

Treinamento de modelos de visão computacional afetiva por crowdsourcing de rótulos de alvo flexível

Peter Washington*, Departamento de Bioengenharia, Universidade de Stanford, peterwashington@stanford.edu

Haik Kalantarian, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, haik.kalantarian@gmail.com

Jack Kent, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, jackkent@stanford.edu

Arman Husic, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, ahusic@stanford.edu

Aaron Kline, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, akline@stanford.edu

Emilie Leblanc, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, emilie.leblanc@stanford.edu

Cathy Hou, Departamento de Ciência da Computação, Universidade de Stanford, cathyhou@stanford.edu

Cezmi Mutlu, Departamento de Engenharia Elétrica, Universidade de Stanford, cezmi@stanford.edu

Kaitlyn Dunlap, Departamento de Pediatria (Medicina de Sistemas), kaiti.dunlap@stanford.edu

Yordan Penev, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, ypenev@stanford.edu

Nate Stockham, Departamento de Neurociência, Universidade de Stanford, stockham@stanford.edu

Brianna Chrisman, Departamento de Bioengenharia, Universidade de Stanford, briannac@stanford.edu

Kelley Paskov, Departamento de Ciência de Dados Biomédicos, Universidade de Stanford, kpaskov@stanford.edu

Jae-Yoon Jung, Departamento de Pediatria (Medicina de Sistemas), Universidade de Stanford, jaeyjung@stanford.edu

Catalin Voss, Departamento de Ciência da Computação, Universidade de Stanford, catalin@cs.stanford.edu

Nick Haber, Graduate School of Education, Stanford University, nhaber@stanford.edu

Dennis P. Wall*, Departamentos de Pediatria (Medicina de Sistemas), Ciência de Dados Biomédicos e Psiquiatria e Comportamental Ciências, Universidade de Stanford, dpwall@stanford.edu

* Autores correspondentes: Peter Washington (peterwashington@stanford.edu), Dennis P. Wall (dpwall@stanford.edu)

Conflitos de interesse: Dr. Dennis P. Wall é o fundador científico da Cognoa, uma empresa focada em saúde pediátrica digital; a abordagem e as descobertas apresentadas neste artigo são independentes/não relacionadas ao Cognoa. Todos os outros autores declararam que não existem interesses concorrentes.

ABSTRATO

Os classificadores de emoções tradicionalmente preveem emoções discretas. No entanto, as expressões de emoção são muitas vezes subjetivas, exigindo assim um método para lidar com rótulos subjetivos. Exploramos o uso de crowdsourcing para adquirir rótulos soft-target confiáveis e avaliamos um classificador de detecção de emoções treinado com esses rótulos. Centramos nosso estudo no conjunto de dados Child Affective Facial Expression (CAFE), uma coleção padrão-ouro de imagens que descrevem expressões faciais pediátricas junto com 100 rótulos humanos por imagem. Para testar a viabilidade do crowdsourcing para gerar esses rótulos, usamos Microworkers para adquirir rótulos para 207 imagens CAFE. Avaliamos tanto os trabalhadores não filtrados quanto os selecionados por meio de um curto processo de filtragem coletiva. Em seguida, treinamos duas versões de classificadores em rótulos CAFE de alvo suave usando as 100 anotações originais fornecidas com o conjunto de dados: (1) um classificador treinado com rótulos codificados one-hot tradicionais e (2) um classificador treinado com rótulos vetoriais representando a distribuição de respostas do anotador CAFE. Comparamos as distribuições de saída softmax resultantes dos dois classificadores com um teste t independente de 2 amostras de distâncias L1 entre a distribuição de probabilidade de saída do classificador e a distribuição de rótulos humanos. Embora a concordância com o CAFE seja fraca para os trabalhadores da multidão não filtrados, a multidão filtrada concorda com os rótulos do CAFE 100% do tempo para muitas emoções. Embora a pontuação F1 para um classificador codificado one-hot seja muito maior (94,33% vs. 78,68%) em relação aos rótulos CAFE de verdade básica, o vetor de probabilidade de saída do classificador treinado por multidão se assemelha mais à distribuição de rótulos humanos ($t=3,2827$, $p=0,0014$). Relatar uma distribuição de probabilidade de emoção que responde pela subjetividade da interpretação humana. Crowdsourcing, incluindo um mecanismo de filtragem suficiente, é uma solução viável para a aquisição de rótulos softtarget.

INTRODUÇÃO

Modelos de aprendizado de máquina que prevêem emoções humanas a partir de imagens de expressões faciais são cada vez mais usados em sistemas interativos [5, 33-34, 51, 83] e aplicações como análise de sentimento multimodal [42, 67-68], saúde [74, 100], e veículos autônomos [23]. O reconhecimento de emoções é tradicionalmente modelado como um problema de classificação, onde o modelo prevê uma categoria de emoção discreta. No entanto, as expressões faciais costumam ser ambíguas [15, 104-107] e muitas vezes não é ideal para um modelo de aprendizado de máquina produzir uma única classe para um rótulo subjetivo. Felizmente, a maioria dos métodos de aprendizado supervisionado produz uma distribuição de probabilidade sobre todas as classes possíveis. Às vezes, o sistema de computação afetiva visualizará essa distribuição para o usuário [33]. Exemplos incluem serviços comerciais de detecção de emoções como o Affectiva [55-56] e displays de veículos autônomos [83]. Em um grande número de casos de uso, no entanto, apenas a classe com a maior probabilidade é visualizada [34, 51].

Embora o paradigma de treinar um modelo com um rótulo codificado one-hot discreto e prever uma distribuição de probabilidade seja razoável quando os dados de treinamento têm rótulos indiscutíveis, as imagens de expressões faciais podem ter rótulos ambíguos ou mesmo vários rótulos corretos simultaneamente, e o rótulo deve idealmente representar essa incerteza inerente. A rotulagem de alvo suave, em que os rótulos de treinamento representam uma distribuição probabilística em vez de um rótulo codificado em um ponto, é uma solução estabelecida para esse problema. O treinamento com rótulos de alvos suaves resulta em classificadores que prevêem distribuições de probabilidade representativas dos rótulos de alvos suaves [3, 25, 101]. Nossa hipótese é que o crowdsourcing pode gerar distribuições que espelham aquelas geradas em um ambiente de laboratório.

Aqui, exploramos o uso do crowdsourcing para adquirir uma distribuição de rótulos para imagens com classes ambíguas ou múltiplas (chamamos de “rótulos subjetivos”). Primeiro descrevemos a aquisição de rótulos de crowdsourcing para quatro imagens representativas que exibimos ao leitor junto com a distribuição de respostas da multidão a

demonstrar os fenômenos de rótulos subjetivos na computação afetiva. Em seguida, coletamos a rotulagem de um subconjunto do conjunto de dados Child Affective Facial Expression (CAFE), uma coleção de imagens emotivas de crianças que convenientemente vem com 100 anotações humanas independentes por imagem. Em seguida, mostramos que a distribuição crowdsourced espelha a distribuição CAFE original, validando a viabilidade do crowdsourcing para gerar uma distribuição confiável e representativa de rótulos humanos para uma imagem. Por fim, comparamos o desempenho de duas versões de uma rede neural convolucional (CNN) treinada no CAFE: uma com vetores codificados onehot tradicionais e a outra com rótulos de alvo suave baseados nas respostas do anotador CAFE. Descobrimos que o classificador treinado com alvos flexíveis resulta em previsões do classificador que refletem muito mais de perto a verdadeira distribuição humana em sujeitos independentes não incluídos no conjunto de treinamento. Esperamos que este trabalho seja útil para designers e desenvolvedores de modelos de aprendizado de máquina para sistemas de computação afetiva que desejam fornecer saídas probabilísticas para o usuário final.

TRABALHO RELATADO

Embora o crowdsourcing e os rótulos de alvo suave tenham sido estudados na computação afetiva, somos os primeiros a explorar a viabilidade de usar o crowdsourcing para adquirir rótulos de alvo suave confiáveis para detecção de emoções por visão computacional. Descrevemos o trabalho relacionado abaixo.

Detecção de emoção facial

A detecção de emoções faciais é um desafio fundamental para o aprendizado de máquina. Para máquinas inteligentes passarem de forma convincente no teste de Turing [80], uma compreensão da emoção humana é crucial. Existe um forte corpo de literatura sobre aprendizado de máquina para detectar o afeto humano a partir de uma variedade de fluxos de dados, incluindo áudio [67-68, 103], texto [5, 48], imagens [75] e vídeo [28, 84- 85]. Aqui, nos concentramos na detecção de emoções baseada em imagens a partir de expressões faciais.

Fundamental para uma abordagem de visão computacional bem-sucedida para computação afetiva é a representação de recursos da imagem, e existem várias abordagens para projetar tais recursos. Uma abordagem comum é extrair pontos-chave faciais e usar uma representação de recursos que consiste nas coordenadas dos pontos-chave [16, 28, 60, 71]. Essa abordagem funciona bem quando o conjunto de dados é pequeno, pois a representação em si é compacta e, portanto, passível de abordagens de aprendizado leve, como regressão logística, máquinas de vetor de suporte e árvores de decisão. Outra abordagem de extração de recursos, CNNs, pode aprender automaticamente mapas de recursos não lineares relevantes. CNNs muitas vezes resultam em desempenho superior a outros métodos quando o conjunto de dados é suficientemente grande [24, 89].

