

Felipe Scherer Vicentin
Gustavo Miller Santos
Pedro Henrique Marques Renó
Me Sérgio Luiz Moral Marques

Uso do aprendizado de máquina na identificação de padrões em ondas cerebrais

Trabalho de Conclusão de Curso
Unicamp - Colégio Técnico de Campinas
DD/MM/2020

Felipe Scherer Vicentin, estudante do 3º ano do curso de Informática no Colégio Técnico de Campinas - Unicamp
Rua Lindolfo Waldemar Masotti, 158 (Jardim dos Manacás), Betel, Paulínia - SP
(19) 98181-2990

<http://lattes.cnpq.br/0601816741122641>

Gustavo Miller Santos, estudante do 3º ano do curso de informática no Colégio Técnico de Campinas - Unicamp

Rua Antônio Faria da Silva, 230, Jardim Maracanã, Sumaré - SP

(19) 99981-2805

<http://lattes.cnpq.br/8032654193175219>

Pedro Henrique Marques Renó, estudante do 3º ano do curso de Informática no Colégio Técnico de Campinas - Unicamp

Rua José Alexandrino Palazzo, 381 (Residencial Barão do café), Barão Geraldo, Campinas - SP

(19) 98814-3785

<http://lattes.cnpq.br/7709289944677407>

Me Sérgio Luiz Moral Marques - Professor do Cotuca

Rua Ludovico Bonato, 121, Jardim Sta. Marcelina, Campinas - SP

(19) 99169-4230 <http://lattes.cnpq.br/6811598469344679>

Resumo

Almeja-se, na presente pesquisa, desenvolver algoritmos de aprendizagem de máquina capazes de identificar estados cerebrais, utilizando dados de eletroencefalografia (EEG). Sendo assim, o objetivo é exploratório, visto que busca compreender a categorização de ondas cerebrais por programas de inteligência artificial (IA). Será feito um experimento audiovisual envolvendo 6 voluntários, no qual, utilizando trechos de filmes e problemas matemáticos, a atividade mental dos participantes será medida com o EEG, e as leituras recebidas então serão usadas para treinar modelos de IA. Cada um deles será gerado a partir de métodos diferentes, como floresta aleatória, SVM e KNN. Em seguida, a precisão deles será medida com um conjunto de dados selecionados aleatoriamente dentre os coletados, e o mais acurado será selecionado e disponibilizado em um *web service*, caracterizando, portanto, o método de engenharia – resolver o problema da categorização de estados mentais a partir da plataforma virtual contendo os algoritmos desenvolvidos. Como resultado, espera-se obter uma alta precisão na identificação de estados mentais. Isso permitiria auxiliar a comunidade científica em futuras pesquisas relacionadas, tendo em mente que há poucos dados e ferramentas de análise de EEG disponíveis publicamente, e o aparelho em si não é muito acessível.

Relevância Social

O projeto visa disponibilizar dados e ferramentas que possam auxiliar pesquisadores na área de neurociências. Por ser uma plataforma genérica, pode ser utilizada para diversas aplicações e estudos com EEG, que podem ter desdobramentos futuros para a sociedade, especialmente àqueles com distúrbios neurológicos ou psicológicos. Como se trata de um projeto que tem como objetivo uma consequência a longo prazo, os resultados específicos são bem imprevisíveis, já que qualquer pesquisador poderá fazer uso dos dados que obtivermos.

Sumário

Introdução	6
Objetivos	7
Objetivos Gerais	7
Objetivos Específicos	7
Hipóteses	8
Local de realização da pesquisa	8
População a ser estudada	8
Garantias éticas	8
Método a ser utilizado	9
Cronograma	10
Orçamento	10
Critérios de inclusão e exclusão dos participantes da pesquisa	11
Riscos e benefícios	11
Critérios de encerramento ou suspensão de pesquisa	11
Resultados do estudo	11
Divulgação dos resultados	12
Referências bibliográficas	12

Introdução

Os eletroencefalógrafos são aparelhos que, por meio de eletrodos e amplificadores, conseguem registrar a atividade cerebral que surge nos neurônios na forma de impulsos elétricos de um modo não invasivo [9]. Uma vez com os dados, a identificação de padrões nos impulsos, comumente chamados de ondas, torna-se possível. Foi com esse mecanismo que cientistas conseguiram, por exemplo, estudar como a mente se comporta durante o sono e dividi-lo em fases [7].

