IJARCCE

ISSN (On-line) 2278-1021
ISSN (impressão) 2319-5940

Revista Internacional de Pesquisa Avançada em Engenharia de Computação e Comunicação

Vol. 10, Edição 6, junho de 2021

DOI 10.17148/IJARCCE.2021.106107

Detecção dos estados afetivos dos alunos em sala de aula usando CNN

Neha Pawar1, Shubhangi Funde2, Revati Kshirsagar3, Vaishnavi Kaulagi4

Faculdade Moderna de Engenharia da Sociedade de Educação Progressiva, Pune1-4

Resumo: Prever o envolvimento emocional do aluno utilizando técnicas de visão computacional é uma tarefa desafiadora. Existem vários trabalhos sobre o reconhecimento do estado afetivo dos alunos com base na visão computacional no ambiente de e-learning, existem trabalhos limitados sobre o reconhecimento do estado afetivo dos alunos no ambiente de sala de aula onde mais de um aluno está presente em um único quadro de imagem. O reconhecimento facial tornou-se um campo atraente na tecnologia baseada em computador. desenvolvimento de aplicações nas últimas décadas. O processo de aprendizagem também evoluiu muito. Porém, a emoção dos alunos costuma ser negligenciada no processo de aprendizagem. Este projeto preocupa-se principalmente em usar a expressão facial para detectar emoções no ambiente de aprendizagem. Existem muitos algoritmos para reconhecimento facial e captura de emoções, dos quais usamos a rede neural convolucional (CNN). A expressão facial capturada será usada no Ambiente de Aprendizagem para analisar o humor do aluno. A arquitetura proposta utiliza as expressões faciais dos alunos para analisar seus estados afetivos. Os resultados experimentais irão prever a probabilidade de estados afetivos dos rostos detectados no ambiente de aprendizagem para a compreensão das emoções durante o processo de aprendizagem, a fim de melhorar o processo de aprendizagem e obtenção de feedback.

Palavras-chave: Detecção facial, Reconhecimento de emoções faciais, Rede neural de convolução, OpenCV, Aprendizado de máquina.

EU. INTRODUÇÃO

Os humanos interagem uns com os outros principalmente através da fala, mas também através de gestos corporais, para enfatizar certas partes do seu discurso e para exibir emoções. Uma das maneiras importantes pelas quais os humanos demonstram emoções é por meio de expressões faciais, que são uma parte muito importante da comunicação. Expressões faciais são as mudanças faciais em resposta aos estados emocionais internos, intenções ou comunicações sociais de uma pessoa. O reconhecimento de emoções faciais é o processo de detecção de emoções humanas a partir de expressões faciais. O cérebro humano reconhece emoções automaticamente e agora foi desenvolvido um software que também pode reconhecer emoções. Essa tecnologia está se tornando cada vez mais precisa e, eventualmente, será capaz de ler emoções tão bem quanto nosso cérebro. A predição automática dos estados afetivos do aluno foram menos exploradas no ambiente de sala de aula. Os estudos existentes prevêem o envolvimento emocional e comportamental dos alunos separadamente em ambientes de e-learning e de sala de aula. O uso da multimodalidade para reconhecimento dos estados afetivos dos alunos foi muito menos explorado. Não foi explorada uma previsão automática do envolvimento dos alunos a nível de grupo. Finalmente, não existe um conjunto de dados padrão para treinar, testar e validar os modelos de aprendizado de máquina/aprendizado profundo em ambientes de sala de aula. Isso nos motivou a propor uma arquitetura de rede neural convolucional para prever automaticamente os estados afetivos dos alunos em sala de aula. Neste artigo, apresentamos uma abordagem baseada em Redes Neurais Convolucionais (CNN) para reconhecimento de expressões faciais. A entrada em nosso sistema é uma imagem; então, usamos a CNN para prever o rótulo da expressão facial, que deve ser um destes rótulos: raiva, felicidade, medo, tristeza, nojo e neutro.

II. TRABALHO RELATADO

Abhilash Dubbaka e Anandha Gopalan [1] propõem um sistema que usará webcam para monitorar rostos de alunos assistindo MOOC (Massive Open Online Course). Este artigo categoriza as expressões faciais e traduz suas expressões no nível de envolvimento dos alunos no ambiente de aprendizagem online. Este artigo explora o uso de webcams para registrar as expressões faciais dos alunos enquanto eles assistem a material de vídeo educacional para analisar seus níveis de envolvimento do aluno.
Redes neurais convolucionais (CNNs) foram treinadas para detectar unidades de ação facial, que foram mapeadas em duas medidas psicológicas, valência (estado emocional).
NIGEL BOSCH e SIDNEY K. D'MELLO, Universidade de Notre Dame JACLYN OCUMPAUGH e RYAN S.
BAKER, Teachers College, Columbia University VALERIE SHUTE, Florida State University [2] usa visão computacional e técnicas de aprendizado de máquina para detectar o estado afetivo do aluno a partir da expressão facial e do movimento corporal bruto durante a interação com um jogo educacional de física.
“Detecção automática dos estados afetivos dos alunos em ambiente de sala de aula usando CNN híbrida” por Ashwin TS , Ram Mohana Reddy Guddeti. [3] Este artigo descreve como a Rede Neural de Convolução pode ser usada para prever o estado afetivo geral de toda a turma. Este artigo usa expressões faciais dos alunos com gestos manuais e postura corporal para