Detecção de emoções com rótulos subjetivos

Paul Ekman postulou que existem sete emoções humanas fundamentais que são universais em todas as culturas e fronteiras geográficas: feliz, triste, surpresa, raiva, medo, nojo e desprezo [21-22]. No entanto, essas expressões não são mutuamente exclusivas. Du et al. discutiu a existência de emoções compostas, ou combinações de emoções existentes para formar novas [17]. Exemplos de emoções compostas incluem “felizmente surpreso”, “terrivelmente surpreso” e “terrivelmente enojado”. Por meio da detecção de smartphones, Zhang et al. descobriram que pares de emoções que são frequentemente apresentados simultaneamente incluem (feliz, surpresa), (triste, nojo) e (triste, medo) [105]. Esta questão foi explorada para o discurso emocional [15, 107]. Embora algumas emoções possam ser expressas em conjunto, outras podem ser singulares, mas ambíguas. A questão da subjetividade nos rótulos de treinamento, seja por

rótulos ambíguos ou múltiplos rótulos corretos, foi documentado nos campos da saúde digital e computação afetiva em particular [54, 57, 69, 82, 104].

O tópico de rótulos subjetivos em conjuntos de dados de computação afetiva contendo dados de fala e áudio foi explorado em trabalhos anteriores. Mower et al. representam rótulos de emoção na granularidade das declarações, representando assim um perfil de tempo de como a emoção dominante na fala muda rapidamente ao longo do tempo [58-59]. Fujioka alterna entre atualizar os parâmetros da rede neural e atualizar os parâmetros de importância da amostra em cada iteração de treinamento [26]. Ando et al. utilizam treinamento de alvo suave, onde os rótulos de emoção são baseados na proporção de anotações humanas em vez da codificação one-hot tradicional [3].

O treinamento de alvo flexível é um método geral de aprendizado de máquina para lidar com rótulos de treinamento subjetivos. Essa abordagem é particularmente desejável quando vários rótulos são adquiridos por imagem. A classificação de rótulos suaves pode ser benéfica porque eles podem levar em conta a subjetividade herdada nos rótulos e são robustos contra ruído aleatório [79]. Uma variação desse método é uma função de perda suave, que consiste em subtrair a distância mínima entre classes da distância máxima dentro da classe [99]. Foi demonstrado que o treinamento de alvo suave e de perda supera o treinamento de alvo rígido (codificação one-hot) quando o objetivo do treinamento é produzir uma distribuição de saída como a distribuição de rótulos de anotadores [65], e esse fenômeno foi observado em vários conjuntos de dados e tarefas [81].

A questão dos rótulos subjetivos também foi explorada na análise de sentimento multimodal, onde o objetivo é prever o sentimento de vários fluxos de dados [42, 67-68], incluindo vídeos enriquecidos por afeto. Chaturvedi et al. criou um classificador difuso para prever o grau em que várias emoções são expressas em uma imagem particular [6]. Outra abordagem é prever a quantidade de valência e excitação exibida em eixos contínuos (regressão) em vez de prever categorias (classificação) [62, 64, 73, 102, 106].

Crowdsourcing com rótulos subjetivos

Existem vários corpos de trabalho que descrevem abordagens para lidar com rótulos de crowdsourcing. Kairam e Heer levantam a hipótese de que existem diferenças intrínsecas, mas válidas, entre trabalhadores de multidão ao rotular pontos de dados e, portanto, categorizar trabalhadores por seus padrões de rotulagem [40]. Outros exemplos de categorização de trabalhadores incluem medir a informatividade da amostra, estratégias de aprendizado ativo e cooperativo e controlar as métricas de confiabilidade do rotulador [69].

Houve outras técnicas de aprendizado estatístico além da rotulagem de alvo flexível discutida acima, que foram bem-sucedidas com rótulos de crowdsourcing. Rodrigues e Pereira adicionam uma “camada de multidão” extra no final de uma arquitetura CNN tradicional treinada para prever as saídas de cada rotulador individualmente e, portanto, os vieses dos trabalhadores da multidão [70].

Crowdsourcing tem sido usado para adquirir rótulos emocionais de imagens. Korovina et al. descobriram que os trabalhadores da multidão rotulando categorias de emoções discretas em uma roda de cores tinham pontuações de concordância baixas (valor Kappa inferior a 0,15) [46] enquanto a consistência entre os trabalhadores ao rotular a valência e a excitação era muito mais forte [45].

MÉTODOS

Adquirindo rótulos de multidão para imagens CAFE

Usamos o conjunto de dados CAFE [52-53], que é o maior conjunto de dados público de imagens frontais representando crianças emocionadas. O CAFE é usado como referência em várias publicações de computação afetiva [52-53] e é o padrão

conjunto de dados de avaliação para computação afetiva pediátrica. O CAFE foi originalmente rotulado por 100 avaliadores humanos não treinados, e a distribuição bruta de 100 rótulos humanos por imagem é fornecida junto com os rótulos de verdade básica. Por exemplo, a primeira imagem no conjunto de dados foi rotulada como "irritada" por 62% dos avaliadores e como "desgostosa" por 25% dos avaliadores. Todas as outras emoções receberam 5% ou menos rótulos. Pode-se supor a partir desses números que a imagem parece "principalmente zangada" com "alguma repulsa". A inspeção manual da imagem revela uma expressão facial que pode ser razoavelmente categorizada como "irritada" ou "desgostosa", dependendo do contexto. (As imagens do CAFE são protegidas por direitos autorais e não podem ser republicadas, por isso nos referimos a esta imagem por seu nome de arquivo no conjunto de dados disponível publicamente: *F-AA-01_052-Angry.jpg*).

Para validar a capacidade do crowdsourcing de produzir uma distribuição confiável de rótulos de verdade, coletamos a tarefa de rotular as imagens CAFE e comparamos a distribuição gerada pela multidão resultante com a distribuição relatada no conjunto de dados CAFE original. Todo o crowdsourcing foi realizado no Microworkers.com, uma plataforma de crowdsourcing semelhante ao Amazon Mechanical Turk [63], mas com um pool de trabalhadores mais representativo globalmente [32]. Cada tarefa consistia em rotular uma das sete categorias de emoção (feliz, triste, surpresa, raiva, medo, nojo e neutra) para um subconjunto de imagens no CAFE. Optamos por limitar os rótulos dos avaliadores a classificações absolutas (representações one-hot) porque queríamos capturar o peso relativo de cada emoção em uma imagem. Como os humanos são notoriamente pobres em quantificar com precisão as contribuições relativas de componentes individuais em representações mistas, especialmente no caso do reconhecimento de emoções humanas [8-9, 27], pedimos a cada avaliador que fornecesse apenas a emoção mais saliente de acordo com sua interpretação. Ao adquirir rótulos de 100 trabalhadores de multidão independentes por imagem, cada um dando seu voto para a emoção mais proeminente, criamos uma representação que descreve a subjetividade entre-sujeitos da emoção expressa na imagem.

Adquirimos rótulos para 131 imagens selecionadas aleatoriamente do CAFE e solicitamos 100 rótulos de multidão por imagem. Verificamos manualmente cada rótulo quanto à exatidão, e trabalhadores com rótulos consistentemente de alta qualidade foram recrutados para tarefas de rotulagem adicionais para 76 imagens separadas. Aqui, "alta qualidade" significa que os autores poderiam potencialmente concordar com o rótulo (por exemplo, um rótulo "feliz" para uma imagem claramente "triste" não seria aceito, mas um rótulo "medo" para uma expressão "terrivelmente surpresa" seria aceito). Nosso objetivo ao excluir trabalhadores sem rótulos consistentes de "alta qualidade" era filtrar os trabalhadores da multidão que estavam respondendo aleatoriamente para receber o pagamento, pois esse é um problema comum no crowdsourcing [2, 4, 10, 49].

Todas as tarefas de crowdsourcing foram aprovadas pelo Institutional Review Board (IRB) da Universidade de Stanford. Todos os trabalhadores foram obrigados a assinar um termo de consentimento eletrônico aprovado pelo CEP antes de participar da tarefa.

Treinamento e teste com distribuições de probabilidade de multidão

Tradicionalmente, modelos multiclasse são treinados com perda categórica de entropia cruzada, onde \sum é o operador de soma, C é o número de aulas, p_{eu} é a probabilidade de verdade fundamental da classe e , e q_{eu} é a previsão do classificador para a classe e .

$$- \sum_{e=1}^C p_{eu} \log(q_{eu})$$

Quando as classes verdadeiras são indiscutíveis, que é a suposição usual para classificação, então a distribuição de probabilidade de verdade básica p_{eu} é uma codificação one-hot (isto é, uma probabilidade de 1 para a classe "verdadeira" e uma probabilidade

de 0 para todas as outras classes). No caso de classes subjetivas em que o verdadeiro rótulo pode consistir em uma combinação ponderada de várias classes, como em conjuntos de dados de emoção em que emoções complexas estão presentes, hipotetizamos que fornecer rótulos de alvo suave em vez de codificações one-hot resultará em previsões de classificador para sujeitos humanos separados que se assemelham à distribuição de resposta do anotador humano.