Atualmente, o uso de EEGs é bem restrito, sendo usados majoritariamente em hospitais para diagnosticar doenças sérias relacionadas ao cérebro, como tumores, Acidentes Vasculares Cerebrais (AVCs) e distúrbios mentais [5]. Contudo, esse equipamento poderia ser usado de maneiras mais simples e muito além do que já é, por exemplo, para reconhecimento de emoções de pacientes [4] ou o quão concentrados esses se encontram [6]. Além disso, as bases de dados disponíveis para pesquisas na área não são numerosas, o que obriga os pesquisadores a fazerem novos testes em cada artigo.

Ao analisar as ondas cerebrais de um EEG em busca de padrões, muita observação é demandada, uma vez que não é trivial achar semelhanças em ondas distintas. É por isso que pesquisadores ao redor do mundo vêm, cada vez mais, usando estratégias de aprendizado de máquina para interpretar os dados obtidos [2, 3, 4, 11]. Essa será a estratégia utilizada na presente pesquisa.

Este trabalho focará no uso de um segmento desses algoritmos chamado “Aprendizado Supervisionado”. Esse ramo consiste em fornecer os dados ao programa com as respostas esperadas e utilizar o resultado do treinamento para classificar dados futuros. O projeto irá comparar as precisões de três tipos de métodos diferentes: K Means Clustering (KNN), Support Vector Machine (SVM) e Florestas Aleatórias.

Tornando cada coluna da tabela de dados em uma dimensão num gráfico e representando cada linha da tabela como um ponto no gráfico, o método de KNN comparará os K pontos mais próximos de um novo ponto X para descobrir sua classificação. A rotulação dominante desses K pontos determinará o resultado atribuído ao novo ponto X.

Utilizando o mesmo tipo de gráfico descrito anteriormente, o SVM criará diversos “hiperplanos” que servirão para indicar quais regiões do gráfico pertencem a cada tipo de classificação. As margens dessas áreas são colocadas na máxima distância entre essas e os pontos de rótulos diferentes que separa. A classificação de um novo ponto qualquer será estabelecida pela região do gráfico que se encontra.

Diferente dos casos anteriores, o método de Florestas Aleatórias utiliza diversas Árvores de Decisão para seus cálculos. Cada uma delas receberá um segmento aleatório do banco de dados, podendo estar faltando tanto linhas quanto colunas desse, e criará diversas verificações que se ramificam e tentam gerar uma rotulação com as informações que possuem. Para estipular a classificação de um novo dado apresentado, o método consultará todas as árvores que o consiste e dará sua resposta com base na decisão majoritária dessas.

Assim que cada um dos melhores modelos de aprendizado de máquina para cada tipo de classificação que desejamos analisar, resultantes do treinamento dos algoritmos com os dados iniciais, estiverem prontos, novos dados poderão ser classificados via um *Web Service*. As leituras dos eletrodos serão mandadas por uma requisição que dará como retorno uma resposta codificada em *JavaScript Object Notation* (JSON), constituída por um relatório gerado a partir das rotulações de cada modelo previamente descrito.

Objetivos

Objetivos Gerais

- Desenvolver aplicação que, a partir de ondas cerebrais, pode classificar os estados mentais do usuário.
- Criar base de dados com as leituras dos eletrodos dos participantes para compartilhamento anônimo com a comunidade científica.

Objetivos Específicos

- Transformar as ondas obtidas em valores numéricos que podem ser interpretados pelo computador;
- Treinar o modelo de aprendizado de máquina para a identificação de estados mentais;
- Criar um *Web Service* que, utilizando o modelo treinado, produzirá os resultados.

Hipóteses

Pretende-se, com esta pesquisa, mostrar que algoritmos de aprendizado de máquina podem identificar estados mentais a partir de ondas cerebrais. Para isso, é necessário que os dados elétricos obtidos com o EEG sejam transformados em informações numéricas passíveis de interpretação computacional.

Uma vez que os dados estiverem refinados e prontos para análise, eles serão apresentados a diversos algoritmos de aprendizagem de máquina, os quais foram descritos na introdução. Dessa maneira, espera-se que os modelos computacionais gerados possam identificar padrões nas informações obtidas, permitindo que estados mentais sejam inferidos.

Com tal modelo estruturado e pronto para uso, ele deverá ser colocado em um *Web Service* disponível publicamente. Desse modo, quaisquer leituras de EEG poderão ser enviadas e, posteriormente, um resultado será retornado contendo a classificação mais provável de estado mental.

Local de realização da pesquisa

O local onde os testes serão realizados é uma sala comum, nas dependências da Universidade Estadual de Campinas (Unicamp) o menos distrativa possível que tenha um bom isolamento acústico e visual. Nessa sala, o participante se encontrará sozinho, para evitar interrupções e influências externas.

