solução utilizada pelos participantes para a identificação das respostas das tarefas, e entender a estratégia da solução adotada por cada participante. Uma análise mais aprofundada, considerando todos participantes, poderia encontrar padrões nas

Figura 74. Visualizações mais utilizadas para solução das tarefas.



Fonte: O autor

Figura 75. Maiores regiões de fixação visual pelos usuário.



Fonte: O autor

estratégias, e essas informações poderiam ser utilizadas para melhoramento da ferramenta, por exemplo, mudança de layout do dashboard, destaque para outras visualizações e identificação de problemas de usabilidade.

Em relação a configuração do Projeto na ferramenta UXmood, todos os participantes conseguiram completar a tarefa, mas destacaram que foi preciso digitar metadados que

poderiam ser automáticos. Além disso, o processo de sincronização, extração dos dados e classificação ocorre logo após o*upload* dos arquivos, como sugestão recomendaram que fosse feito em *background*, liberando a ferramenta para outra tarefa, ou em fases independentes, gerando os arquivos para visualização para serem carregados.

Sobre as questões abertas do questionário considerando o uso dos componentes de interação e técnicas de visualização de dados, as Tabelas 5 e 6 mostram os principais aspectos positivos e negativos relatados pelos participantes.

É possível perceber que técnicas de visualização como scanpath, scatterplot e emojiText sofrem com oclusão de dados devido apresentarem muitos elementos, principalmente os dois últimos, possíveis soluções para esse problema podem ser a adoção de visões por tarefas (como no scanpath), utilização de técnicas de interação como zoom e filtro para reduzir o conjunto de dados em subconjuntos e utilizar técnicas de distorção para separar ainda mais os itens visuais.

Tabela 5. Dados qualitativos dos recursos de interação e visualizações stackedbar e gantt chart.

Pontos	**Interação**	**Stackedbar**	**Gantt Chart**
Positivos	Filtro e detalhes sob demanda fácil e útil;	Fácil identificação; Boa organização;	Facilidade de identificação das variações temporais; Fácil representação visual;
Negativos	Faltou informações no detalhes sob demanda; Demora ao recarregar;	Barra de progresso do tempo de difícil entendimento;	Muitos elementos; Esquecimento por estar oculto; Representação de muitos dados; Dificuldade nas corres;

A Tabela 5 apresenta informações sobre os recursos de interação e um destaque foi o filtro que teve comentários positivos sobre a facilidade de uso e aplicação em todas as técnicas de visualização utilizadas.

Tabela 6. Dados qualitativos das visualizações wordcloud, scanpath, scatterplot e EmojiText.

Pontos	**Wordcloud**	**Scanpath**	**Scatterplot**	**EmojiText**
Positivos	Boa representação da frequência;	Ótimo para análise em tempo real;	Facilidade com filtros;	Fácil de perceber as frases;
Negativos	Dificuldade de formar frases;	Forte Oclusão;	Confusão com cores; Oclusão;	Ícones pequenos; Forte oclusão;

A Tabela 6 apresenta dados qualitativos das técnicas wordcloud e emojiText. Em comparação as duas técnicas, alguns usuários comentaram que o wordcloud tem uma boa representação da frequência e que o EmojiText é fácil de perceber as frases.

De maneira geral, a ferramenta apresentou boa usabilidade e demonstrou facilitar as análises de usabilidade e experiência do usuário com o uso dos recursos de reprise, filtro, detalhes sob demanda e configurações.

7 Considerações Finais e Trabalhos Futuros

O modelo desenvolvido por esta tese objetiva abstrair a complexidade da geração (classificação) de dados de sentimento a partir de vários tipos de mídia ou dados, bem como abstrair a complexidade de criar cenários de análise em conjunto com outros dados do domínio de um problema.

A necessidade e a aplicação de dados de sentimentos em várias áreas é evidenciada ao longo do texto, bem como a complexidade de seu uso por usuários que não são especialistas em técnicas inteligentes e de visualização de dados. Além disso, esses dados geralmente são caracterizado por um grande volume e muitas vezes se apresentam em forma não estruturada, o pré-processamento, integração e visualização com outros dados também é um desafio. Assim, esse contexto desafiador que motivou a criação e aplicação deste modelo.

A área de análise de sentimentos tem sido mais frequentemente aplicada em diversas áreas mais recentemente, em função dos avanços ocorridos nas técnicas de aprendizado de máquina. No passado, eram utilizadas técnicas de aprendizado mais tradicionais, como SVM e aplicadas para um único tipo de mídia, principalmente texto. Atualmente, as redes sociais, por exemplo, tem proporcionado os mais variados tipos de dados, tais como: texto, imagem, vídeo, emojis, etc, utilizar técnicas de análise mais robustas e de forma conjunta é um desafio atual. Pesquisas recentes têm apontado na direção da utilização de técnicas de aprendizado profundo e suas diversas variantes para atacar esse problema e a utilização de técnicas de fusão para sua análise multimodal de análise de sentimentos, ainda são questões em aberto que requerem maior investigação.

A implementação do modelo proposto, seguiu as diretrizes de investigar a viabilidade de utilização de redes neurais convolucionais (CNN) para análise de sentimento em vídeo, áudio e texto, com o uso de estratégia de fusão em nível de decisão para análise de sentimentos multimodal. Essa estratégia de fusão combina através de pesos os resultados da análise de sentimentos unimodal das diferentes mídias. Os resultados apresentados da aplicação das CNNs para análise de sentimentos se mostraram compatíveis com resultados apresentados na literatura, apesar de não serem os melhores, são viáveis de aplicação nesse cenário, e requerem mais investigações para aprimorar os resultados obtidos. Em relação visualização dos dados de sentimento, as diretrizes foram a utilização de de múltiplas visões de dados, para diminuir a complexidade de codificação em uma única visualização de dados, e a natureza diferente dos dados requer também visualizações diferentes. Os sentimentos, emoções ou polaridades, deveriam ser codificados pela cor e permear todas as visualizações disponíveis. Além disso, a utilização de dados temporais e geoespaciais

devem ser consideradas para uso.

Como cenário de uso, a implementação do modelo, denominada UXmood, foi direcionada para dados sobre avaliação de usabilidade e experiência de usuários na interação com um software. Além de dados de vídeo, áudio e texto transcrito, o conjunto de dados apresentavam dados de rastreamento ocular dos usuários, dados de rastreamento do mouse, erros e acertos de tarefas realizadas, tempo gasto para realizar as tarefas, etc.

De maneira geral, o usuário precisa apenas carregar na ferramenta os arquivos de dados que serão utilizados na análise visual, além de informar metadados sobre os arquivos, e criar a estrutura do processo de avaliação, caracterizada por Projeto -> Teste -> Tarefa, onde um teste equivale a um participante do teste, em outras palavras, pode haver diversos projetos, cada projeto tem participantes, e participantes realizam tarefas. A criação dessa estrutura e a inserção desses dados ao sistema foi alvo de avaliação junto a usuários, onde resumiram que foi um procedimento fácil, porém que poderia ser mais automatizado, principalmente em relação a metadados de arquivos, mas ressaltando a abstração da complexidade do processo de gerar dados de sentimento.

Para análise visual dos dados gerados, um dashboard foi criado com múltiplas visões de dados coordenadas, seguindo as diretrizes mínimas estipulada. As principais visualizações que compõem o *dashboard* são: *Gantt Chart* e gráfico de barras empilhadas para representações de emoções e polarizações, gráfico de dispersão para representação de dados sobre rastreamento ocular, EmojiText e *wordcloud* para a representação de texto. Todas as visões são permeadas de informações sobre emoções ou polaridade sugeridas pelo UXmood, com base em classificadores de sentimentos da literatura. Uma recomendação automática de sentimentos torna-se valiosa uma vez que os participantes dos testes podem ter dificuldade em expressar seus sentimentos, mas a coleta desses dados geralmente ocorre no final dos testes por meio de questionários, em vez de durante a avaliação pela interação real. A dedução automática de sentimentos dos dados coletados durante a interação ajuda a encontrar informações mais precisas sobre a experiência do usuário.

Estudos de usabilidade foram conduzidos para avaliar o design e a usabilidade da ferramenta em duas etapas. A primeira etapa aplicou um questionário on-line para 150 pessoas que avaliaram a capacidade representativa de dados nas visualizações de informações escolhidas. A Escala de Usabilidade (SUS) foi utilizada no questionário proposto, bem como perguntas abertas sobre pontos positivos e negativos para avaliar quantitativamente e qualitativamente cada técnica de visualização. Analisando os dados coletados foi possível melhorar a representação dos dados nas visualizações para próxima etapa de avaliação. A segunda etapa realizou teste de usabilidade com participantes com experiência na condução em avaliações de usabilidade ou experiência do usuário, com realização de tarefas, aplicação do SUS, e questionários por perguntas abertas sobre características da UXmood. A ferramenta apresentou bons resultados na escala SUS,

e pontos positivos na análise dos dados e baixa complexidade para classificação e uso de dados de sentimentos de acordo com participantes do teste, mas ainda há pontos de melhoria, relativos a oclusão de dados em função da quantidade de dados a serem visualizados, um problema comum na área de visualização, aqui abre-se uma possibilidade de avaliar outras técnicas que possam mitigar essa questão.

Em termos de aprendizado de máquina, modelos inteligentes mais complexos podem ser treinados em função das entradas multimodais. Neste trabalho a avaliação foi unimodal, o que dá margem para uma avaliação em termos multimodais, em treino e avaliação. A aquisição de novas bases de imagens e vídeos permite essa nova avaliação. Modelos com várias entradas, com extração de áudio e texto baseado em um arquivo de vídeo podem ser treinados, assim como as funções objetivo devem ser configuradas para tal (JOHNSON; KARPATHY; FEI-FEI, 2016).

