DETECÇÃO DA DOENÇA DE PARKINSON: UMA PROPOSTA BASEADA NO MODELO DE REGRESSÃO LOGÍSTICA REGULARIZADA

Mota, I. L.[1]
Guimarães, A. A. P.[2]

RESUMO

A Doença de Parkinson (DP) é uma doença degenerativa do sistema nervoso central, crônica e progressiva. Causada por uma redução intensa da produção de dopamina, um neurotransmissor. A primeira manifestação notada é o tremor, porém antes disso, outros sentidos já sofrem alterações que nem sempre são observadas com facilidade, como por exemplo, a fala. Apesar de ser conhecida pela capacidade motora, como o tremor de repouso, dificuldade em realizar movimentos voluntários a DP possui sintomas não motores, dentre eles: distúrbios autonômicos, distúrbios de sono e sintomas neuropsiquiátricos. Neste contexto, muitos métodos de inteligência computacional baseados em Machine Learning têm sido aplicados para o diagnóstico da doença. Este trabalho objetiva desenvolver um classificador, baseado no modelo de Regressão Logística Regularizada (RL-Reg), que auxiliará o clínico especialista no diagnóstico da Doença de Parkinson, tendo como base de dados amostras de voz de pacientes portadores e não portadores de Doença de Parkinson. Por fim, promissoramente, vislumbra-se que o simulador seja inserido à sociedade como uma alternativa eficiente às ferramentas existentes no auxílio ao prognóstico de DP.

Palavras-chave: Regressão Logística Regularizada; Doença de Parkinson; Machine learning.

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (Unilab), Instituto de Engenharias e Desenvolvimento Sustentável (IEDS), Discente, ithalo@aluni.unilab.edu.br[1]
Universidade da Integração Internacional da Lusofonia Afro-Brasileira, CE, Docente, alisson@unilab.edu.br[2]

ISSN: 2447-6161

INTRODUÇÃO

Apesar de inúmeros progressos obtidos nos últimos anos na pesquisa da doença de Parkinson (DP), de acordo com Teive (2005), pesquisas demonstram que a DP deve ser considerada como uma enfermidade neurodegenerativa, progressiva. Causada por uma redução intensa da produção de dopamina, A dopamina ajuda na realização dos movimentos voluntários do corpo de forma automática, ou seja, não é necessário pensar em cada movimento que os músculos realizam, graças à presença dessa substância no cérebro. A falta de dopamina, particularmente na região encefálica chamada substância negra, causa a perda do controle motor do indivíduo. Entretanto, a etiologia da doença ainda é uma incógnita. Apesar de ser conhecida pela capacidade motora, como o tremor de repouso, tenha dificuldade em realizar movimentos voluntários a DP possui sintomas não motores, dentre eles: distúrbios autonômicos, distúrbios de sono, sintomas neuropsiquiátricos (cognitivo, comportamento e humor). Dito isso, a moléstia de Parkinson apesar de bem característica não possui um diagnóstico difícil, porém pode ser demorado, levando até anos. Dessa forma, a tecnologia tende a otimizar e acelerar esse processo, através das análises de parâmetros e padrões. O método de machine learning é bastante utilizado para fins medicinais, pois facilita a detecção, previsão e diagnóstico de doenças pela “simples” análises de dados. A regressão logística, por exemplo, é um método de otimização matemática que encontra relações entre dados a partir de técnicas de análise desses dados. Essas respostas geralmente são binárias, ou seja sim ou não.

METODOLOGIA

Com os dados públicos disponibilizados em UCI Machine Learning Repository: Parkinson Speech Dataset with Multiple Types of Sound Recordings Data Set, começou-se a elaboração do algoritmo que interpretasse as variáveis. Os dados pertencem a 20 pacientes diagnosticados com Parkinson (6 mulheres e 14 homens) e 20 pessoas saudáveis (10 homens e 10 mulheres) que recorreram ao Departamento de Neurologia da Faculdade de Medicina de Cerrahpasa, na Universidade de Istambul.