População a ser estudada

Seis participantes comporão o experimento, esses deverão ter de 15 a 18 anos de idade, de ambos os sexos, sem histórico pessoal de transtornos neurológicos confirmados, com capacidade visual e auditiva normal e sem quadros de seborreia aguda ou infecção no couro cabeludo. Eles serão alunos voluntários da instituição de ensino “Colégio Técnico de Campinas”, sem quaisquer outros critérios de seleção, visto que não influenciam na aplicação a ser desenvolvida. Os próprios participantes informarão se cumprem com os requisitos aqui mencionados.

Garantias éticas

Todos os dados coletados ficarão armazenados com o pesquisador responsável, e somente a equipe de pesquisa terá acesso às informações coletadas, as quais serão identificadas por um número para preservar a anonimidade dos participantes.

O participante poderá sair do experimento a qualquer momento, para isso basta comunicar o pesquisador e não haverá nenhum tipo de prejuízo. Os equipamentos utilizados serão testados previamente, e portanto serão seguros para uso.

Uma semana antes do experimento, será cedida a cada voluntário uma cópia física do Termo de Consentimento Livre e Esclarecido (TCLE) para maiores ou menores, dependendo da idade. Na segunda situação, os participantes serão instruídos a preencher o documento com auxílio de um responsável legal.

Método a ser utilizado

A pesquisa utilizará o método de engenharia, no qual um problema é analisado e uma solução, neste caso na forma de software, é desenvolvida. Ela tem objetivo exploratório, isto é, contribui para o esclarecimento de uma questão pouco investigada.

O projeto será realizado em duas partes, que têm o objetivo de registrar as leituras de EEG do voluntário em variados estados mentais. Os dois testes juntos, que serão realizados no mesmo dia, durarão cerca de 75 minutos.

No primeiro teste, serão feitas algumas perguntas sobre se o voluntário é destro ou canhoto, sua idade e sexo. Depois disso, os pesquisadores auxiliarão o participante a posicionar e, posteriormente, retirar o eletroencefalógrafo corretamente de sua cabeça, similarmente ao uso de um capacete comum. Além disso, eles solicitarão para assistir um vídeo com o conteúdo descrito no próximo parágrafo.

Serão usados os seguintes espaços de tempo para mostrar trechos de filmes: cinco segundos para indicar que o trecho começará, quatro minutos para o próprio conteúdo e os últimos quinze segundos para um momento de descanso. Esse processo será repetido para cada um dos trechos que serão mostrados, num total de dez vezes.

No segundo teste, em intervalos constantes de quatro minutos, o participante será submetido a duas situações distintas para coletar dados sobre seus níveis de concentração. A primeira situação consiste em ouvir uma música calma e a segunda em resolver alguns problemas de matemática. A cada troca de situação, ele terá quinze segundos de descanso e, antes da situação posterior começar, cinco segundos de preparo. Esses intervalos de tempo são exatamente os mesmos da etapa anterior do experimento.

Os materiais que serão usados são: um eletroencefalógrafo que fará as leituras das ondas cerebrais do voluntário, um computador que mostrará as mídias necessárias para induzir os estados mentais que interessam e um computador que receberá os dados lidos para análise posterior.

A ordem desses dois testes pode variar para demonstrar que esse fator não influencia nas análises finais. Além disso, entre os dois testes, o participante terá cinco minutos de descanso.

Assim que as informações forem obtidas, elas serão analisadas automaticamente por algoritmos de aprendizagem de máquina, explicados previamente na introdução. Uma vez que os modelos de inteligência artificial forem gerados a partir da análise, suas precisões serão testadas utilizando um conjunto de dados selecionados aleatoriamente das amostras. Posteriormente, o melhor modelo será colocado em um *Web Service* público, permitindo assim que análises com novas leituras sejam feitas por outros pesquisadores.

Cronograma

Mês	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov
Referencial Bibliográfico	X	X			X	X				
Plano de Pesquisa		X	X							
Avaliação do Comitê de Ética em Pesquisa		X	X							
Pitch			X							
Protótipo do web service			X	X						
Protótipo do website				X	X					
Obtenção de dados do EEG					X	X	X	X		
Análise dos resultados							X	X		
Conexão dos dados com web service e web site							X	X		
Testes finais									X	X
Elaboração dos documentos finais									X	X

Orçamento

Não haverá custos do projeto, visto que o equipamento utilizado será cedido por empréstimo pelo instituto BRAINN, da Universidade Estadual de Campinas. A impressão dos documentos físicos, TCLEs, será bancados pelos pesquisadores.