Além disso, novas estratégias de treino podem ser adotadas, realizar o pré-treino em outras bases de sentimentos ao invés da base do ImageNet pode melhorar a acurácia em relação aos classificadores. Bases de faces e de áudio podem ser utilizadas para realizar o pré-treino. O treino com as MFCCs no áudio obteve bons resultados, mas uma nova abordagem para ser avaliada é a transformação das $waveforms$ dos sons em imagem, para classificação por CNNs 2D, reutilizando até as mesmas redes utilizadas pelos classificadores de imagem.

Além das estratégias de treino, a mudança das arquiteturas clássicas de CNN para variação de processamento de sequência podem ser feitas. Vídeo, áudio e texto possuem variações com CNNs 3D, RNN e Transformers (WOLF et al., 2019), respectivamente, que possuem resultados do estado da arte para tarefas sequenciais nessas mídias. Adicionalmente, a utilização de novas bases de dados, com anotações (pontos de interesses) das faces (BUSSO et al., 2008) pode ser utilizada como forma de apoio para classificação dos sentimentos. Além de oferecer mais informação para o classificador, também pode ser utilizada para interpretação dos modelos inteligentes.

Como contribuições desta tese pode-se destacar: um modelo para integrar dados de análise de sentimentos multimodal à dados de interação de usuários de forma visual pelo especialistas em usabilidade e UX, uma ferramenta computacional para abstrair a complexidade de classificação de dados de sentimento e uma painel de visualização coordenadas para analisar esse dados. Além disso, a ferramenta disponibilizada permite a análise desses dados a outros especialista para avaliação a um baixo custo de tempo e recursos. Por fim, como produto desta tese foi proposta uma nova visualização para texto associando à dados de análise de sentimentos.

Algumas limitações que podem ser destacadas no desenvolvimento desta tese são: realizar um estudo mais abrangente sobre classificadores de análise de sentimentos visando melhor acurácia dos resultados, uma vez que o foco foi voltado para a integração de

dados de análise de sentimentos multimodais com dados de interação. Apesar dos esforços de re-design da ferramenta e das técnicas de visualização de dados, como consequência da primeira etapa de avaliação, a avaliação com especialistas apresentou ainda novas necessidades de re-design que deverão ser alvo de trabalhos futuros. Além de avaliar novas técnicas de visualização de informação. Foi pouco explorado nesta tese as visualizações para mais de um teste, por exemplo, a ferramenta não combina dados dos rastreadores de olhos e mouse de mais de um teste, explorar cenários de análise para múltiplos testes de participantes diferentes é uma direção futura para auxiliar na descoberta de padrões em estratégias de interação ou de comportamento entre os participantes. Outro ponto importante foi o carregamento inicial dos dados e metadados que devido ao tempo de *upload* e processamento cria uma expectativa dos usuários uma vez que este é um sistema projetado para web. Outra limitação, o modelo não utiliza dados de sensores fisiológicos para os classificadores de análise de sentimentos.

Como trabalhos futuros, pode-se conduzir novas avaliações na ferramenta com especialistas em visualização e UX, incluindo novas técnicas de visualização, visualização para representar as expressões faciais, novas configurações e interações para os especialistas, buscando automatizar métricas e metodologias de UX e usabilidade que possam ser integradas ao modelo, utilização de dados de outros sensores - como eletrocardiografia (ECG) - para análise de sentimentos, aplicação da ferramenta para análise de dados de outros contextos de avaliação, por exemplo, de produtos, televisão, carros, *smartphones*, etc. Além disso, poderão ser avaliados outras formas de sincronização e pré-processamento mais automatizados e a expansão do suporte a diferentes idioma, novas configurações e interações para o *dashboard*.

Referências

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANé, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIéGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>. Citado na página 84.

ALBAUM, G. The likert scale revisited. *Market Research Society. Journal.*, SAGE Publications Sage UK: London, England, v. 39, n. 2, p. 1–21, 1997. Citado na página 91.

ALSBURY, Q.; BECERRA, D. *Displaying stacked bar charts in a limited display area*. [S.l.]: Google Patents, 2012. US Patent 8,239,765. Citado na página 76.

BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: *Lrec*. [S.l.: s.n.], 2010. v. 10, p. 2200–2204. Citado na página 85.

BAKIU, E.; GUZMAN, E. Which feature is unusable? detecting usability and user experience issues from user reviews. In: IEEE. *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*. [S.l.], 2017. p. 182–187. Citado 2 vezes nas páginas 65 e 66.

BAKIU, E.; GUZMAN, E. Which feature is unusable? detecting usability and user experience issues from user reviews. In: IEEE. *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*. [S.l.], 2017. p. 182–187. Citado na página 65.

BASRI, N. H.; ADNAN, W. A. W.; BAHARIN, H. E-participation service in malaysian e-government website: the user experience evaluation. In: ACM. *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning*. [S.l.], 2019. p. 342–346. Citado 2 vezes nas páginas 65 e 66.

BAUMAN, K.; LIU, B.; TUZHILIN, A. Recommending items with conditions enhancing user experiences based on sentiment analysis of reviews. In: *CBRecSys@ RecSys*. [S.l.: s.n.], 2016. p. 19–22. Citado 3 vezes nas páginas 64, 65 e 66.

BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. In: *Brazilian Symposium on Multimedia and the Web (Webmedia), Manaus, Brasil*. [S.l.: s.n.], 2015. Citado 2 vezes nas páginas 23 e 24.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 35, n. 8, p. 1798–1828, 2013. Citado 3 vezes nas páginas 28, 29 e 58.

BISHOP, C. M. *Pattern recognition and machine learning*. [S.l.]: springer, 2006. Citado 2 vezes nas páginas 27 e 28.

BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, IEEE, v. 17, n. 12, p. 2301–2309, 2011. Citado na página 69.

BOTTOU, L. Large-scale machine learning with stochastic gradient descent. In: *Proceedings of COMPSTAT'2010*. [S.l.]: Springer, 2010. p. 177–186. Citado na página 32.

BRADLEY, M. M.; LANG, P. J. *Affective norms for English words (ANEW): Instruction manual and affective ratings*. [S.l.], 1999. Citado na página 24.

BROAD, C. D. Emotion and sentiment. *The Journal of Aesthetics and Art Criticism*, JSTOR, v. 13, n. 2, p. 203–214, 1954. Citado na página 22.

BROOKE, J. et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, London–, v. 189, n. 194, p. 4–7, 1996. Citado na página 89.

BROOKE, J. et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, London–, v. 189, n. 194, p. 4–7, 1996. Citado 3 vezes nas páginas 90, 91 e 99.

BUSSO, C.; BULUT, M.; LEE, C.-C.; KAZEMZADEH, A.; MOWER, E.; KIM, S.; CHANG, J. N.; LEE, S.; NARAYANAN, S. S. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, Springer, v. 42, n. 4, p. 335, 2008. Citado na página 107.

CAMBRIA, E.; HAZARIKA, D.; PORIA, S.; HUSSAIN, A.; SUBRAMANYAM, R. Benchmarking multimodal sentiment analysis. In: SPRINGER. *International Conference on Computational Linguistics and Intelligent Text Processing*. [S.l.], 2017. p. 166–179. Citado 2 vezes nas páginas 49 e 84.

CARD, S.; MACKINLAY, J.; SHNEIDERMAN, B. Information visualization. *Human-computer interaction: Design issues, solutions, and applications*, CRC press, v. 181, 2009. Citado na página 33.

CATTELL, R. B. Sentiment or attitude? the core of a terminology problem in personality research. *Character & Personality; A Quarterly for Psychodiagnostic & Allied Studies*, Blackwell Publishing, 1940. Citado na página 23.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1251–1258. Citado 2 vezes nas páginas 68 e 83.

CHOLLET, F. et al. *Keras*. 2015. <https://keras.io>. Citado na página 84.

CLARK, W. *The Gantt chart: A working tool of management*. [S.l.]: Ronald Press, 1922. Citado na página 76.

CORNEANU, C. A.; SIMÓN, M. O.; COHN, J. F.; GUERRERO, S. E. Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 38, n. 8, p. 1548–1568, 2016. Citado 4 vezes nas páginas 47, 48, 49 e 84.

CORRÊA, A. G.; BORBA, E. Z.; LOPES, R.; ZUFFO, M. K.; ARAUJO, A.; KOPPER, R. User experience evaluation with archaeometry interactive tools in virtual reality environment. In: IEEE. *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. [S.l.], 2017. p. 217–218. Citado na página 65.

CUI, W.; WU, Y.; LIU, S.; WEI, F.; ZHOU, M. X.; QU, H. Context preserving dynamic word cloud visualization. In: IEEE. *2010 IEEE Pacific Visualization Symposium (PacificVis)*. [S.l.], 2010. p. 121–128. Citado na página 76.

DODD, C.; ATHAUDA, R.; ADAM, M. Designing user interfaces for the elderly: a systematic literature review. In: *Proceedings of the Australasian Conference on Information Systems*. [S.l.: s.n.], 2017. p. 1–11. Citado 2 vezes nas páginas 65 e 66.

DOVE, G.; HALSKOV, K.; FORLIZZI, J.; ZIMMERMAN, J. Ux design innovation: Challenges for working with machine learning as a design material. In: ACM. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. [S.l.], 2017. p. 278–288. Citado na página 18.