Treinamos um modelo de aprendizado de máquina usando dois conjuntos de rótulos de imagem: (1) os rótulos CAFE originais como vetores codificados onehot e (2) vetores de alvo suave representando a distribuição de 100 respostas humanas do conjunto de dados CAFE original. Recolhemos todas as imagens de 5 crianças selecionadas aleatoriamente do CAFE (F-AA-01, F-EA-39, M-LA-08, M-AA-11 e F-LA-13, correspondendo a uma mulher africana americano, uma americana européia, um latino-americano masculino, um afro-americano masculino e uma latino-americana feminina) e os usei como imagens de conjunto de teste. As demais imagens foram usadas para treinar o classificador. 1.141 imagens (196 com raiva, 180 com nojo, 135 com medo, 206 felizes, 222 neutras, 103 tristes e 99 surpresas) foram usadas no conjunto de trem e 51 imagens (9 com raiva, 11 com nojo, 5 com medo, 9 felizes, 8 neutras, 5 triste,

Transferimos o aprendizado em uma CNN ResNet-152 [31] pré-treinada no ImageNet [13]. Treinamos cada rede neural usando a estrutura Keras [7] com um back-end TensorFlow [1] por 100 épocas com um tamanho de lote de 16 e uma taxa de aprendizado de 0,0003 usando a otimização de Adam [43]. Para aumentar a generalização do processo de treinamento e reduzir o overfitting, aplicamos as seguintes estratégias de aumento de dados: uma faixa de rotação de 7 graus, uma faixa de zoom de 15%, uma faixa de cisalhamento de 5%, uma faixa de brilho de 70% a 130%, e inversão horizontal.

RESULTADOS

Demonstração de Emoções Subjetivas

Os métodos descritos aqui não são específicos do CAFE. Nós nos concentramos no CAFE neste artigo como um estudo de caso de um conjunto de dados de computação afetiva popular e como um conjunto de dados que fornece rótulos de verdade para muitos anotadores humanos (100) por imagem. No entanto, as imagens do CAFE estão sujeitas a direitos autorais e não podem ser republicadas. Para fornecer ao leitor exemplos visuais de expressões faciais com um grande número de anotações de multidão por imagem, exibimos imagens gratuitas para republicação na Figura 1. Para cada imagem, adquirimos 200 rótulos de crowdsourcing da Microworkers.com, conforme descrito acima.

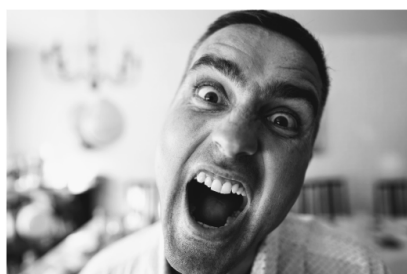
A Figura 1A mostra uma imagem que pode ser rotulada como zangada ou enojada, e a Figura 1C mostra uma imagem que é possivelmente zangada, medrosa, surpresa ou alguma combinação dos 3. Mais contexto é necessário para obter total confiança sobre as verdadeiras classes. A Figura 1B apresenta uma emoção composta, onde o indivíduo aparenta estar “terrivelmente surpreso”. Atribuir apenas uma única categoria à imagem seria enganoso. A Figura 1D descreve uma situação em que não está claro se o rosto neutro do indivíduo parece triste ou se ele está fazendo uma cara triste. Em casos como esse, provavelmente seria necessário um modelo personalizado de reconhecimento de emoções.



Anger: 55%, Disgust: 36%
(Original Label: Anger)



Fear: 42.5%, Surprise: 49.5%
(Original Label: Surprise)



Anger: 27%, Fear: 22%, Surprise: 34%
(Original Label: Anger)



Neutral: 58%, Sad: 39.5%
(Original Label: Neutral)

cada imagem acima. As porcentagens de rótulos para cada categoria de emoção são exibidas abaixo da imagem para todas as emoções que recebem pelo menos 10% dos votos. (A) Esta expressão pode ser raiva ou nojo. (B) Esta expressão é uma expressão composta de medo e surpresa. (C) Dependendo do contexto, esta imagem pode ser raiva, medo ou surpresa (ou alguma combinação). (D) Não está claro se esse rosto é neutro ou triste, destacando a necessidade de técnicas personalizadas de reconhecimento de emoções.

Também quantificamos a subjetividade das imagens no CAFE. Medimos o número de imagens com 80% de anotações representadas com os N principais rótulos mais frequentes para N variando de 1 a 5 inclusive (histograma na Figura 2). Vemos que, embora muitas emoções não contenham muita subjetividade (N=1), a maioria das imagens são ambíguas ou compostas por 2 ou mais emoções. Quando o corte é aumentado para 90% (Figura 3), o número de rótulos subjetivos aumenta ainda mais.

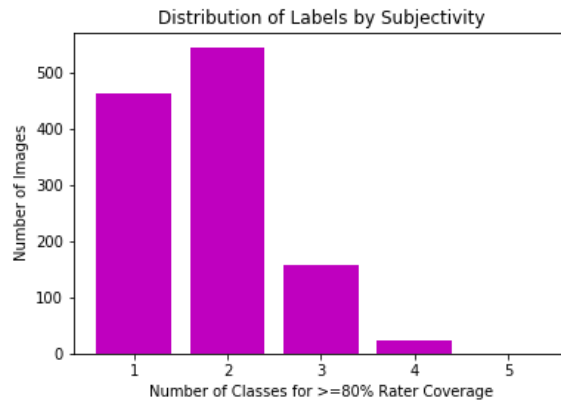


Figura 2: Distribuição dos rótulos por subjetividade. Histograma do número de classes mais votadas necessárias para atingir uma cobertura maior ou igual a 80% do avaliador para cada imagem.

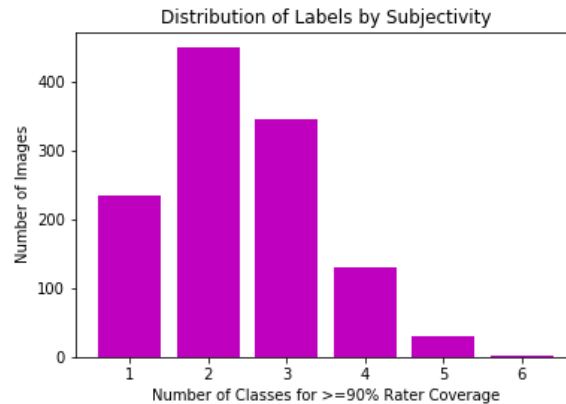


Figura 3: Distribuição dos rótulos por subjetividade. Histograma do número de classes mais votadas necessárias para atingir uma cobertura maior ou igual a 90% do avaliador para cada imagem.

Comparação de rótulos CAFE e desempenho da multidão

Ao olhar para o rótulo de consenso majoritário, a multidão filtrada concordou com os rótulos CAFE 100% do tempo para feliz, neutro, triste e surpresa. Houve concordância de 90% para repulsa, concordância de 75% para raiva e concordância de 50% para medo. Ao combinar rótulos comumente confusos em uma classe (“raiva + nojo” e “medo + surpresa”), a multidão filtrada concordou com os rótulos CAFE 100% do tempo para feliz, neutro, triste e “medo + surpresa” e 88,8% para “raiva + nojo”.

Tabela 1: Distribuição do anotador original do CAFE vs. *filtrado* distribuição do trabalhador para o assunto F-AA-15 no CAFE.

Imagem	Distribuição CAFE Labeler (contar)	Distribuição de rótulos colaborativos (contar)
Raiva, Desgosto, Medo, Feliz, Neutro, Triste, Surpreso		
9990-angry_F-AA-15.jpg	30, 37, 15, 8, 0, 8, 2	7, 3, 0, 4, 0, 0, 0
10108-angryopen_F-AA-15.jpg 10194-	29, 6, 35, 1, 1, 23, 5	2, 2, 4, 0, 0, 6, 0
disgust_F-AA-15.jpg 10288-	3, 86, 3, 2, 1, 5, 0	2, 10, 0, 0, 0, 2, 0
disgustwithtongue_F-AA-15.jpg 10383-	3, 91, 0, 3, 2, 0, 1	1, 6, 1, 5, 0, 0, 1
fearful_F-AA-15.jpg	2, 1, 82, 2, 1, 6, 6	0, 1, 10, 0, 0, 0, 3
10461-fearfulopen_F-AA-15.jpg	2, 3, 58, 2, 3, 1, 31	0, 0, 5, 0, 0, 0, 9
10526-happy_F-AA-15.jpg 10739-	1, 0, 0, 96, 2, 1, 0	0, 0, 0, 14, 0, 0, 0
neutral_F-AA-15.jpg 10867-	1, 0, 1, 1, 89, 7, 1	0, 0, 0, 0, 14, 0, 0
neutralopen_F-AA-15.jpg 10967-	2, 2, 10, 1, 33, 0, 52	0, 0, 0, 0, 7, 0, 7
sad_F-AA-15.jpg	3, 3, 6, 1, 2, 85, 0, 0	2, 0, 0, 0, 0, 12, 0
11027-sadopen_F-AA-15.jpg	5, 22, 0, 0, 72, 1, 1	0, 0, 3, 0, 0, 11, 0
11079-surprise_F-AA-15.jpg	0, 23, 0, 2, 0, 74	0, 0, 1, 0, 0, 0, 13

Por outro lado, os trabalhadores da multidão não filtrados não concordaram tão fortemente com os rótulos do CAFE ao olhar para o consenso da maioria, destacando a necessidade de medidas de controle de qualidade ao fazer anotações de emoções em crowdsourcing. Houve 100% de concordância para surpreso, 93,3% de concordância para feliz, 83,3% de concordância para triste, 76,9% de concordância para nojo, 64,3% de concordância para zangado, 61,5% de concordância para neutro e 30,8% de concordância para medroso.