Para cada paciente foram utilizados múltiplos tipos sons (26 para cada) variando entre vogais, números, palavras e pequenas sentenças. A partir desses sons foram retiradas 26 atributos a serem estudados, são eles:

(1~5): Jitter (local), Jitter (local, absolute), Jitter (rap), Jitter (ppq5), Jitter (ddp) - medições de variação e instabilidade de frequência;

(6~11): Shimmer (local), Shimmer (local, dB), Shimmer (apq3), Shimmer (apq5), Shimmer (apq11), Shimmer (dda) - uma medição de estabilidade da amplitude;

(12~14): AC, NTH, HTN - autocorrelação e proporção de ruído para componentes tonais na voz;

(15~19): tom mediano, tom médio, desvio padrão, tom mínimo, tom máximo;

(20~23): número de pulsos, número de períodos, período médio, desvio padrão do período;

(24~26): fração de quadros não vozeados localmente, número de “quebras” de voz, grau de “quebras” de voz;

A coleta desses atributos não é invasiva com os pacientes já que são simples testes de voz. Os arquivos dispunham de textos contendo as informações citadas anteriormente em valores numéricos. Dessa forma, deu-se início a criação do algoritmo. O método utilizado foi o de regressão logística.

Por ser um método de otimização, é uma ótima maneira de aplicar os dados, em formato de matriz, para chegar em um resultado. Desse modo, há no total 1040 amostras com 26 atributos cada uma. Dentre esses valores, 80% (832) foram utilizados para o treinamento do algoritmo, enquanto os 20% (208) restantes para os testes. Houve uma pequena adaptação na organização dos dados para a melhor leitura do algoritmo.

ISSN: 2447-6161

RESULTADOS E DISCUSSÃO

O algoritmo tem como principal objetivo aprender com os dados e conseguir armazenar essas informações para separar o que são características de Parkinson e o que não é. Dessa forma, classificando em 0 e 1, se há parkinson ou não.

Ao rodar o programa ele seleciona aleatoriamente 80% dos valores e aprende sobre eles até conseguir compreender e encontrar o resultado correto desses valores para depois prever os 20% restantes.

Pode-se utilizar seeds para manipular essa aleatoriedade, assim rodando o algoritmo ele escolherá sempre os mesmos valores. A partir da figura 1 é perceptível que em 13 épocas o algoritmo convergiu, isto é, chegou ao parâmetro de regularização.

Figura 1: Convergência do algoritmo

Além disso, também há um relatório do próprio produzido pelo próprio algoritmo, segue o trecho da fase de aprendizado:

***** Resultados na etapa de treinamento ******

A precisão do algoritmo na fase de treinamento é de: 100.00%

Número de épocas: 13

Norma do gradiente da função objetivo na última iteração (época 13): 0.000384

Verdadeiro positivo (VP): 422

Verdadeiro negativo (VN): 410

Falso positivo (FP): 0

Falso negativo (FN): 0

É notável, a partir desses resultados, que a precisão do algoritmo está muito boa, pelo menos na fase de treinamento. Que ele converge em 13 épocas com um parâmetro de 0.000384. Em relação aos verdadeiros positivos e negativos, isso significa que de fato ele acertou o que era falso e o que era positivo, do contrário os resultados estariam errados, um paciente saudável sendo acusado de estar doente ou o inverso.

O trecho do resultado está a seguir:

***** Resultados na etapa de teste ******

A precisão do algoritmo na fase de teste é de: 100.00%

Verdadeiro positivo (VP): 98

Verdadeiro negativo (VN): 110

Falso positivo (FP): 0

Falso negativo (FN): 0

OBS: Houve 0 erro(s) em 208 objetos analisados

Nessa parte o algoritmo mostrou um ótimo resultado, obtendo 100% de acerto na fase de teste. Que é a parte principal, onde ele analisa sem saber as respostas e somente depois as verifica.

ISSN: 2447-6161

CONCLUSÕES

Os resultados se mostraram promissores, com um ótimo valor na fase de teste, o que significa que além do bom desempenho do código, os dados foram bem coletados. O resultado de 100% é surpreendente. Por esse motivo, podemos continuar trabalhando e aprimorando esse código para aplicarmos realmente com novas pessoas.
Outrossim, com mais testes e melhorias, conseguiremos ajudar no processo diagnóstico de pessoas com DP, acelerar o tratamento e consequentemente melhorar a vida dessas pessoas.

AGRADECIMENTOS

A FUNCAP e PIBIC-UNILAB pela oportunidade de participar desse projeto como bolsista. Ao meu orientador Prof. Dr. Alisson Guimarães pela excelente condução no projeto, por todo suporte e apoio.

REFERÊNCIAS

Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O., 'Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings', IEEE Journal of Biomedical and Health Informatics, vol. 17(4), pp. 828-834, 2013;
Teive HAG. Etiopatogenia da Doença de Parkinson. Rev Neurociencias 2005; 13(4): 201-214;

ISSN: 2447-6161