Dudzik, B.; Jansen, M.; Burger, F.; Kaptein, F.; Broekens, J.; Heylen, D. K. J.; Hung, H.; Neerincx, M. A.; Truong, K. P. Context in human emotion perception for automatic affect detection: A survey of audiovisual databases. In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. [S.l.: s.n.], 2019. p. 206–212. ISSN 2156-8103. Citado na página 47.

DÜNSER, A.; BILLINGHURST, M. Evaluating augmented reality systems. In: *Handbook of augmented reality*. [S.l.]: Springer, 2011. p. 289–307. Citado na página 65.

Ebrahimi, M.; Yazdavar, A. H.; Sheth, A. Challenges of sentiment analysis for dynamic events. *IEEE Intelligent Systems*, v. 32, n. 5, p. 70–75, Sep. 2017. ISSN 1941-1294. Citado 2 vezes nas páginas 16 e 17.

EKMAN, P. Basic emotions. *Handbook of cognition and emotion*, v. 98, n. 45-60, p. 16, 1999. Citado 2 vezes nas páginas 23 e 81.

FAN, M.; WU, K.; ZHAO, J.; LI, Y.; WEI, W.; TRUONG, K. N. Vista: Integrating machine intelligence with visualization to support the investigation of think-aloud sessions. *IEEE transactions on visualization and computer graphics*, IEEE, 2019. Citado na página 65.

FEW, S. *Information dashboard design*. [S.l.]: O'reilly Sebastopol, CA, 2006. Citado na página 33.

FIGUEIRAS, A. A review of visualization assessment in terms of user performance and experience. In: IEEE. *2018 22nd International Conference Information Visualisation (IV)*. [S.l.], 2018. p. 145–152. Citado 2 vezes nas páginas 65 e 66.

Fong, S.; Zhuang, Y.; Li, J.; Khoury, R. Sentiment analysis of online news using mallet. In: *2013 International Symposium on Computational and Business Intelligence*. [S.l.: s.n.], 2013. p. 301–304. Citado na página 16.

FOREMAN-MACKEY, D. corner. py: Scatterplot matrices in python. *The Journal of Open Source Software*, v. 1, 2016. Citado na página 76.

FORTIN-CÔTÉ, A.; BEAUDIN-GAGNON, N.; CAMPEAU-LECOURS, A.; TREMBLAY, S.; JACKSON, P. L. Affective computing out-of-the-lab: The cost of low cost. In: IEEE. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. [S.l.], 2019. p. 4137–4142. Citado na página 23.

FRANCO, R. Y. da S.; LIMA, R. Santos do A. D.; PAIXÃO, M.; SANTOS, C. G. Resque dos; MEIGUINS, B. S. et al. Uxmood—a sentiment analysis and information visualization tool to support the evaluation of usability and user experience. *Information*, Multidisciplinary Digital Publishing Institute, v. 10, n. 12, p. 366, 2019. Citado na página 80.

FRIEDENBERG, J.; SILVERMAN, G. *Cognitive science: An introduction to the study of mind*. [S.l.]: Sage, 2011. Citado na página 23.

FUCHS, J.; FISCHER, F.; MANSMANN, F.; BERTINI, E.; ISENBERG, P. Evaluation of alternative glyph designs for time series data in a small multiple setting. In: ACM. *Proceedings of the SIGCHI conference on human factors in computing systems*. [S.l.], 2013. p. 3237–3246. Citado na página 76.

FULSE, S.; SUGANDHI, R.; MAHAJAN, A. A survey on multimodal sentiment analysis. *International Journal of Engineering Research and Technology*, v. 3, n. 11, p. 1233–1238, 2014. Citado 2 vezes nas páginas 47 e 48.

GIANNAKOPOULOS, T.; PIKRAKIS, A.; THEODORIDIS, S. A dimensional approach to emotion recognition of speech from movies. In: IEEE. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.], 2009. p. 65–68. Citado na página 85.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Citado 3 vezes nas páginas 27, 28 e 31.

GORDON, S. L. The sociology of sentiments and emotion. In: *Social psychology*. [S.l.]: Routledge, 2017. p. 562–592. Citado 2 vezes nas páginas 22 e 23.

Hakak, N. M.; Mohd, M.; Kirmani, M.; Mohd, M. Emotion analysis: A survey. In: *2017 International Conference on Computer, Communications and Electronics (Comptelix)*. [S.l.: s.n.], 2017. p. 397–402. ISSN null. Citado na página 47.

HARTSON, R.; PYLA, P. S. *The UX Book: Process and guidelines for ensuring a quality user experience*. [S.l.]: Elsevier, 2012. Citado na página 63.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Identity mappings in deep residual networks. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 630–645. Citado na página 83.

HEDEGAARD, S.; SIMONSEN, J. G. Extracting usability and user experience information from online user reviews. In: ACM. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. [S.l.], 2013. p. 2089–2098. Citado 2 vezes nas páginas 64 e 65.

HOLOVATY, A.; KAPLAN-MOSS, J. *The definitive guide to Django: Web development done right*. [S.l.]: Apress, 2009. Citado na página 69.

HOLTZ, Y.; HEALY, C. From data to viz leads you to the most appropriate graph for your data. *DataViz*, DV press, 2018. Citado 2 vezes nas páginas 37 e 76.

HUSSAIN, J.; KHAN, W. A.; HUR, T.; BILAL, H. S. M.; BANG, J.; HASSAN, A. U.; AFZAL, M.; LEE, S. A multimodal deep log-based user experience (ux) platform for ux evaluation. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 5, p. 1622, 2018. Citado 5 vezes nas páginas 35, 63, 64, 65 e 66.

HUSSEIN, D. M. E.-D. M. A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, Elsevier, v. 30, n. 4, p. 330–338, 2018. Citado 4 vezes nas páginas 16, 19, 56 e 57.

JAIN, V. K.; KUMAR, S.; FERNANDES, S. L. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. *Journal of computational science*, Elsevier, v. 21, p. 316–326, 2017. Citado na página 69.

JOHNSON, J.; KARPATHY, A.; FEI-FEI, L. Densecap: Fully convolutional localization networks for dense captioning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 4565–4574. Citado na página 107.

KARDUNI, A.; CHO, I.; WESSLEN, R.; SANTHANAM, S.; VOLKOVA, S.; ARENDT, D. L.; SHAIKH, S.; DOU, W. Vulnerable to misinformation? verifi! In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. [S.l.: s.n.], 2019. p. 312–323. Citado na página 44.

KARPATHY, A. *What I learned from competing against a ConvNet on ImageNet*. 2014. <http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/>. Acessado em: 2019-11-10. Citado na página 27.

KARPATHY, A. *Stanford CS231n Course Notes*. 2017. <http://cs231n.github.io/neural-networks-1/>. Acessado em: 2019-11-10. Citado na página 31.

KAUR, R.; KAUTISH, S. Multimodal sentiment analysis: A survey and comparison. *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, IGI Global, v. 10, n. 2, p. 38–58, 2019. Citado 4 vezes nas páginas 16, 19, 56 e 60.

KERRACHER, N.; KENNEDY, J.; CHALMERS, K. A task taxonomy for temporal graph visualisation. *IEEE transactions on visualization and computer graphics*, IEEE, v. 21, n. 10, p. 1160–1172, 2015. Citado na página 33.

KHAREL, S.; FERNSTROM, M. Quantifying user experience of mobile applications using a sentimental analysis approach. In: *6th RD NUI GALWAY'UL Alliance postgraduate research day 2016*. [S.l.: s.n.], 2016. p. 19–16. Citado 2 vezes nas páginas 65 e 66.

Khozyainov, I.; Pyshkin, E.; Klyuev, V. Spelling out opinions: Difficult cases of sentiment analysis. In: *2013 International Joint Conference on Awareness Science and Technology Ubi-Media Computing (iCAST 2013 UMEDIA 2013)*. [S.l.: s.n.], 2013. p. 231–237. ISSN null. Citado na página 16.

Khozyainov, I.; Pyshkin, E.; Klyuev, V. Spelling out opinions: Difficult cases of sentiment analysis. In: *2013 International Joint Conference on Awareness Science and Technology Ubi-Media Computing (iCAST 2013 UMEDIA 2013)*. [S.l.: s.n.], 2013. p. 231–237. ISSN null. Citado na página 17.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105. Citado na página 29.

KUCHER, K.; KERREN, A. Text visualization techniques: Taxonomy, visual survey, and community insights. In: IEEE. *2015 IEEE Pacific Visualization Symposium (PacificVis)*. [S.l.], 2015. p. 117–121. Citado na página 36.

KUCHER, K.; PARADIS, C.; KERREN, A. The state of the art in sentiment visualization. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2018. v. 37, n. 1, p. 71–96. Citado 3 vezes nas páginas 17, 19 e 76.

KUCHER, K.; PARADIS, C.; KERREN, A. The state of the art in sentiment visualization. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2018. v. 37, n. 1, p. 71–96. Citado 2 vezes nas páginas 36 e 44.

KWIATKOWSKA, J.; SZÓSTEK, A.; LAMAS, D. Design and business gaps: from literature to practice. In: ACM. *Proceedings of the 2014 Mulitmedia, Interaction, Design and Innovation International Conference on Multimedia, Interaction, Design and Innovation*. [S.l.], 2014. p. 1–7. Citado 2 vezes nas páginas 16 e 63.

