

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA A IDENTIFICAÇÃO DE ANORMALIDADES EM ELETROCARDIOGRAMAS

Moacyr Francischetti Corrêa
Unimar – Universidade de Marília
moacyr@spacnet.com.br

RESUMO

Este estudo compara o desempenho de quatro algoritmos de aprendizado de máquina na classificação de batidas cardíacas em exames de eletrocardiograma. Os algoritmos avaliados são: Árvore de Decisão, Máquinas de Vetores de Suporte (SVM), K-Nearest Neighbors (KNN) e Naive Bayes. Os dados utilizados são do MIT-BIH Arrhythmia Database.

Palavras-Chave: Aprendizado de Máquina, Classificadores, Árvore de decisão, Máquinas de Vetores de Suporte, SVM, K-Nearest Neighbors, KNN, Naive Bayes. MIT-BIH Arrhythmia Database.

INTRODUÇÃO

A análise de eletrocardiogramas (ECG) é uma ferramenta essencial na medicina moderna, permitindo a detecção de uma ampla gama de anormalidades cardíacas. A classificação automática das batidas cardíacas pode auxiliar na identificação de condições cardíacas potencialmente perigosas, permitindo intervenções médicas mais rápidas e eficazes. Neste estudo, exploramos a aplicação de várias técnicas de aprendizado de máquina para classificar batidas cardíacas em exames de ECG.

As batidas são classificadas em cinco categorias:

- **Batidas normais:** Estas são batidas cardíacas que seguem o padrão normal e regular do coração. São o tipo mais comum de batidas e não indicam nenhuma condição médica adversa.
- **Batidas de Ectopia Supraventricular:** As batidas de Ectopia Supraventricular são batidas extras que se originam nas câmaras superiores do coração, conhecidas como átrios. Elas podem ocorrer isoladamente ou em padrões irregulares. Embora geralmente sejam inofensivas, as batidas de Ectopia Supraventricular podem ser desconfortáveis e causar sintomas como palpitações.
- **Batidas de Ectopia Ventricular:** As batidas de Ectopia Ventricular são batidas extras que se originam nas câmaras inferiores do coração, conhecidas como ventrículos. Elas podem ocorrer ocasionalmente em pessoas saudáveis, mas também podem ser um sinal de doença cardíaca ou danos no coração.
- **Batidas de fusão:** As batidas de fusão ocorrem quando uma batida normal e uma batida ectópica (extra) se fundem para formar uma única batida. Elas são geralmente raras e podem ser causadas por condições como a doença arterial coronariana.
- **Batidas Inclassificáveis:** Estas são batidas que não se encaixam claramente em nenhuma das categorias acima. Elas podem ser o resultado de artefatos no ECG, ruído ou outras irregularidades que tornam difícil a classificação.

Foi necessário realizar a normalização dos dados, pois a quantidade de “batidas normais” é muito maior que as outras classes, o que poderia levar a um viés indesejado.

METODOLOGIA

Os dados utilizados neste estudo são do MIT-BIH Arrhythmia Database, um recurso amplamente utilizado na pesquisa de arritmias cardíacas. Este banco de dados contém 48 trechos de meia hora de gravações ambulatoriais de ECG de dois canais, obtidos de 47 indivíduos estudados pelo Laboratório de Arritmia BIH entre 1975 e 1979.

Os algoritmos foram avaliados com base em várias métricas:

- **Acurácia:** Esta é a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões. É uma medida útil quando as classes são bem balanceadas.
- **Precisão:** Esta é a proporção de previsões positivas que foram corretas. É uma medida útil quando o custo de Falso Positivo é alto.
- **Recall:** Esta é a proporção de observações positivas reais que foram corretamente previstas. É uma medida útil quando o custo de Falso Negativo é alto.
- **f1-Score:** Esta é a média harmônica de Precisão e Recall e tenta encontrar o equilíbrio entre essas duas métricas.

RESULTADOS

Os resultados obtidos pelos quatro algoritmos são apresentados na tabela abaixo:

Classificador	Acurácia	Precisão	Recall	f1-Score	Matriz de Confusão				
					0	1	2	3	4
Árvore de Decisão	0,98575	0,98581	0,98595	0,98565	3762	122	75	25	34
					0	4015	0	0	0
					15	0	3981	2	4
					0	0	0	3987	0
					7	1	0	0	3970
SVM	0,91965	0,9198	0,92364	0,92066	3576	159	69	176	38
					505	3476	10	15	9
					175	30	3660	123	14
					117	10	30	3824	6
					100	4	13	4	3857
KNN	0,97555	0,97562	0,97588	0,97538	3669	190	82	53	24
					10	4005	0	0	0
					40	8	3913	28	13
					0	0	0	3987	0
					19	4	17	1	3937
Naive Bayes	0,4001	0,40134	0,6013	0,36425	452	31	263	189	3083
					249	551	220	166	2829
					349	40	779	39	2795
					11	15	38	2273	1650
					0	0	15	16	3947

0: Batidas normais
1: Batidas de Ectopia Supraventricular
2: Batidas de Ectopia Ventricular
3: Batidas de fusão
4: Batidas Inclassificáveis

CONCLUSÕES

Este estudo apresentou uma análise comparativa do desempenho de quatro algoritmos de aprendizado de máquina na classificação de batidas cardíacas em exames de eletrocardiograma. Os resultados mostraram que a Árvore de Decisão e o KNN superaram os outros algoritmos em todas as métricas avaliadas, demonstrando serem métodos eficazes para essa tarefa específica.

A Árvore de Decisão, com sua estrutura de regras simples e interpretabilidade, provou ser um método robusto para a classificação de batidas cardíacas. Isso sugere que a estrutura dos dados pode ser bem capturada por um modelo baseado em regras.

O KNN também apresentou um desempenho impressionante, o que indica que a natureza dos dados pode ser tal que batidas cardíacas semelhantes tendem a ser agrupadas juntas no espaço de características. Isso faz do KNN uma boa escolha para essa tarefa de classificação.

Por outro lado, o algoritmo Naive Bayes apresentou o desempenho mais baixo em todas as métricas. Isso pode ser devido à sua suposição de independência entre os recursos, que pode não ser válida para este conjunto de dados.

Os resultados deste estudo destacam a importância da escolha do algoritmo de aprendizado de máquina adequado para a tarefa de classificação de batidas cardíacas em exames de ECG. Cada algoritmo tem suas próprias forças e fraquezas, e a eficácia de um algoritmo pode depender fortemente da natureza dos dados.

Além disso, este estudo ressalta a importância da normalização dos dados ao lidar com classes desequilibradas. Sem a normalização, o modelo pode ser excessivamente influenciado pela classe majoritária – nesse caso, batidas normais – e ter um desempenho ruim na classificação das classes minoritárias.

Concluindo, este estudo fornece insights valiosos sobre a aplicação de técnicas de aprendizado de máquina para a classificação de batidas cardíacas em exames de ECG. Os resultados podem servir como um ponto de partida para pesquisas futuras nesta área, com o objetivo de desenvolver sistemas de diagnóstico cardíaco mais precisos e eficazes.

REFERÊNCIAS

- Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003). *Circulation*. 101(23):e215-e220.
- Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng in Med and Biol* 20(3):45-50 (May-June 2001). (PMID: 11446209)
- Kotsiantis SB. Decision trees: a recent overview. *Artif Intell Rev*. 2013; 39:261–283.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995; 20:273–297.
- Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. 1967; 13:21–27.
- Rish I. An empirical study of the naive Bayes classifier. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*. Vol. 3. 2001. p. 41–46.