Tabela 2: Distribuição do anotador original do CAFE vs. *não filtrado* distribuição coletiva de aulas para a matéria F-AA-93 no CAFE.

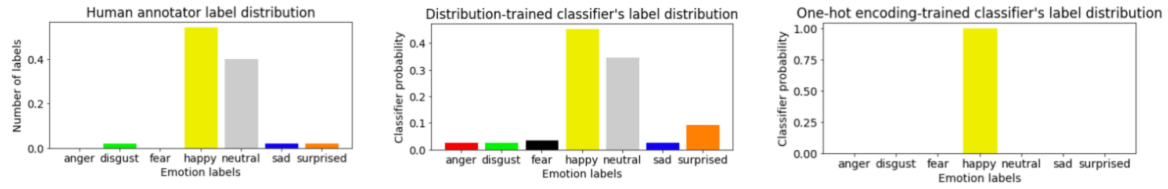
Imagem	Distribuição CAFE Labeler (contar)	Distribuição de rótulos colaborativos (contar)
Raiva, Desgosto, Medo, Feliz, Neutro, Triste, Surpreso		
9979-angry_F-AA-03.jpg	89, 4, 0, 0, 0, 4, 3	78, 31, 2, 8, 3, 2, 3
10100-angryopen_F-AA-03.jpg 10184-	16, 0, 36, 5, 1, 2, 40	17, 3, 38, 15, 0, 0, 54
disgust_F-AA-03.jpg 10280-	17, 41, 2, 12, 19, 8, 1	12, 75, 1, 10, 25, 3, 1
disgustwithtongue_F-AA-03.jpg 10375-	19, 77, 0, 0, 2, 1, 1	19, 85, 2, 13, 6, 0, 2
fearful_F-AA-03.jpg	2, 4, 49, 13, 2, 3, 27	10, 12, 27, 15, 16, 3, 44
10454-fearfulopen_F-AA-03.jpg	0, 1, 27, 4, 1, 0, 67, 1	1, 0, 43, 3, 0, 0, 80
10515-happy_F-AA-03.jpg 10635-	0, 0, 98, 1, 0, 0	0, 4, 0, 113, 8, 1, 1
happyopen_F-AA-03.jpg 10730-	1, 2, 0, 94, 1, 0, 2	0, 0, 0, 126, 0, 0, 1
neutral_F-AA-03.jpg 10858-	3, 2, 0, 2, 73, 19, 1	2, 1, 0, 0, 77, 47, 0
neutralopen_F-AA-03.jpg 10960-	4, 4, 4, 2, 17, 3, 66	1, 7, 21, 1, 10, 5, 82
sad_F-AA-03.jpg	2, 1, 2, 1, 2, 92, 0	8, 9, 3, 0, 4, 103, 0
11021-sadopen_F-AA-03.jpg	1, 6, 21, 15, 13, 30, 14	9, 21, 24, 14, 12, 43, 4
11068-surprise_F-AA-03.jpg	2, 1, 13, 20, 1, 0, 63	1, 1, 12, 30, 0, 1, 82

As Tabelas 1 e 2 comparam a distribuição de rótulos dos rotuladores CAFE originais, bem como os crowd workers filtrados e não filtrados (respectivamente) para uma única imagem. Nos casos filtrados e não filtrados, as distribuições se espelham qualitativamente em termos de seus picos. Em todos os casos, os picos que aparecem na distribuição do anotador CAFE também aparecem na distribuição da multidão e vice-versa. No entanto, essas distribuições são ruidosas e a relação entre os picos não pode ser garantida (por exemplo, se “raiva” tiver mais rótulos do que “nojo” para anotadores do CAFE, “nojo” pode ter mais classificações para anotadores de multidão). As distribuições de multidão geradas devem, portanto, ser consideradas como uma aproximação ruidosa da verdadeira distribuição de probabilidade, e trabalhos adicionais devem levar em consideração esse ruído na representação do rótulo.

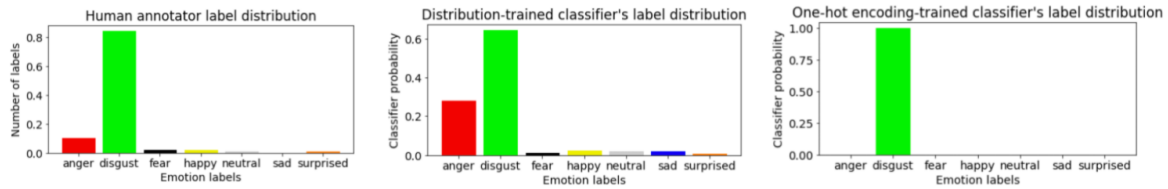
Treinamento e teste com distribuições de probabilidade de multidão

Avaliamos os modelos com pontuação F1 em vez de precisão porque o CAFE não é um conjunto de dados balanceado. Ao treinar com os rótulos codificados one-hot, a pontuação F1 no conjunto de teste retido é de 94,33%. Enfatizamos que esse alto desempenho é enganoso devido à ambigüidade dos rótulos de verdade básica. Ao treinar com vetores que representam a distribuição de rótulos humanos, a pontuação F1 no conjunto de teste realizado é de 78,68%. Embora a pontuação F1 seja menor ao treinar com rótulos de distribuição humanos, a distribuição de previsões de emoção se assemelha muito mais à distribuição de rótulos humanos para o classificador treinado em distribuição. Para muitas aplicações de computação afetiva, ter uma distribuição de rótulos representativa é mais importante do que precisão absoluta. A distância L1 média entre a distribuição de rótulo humano para o conjunto de teste e o classificador treinado em distribuição é 0,3727 (DP=0,3000); a distância L1 média entre a distribuição de rótulo humano e o classificador treinado em codificação one-hot é 0,6078 (DP=0,4143). A diferença nas distâncias L1 entre esses dois grupos é estatisticamente significativa de acordo com um teste t independente de 2 amostras ($t=3,2827$, $p=0,0014$). Para visualizar essa diferença, a Figura 4 compara a distribuição da emoção verdadeira com a distribuição emitida de cada um dos dois classificadores para 3 imagens representativas no conjunto de teste com rótulos subjetivos. A diferença nas distâncias L1 entre esses dois grupos é estatisticamente significativa de acordo com um teste t independente de 2 amostras ($t=3,2827$, $p=0,0014$). Para visualizar essa diferença, a Figura 4 compara a distribuição da emoção verdadeira com a distribuição emitida de cada um dos dois classificadores para 3 imagens representativas no conjunto de teste com rótulos subjetivos. A diferença nas distâncias L1 entre esses dois grupos é estatisticamente significativa de acordo com um teste t independente de 2 amostras ($t=3,2827$, $p=0,0014$). Para visualizar essa diferença, a Figura 4 compara a distribuição da emoção verdadeira com a distribuição emitida de cada um dos dois classificadores para 3 imagens representativas no conjunto de teste com rótulos subjetivos.

Official label: Neutral



Official label: Disgust



Official label: Neutral

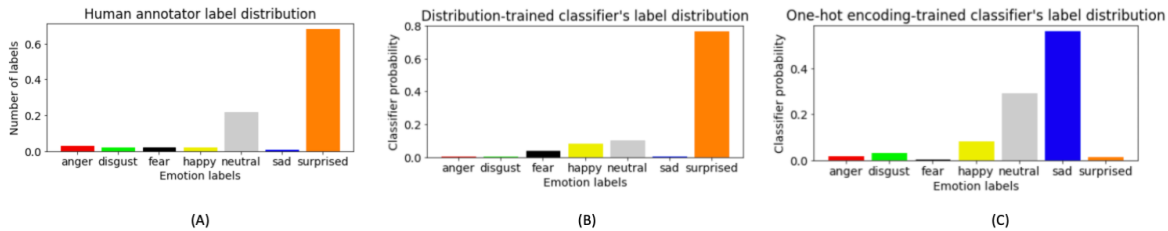


Figura 4: Comparação da distribuição de rótulos humanos (A) e a probabilidade prevista de um classificador treinado com um vetor de probabilidade representando a variedade de rótulos humanos (B) versus um vetor codificado one-hot para imagens com rótulos de emoção subjetiva (C).

DISCUSSÃO

Designers de interação e desenvolvedores de sistemas de computação afetiva devem considerar se alvos fáceis ou difíceis é a representação de rótulo mais apropriada para treinar um classificador de visão computacional afetiva para um aplicativo e conjunto de dados específicos. Modelos de visão computacional afetiva que são otimizados para entender a gama potencialmente diversificada de interpretações humanas de emoções podem ser usados em várias aplicações de sistemas interativos, como sistemas baseados em IA que auxiliam indivíduos com autismo e outros atrasos no desenvolvimento [11-12, 14, 29, 35-39, 44, 66, 72, 84-85, 95-96], sistemas de e-learning [41, 78] ou ferramentas de triagem de diagnóstico em casa para condições psiquiátricas [18-20, 30, 47, 50, 76-77, 86-88, 90-94, 97].