LAW, E. L.-C.; ROTO, V.; HASSENZAHL, M.; VERMEEREN, A. P.; KORT, J. Understanding, scoping and defining user experience: a survey approach. In: ACM. *Proceedings of the SIGCHI conference on human factors in computing systems*. [S.l.], 2009. p. 719–728. Citado na página 66.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015. Citado 4 vezes nas páginas 29, 30, 31 e 57.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Taipei, Taiwan, v. 86, n. 11, p. 2278–2324, 1998. Citado 2 vezes nas páginas 31 e 32.

LEE, B.; PLAISANT, C.; PARR, C. S.; FEKETE, J.-D.; HENRY, N. Task taxonomy for graph visualization. In: ACM. *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. [S.l.], 2006. p. 1–5. Citado na página 33.

LEI, X.; JIANG, X.; WANG, C. Design and implementation of a real-time video stream analysis system based on ffmpeg. In: IEEE. *2013 Fourth World Congress on Software Engineering*. [S.l.], 2013. p. 212–216. Citado na página 69.

LI, Y.; YUAN, Y. Convergence analysis of two-layer neural networks with relu activation. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017. p. 597–607. Citado na página 30.

LIU, B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012. Citado na página 24.

LIU, B. et al. Sentiment analysis and subjectivity. *Handbook of natural language processing*, v. 2, n. 2010, p. 627–666, 2010. Citado na página 24.

LIU, X.; XU, A.; GOU, L.; LIU, H.; AKKIRAJU, R.; SHEN, H.-W. Socialbrands: Visual analysis of public perceptions of brands on social media. In: IEEE. *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. [S.l.], 2016. p. 71–80. Citado na página 37.

LIVINGSTONE, S. R.; RUSSO, F. A. The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english. *PloS one*, Public Library of Science, v. 13, n. 5, p. e0196391, 2018. Citado na página 83.

MALGAONKAR, S.; LEE, C. W.; LICORISH, S. A.; SAVARIMUTHU, B. T. R.; TAHIR, A. Appsent a tool that analyzes app reviews. *arXiv preprint arXiv:1907.10191*, 2019. Citado na página 65.

MATTHEWS, T.; JUDGE, T.; WHITTAKER, S. How do designers and user experience professionals actually perceive and use personas? In: ACM. *Proceedings of the SIGCHI conference on human factors in computing systems*. [S.l.], 2012. p. 1219–1228. Citado na página 63.

MCFEE, B.; RAFFEL, C.; LIANG, D.; ELLIS, D. P.; MCVICAR, M.; BATTENBERG, E.; NIETO, O. librosa: Audio and music signal analysis in python. In: *Proceedings of the 14th python in science conference*. [S.l.: s.n.], 2015. v. 8. Citado na página 85.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, Elsevier, v. 5, n. 4, p. 1093–1113, 2014. Citado 3 vezes nas páginas 16, 24 e 56.

MEHRABIAN, A.; FRIAR, J. T. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology*, American Psychological Association, v. 33, n. 3, p. 330, 1969. Citado 2 vezes nas páginas 69 e 85.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, Wiley Online Library, v. 29, n. 3, p. 436–465, 2013. Citado 2 vezes nas páginas 22 e 23.

MOIZER, J.; LEAN, J.; DELL'AQUILA, E.; WALSH, P.; KEARY, A. A.; O'BYRNE, D.; FERDINANDO, A. D.; MIGLINO, O.; FRIEDRICH, R.; ASPERGES, R. et al. An approach to evaluating the user experience of serious games. *Computers & Education*, Elsevier, v. 136, p. 141–151, 2019. Citado 2 vezes nas páginas 65 e 66.

MUKESHIMANA, M.; BAN, X.; KARANI, N.; LIU, R. Multimodal emotion recognition for human-computer interaction: A survey. *System*, v. 9, p. 10, 2017. Citado 4 vezes nas páginas 28, 30, 47 e 49.

Munezero, M.; Montero, C. S.; Sutinen, E.; Pajunen, J. Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Transactions on Affective Computing*, v. 5, n. 2, p. 101–111, April 2014. ISSN 2371-9850. Citado na página 22.

MUNIM, K. M.; ISLAM, I.; KHATUN, M.; KARIM, M. M.; ISLAM, M. N. Towards developing a tool for ux evaluation using facial expression. In: IEEE. *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*. [S.l.], 2017. p. 1–6. Citado 3 vezes nas páginas 64, 65 e 66.

MURRAY, H. A.; MORGAN, C. D. A clinical study of sentiments (i & ii). *Genetic Psychology Monographs*, Heldref Publications, 1945. Citado na página 22.

NETO, J. R. T.; FILHO, P. G.; MANO, L. Y.; UEYAMA, J. Verbo: voice emotion recognition database in portuguese language. *Journal of Computer Science*, v. 14, n. 11, p. 1420–1430, 2018. Citado 2 vezes nas páginas 82 e 85.

NGUYEN, V. D.; VARGHESE, B.; BARKER, A. The royal birth of 2013: Analysing and visualising public sentiment in the uk using twitter. In: IEEE. *2013 IEEE International Conference on Big Data*. [S.l.], 2013. p. 46–54. Citado na página 45.

NIELSEN, J.; LANDAUER, T. K. A mathematical model of the finding of usability problems. In: *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 1993. (CHI '93), p. 206–213. ISBN 0-89791-575-5. Disponível em: <http://doi.acm.org/10.1145/169059.169166>. Citado na página 99.

OELKE, D.; HAO, M.; ROHRDANTZ, C.; KEIM, D. A.; DAYAL, U.; HAUG, L.-E.; JANETZKO, H. Visual opinion analysis of customer feedback data. In: IEEE. *2009 IEEE Symposium on Visual Analytics Science and Technology*. [S.l.], 2009. p. 187–194. Citado na página 41.

PANG, B.; LEE, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. [S.l.], 2004. p. 271. Citado 2 vezes nas páginas 23 e 24.

PARAMITHA, A. I. I.; DANTES, G. R.; INDRAWAN, G. The evaluation of web based academic progress information system using heuristic evaluation and user experience questionnaire (ueq). In: IEEE. *2018 Third International Conference on Informatics and Computing (ICIC)*. [S.l.], 2018. p. 1–6. Citado na página 65.

Patil, H. P.; Atique, M. Sentiment analysis for social media: A survey. In: *2015 2nd International Conference on Information Science and Security (ICISS)*. [S.l.: s.n.], 2015. p. 1–4. ISSN null. Citado na página 16.

PHAM, H. Pyaudio: Portaudio v19 python bindings. *URL: https://people. csail. mit. edu/hubert/pyaudio*, 2006. Citado na página 69.

PICARD, R. W. *Affective computing*. [S.l.]: MIT press, 2000. Citado na página 23.

PICARD, R. W. Affective computing: challenges. *International Journal of Human-Computer Studies*, Elsevier, v. 59, n. 1-2, p. 55–64, 2003. Citado na página 23.

PLAISANT, C. The challenge of information visualization evaluation. In: ACM. *Proceedings of the working conference on Advanced visual interfaces*. [S.l.], 2004. p. 109–116. Citado na página 33.

PLANET, S.; IRIONDO, I. Comparison between decision-level and feature-level fusion of acoustic and linguistic features for spontaneous emotion recognition. In: IEEE. *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)*. [S.l.], 2012. p. 1–6. Citado na página 25.

PORAT, T.; SCHCLAR, A.; SHAPIRA, B. Mate: a mobile analysis tool for usability experts. In: ACM. *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. [S.l.], 2013. p. 265–270. Citado 3 vezes nas páginas 64, 65 e 66.

PORIA, S.; CAMBRIA, E.; BAJPAI, R.; HUSSAIN, A. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, Elsevier, v. 37, p. 98–125, 2017. Citado 4 vezes nas páginas 26, 30, 47 e 49.

PORIA, S.; HUSSAIN, A.; CAMBRIA, E. *Multimodal Sentiment Analysis*. [S.l.]: Springer, 2018. v. 8. Citado 12 vezes nas páginas 16, 19, 24, 25, 26, 48, 49, 55, 56, 57, 60 e 84.

RAMACHANDRAM, D.; TAYLOR, G. W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, IEEE, v. 34, n. 6, p. 96–108, 2017. Citado 3 vezes nas páginas 47, 48 e 50.

ROBINSON, P.; KALIOUBY, R. E. *Computation of emotions in man and machines*. [S.l.]: The Royal Society, 2009. Citado na página 16.

RUDOVIC, O.; PAVLOVIC, V.; PANTIC, M. Context-sensitive dynamic ordinal regression for intensity estimation of facial action units. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 37, n. 5, p. 944–958, 2014. Citado na página 24.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, n. 3, p. 211–252, 2015. Citado 3 vezes nas páginas 30, 83 e 84.

SÁNCHEZ-ADAME, L. M.; MENDOZA, S.; GONZÁLEZ-BELTRÁN, B. A.; RODRÍGUEZ, J.; VIVEROS, A. M. Aux and ux evaluation of user tools in social networks. In: IEEE. *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. [S.l.], 2018. p. 104–111. Citado 2 vezes nas páginas 65 e 66.

SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 4510–4520. Citado na página 83.

SAURO, J.; LEWIS, J. R. *Quantifying the User Experience: Practical Statistics for User Research*. 1st. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2012. ISBN 9780123849687, 9780123849694. Citado 3 vezes nas páginas 91, 98 e 99.

SCHERER, K.; WALLBOTT, H. *'s,"international survey on emotion antecedents and reactions (isear)"*. 1990. Citado na página 85.

SCHERER, K. R. et al. Psychological models of emotion. *The neuropsychology of emotion*, v. 137, n. 3, p. 137–162, 2000. Citado na página 22.