IJARCCE

ISSN (On-line) 2278-1021
ISSN (impressão) 2319-5940

Revista Internacional de Pesquisa Avançada em Engenharia de Computação e Comunicação

Vol. 10, Edição 6, junho de 2021

DOI 10.17148/IJARCCE.2021.106107

analisar seus estados afetivos.

Em 2016, Pramerdorfer e Kampel obtiveram o estado da arte, que é de 75,2% de precisão no FER2013, usando Redes Neurais Convolucionais (CNNs) [4]. Os autores utilizaram um conjunto de CNNs utilizando VGG, Inception e ResNet com profundidades de 10, 16 e 33, com parâmetros de 1,8m, 1,6m e 5,3m, respectivamente. Os autores usaram as imagens faciais fornecidas no conjunto de dados e, para correção de iluminação, usaram a equalização do histograma. Eles realizaram espelhamento horizontal para aumento de dados de treinamento e cortaram imagens aleatoriamente no tamanho de 48 x 48 pixels. Eles também treinaram a arquitetura por até 300 épocas e usaram gradiente descendente estocástico para otimizar a perda de entropia cruzada, com um valor de momento de 0,9. Os demais parâmetros foram fixos, como taxa de aprendizagem com 0,1, tamanho do lote com 128 e redução de peso com 0,0001.

Zhang et al. [5] usaram uma rede siamesa para introduzir um método para compreender comportamentos de relações sociais a partir de imagens e alcançaram uma precisão de teste de 75,1% no desafiador conjunto de dados de expressão facial Kaggle. Os autores usaram vários conjuntos de dados, com vários rótulos, para aumentar os dados de treinamento; eles também introduziram um método de extração de recursos e registro baseado em patch, além de trabalhar na integração de recursos por meio de fusão inicial.

Kim et al. [6] propuseram um conjunto de CNNs e demonstraram que durante o treinamento e teste é vantajoso usar formas registradas e não registradas de determinadas imagens faciais. Os autores alcançaram uma precisão de teste de 73,73% no conjunto de dados FER2013. Eles também conduziram o Intraface para um alinhamento 2-D convencional, que está disponível publicamente para detector de pontos de referência, e realizaram a normalização da iluminação. Para evitar o erro de registro, eles realizaram o registro seletivamente, com base nos resultados da detecção de pontos de referência faciais.

III. CONJUNTO DE DADOS USADO

O conjunto de dados é baixado do site: https://www.kaggle.com/msambare/fer2013. Os dados consistem em imagens de rostos em escala de cinza de 48x48 pixels. O conjunto de dados oferece sete categorias: [Irritado, Nojento, Medo, Feliz, Triste, Surpresa, Neutro].

O conjunto de treinamento consiste em 28.709 exemplos. O conjunto de testes público consiste em 3.589 exemplos.