Existem várias limitações deste trabalho. Este estudo foi realizado em um único conjunto de dados. Para que esses resultados sejam generalizados para outros tipos de imagens, inclusive para geração de rótulos de alvo suave de crowdsourcing em domínios fora do reconhecimento de emoções, outros conjuntos de dados devem ser explorados. Outra limitação é que não registramos ou consideramos possíveis vieses nas etapas de controle de qualidade para filtrar a multidão. Um estudo mais aprofundado sobre como diferentes mecanismos de qualidade de multidão afetam o resultado seria interessante, pois a qualidade do rótulo de dados pode afetar drasticamente de um

algoritmo de aprendizado de máquina. Uma limitação final é que não tínhamos um método confiável para separar emoções compostas de rótulos ambíguos.

Ao adquirir rótulos de 100 trabalhadores de multidão independentes por imagem, cada um dando seu voto para a emoção mais proeminente, criamos uma representação que descreve a subjetividade entre-sujeitos da emoção expressa na imagem. Essa representação ofusca notavelmente a subjetividade intrassujeito, e uma alternativa que deve ser estudada em trabalhos futuros é solicitar a cada avaliador que forneça seleções múltiplas por meio de uma escala semântica, como em Korovina et al. [45-46].

CONCLUSÃO

Para muitas aplicações de computação afetiva, relatar uma distribuição de probabilidade de emoção que representa a subjetividade da interpretação humana pode ser mais importante do que as métricas tradicionais de aprendizado de máquina. Crowdsourcing é uma solução viável para a aquisição de rótulos soft-target, desde que haja um mecanismo de filtragem suficiente para selecionar crowd workers confiáveis.

AGRADECIMENTOS

Este trabalho foi financiado em parte por fundos para DPW do National Institutes of Health (1R01EB025025-01, 1R21HD091500-01, 1R01LM013083, 1R01LM013364), National Science Foundation (Prêmio 2014232), The Hartwell Foundation, Bill and Melinda Gates Foundation, Coulter Foundation, Lucile Packard Foundation, Weston Havens Foundation e subsídios do programa de Stanford's Human Centered Artificial Intelligence Program, Precision Health and Integrated Diagnostics Center (PHIND) de Stanford, Beckman Center de Stanford, Bio-X Center de Stanford, Predictives and Diagnostics Accelerator (SPADA) Spectrum, Spark Program in Translational Research de Stanford, Stanford mediaX e Neuroscience: Translate Program do Instituto de Neurociências Wu Tsai de Stanford. Também reconhecemos o generoso apoio de David Orr, Imma Calvo, Bobby Dekesyer e Peter Sullivan. PW

CONFORMIDADE COM NORMAS ÉTICAS

Financiamento: Este estudo foi financiado pelo National Institutes of Health (1R01EB025025-01, 1R21HD091500-01, 1R01LM013083, 1R01LM013364), National Science Foundation (Award 2014232), The Hartwell Foundation, Bill and Melinda Gates Foundation, Coulter Foundation, Lucile Packard Foundation, da Weston Havens Foundation e doações do programa de Stanford's Human Centered Artificial Intelligence Program, Stanford's Precision Health and Integrated Diagnostics Center (PHIND), Stanford's Beckman Center, Stanford's Bio-X Center, Predictives and Diagnostics Accelerator (SPADA) Spectrum, Stanford's Spark Program in Translational Research, Stanford mediaX e Neurociência do Instituto de Neurociências Wu Tsai de Stanford: Programa Tradutor. Também reconhecemos o generoso apoio de David Orr, Imma Calvo, Bobby Dekesyer e Peter Sullivan. PW

Consentimento informado: O consentimento informado foi obtido de todos os participantes individuais (trabalhadores da multidão em Microworkers.com) incluídos no estudo.

Aprovação ética: Todos os procedimentos realizados em estudos envolvendo participantes humanos estavam de acordo com os padrões éticos da instituição (aprovados pelo Conselho de Revisão Institucional da Universidade de Stanford) e com a declaração de Helsinki de 1964 e suas emendas posteriores ou padrões éticos comparáveis.

REFERÊNCIAS

- [1] Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin et al. "Tensorflow: um sistema para aprendizado de máquina em larga escala." *Dentro 12º Simpósio {USENIX} sobre design e implementação de sistemas operacionais ({OSDI} 16)*, pp. 265-283. 2016.
- [2] Allahbakhsh, Mohammad, Boualem Benatallah, Aleksandar Ignjatovic, Hamid Reza Motahari-Nezhad, Elisa Bertino e Schahram Dustdar. "Controle de qualidade em sistemas de crowdsourcing: problemas e direções." *IEEE Internet Computing* 17, não. 2 (2013): 76-81.
- [3] Ando, Atsushi, Satoshi Kobashikawa, Hosana Kamiyama, Ryo Masumura, Yusuke Ijima e Yushi Aono. "Treinamento de alvo suave com expressões emocionais ambíguas para classificação de emoção de fala baseada em DNN." *Dentro 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4964-4968. IEEE, 2018.
- [4] Buchholz, Sabine e Javier Latorre. "Testes de preferência de crowdsourcing e como detectar trapaças." *Dentro Décima Segunda Conferência Anual da International Speech Communication Association*. 2011.
- [5] Cambria, Erik, Dipankar Das, Sivaji Bandyopadhyay e Antonio Feraco. "Computação afetiva e análise de sentimento." *Dentro Um guia prático para análise de sentimento*, pp. 1-10. Springer, Cham, 2017.
- [6] Chaturvedi, Iti, Ranjan Satapathy, Sandro Cavallari e Erik Cambria. "Raciocínio de senso comum difuso para análise de sentimento multimodal." *Letras de Reconhecimento de Padrão* 125 (2019): 264-270.
- [7] Chollet, François. "Keras: a biblioteca de aprendizado profundo do python." *ascl*(2018): ascl-1806.
- [8] Coolican, Jamesie, Gail A. Eskes, Patricia A. McMullen e Erin Lecky. "Vieses de percepção no processamento de identidade facial e emoção." *Cérebro e Cognição* 66, n.º. 2 (2008): 176-187.
- [9] Coren, Stanley e James A. Russell. "O domínio relativo de diferentes expressões faciais de emoção sob condições de ambigüidade perceptiva." *Cognição e Emoção* 6, não. 5 (1992): 339-356.
- [10] Daniel, Florian, Pavel Kucheraev, Cinzia Cappiello, Boualem Benatallah e Mohammad Allahbakhsh. "Controle de qualidade em crowdsourcing: uma pesquisa de atributos de qualidade, técnicas de avaliação e ações de garantia." *Levantamentos de Computação ACM (CSUR)* 51, n.º. 1 (2018): 1-40.
- [11] Daniels, Jena, Jessey N. Schwartz, Catalin Voss, Nick Haber, Azar Fazel, Aaron Kline, Peter Washington, Carl Feinstein, Terry Winograd e Dennis P. Wall. "Estudo exploratório examinando a viabilidade em casa de uma ferramenta vestível para aprendizagem socioafetiva em crianças com autismo." *NPJ medicina digital* 1, não. 1 (2018): 1-10.
- [12] Daniels, Jena, Nick Haber, Catalin Voss, Jessey Schwartz, Serena Tamura, Azar Fazel, Aaron Kline et al. "Teste de viabilidade de uma ajuda comportamental vestível para aprendizagem social em crianças com autismo." *Informática clínica aplicada* 9, não. 1 (2018): 129.
- [13] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li e Li Fei-Fei. "Imagenet: Um banco de dados de imagem hierárquica em larga escala." *Dentro Conferência IEEE 2009 sobre visão computacional e reconhecimento de padrões*, pp. 248-255. IEEE, 2009.
- [14] Deriso, David, Joshua Susskind, Lauren Krieger e Marian Bartlett. "Espelho de emoção: uma nova intervenção para o autismo baseada no reconhecimento de expressões em tempo real." *Dentro Conferência Europeia sobre Visão Computacional*, pp. 671-674. Springer, Berlim, Heidelberg, 2012.
- [15] Devillers, Laurence, Laurence Vidrascu e Lori Lamel. "Desafios na anotação de emoções da vida real e detecção baseada em aprendizado de máquina." *Redes neurais* 18, não. 4 (2005): 407-422.
- [16] Dinculescu, Adrian, Andra Băltoiu, Carmen Strungaru, Livia Petrescu, Cristian Vizitiu, Alexandru Mandu, Nicoară Talpeş e Vlad Văleanu. "Identificação automática de marcos faciais antropológicos para detecção de emoções." *Dentro 2019 9ª Conferência Internacional sobre Avanços Recentes em Tecnologias Espaciais (RAST)*, pp. 585-590. IEEE, 2019.
- [17] Du, Shichuan, Yong Tao e Aleix M. Martinez. "Expressões faciais compostas de emoção." *Anais da Academia Nacional de Ciências* 111, n.º. 15 (2014): E1454-E1462.
- [18] Duda, Marlena, Jena Daniels e Dennis P. Wall. "Avaliação clínica de uma avaliação de risco de autismo nova e móvel." *Jornal de autismo e transtornos do desenvolvimento* 46, n.º. 6 (2016): 1953-1961.
- [19] Duda, M., N. Haber, J. Daniels e DP Wall. "Validação colaborativa de um sistema de classificação de aprendizado de máquina para autismo e TDAH." *psiquiatria translacional* 7, não. 5 (2017): e1133-e1133.
- [20] Duda, M., R. Ma, N. Haber e DP Wall. "Uso de aprendizado de máquina para distinção comportamental de autismo e TDAH." *psiquiatria translacional* 6, não. 2 (2016): e732-e732.
- [21] EKMAN, Paul. "Existem emoções básicas?" (1992): 550.
- [22] EKMAN, Paul. "Emoções básicas." *Manual de cognição e emoção* 98, n.º. 45-60 (1999): 16.
- [23] Eyben, Florian, Martin Wöllmer, Tony Poitschke, Björn Schuller, Christoph Blaschke, Berthold Färber e Nhu Nguyen-Thien. "Emoção na estrada — necessidade, aceitação e viabilidade da computação afetiva no carro." *Avanços na interação humano-computador* 2010 (2010).
- [24] Fan, Yingruo, Jacqueline CK Lam e Victor OK Li. "Rede neural convolucional de conjunto multirregional para expressão facial