SCHULLER, B.; BATLINER, A.; STEIDL, S.; SEPPI, D. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, Elsevier, v. 53, n. 9-10, p. 1062–1087, 2011. Citado na página 24.

SCHULZ, H.-J.; NOCKE, T.; HEITZLER, M.; SCHUMANN, H. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, IEEE, v. 19, n. 12, p. 2366–2375, 2013. Citado na página 33.

SCHUSTER, M. Speech recognition for mobile devices at google. In: SPRINGER. *Pacific Rim International Conference on Artificial Intelligence*. [S.l.], 2010. p. 8–10. Citado na página 85.

SEGUIN, J. A.; SCHARFF, A.; PEDERSEN, K. Triptech: A method for evaluating early design concepts. In: ACM. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. [S.l.], 2019. p. CS24. Citado 2 vezes nas páginas 65 e 66.

SENG, J. K. P.; ANG, K. L.-M. Multimodal emotion and sentiment modeling from unstructured big data: Challenges, architecture, & techniques. *IEEE Access*, IEEE, v. 7, p. 90982–90998, 2019. Citado 2 vezes nas páginas 47 e 48.

SETCHI, R.; ASIKHIA, O. K. Exploring user experience with image schemas, sentiments, and semantics. *IEEE Transactions on Affective Computing*, IEEE, v. 10, n. 2, p. 182–195, 2017. Citado 3 vezes nas páginas 64, 65 e 66.

SHNEIDERMAN, B.; PLAISANT, C. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: ACM. *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. [S.l.], 2006. p. 1–7. Citado 3 vezes nas páginas 33, 74 e 76.

SHNEIDERMAN, B.; WATTENBERG, M. Ordered treemap layouts. In: IEEE. *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.* [S.l.], 2001. p. 73–78. Citado na página 86.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado 2 vezes nas páginas 31 e 83.

Soares, A. G. M.; dos Santos, D. H.; Barbosa, C. L. R.; Gonçalves, A. S.; dos Santos, C. G. R.; Meiguins, B. S.; Miranda, E. T. C. Visualizing multidimensional data in treemaps with adaptive glyphs. In: *2018 22nd International Conference Information Visualisation (IV)*. [S.l.: s.n.], 2018. p. 58–63. Citado 2 vezes nas páginas 86 e 87.

SOLEYMANI, M.; GARCIA, D.; JOU, B.; SCHULLER, B.; CHANG, S.-F.; PANTIC, M. A survey of multimodal sentiment analysis. *Image and Vision Computing*, Elsevier, v. 65, p. 3–14, 2017. Citado 2 vezes nas páginas 16 e 57.

SOLEYMANI, M.; GARCIA, D.; JOU, B.; SCHULLER, B.; CHANG, S.-F.; PANTIC, M. A survey of multimodal sentiment analysis. *Image and Vision Computing*, Elsevier, v. 65, p. 3–14, 2017. Citado 2 vezes nas páginas 47 e 48.

SOMMERVILLE, I. *Software Engineering*. 9th. ed. USA: Addison-Wesley Publishing Company, 2010. ISBN 0137035152, 9780137035151. Citado na página 66.

SOUZA, K. E.; SERUFFO, M. C.; MELLO, H. D. D.; SOUZA, D. D. S.; VELLASCO, M. M. User experience evaluation using mouse tracking and artificial intelligence. *IEEE Access*, IEEE, v. 7, p. 96506–96515, 2019. Citado na página 65.

STARK, L.; ELLIS, S.; INOUE, H.; FREKSA, C.; ZEEVI, J. Cognitive models direct scanpath eye movements. In: *XII International Conference on Medical and Biological Engineering, Jerusalem, Israel*. [S.l.: s.n.], 1979. Citado na página 76.

STRENGELL, N.; SIGG, S. Local emotions-using social media to understand human-environment interaction in cities. In: IEEE. *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. [S.l.], 2018. p. 615–620. Citado na página 42.

SYAHRIR, S.; SFENRIANTO, S. User experience questioner and heuristics evaluation in online learning environment. *Journal of Theoretical and Applied Information Technology*, v. 97, p. 1071–1081, 02 2019. Citado 2 vezes nas páginas 65 e 66.

SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence*. [S.l.: s.n.], 2017. Citado 2 vezes nas páginas 31 e 83.

TAO, J.; TAN, T. *Affective information processing*. [S.l.]: Springer, 2009. Citado na página 16.

TRIPATHI, S.; BEIGI, H. Multi-modal emotion recognition on iemocap dataset using deep learning. *arXiv preprint arXiv:1804.05788*, 2018. Citado na página 85.

VASA, R.; HOON, L.; MOUZAKIS, K.; NOGUCHI, A. A preliminary analysis of mobile app user reviews. In: ACM. *Proceedings of the 24th Australian Computer-Human Interaction Conference*. [S.l.], 2012. p. 241–244. Citado 2 vezes nas páginas 64 e 66.

VINYALS, O.; BABUSCHKIN, I.; CZARNECKI, W. M.; MATHIEU, M.; DUDZIK, A.; CHUNG, J.; CHOI, D. H.; POWELL, R.; EWALDS, T.; GEORGIEV, P. et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, Nature Publishing Group, v. 575, n. 7782, p. 350–354, 2019. Citado na página 31.

VOIGT, C.; KIESLINGER, B.; SCHÄFER, T. User experiences around sentiment analyses, facilitating workplace learning. In: SPRINGER. *International Conference on Social Computing and Social Media*. [S.l.], 2017. p. 312–324. Citado 2 vezes nas páginas 64 e 66.

VOIGT, C.; KIESLINGER, B.; SCHÄFER, T. User experiences around sentiment analyses, facilitating workplace learning. In: SPRINGER. *International Conference on Social Computing and Social Media*. [S.l.], 2017. p. 312–324. Citado na página 65.

Wagner, J.; Andre, E.; Lingenfelser, F.; Kim, J. Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE Transactions on Affective Computing*, v. 2, n. 4, p. 206–218, Oct 2011. ISSN 2371-9850. Citado na página 25.

WANG, F. Y.; SALLABERRY, A.; KLEIN, K.; TAKATSUKA, M.; ROCHE, M. Senticompass: Interactive visualization for exploring and comparing the sentiments of time-varying twitter data. In: IEEE. *2015 IEEE Pacific Visualization Symposium (PacificVis)*. [S.l.], 2015. p. 129–133. Citado 2 vezes nas páginas 45 e 46.

WANG, Y.; HALEEM, H.; SHI, C.; WU, Y.; ZHAO, X.; FU, S.; QU, H. Towards easy comparison of local businesses using online reviews. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 2018. v. 37, n. 3, p. 63–74. Citado na página 38.

WARD, M. O.; GRINSTEIN, G.; KEIM, D. *Interactive data visualization: foundations, techniques, and applications*. [S.l.]: AK Peters/CRC Press, 2015. Citado na página 34.

WOLF, T.; DEBUT, L.; SANH, V.; CHAUMOND, J.; DELANGUE, C.; MOI, A.; CISTAC, P.; RAULT, T.; LOUF, R.; FUNTOWICZ, M.; BREW, J. Huggingface's transformers: State-of-the-art natural language processing. *ArXiv*, abs/1910.03771, 2019. Citado na página 107.

YADAV, S. K.; BHUSHAN, M.; GUPTA, S. Multimodal sentiment analysis: Sentiment analysis using audiovisual format. In: IEEE. *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.], 2015. p. 1415–1419. Citado 2 vezes nas páginas 64 e 65.

YAN, W.-J.; WU, Q.; LIANG, J.; CHEN, Y.-H.; FU, X. How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior*, Springer, v. 37, n. 4, p. 217–230, 2013. Citado na página 83.

YANG, Q.; SCUITO, A.; ZIMMERMAN, J.; FORLIZZI, J.; STEINFELD, A. Investigating how experienced ux designers effectively work with machine learning. In: ACM. *Proceedings of the 2018 Designing Interactive Systems Conference*. [S.l.], 2018. p. 585–596. Citado na página 18.

YU, M.; ZHOU, R.; WANG, H.; ZHAO, W. An evaluation for vr glasses system user experience: The influence factors of interactive operation and motion sickness. *Applied ergonomics*, Elsevier, v. 74, p. 206–213, 2019. Citado 2 vezes nas páginas 65 e 66.

ZADEH, A.; CHEN, M.; PORIA, S.; CAMBRIA, E.; MORENCY, L.-P. Tensor fusion network for multimodal sentiment analysis. *arXiv preprint arXiv:1707.07250*, 2017. Citado 3 vezes nas páginas 16, 19 e 26.

ZHANG, C.; LIU, Y.; WANG, C. Time-space varying visual analysis of micro-blog sentiment. In: ACM. *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction*. [S.l.], 2013. p. 64–71. Citado na página 43.

ZHANG, L.; WANG, S.; LIU, B. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 8, n. 4, p. e1253, 2018. Citado 2 vezes nas páginas 48 e 49.