Critérios de inclusão e exclusão dos participantes da pesquisa

A pesquisa será feita com pessoas voluntárias e não haverá especificidades, como cor, sexo, gênero, entre outros. Eles serão do Colégio Técnico de Campinas, com idades de 15 a 18 anos.

É exigido que os participantes tenham capacidade de audição e visão, pois serão expostos a trechos de filmes no experimento. Também espera-se que os voluntários não tenham histórico pessoal de transtornos neurológicos que possam afetar os resultados, como epilepsia. Não serão selecionados voluntários com seborreia grave ou infecção do couro cabeludo, de forma a evitar que o uso do eletroencefalógrafo cause danos à saúde.

Riscos e benefícios

Não há nenhum risco previsível nos procedimentos que serão utilizados. Todos os vídeos escolhidos e que serão usados foram retirados de filmes exibidos publicamente e conforme a classificação indicativa recomendável para a faixa etária dos participantes do estudo. Portanto, não apresentam conteúdo psicologicamente danoso. No entanto, o participante poderá ter o desconforto de ter que despender seu tempo para participar da pesquisa.

Os benefícios para a sociedade deste estudo é que num futuro onde EEGs são mais acessíveis à população, os resultados obtidos, que, embora simples, podem ser tão úteis quanto aplicativos usados para monitoramento corporal durante atividades físicas ou ferramentas que auxiliam o aprendizado e ensino. Entretanto, os participantes não terão nenhum benefício direto com a participação no estudo.

Critérios de encerramento ou suspensão de pesquisa

A pesquisa será encerrada assim que for criada uma aplicação teste para que seja possível testar os resultados obtidos, previsto até o fim de novembro de 2020. A suspensão da pesquisa ocorrerá se não for conseguido um aparelho de EEG até setembro de 2020, ou se não for possível realizar o experimento por ocasião da pandemia de Covid-19.

Resultados do estudo

Será gerado um banco de dados público a partir das leituras feitas pelo equipamento, em conjunto com um *Web Service* que permita o acesso aos algoritmos treinados de identificação. Para testá-los, será desenvolvido um protótipo

de aplicação, o qual poderá prever o estado mental do usuário, caso a hipótese esteja correta.

Com isso, espera-se que a plataforma com os dados e algoritmos possa contribuir para a pesquisa científica no uso de EEG.

Divulgação dos resultados

Como a finalidade da pesquisa é disponibilizar os dados obtidos e os algoritmos desenvolvidos, ela é feita com o objetivo de ser divulgada diretamente. Nessa divulgação, as leituras do EEG serão anônimas, sem qualquer maneira de identificar os participantes.

Referências bibliográficas

- [1] AMIN, Hafeez U. et al. Classification of EEG Signals Based on Pattern Recognition Approach, Front. Comput. Neurosci. Nov. 2017.
- [2] BIRD, Jordan J. et al. A Deep Evolutionary Approach to Bioinspired Classifier Optimisation for Brain-Machine Interaction. Complexity, Mar. 2019.
- [3] BIRD, Jordan J. et al. A study on mental state classification using eeg-based brain-machine interface. 9th International Conference on Intelligent Systems, Set. 2018.
- [4] BIRD, Jordan J. et al. Mental emotional sentiment classification with an EEG-based brain-machine Interface. The International Conference on Digital Image & Signal Processing (DISP'19), Abr. 2019.
- [5] CAMPISI, Patrizio. Brain Waves for Automatic Biometric-Based User Recognition. IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, v. 9, n. 5, p. 782-800, Mai. 2014.
- [6] HUANG, Ruey S. et al. EEG Pattern Recognition - Arousal States Detection and Classification. Proceedings of International Conference on Neural Networks (ICNN'96), Washington, DC, vol.2, p. 641-646, Jun. 1996.
- [7] MARTIN, W.B. et al. Pattern recognition of EEG-EOG as a technique for all-night sleep stage scoring. Science Direct, v. 32, n. 4, p. 417-427, Abr. 1972.
- [8] PENNY, William D. et al. EEG-Based Communication: A Pattern Recognition Approach. IEEE TRANSACTIONS ON REHABILITATION ENGINEERING, v. 8, n. 2, p. 214-215, Jun. 2000.
- [9] TEPLAN, Michal. "Fundamentals of EEG measurement.". Measurement science review. v. 2, n. 2, p. 1-11, Jul. 2002.

- [10] WANG, Hee Lin; CHEONG, Loong-Fah. "Affective understanding in film". IEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, v.16, n.6, p. 689-704, Jun. 2006.
- [11] WANG, Xiao-Wei. et al. Emotional state classification from EEG data using machine learning approach. Science Direct, v. 129, p. 94-106.