- reconhecimento." Em *Conferência Internacional sobre Redes Neurais Artificiais*, pp. 84-94. Springer, Cham, 2018.
- [25] Fang, Xi, Jiancheng Yang e Bingbing Ni. "Refinaria de rótulos estocásticos: em direção a uma melhor distribuição de rótulos alvo." Dentro *2020 25ª Conferência Internacional sobre Reconhecimento de Padrões (ICPR)*, pp. 9115-9121. IEEE, 2021. See More
- [26] Fujioka, Takuya, Dario Bertero, Takeshi Homma e Kenji Nagamatsu. "Lidando com a ambigüidade dos rótulos emocionais por meio do meta-aprendizado." *arXiv pré-impressão arXiv:1911.02216*(2019).
- [27] Gray, Katie LH, Wendy J. Adams, Nicholas Hedger, Kristiana E. Newton e Matthew Garner. "Rostos e consciência: fatores de baixo nível, não emocionais, determinam o domínio perceptivo." *Emoção*13, não. 3 (2013): 537.
- [28] Haber, Nick, Catalin Voss, Azar Fazel, Terry Winograd e Dennis P. Wall. "Uma abordagem prática para subtração de recursos neutros em tempo real para reconhecimento de expressões faciais." Dentro *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1-9. IEEE, 2016.
- [29] Haber, Nick, Catalin Voss e Dennis Wall. "Tornando as emoções transparentes: o Google Glass ajuda crianças autistas a entender expressões faciais por meio da terapia de realidade aumentada." *Espectro IEEE*57, n°. 4 (2020): 46-52.
- [30] Halim, Abbas, Garberson Ford, Stuart Liu-Mayo, Eric Glover e Dennis P. Wall. "Abordagem de IA multimodular para simplificar o diagnóstico de autismo em crianças pequenas." *Relatórios Científicos (Nature Publisher Group)*10, não. 1 (2020).
- [31] Ele, Kaiming, Xiangyu Zhang, Shaoqing Ren e Jian Sun. "Aprendizado residual profundo para reconhecimento de imagem." Dentro *Anais da Conferência IEEE sobre Visão Computacional e Reconhecimento de Padrões*, pp. 770-778. 2016.
- [32] Hirth, Matthias, Tobias Hoßfeld e Phuoc Tran-Gia. "Anatomia de uma plataforma de crowdsourcing usando o exemplo de microworkers. com." Dentro *2011 Quinta conferência internacional sobre serviços móveis e de internet inovadores em computação ubíqua*, pp. 322-329. IEEE, 2011.
- [33] Hupont, Isabelle, Sandra Baldassarri, Eva Cerezo e Rafael Del-Hoyo. "Visualização avançada do efeito humano." Dentro *Conferência Internacional IEEE 2013 sobre Sistemas, Homem e Cibernética*, pp. 2700-2705. IEEE, 2013.
- [34] Jerauld, Robert. "Sistema de detecção e feedback de emoções vestíveis." Patente dos EUA 9.019.174, emitida em 28 de abril de 2015.
- [35] Kalantarian, Haik, Khaled Jedoui, Peter Washington e Dennis P. Wall. "Um jogo para celular para rotulagem automática de imagens." *transações IEEE em jogos*12, não. 2 (2018): 213-218.
- [36] Kalantarian, Haik, Khaled Jedoui, Peter Washington, Qandeel Tariq, Kaiti Dunlap, Jessey Schwartz e Dennis P. Wall. "Rotular imagens com emoção facial e potencial para cuidados de saúde pediátricos." *Inteligência artificial na medicina*98 (2019): 77-86.
- [37] Kalantarian, Haik, Khaled Jedoui, Kaitlyn Dunlap, Jessey Schwartz, Peter Washington, Arman Husic, Qandeel Tariq, Michael Ning, Aaron Kline e Dennis Paul Wall. "O desempenho de classificadores de emoção para crianças com autismo relatado pelos pais: estudo de viabilidade quantitativo." *saúde mental JMIR*7, não. 4 (2020): e13174.
- [38] Kalantarian, Haik, Peter Washington, Jessey Schwartz, Jena Daniels, Nick Haber e Dennis P. Wall. "Adivinha?." *Revista de pesquisa em informática em saúde*3, não. 1 (2019): 43-66.
- [39] Kalantarian, Haik, Peter Washington, Jessey Schwartz, Jena Daniels, Nick Haber e Dennis Wall. "Um sistema móvel gamificado para vídeo de crowdsourcing para pesquisa sobre autismo." Dentro *Conferência internacional IEEE 2018 sobre informática em saúde (ICHI)*, pp. 350-352. IEEE, 2018.
- [40] Kairam, Sanjay e Jeffrey Heer. "Multidões de separação: Caracterizando interpretações divergentes em tarefas de anotação de crowdsourcing." Dentro *Anais da 19ª Conferência da ACM sobre Trabalho Cooperativo e Computação Social Suportados por Computador*, pp. 1637-1648. 2016.
- [41] Kaiser, Robin e Karina Oertel. "Emoções em HCI: um sistema de e-learning afetivo." Dentro *Anais do workshop HCSNet sobre o uso da visão na interação humano-computador-Volume 56*, pp. 105-106. Australian Computer Society, Inc., 2006.
- [42] Kaur, Ramandeep e Sandeep Kautish. "Análise de sentimento multimodal: uma pesquisa e comparação." *Jornal Internacional de Ciência, Gestão, Engenharia e Tecnologia de Serviços (IJSSMET)*10, não. 2 (2019): 38-58.
- [43] Kingma, Diederik P. e Jimmy Ba. "Adam: Um método para otimização estocástica." *arXiv pré-impressão arXiv:1412.6980*(2014).
- [44] Kline, Aaron, Catalin Voss, Peter Washington, Nick Haber, Hessey Schwartz, Qandeel Tariq, Terry Winograd, Carl Feinstein e Dennis P. Parede. "Vidro superpoderoso." *GetMobile: Computação e Comunicações Móveis*23, não. 2 (2019): 35-38.
- [45] Korovina, Olga, Marcos Baez e Fabio Casati. "Confiabilidade do crowdsourcing como um método para coletar rótulos de emoções em fotos." *Notas de pesquisa do BMC*12, não. 1 (2019): 1-6.
- [46] Korovina, Olga, Fabio Casati, Radoslaw Nielek, Marcos Baez e Olga Berestneva. "Investigando o crowdsourcing como um método para coletar rótulos de emoção para imagens." Dentro *Resumos Estendidos da Conferência CHI de 2018 sobre Fatores Humanos em Sistemas de Computação*, pp. 1-6. 2018.
- [47] Kosmicki, JA, V. Sochat, M. Duda e DP Wall. "Procurando um conjunto mínimo de comportamentos para detecção de autismo por meio de aprendizado de máquina baseado em seleção de recursos." *psiquiatria translacional*5, não. 2 (2015): e514-e514.
- [48] Kratzwald, Bernhard, Suzana Illic, Mathias Kraus, Stefan Feuerriegel e Helmut Prendinger. "Aprendizado profundo para computação afetiva: reconhecimento de emoções baseado em texto no suporte à decisão." *Sistemas de Suporte à Decisão*115 (2018): 24-35.
- [49] Locação, Mateus. "Sobre controle de qualidade e aprendizado de máquina em crowdsourcing." *Computação Humana*11, não. 11 (2011).
- [50] Leblanc, Emilie, Peter Washington, Maya Varma, Kaitlyn Dunlap, Yordan Penev, Aaron Kline e Dennis P. Wall. "Os métodos de substituição de recursos permitem uma análise confiável de vídeo doméstico para detecção de autismo por aprendizado de máquina." *Relatórios científicos*10, não. 1 (2020): 1-11.
- [51] Liu, Runkeng, Joseph P. Salisbury, Arshya Vahabzadeh e Ned T. Sahin. "Viabilidade de um sistema de smartglasses de realidade aumentada focado no autismo para comunicação social e treinamento comportamental." *Fronteiras em pediatria*5 (2017): 145.
- [52] LoBue, Vanessa, Lewis Baker e Cat Thrasher. "Através dos olhos de uma criança: identificação de expressões emocionais de crianças em idade pré-escolar do conjunto de expressão facial afetiva infantil (CAFE)." *Cognição e Emoção*32, n°. 5 (2018): 1122-1130.