APÊNDICE A . PARECER DO COMITÉ DE ÉTICA

DADOS DA VERSÃO DO PROJETO DE PESQUISA

Título da Pesquisa: Avaliações Quantitativas e Qualitativas em Técnicas de Visualização da Informação
Pesquisador Responsável: BIANCHI SERIQUE MEIGUINS
Área Temática:
Versão: 1
CAAE: 15522319.2.0000.0018
Submetido em: 07/06/2019
Instituição Proponente: Universidade Federal do Pará
Situação da Versão do Projeto: Aprovado
Localização atual da Versão do Projeto: Pesquisador Responsável
Patrocinador Principal: Financiamento Próprio

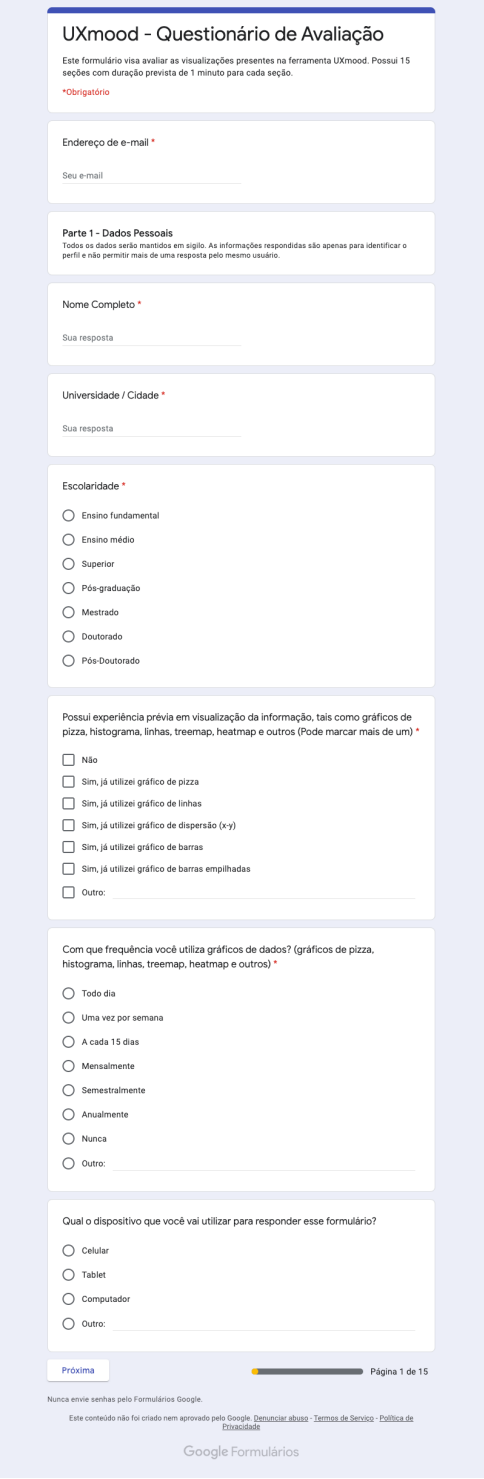
Comprovante de Recepção: PB_COMPROVANTE_RECEPCAO_1374065

DOCUMENTOS DO PROJETO DE PESQUISA

LISTA DE APRECIAÇÕES DO PROJETO

Apreciação	Pesquisador Responsável	Versão	Submissão	Modificação	Situação	Exclusiva do Centro Coord.	Ações
PO	BIANCHI SERIQUE MEIGUINS	1	07/06/2019	03/10/2019	Aprovado	Não	

APÊNDICE B . Questionário On-line



UXmood - Questionário de Avaliação

Este formulário visa avaliar as visualizações presentes na ferramenta UXmood. Possui 15 seções com duração prevista de 1 minuto para cada seção.

*Obrigatório

Endereço de e-mail *

Seu e-mail

Parte 1 - Dados Pessoais

Todos os dados serão mantidos em sigilo. As informações respondidas são apenas para identificar o perfil e não permitir mais de uma resposta pelo mesmo usuário.

Nome Completo *

Sua resposta

Universidade / Cidade *

Sua resposta

Escolaridade *

- Ensino fundamental
- Ensino médio
- Superior
- Pós-graduação
- Mestrado
- Doutorado
- Pós-Doutorado

Possui experiência prévia em visualização da informação, tais como gráficos de pizza, histograma, linhas, treemap, heatmap e outros (Pode marcar mais de um) *

- Não
- Sim, já utilizei gráfico de pizza
- Sim, já utilizei gráfico de linhas
- Sim, já utilizei gráfico de dispersão (x-y)
- Sim, já utilizei gráfico de barras
- Sim, já utilizei gráfico de barras empilhadas
- Outro:

Com que frequência você utiliza gráficos de dados? (gráficos de pizza, histograma, linhas, treemap, heatmap e outros) *

- Todo dia
- Uma vez por semana
- A cada 15 dias
- Mensalmente
- Semestralmente
- Anualmente
- Nunca
- Outro:

Qual o dispositivo que você vai utilizar para responder esse formulário?

- Celular
- Tablet
- Computador
- Outro:

Próxima

Página 1 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

Informações Gerais

A visualização de dados tem sido amplamente utilizada para comunicar resultados nas mais diversas áreas, impulsionada pela crescente quantidade de dados que devem ser analisados. Alguns exemplos de técnicas de visualização utilizadas em user experience (UX) são gráficos de barras ou histogramas para representar dados quantitativos sobre a interação do participante; gráficos para representar trajetória de interação do usuário com o mouse ou o rastreamento do olhar; mapas de calor para apresentar o tempo ou a quantidade de vezes que o participante olhou para determinada área de interesse. Para efetuar a análise de sentimentos também são utilizados frequentes dashboards com velocímetros, gráficos de linhas e histogramas a fim de representar informações quantitativas sobre as análises, wordcloud para representar a frequência por tamanho e qualidade (sentimento ou polarização) pela cor, radar para identificar e comparar informações quantitativas, scatterplots para correlações, técnicas hierárquicas para agrupamento e correlação de dados, representações temporais, entre outras técnicas.

Legenda de emoções usada para representar ícones e cores em todos as questões.

Voltar Próxima Página 2 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

Gantt Chart - Apresentação

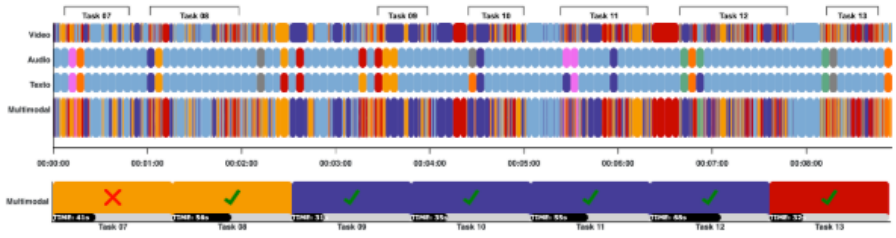
Nesta sessão será apresentada a visualização Gantt Chart. O diagrama de Gantt é um gráfico usado para ilustrar a mudança das emoções durante um intervalo de tempo. Os intervalos de tempo representam o início e o fim de cada emoção e aparecem como barras coloridas sobre o eixo horizontal do gráfico.

O Gantt Chart funciona como uma timeline de sentimentos em que apresenta o tempo e os sentimentos de cada entrada da análise de vídeo, áudio e texto (transcrições do áudio), e uma análise combinada (multimodal). A figura abaixo apresenta os sentimentos classificados em emoções (cores), destacando o tempo e a duração de cada tarefa.

Outra abordagem do Gantt Chart é a análise em conjunto de participantes de um mesmo teste ou conjunto de tarefas, ou mesmo a comparação entre participantes de testes diferentes. A figura a seguir apresenta os sentimentos com a maior duração classificados em emoções (cores), porém, nessa visualização são considerados diversos usuários de um mesmo teste, ou seja, que executaram as mesmas tarefas. Ao final, é apresentado uma média dos sentimentos e tempo, agrupados por tarefas.

Tarefa exemplo: Qual o sentimento predominante com relação aos dados de áudio?
Resposta: Neutral.

Gantt Chart como timeline de emoções



Voltar | Próxima

Página 3 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

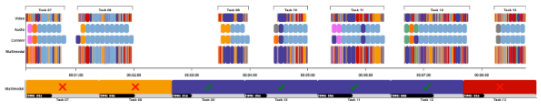
Google Formulários

UXmood - Questionário de Avaliação

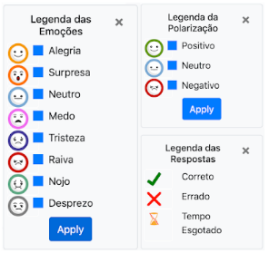
*Obrigatório

Gantt Chart - Tarefas

Gantt Chart como timeline de emoções



Legenda de emoções usada para representar ícones e cores em todos as questões.



1) Existe alguma tarefa em que o usuário errou a resposta da tarefa e mesmo assim apresentou emoções predominantemente positivas? Quais as tarefas? *

- ◯ T7
- ◯ T7 e T8
- ◯ T13
- ◯ nenhuma

2) Qual a emoção predominante na tarefa que demorou mais tempo? *

- ◯ Tristeza
- ◯ Alegria
- ◯ Neutro
- ◯ Raiva

3) Quais os sentimentos predominantes nas tarefas que obtiveram respostas erradas? *

- ◯ Tristeza
- ◯ Alegria
- ◯ Neutro
- ◯ Raiva

Voltar Próxima Página 4 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

*Obrigatório

Gantt Chart - Avaliação

Se eu precisasse fazer mais análises de sentimentos gostaria de usar o Gantt chart? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Gantt chart foi desnecessariamente complexo? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Gantt chart tornou a interpretação dos dados mais fácil? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu precisaria do apoio de um responsável técnico para poder usar o Gantt chart? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Os vários dados no Gantt chart estavam bem representados? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Havia muita inconsistência no Gantt chart? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

A maioria das pessoas aprenderia a usar o Gantt chart muito rapidamente? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Gantt chart foi muito complicado de usar? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

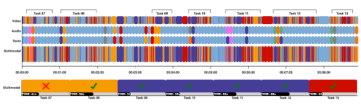
Eu me senti muito confiante usando o Gantt chart? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu precisava aprender muitas coisas antes de começar o Gantt chart? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Gantt Chart como timeline de emoções



Faça comentários sobre o Gantt chart citando pontos positivos e negativos.