4. ARQUITETURA DE SISTEMA

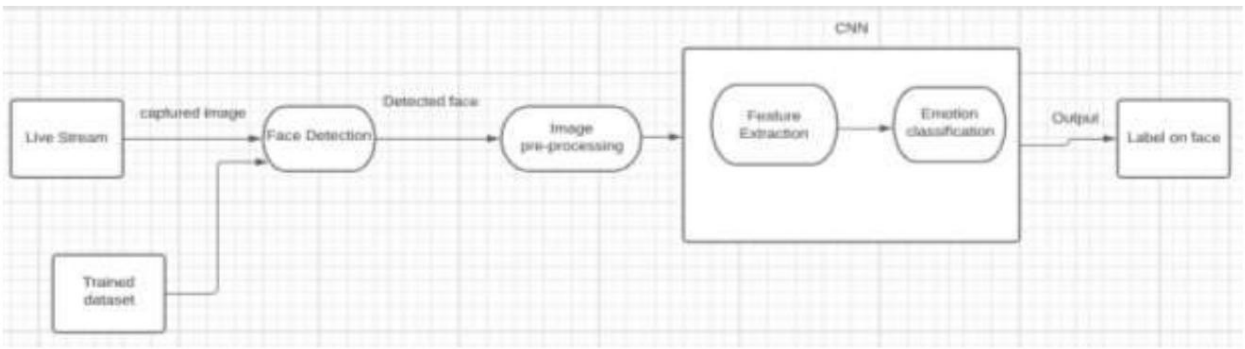

Diagrama de arquitetura

O diagrama de arquitetura acima mostra o modelo proposto para os estados afetivos do aluno no ambiente de sala de aula. Na transmissão ao vivo, os dados serão considerados como entrada. A imagem será fornecida como entrada para o próximo processo, ou seja, detecção de rosto. De toda a imagem capturada, apenas o rosto será detectado após a digitalização. O rosto detectado será inserido no próximo processo onde a imagem será redimensionada e os valores RGB serão convertidos em valores de cinza. Esses valores serão usados pela CNN para extrair características e classificar emoções. Para o modelo pré-treinado os valores serão correspondidos e a categoria para a qual os valores têm maior inclinação será considerada o estado emocional.

A. Pré-processando

O pré-processamento pode ser usado para melhorar o desempenho do sistema FER e pode ser feito antes do processo de extração de recursos. O pré-processamento de imagens possui diversos processos, como detecção e alinhamento de faces, correção de iluminação, pose, oclusão e aumento de dados.

No conjunto de dados FER2013, os rostos são registrados automaticamente, de forma que tenham requisitos de espaço semelhantes e fiquem mais ou menos centralizados nas imagens. a detecção de rosto é feita usando o classificador Haar Cascade [11]. Quando as imagens são capturadas em vários tipos de luz, as características de expressão às vezes são detectadas de forma imprecisa e, portanto, a taxa de reconhecimento de expressão pode ser baixa e dificultar a extração de características.

B. Extração de recursos

A extração de características faciais requer a tradução dos dados de entrada em um conjunto de características. Usando a extração de recursos,

IJARCCE

ISSN (On-line) 2278-1021
ISSN (impressão) 2319-5940

Revista Internacional de Pesquisa Avançada em Engenharia de Computação e Comunicação

Vol. 10, Edição 6, junho de 2021

DOI 10.17148/IJARCCE.2021.106107

os pesquisadores podem reduzir uma imensa quantidade de dados a um conjunto relativamente pequeno, o que permite uma computação mais rápida. Aplicamos o detector de pontos de referência facial dlib pré-treinado no conjunto de dados iBUG 300-W [12], [13], [14] para extração de recursos e extraímos as oito partes mais proeminentes de um rosto, incluindo ambas as sobrancelhas, ambos os olhos, o nariz , os contornos internos e externos da boca e da mandíbula. No qual extraímos os olhos direito e esquerdo, nariz e contorno interno e externo da boca, que estão marcados com a cor amarela.

C. Arquitetura CNN

As CNNs têm sido amplamente utilizadas em uma variedade de aplicações de visão computacional, incluindo FER. No início do século 21, vários estudos da literatura FER [15], [16] determinaram que as CNNs funcionam bem em mudanças de localização de face, bem como em variações de escala. Descobriu-se também que eles funcionam melhor do que o perceptron multicamadas (MLP) ao observar variações de pose de rosto não vistas anteriormente. Os pesquisadores usaram a CNN para ajudar a resolver vários problemas de reconhecimento de expressões faciais, como tradução, rotação, independência de assunto e invariância de escala [17]. Nosso modelo foi treinado usando as seguintes características:

- seis camadas convolucionais usando “RELU” como função de ativação;
- três max-pooling: dos quais os dois primeiros usam o tamanho da piscina (3,3) e a passada (2,2), e o terceiro usa o tamanho da piscina (2,2) e a passada (2,2); cada pooling máximo é seguido por cada duas camadas convolucionais; dois desistem com valor 0,2;
- uma camada achatada e duas camadas densas: uma camada densa com “RELU” e outra com “Softmax” como função de ativação;
- os parâmetros totais e os parâmetros treináveis são 1,2 milhão, respectivamente.

V. CONCLUSÃO

O projeto explorará os estados afetivos do aluno no ambiente de sala de aula, uma vez que tanto o envolvimento emocional quanto o comportamental são considerados para prever os estados afetivos do aluno com base em emoções como engajado, entediado, feliz, neutro, etc.

REFERÊNCIAS

[1] Conferência IEEE por Abhilash Dubbaka Anandha Gopalan, Departamento de Computação Imperial College London Londres, Reino Unido abhilash., “Detectando o envolvimento do aluno em MOOCs usando reconhecimento automático de expressão facial”.

[2] Transações ACM por Nigel bosch Sidney K. D mello Ryan Baker. “Usando vídeo para detectar automaticamente o efeito do aluno em computadores habilitados salas de aula.”

[3] ScienceDirect por Ashwin TS, Ram Mohana Reddy Guddeti. “Detecção automática dos estados afetivos dos alunos no ambiente de sala de aula usando CNN híbrida.”

[4] Outro site usado para coleta de conjuntos de dados: https://www.kaggle.com/c/challenges-in representação-learning-facial-expression reconhecimento-desafio/dados