- [53] LoBue, Vanessa e Cat Thrasher. "The Child Affective Facial Expression (CAFE) set: Validade e confiabilidade de adultos não treinados." *Fronteiras da psicologia* 5 (2015): 1532.
- [54] Lotfian, Reza e Carlos Busso. "Amostragem excessiva de dados de fala emocional com base em avaliações subjetivas fornecidas por vários indivíduos." *Transações IEEE em Computação Afetiva* (2019).
- [55] Magdin, Martin e F. Prikler. "Reconhecimento de expressões faciais em tempo real usando webcam e SDK afetivo." *IJIMA* 5, não. 1 (2018): 7-15.
- [56] McDuff, Daniel, Rana Kaliouby, Thibaud Senechal, May Amr, Jeffrey Cohn e Rosalind Picard. "Conjunto de dados de expressão facial Affective-mit (am-fed): expressões faciais naturalísticas e espontâneas coletadas." *Dentro Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 881-888. 2013.
- [57] McFarland, Dennis J., Muhammad A. Parvaz, William A. Sarnacki, Rita Z. Goldstein e Jonathan R. Wolpaw. "Previsão de classificações subjetivas de imagens emocionais por recursos de EEG." *Jornal de engenharia neural* 4, não. 1 (2016): 016009.
- [58] Mower, Emily, Maja J. Matarić e Shrikanth Narayanan. "Uma estrutura para classificação automática de emoções humanas usando perfis emocionais." *Transações IEEE sobre processamento de áudio, fala e linguagem* 19, não. 5 (2010): 1057-1070.
- [59] Mower, Emily, Angeliki Metallinou, Chi-Chun Lee, Abe Kazemzadeh, Carlos Busso, Sungbok Lee e Shrikanth Narayanan. "Interpretar expressões emocionais ambíguas." *Dentro 2009 3ª Conferência Internacional sobre Computação Afetiva e Interação Inteligente e Workshops*, pp. 1-8. IEEE, 2009.
- [60] Nguyen, Binh T., Minh H. Trinh, Tan V. Phan e Hien D. Nguyen. "Uma detecção de emoção eficiente em tempo real usando câmeras e marcos faciais." *Dentro 2017 sétima conferência internacional sobre ciência e tecnologia da informação (ICTST)*, pp. 251-255. IEEE, 2017.
- [61] Ning, Michael, Jena Daniels, Jesse Schwartz, Kaitlyn Dunlap, Peter Washington, Haik Kalantarian, Michael Du e Dennis P. Wall. "Identificação e quantificação de lacunas no acesso a recursos de autismo nos Estados Unidos: um estudo infodemiológico." *Jornal de pesquisa médica na Internet* 21, não. 7 (2019): e13094.
- [62] Nicolaou, Mihalís A., Hatice Gunes e Maja Pantic. "Previsão contínua de afeto espontâneo a partir de múltiplas pistas e modalidades no espaço de excitação-valência." *Transações IEEE em Computação Afetiva* 2, não. 2 (2011): 92-105.
- [63] Paolacci, Gabriele, Jesse Chandler e Panagiotis G. Ipeirotis. "Executando experimentos no Amazon Mechanical Turk." *Julgamento e tomada de decisão* 5, não. 5 (2010): 411-419.
- [64] Parthasarathy, Srinivas e Carlos Busso. "Prevendo Conjuntamente Arousal, Valência e Dominância com Multi-Task Learning." *Dentro Interdiscurso*, vol. 2017, pp. 1103-1107. 2017.
- [65] Peterson, Joshua C., Ruairidh M. Battleday, Thomas L. Griffiths e Olga Russakovsky. "A incerteza humana torna a classificação mais robusta." *Dentro Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9617-9626. 2019.
- [66] Ploggia, Giovanni, Roberta Igliozzi, Marcello Ferro, Arti Ahluwalia, Filippo Muratori e Danilo De Rossi. "Um andróide para melhorar as habilidades sociais e o reconhecimento de emoções em pessoas com autismo." *Transações IEEE em Sistemas Neurais e Engenharia de Reabilitação* 13, não. 4 (2005): 507-515.
- [67] Poria, Soujanya, Erik Cambria e Alexander Gelbukh. "Recursos textuais de redes neurais convolucionais profundas e aprendizado de kernel múltiplo para análise de sentimento multimodal em nível de enunciado." *Dentro Anais da conferência de 2015 sobre métodos empíricos no processamento de linguagem natural*, pp. 2539-2544. 2015.
- [68] Poria, Soujanya, Erik Cambria, Newton Howard, Guang-Bin Huang e Amir Hussain. "Fusão de pistas de áudio, visuais e textuais para análise de sentimento de conteúdo multimodal." *neurocomputação* 174 (2016): 50-59.
- [69] Rizos, Georgios e Björn W. Schuller. "Average Jane, Where Art Thou?—Recentes avenidas em aprendizado de máquina eficiente sob a incerteza da subjetividade." *Dentro Conferência Internacional sobre Processamento de Informação e Gestão de Incerteza em Sistemas Baseados em Conhecimento*, pp. 42-55. Springer, Cham, 2020.
- [70] Rodrigues, Filipe e Francisco Pereira. "Aprendizagem profunda com multidões." *Dentro Anais da Conferência AAAI sobre Inteligência Artificial*, vol. 32, n°. 1. 2018.
- [71] Sharma, Mukta, Anand Singh Jalal e Aamir Khan. "Reconhecimento de emoção usando expressão facial, combinando descritor de pontos-chave e recursos de textura." *Ferramentas e aplicativos multimídia* 78, n°. 12 (2019): 16195-16219.
- [72] Smitha, Kavallur Gopi e A. Prasad Vinod. "Sistema de reconhecimento de emoções faciais para crianças autistas: um estudo viável baseado na implementação de FPGA." *Engenharia médica e biológica e computação* 53, n°. 11 (2015): 1221-1229.
- [73] Stappen, Lukas, Alice Baird, Erik Cambria e Björn W. Schuller. "Análise de sentimento e reconhecimento de tópico em transcrições de vídeo." *Sistemas Inteligentes IEEE* 36, n°. 2 (2021): 88-95.
- [74] Tahir, Madiha, Abdallah Tubaishat, Feras Al-Obeidat, Babar Shah, Zahid Halim e Muhammad Waqas. "Um novo algoritmo genético caótico binário para seleção de recursos e sua utilidade em computação afetiva e assistência médica." *Computação Neural e Aplicações* (2020): 1-22.
- [75] Tao, Jianhua e Tieniu Tan. "Computação afetiva: uma revisão." *Dentro Conferência Internacional sobre Computação Afetiva e Interação Inteligente*, pp. 981-995. Springer, Berlim, Heidelberg, 2005.
- [76] Tariq, Qandeel, Scott Lanyon Fleming, Jesse Nicole Schwartz, Kaitlyn Dunlap, Conor Corbin, Peter Washington, Haik Kalantarian, Naila Z. Khan, Gary L. Darmstadt e Dennis Paul Wall. "Detectando atraso no desenvolvimento e autismo por meio de modelos de aprendizado de máquina usando vídeos caseiros de crianças de Bangladesh: estudo de desenvolvimento e validação." *Jornal de pesquisa médica na Internet* 21, não. 4 (2019): e13822.
- [77] Tariq, Qandeel, Jena Daniels, Jesse Nicole Schwartz, Peter Washington, Haik Kalantarian e Dennis Paul Wall. "Detecção móvel de autismo por meio de aprendizado de máquina em vídeo doméstico: um estudo de validação e desenvolvimento prospectivo." *medicamento PLoS* 15, não. 11 (2018): e1002705.