Sua resposta

Voltar Próxima Página 5 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

Wordcloud - Apresentação

A Wordcloud apresenta as principais palavras ditas pelo voluntário durante a execução das tarefas no teste. Essas palavras levam em consideração a emoção do voluntário durante o contexto em que foi citada. Por exemplo, na Figura abaixo a palavra "região" foi citada em oito frases classificadas, em sua grande maioria, como Disgust, por isso, essa palavra é associada a este sentimento.

Tarefa exemplo: Qual sentimento com menor predominância na figura abaixo?
Resposta: Nojo.

Wordcloud de frases ditas durante um teste

Voltar | Próxima

Página 6 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

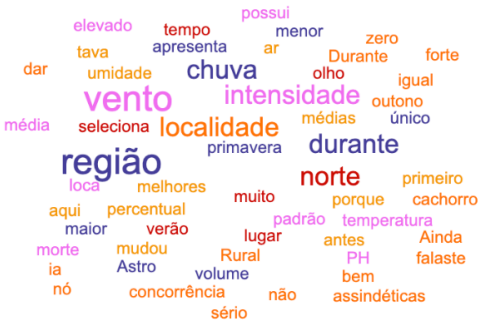
UXmood - Questionário de Avaliação

*Obrigatório

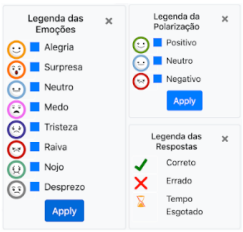
Wordcloud - Tarefas

Wordcloud de frases

Wordcloud de frases ditas durante um teste



Legenda de emoções usada para representar ícones e cores em todos as questões.



1) Qual o sentimento predominante nesta visualização? *

- ◯ Surpresa
- ◯ Medo
- ◯ Tristeza
- ◯ Alegria

2) Qual a palavra mais frequente? *

- ◯ Região
- ◯ Intensidade
- ◯ Vento
- ◯ Norte

3) Qual o sentimento predominante das palavras mais frequentes? *

- ◯ Alegria
- ◯ Tristeza
- ◯ Medo
- ◯ Raiva

Voltar Próxima Página 7 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

*Obrigatório

Wordcloud - Avaliação

Se eu precisasse fazer mais análises de sentimentos gostaria de usar o Wordcloud? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Wordcloud foi desnecessariamente complexo? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Wordcloud tornou a interpretação dos dados mais fácil? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu precisaria do apoio de um responsável técnico para poder usar o Wordcloud? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Os vários dados no Wordcloud estavam bem representados? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Havia muita inconsistência no Wordcloud? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

A maioria das pessoas aprenderia a usar o Wordcloud muito rapidamente? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Wordcloud foi muito complicado de usar? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu me senti muito confiante usando o Wordcloud? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu precisava aprender muitas coisas antes de começar o Wordcloud? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Wordcloud de um texto

Faça comentários sobre o Wordcloud citando pontos positivos e negativos.

Sua resposta

Voltar Próxima Página 8 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

Scanpath / Scatterplot - Apresentação

É o caminho seguido pelos olhos ao ler um documento ou observar uma cena.

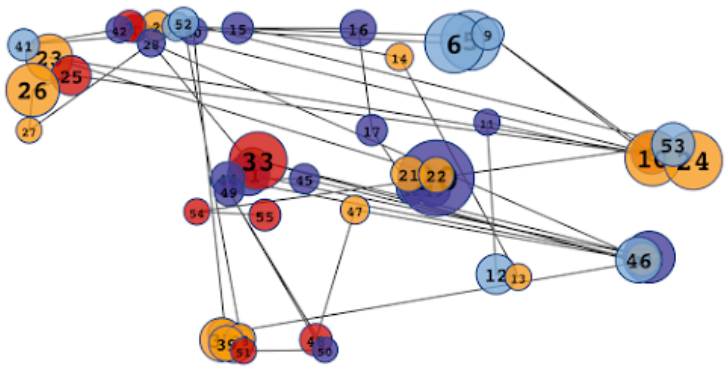
Grafos e Scatterplot são utilizados para apresentar dados referentes ao rastreamento do olhar dos participantes. São geradas duas visualizações: Scanpath das fixações do voluntário durante o teste e Scatterplot dos pontos olhados na interface.

O Scanpath apresenta a ordem de eventos de olhar do participante. O tempo de fixação do olhar é demonstrado com o tamanho do círculo naquele ponto e a classificação do sentimento é associada à cor do círculo. O rastreamento do olhar é desenhado em uma camada por cima do vídeo da tela com as interações do participantes. Esses pontos são animados e sincronizados com o vídeo do voluntário, de modo a permitir que o analista consiga saber para onde o usuário olhou e qual a emoção dele ao olhar esse determinado ponto.

Tarefa exemplo (Scanpath): Qual dos sentimentos possui menor tempo de fixação na visualização 1?
Resposta: Raiva.

Tarefa exemplo (ScatterPlot): Na área de legenda do gif, qual o sentimento predominante no segundo momento que o usuário observa?
Resposta: Baixa.

Visualização 1



Visualização 2

Voltar Próxima Página 9 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

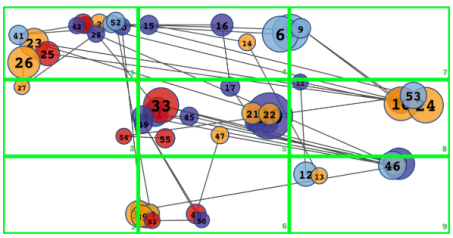
Google Formulários

UXmood - Questionário de Avaliação

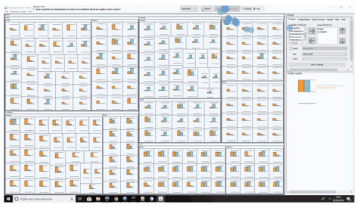
*Obrigatório

Scanpath / Scatterplot - Tarefas

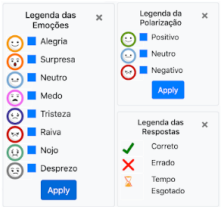
Scanpath de dados do rastreador de olhos



Scatterplot de rastreador de olhos



Legenda de emoções usada para representar ícones e cores em todos as questões.



1) Qual o sentimento associado à fixação de olhar mais longa? *

- ◯ Alegria
- ◯ Tristeza
- ◯ Raiva
- ◯ Surpresa

2) Qual a área da tela com maior frequência de tristeza? *

- ◯ 1
- ◯ 4
- ◯ 5
- ◯ 8

3) Qual a emoção quando o usuário olhou para a legenda? *

- ◯ Alegria
- ◯ Tristeza
- ◯ Raiva
- ◯ Surpresa

Voltar | Próxima | Página 10 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

*Obrigatório

Scanpath / Scatterplot - Avaliação

Se eu precisasse fazer mais análises de sentimentos gostaria de usar o Scanpath / Scatterplot? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Scanpath / Scatterplot foi desnecessariamente complexo? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Scanpath / Scatterplot tornou a interpretação dos dados mais fácil? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu precisaria do apoio de um responsável técnico para poder usar o Scanpath / Scatterplot? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Os vários dados no Scanpath / Scatterplot estavam bem representados? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Havia muita inconsistência no Scanpath / Scatterplot? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

A maioria das pessoas aprenderia a usar o Scanplat / Scatterplot muito rapidamente? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O Scanpath / Scatterplot foi muito complicado de usar? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

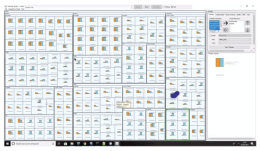
Eu me senti muito confiante usando o Scanpath / Scatterplot? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

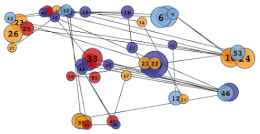
Eu precisava aprender muitas coisas antes de começar o Scanplath / Scatterplot? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Visualização 2



Visualização 1



Faça comentários sobre o Scanpath / Scatterplot citando pontos positivos e negativos.

Sua resposta

Voltar Próxima Página 11 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

EmojiText - Apresentação

EmojiText é uma técnica de visualização da informação desenvolvida para representar a fala humana através da transcrição do texto falado e com dados de análise de sentimentos. Essa visualização utiliza apenas as palavras-chaves para representar uma frase, mantendo preservado o sentido, contexto e a emoção ou polarização associada a cada palavra da frase, ou seja, apenas são exibidos os substantivos, verbos, adjetivos e pronomes, mantendo o sentido original da frase e removidas as stopwords (conectivos) para criar uma visualização mais intuitiva.

Tarefa exemplo: Quais sentimentos estão atrelados a palavra "região".
Resposta: Felicidade e Surpresa.

EmojiText de um texto

Voltar | Próxima

Página 12 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

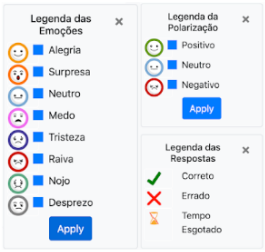
UXmood - Questionário de Avaliação

*Obrigatório

EmojiText - Tarefas

EmojiText de frases

Legenda de emoções usada para representar ícones e cores em todos as questões.



1) Quantas frases são exibidas na visualização? *

- ◯ 9
- ◯ 8
- ◯ 7
- ◯ 6
- ◯ 5

2) Quais a emoção predominante nas frases que envolvem a palavra vento? *

- ◯ Medo
- ◯ Raiva
- ◯ Alegria
- ◯ Surpresa

3) Qual é a emoção da frase que fala sobre a "localidade"? *

- ◯ Tristeza
- ◯ Raiva
- ◯ Alegria
- ◯ Surpresa

Voltar Próxima Página 13 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

UXmood - Questionário de Avaliação

*Obrigatório

EmojiText - Avaliação

Se eu precisasse fazer mais análises de sentimentos gostaria de usar o EmojiText? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O EmojiText foi desnecessariamente complexo? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O EmojiText tornou a interpretação dos dados mais fácil? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu precisaria do apoio de um responsável técnico para poder usar o EmojiText? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Os vários dados do EmojiText estavam bem representados? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Havia muita inconsistência no EmojiText? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

A maioria das pessoas aprenderia a usar o EmojiText muito rapidamente? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

O EmojiText foi muito complicado de usar? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu me senti muito confiante usando o EmojiText? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

Eu precisava aprender muitas coisas antes de começar o EmojiText? *

	1	2	3	4	5	
Discordo totalmente	○	○	○	○	○	Concordo totalmente

EmojiText de frases

Faça comentários sobre o EmojiText citando pontos positivos e negativos.

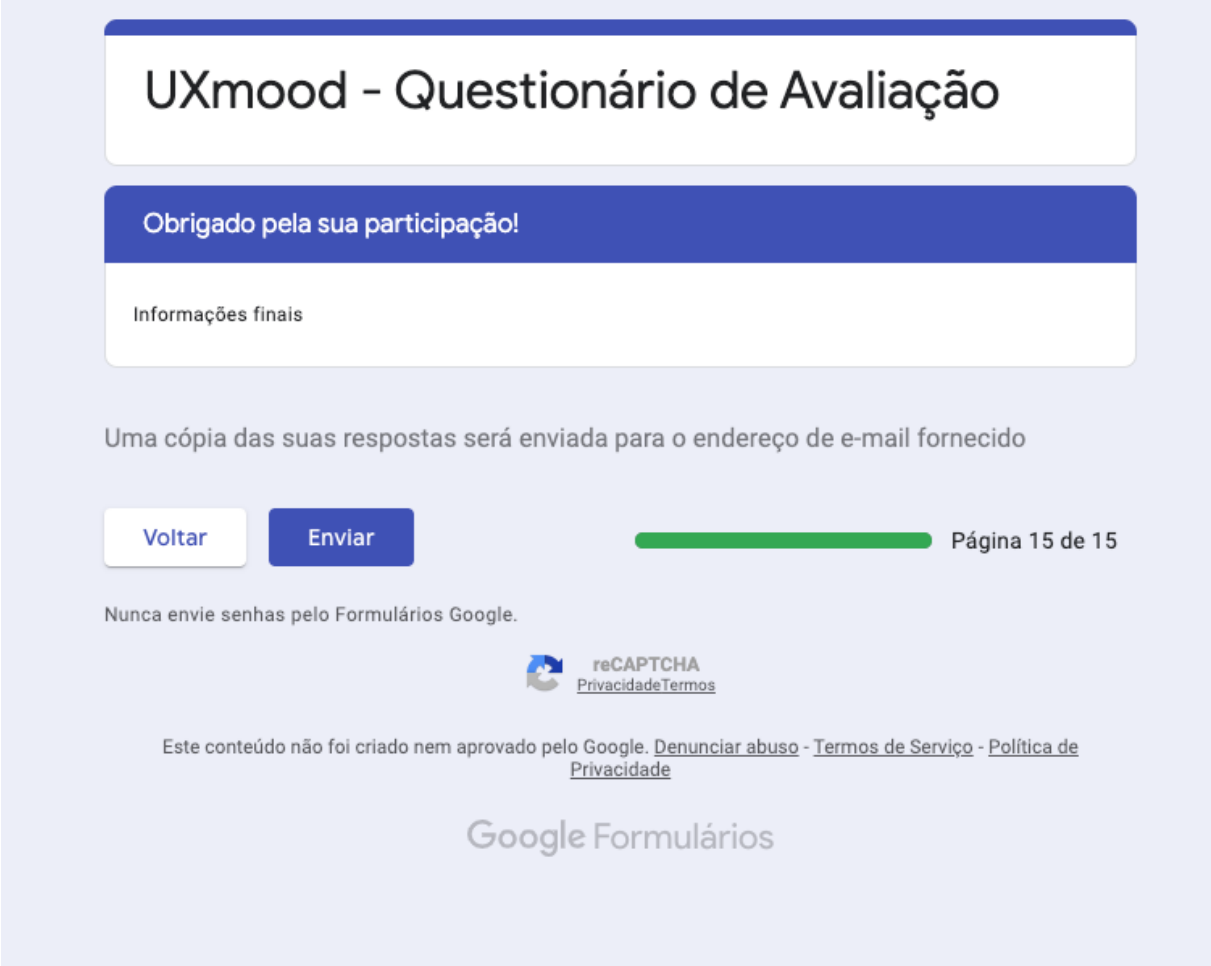
Sua resposta

Voltar Próxima Página 14 de 15

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários


UXmood - Questionário de Avaliação
Obrigado pela sua participação!
Informações finais
Uma cópia das suas respostas será enviada para o endereço de e-mail fornecido
Voltar
Enviar
Página 15 de 15
Nunca envie senhas pelo Formulários Google.
reCAPTCHA
PrivacidadeTermos
Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade
Google Formulários

APÊNDICE C. TCLE - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Programa de Pós-graduação em Ciência da Computação

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Este é um convite para você participar da pesquisa: **"UXmood - Uma Abordagem Baseada em Análise de Sentimentos e Visualização de Dados para Suporte à Avaliação da Usabilidade e da Experiência do Usuário".**

O objetivo desta pesquisa é desenvolver um teste para avaliar as estratégias de exploração de informação visual. Caso concorde com a participação de sua participação nesta pesquisa, você deverá responder a alguns questionários e deverá participar de algumas sessões de avaliação como segue abaixo:

1 - Análise da exploração visual com o tarefas e perguntas, onde estímulos visuais são apresentados em um monitor de computador e você indica com a resposta correta para as perguntas.

Todo o esforço será feito para que as avaliações sejam realizadas de maneira rápida e ágil. Dentre os prováveis benefícios da pesquisa está a satisfação de contribuir para o conhecimento científico sobre a mente humana e também para o desenvolvimento de instrumentos de avaliação adaptados para a realidade brasileira. Os riscos da participação na pesquisa são mínimos e estão relacionados com o cansaço e ansiedade naturais durante a realização de testes. A identidade e os dados pessoais dos participantes serão mantidos em sigilo e não serão divulgados. Caso haja necessidade de identificação dos participantes em publicações e na divulgação dos resultados do trabalho, utilizaremos as iniciais dos nomes ou algum outro código alfanumérico.

Caso você concorde em participar da pesquisa, receberá um questionário (instrumento da pesquisa) que poderá será lido em voz alta por um dos pesquisadores se for necessário. O preenchimento do questionário pode ser feito de forma manual individualmente e entregues junto com este termo no centro ou oralmente para registro por um dos pesquisadores. Após o recebimento, os questionários serão recolhidos e postos em envelopes lacrados para análise posterior dos dados.

A coleta dos dados será feita através observação *in loco* do desempenho dos participantes nos testes do comportamento de exploração visual. Os registros serão automaticamente registrados pelo computador. Os dados obtidos em arquivos digitais com acesso protegido por senha.

Durante todo o período da pesquisa você poderá tirar suas dúvidas ligando para os números de telefone discriminados no final deste documento. Você tem o direito de recusar a participação na pesquisa em qualquer fase da mesma, sem nenhum prejuízo. Os dados que você irá nos fornecer serão confidenciais e serão divulgados apenas em congressos ou publicações científicas. A privacidade do participante do estudo será

preservada em todas as ocasiões. Esses dados serão guardados pelo pesquisador responsável por essa pesquisa em local seguro e por um período de 5 anos. Você receberá uma via do TCLE assinada por ambas as partes e terá acesso ao registro do consentimento, sempre que solicitado.

Não haverá nenhuma despesa adicional ao participante do estudo e também não haverá compensação financeira relacionada à sua participação. Em caso de danos pessoais diretamente decorrentes do estudo assegura-se que o participante possui direito a assistência imediata e gratuita por conta dos pesquisadores.

Caso haja ainda alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com os responsáveis Roberto Yuri da Silva Franco roberto.yuri.franco@gmail.com, Rafael Paixão rafaelpx8812@gmail.com ou Bianchi Serique Meiguins bianchi.serique@gmail.com.

Consentimento Livre e Esclarecido

Após ter sido esclarecido sobre os objetivos, importância e o modo como os dados serão coletados nessa pesquisa, além de conhecer os riscos, desconfortos e benefícios da mesma e ter ficado ciente de todos os meus direitos, concordo em participar da pesquisa **"UXmood - Uma Abordagem Baseada em Análise de Sentimentos e Visualização de Dados para Suporte à Avaliação da Usabilidade e da Experiência do Usuário"** e autorizo a divulgação das informações por mim fornecidas em congressos e/ou publicações científicas, desde que nenhum dado possa me identificar.

Belém, _____/_____/_____

_____

Assinatura do participante da pesquisa