- [78] Thiam, Patrick, Sascha Meudt, Markus Kächele, Günther Palm e Friedhelm Schwenker. "Detecção de eventos emocionais utilizando métodos de vetores de suporte em um cenário de HCI de aprendizado ativo." *DentroAnais do workshop de 2014 sobre representação e modelagem de emoções em sistemas de interação humano-computador*, pp. 31-36. 2014.
- [79] Thiel, cristão. "A classificação em rótulos macios é robusta contra o ruído do rótulo." *DentroConferência Internacional sobre Sistemas de Informação e Engenharia Baseados em Conhecimento e Inteligentes*, pp. 65-73. Springer, Berlin, Heidelberg, 2008.
- [80] Turing, Alan M. "Máquinas de computação e inteligência." *DentroAnalisando o teste de turing*, pp. 23-65. Springer, Dordrecht, 2009.
- [81] Uma, Alexandra, Tommaso Fornaciari, Dirk Hovy, Silviu Paun, Barbara Plank e Massimo Poesio. "Um caso para funções de perda suave." *DentroAnais da Conferência AAAI sobre Computação Humana e Crowdsourcing*, vol. 8, não. 1, pp. 173-177. 2020.
- [82] Villon, Olivier e Christine Lisetti. "No sentido de reconhecer a emoção subjetiva do indivíduo a partir de sinais fisiológicos na aplicação prática." *Dentro Vigésimo Simpósio Internacional IEEE sobre Sistemas Médicos Baseados em Computador (CBMS'07)*, pp. 357-362. IEEE, 2007.
- [83] Völkel, Sarah Theres, Julia Graefe, Ramona Schödel, Renate Häuslschmid, Clemens Stachl, Quay Au e Heinrich Hussmann. "Eu dirijo meu carro e meus estados me conduzem: visualizando os estados emocionais e físicos do motorista." *DentroAnais Adjuntos da 10ª Conferência Internacional sobre Interfaces de Usuário Automotivo e Aplicativos Veiculares Interativos*, pp. 198-203. 2018.
- [84] Voss, Catalin, Jessey Schwartz, Jena Daniels, Aaron Kline, Nick Haber, Peter Washington, Qandeel Tariq et al. "Efeito da intervenção digital vestível para melhorar a socialização em crianças com transtorno do espectro do autismo: um ensaio clínico randomizado." *JAMA pediatria* 173, n°. 5 (2019): 446-454.
- [85] Voss, Catalin, Peter Washington, Nick Haber, Aaron Kline, Jena Daniels, Azar Fazel, Titas De et al. "Vidro de superpotência: fornecendo dicas sociais discretas em tempo real em sistemas vestíveis." *DentroAnais da Conferência Conjunta Internacional ACM 2016 sobre Computação Pervasiva e Ubíqua: Adjunto*, pp. 1218-1226. 2016.
- [86] Wall, Dennis Paul, J. Kosmicki, TF Deluca, E. Harstad e Vincent Alfred Fusaro. "Uso de aprendizado de máquina para encurtar triagem e diagnóstico de autismo baseados em observação." *psiquiatria translacional* 2, não. 4 (2012): e100-e100.
- [87] Washington, Peter, Emilie Leblanc, Kaitlyn Dunlap, Yordan Penev, Aaron Kline, Kelley Paskov, Min Woo Sun et al. "Telemedicina de precisão por meio do aprendizado de máquina colaborativo: testando a variabilidade de trabalhadores da multidão para reconhecimento de recursos de autismo baseado em vídeo." *Revista de medicina personalizada* 10, não. 3 (2020): 86.
- [88] Washington, Peter, Natalie Park, Parishkrita Srivastava, Catalin Voss, Aaron Kline, Maya Varma, Qandeel Tariq et al. "Diagnóstico orientado por dados e o potencial da inteligência artificial móvel para fenotipagem terapêutica digital em psiquiatria computacional." *Psiquiatria Biológica: Neurociência Cognitiva e Neuroimagem* (2019).
- [89] Washington, Peter, Haik Kalantarian, Jack Kent, Arman Husic, Aaron Kline, Emilie Leblanc, Cathy Hou e outros. "Treinando um Classificador de Detecção de Emoção usando Quadros de um Jogo Terapêutico Móvel para Crianças com Distúrbios do Desenvolvimento." *arXiv pré-impressão arXiv:2012.08678* (2020).
- [90] Washington, Peter, Haik Kalantarian, Qandeel Tariq, Jessey Schwartz, Kaitlyn Dunlap, Brianna Chrisman, Maya Varma et al. "Validade da triagem online para autismo: estudo de crowdsourcing comparando tarefas de diagnóstico pagas e não pagas." *Jornal de pesquisa médica na Internet* 21, não. 5 (2019): e13668.
- [91] Washington, Peter, Aaron Kline, Onur Cezmi Mutlu, Emilie Leblanc, Cathy Hou, Nate Stockham, Kelley Paskov, Brianna Chrisman e Dennis P. Wall. "Reconhecimento de atividades com câmeras em movimento e alguns exemplos de treinamento: aplicações para detecção de headbanging relacionado ao autismo." *arXiv pré-impressão arXiv:2101.03478* (2021).
- [92] Washington, Peter, Emilie Leblanc, Kaitlyn Dunlap, Yordan Penev, Maya Varma, Jae-Yoon Jung, Brianna Chrisman et al. "Seleção de trabalhadores de multidão confiáveis para diagnóstico telemedico de transtorno do espectro do autismo pediátrico." *DentroBIOCOMPUTAÇÃO 2021: Anais do Simpósio do Pacífico*, pp. 14-25. 2020.
- [93] Washington, Peter, Kelley Marie Paskov, Haik Kalantarian, Nathaniel Stockham, Catalin Voss, Aaron Kline, Ritik Patnaik et al. "Seleção de recursos e redução de dimensão de dados de autismo social." *DentroPac Symp Biocomputador*, vol. 25, pp. 707-718. 2020.
- [94] Washington, Peter, Qandeel Tariq, Emilie Leblanc, Brianna Chrisman, Kaitlyn Dunlap, Aaron Kline, Haik Kalantarian e outros. "Tagging de recursos de privacidade preservada por crowdsourcing de vídeos caseiros curtos para detecção de ASD de aprendizado de máquina." *Relatórios científicos* 11, não. 1 (2021): 1-11.
- [95] Washington, Peter, Catalin Voss, Nick Haber, Serena Tanaka, Jena Daniels, Carl Feinstein, Terry Winograd e Dennis Wall. "Um dispositivo de interação social vestível para crianças com autismo." *DentroAnais da 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2348-2354. 2016.
- [96] Washington, Peter, Catalin Voss, Aaron Kline, Nick Haber, Jena Daniels, Azar Fazel, Titas De, Carl Feinstein, Terry Winograd e Dennis Wall. "SuperpowerGlass: uma ajuda vestível para a terapia em casa de crianças com autismo." *Anais do ACM sobre tecnologias interativas, móveis, vestíveis e ubíquas* 1, não. 3 (2017): 1-22.
- [97] Washington, Peter, Serena Yeung, Bethany Percha, Nicholas Tatonetti, Jan Liphardt e Dennis P. Wall. "Obtendo soluções de dados biomédicos confiáveis." *DentroBIOCOMPUTAÇÃO 2021: Anais do Simpósio do Pacífico*, pp. 1-13. 2020.
- [98] White, Susan W., Lynn Abbott, Andrea Trubanova Wieckowski, Nicole N. Capriola-Hall, Sherin Aly e Amira Youssef. "Viabilidade de treinamento automatizado para expressão e reconhecimento de emoções faciais no autismo." *terapia comportamental* 49, n°. 6 (2018): 881-888.
- [99] Yang, Zhao, Tie Liu, Jiehao Liu, Li Wang e Sai Zhao. "Uma nova função de perda de margem suave para aprendizagem de incorporação discriminativa profunda." *Acesso IEEE* (2020): 202785-202794.
- [100] Yannakakis, Georgios N. "Aprimorando os cuidados de saúde por meio da computação afetiva." (2018).
- [101] Yin, Da, Xiao Liu, Xiuyu Wu e Baobao Chang. "Uma estratégia de soft label para classificação de sentimento no nível de destino." *DentroAnais do Décimo Workshop sobre Abordagens Computacionais para Subjetividade, Sentimento e Análise de Mídia Social*, pp. 6-15. 2019.

- [102] Yu, Liang-Chih, Jin Wang, K. Robert Lai e Xue-jie Zhang. "Prevendo classificações de ativação de valência de palavras usando um método gráfico ponderado." *Dentro Anais da 53ª Reunião Anual da Associação de Linguística Computacional e da 7ª Conferência Conjunta Internacional sobre Processamento de Linguagem Natural (Volume 2: Short Papers)*, pp. 788-793. 2015.
- [103] Zeng, Zhihong, Jilin Tu, Ming Liu, Thomas S. Huang, Brian Pianfetti, Dan Roth e Stephen Levinson. "Reconhecimento de afeto audiovisual." *Transações IEEE em multimídia* 9, não. 2 (2007): 424-428.
- [104] Zhang, Lin, Steffen Walter, Xueyao Ma, Philipp Werner, Ayoub Al-Hamadi, Harald C. Traue e Sascha Gruss. "'BioVid Emo DB': Um banco de dados multimodal para análises de emoções validado por avaliações subjetivas." *Dentro Série de Simpósios IEEE 2016 sobre Inteligência Computacional (SSCI)*, pp. 1-6. IEEE, 2016.
- [105] Zhang, Xiao, Wenzhong Li, Xu Chen e Sanglu Lu. "Moodexplorer: Rumo à detecção de emoções compostas por meio da detecção de smartphones." *Anais do ACM sobre tecnologias interativas, móveis, vestíveis e ubíquas* 1, não. 4 (2018): 1-30.
- [106] Zhao, Sicheng, Hongxun Yao e Xiaolei Jiang. "Prevendo distribuição de probabilidade contínua de emoções de imagem no espaço de ativação de valência." *Dentro Actas da 23ª Conferência Internacional da ACM sobre Multimídia*, pp. 879-882. 2015.
- [107] Zhou, Ying, Xuefeng Liang, Yu Gu, Yifei Yin e Longshan Yao. "Aprendizagem Interativa Multiclassificadora para Reconhecimento Ambíguo de Emoções de Fala." *arXiv pré-impressão arXiv:2012.05429*(2020).