

Predição de sepse em unidade de terapia intensiva: uma abordagem de aprendizado de máquina

Diogo Hideki Shiraishi¹, Ernesto Cordeiro Marujo²

¹Data Science Team – Itaú Unibanco
São Paulo, Brasil

²Instituto Tecnológico de Aeronáutica
São José dos Campos, Brasil

diogohshiraishi@gmail.com, marujo@ita.br

Abstract. *Sepsis is a life-threatening organ dysfunction and a leading cause of mortality in hospital Intensive Care Unit (ICU). Early diagnosis of sepsis is fundamental, considering immediate medical intervention improves patient outcomes. The initiation of the treatment protocol for sepsis at the right time reduces mortality. We aim our study to predict sepsis six hours prior to the development of symptoms that would render a clinical diagnosis according to the Sepsis-3 guidelines.*

Resumo. *A sepse é uma disfunção orgânica com risco de vida e uma das principais causas de mortalidade em Unidade de Terapia Intensiva (UTI). O diagnóstico precoce da sepse é fundamental, considerando que a pronta intervenção médica melhora o desfecho do paciente. O início do protocolo de tratamento de sepse no tempo certo reduz a mortalidade. Nosso objetivo é prever a sepse seis horas antes do desenvolvimento dos sintomas que levariam a um diagnóstico clínico de acordo com as diretrizes da Sepsis-3.*

1. Introdução

A sepse é uma condição caracterizada pela resposta exacerbada do organismo a uma infecção [Singer et al. 2016]. Órgãos e tecidos do próprio indivíduo são prejudicados, levando à disfunção orgânica. Um estágio mais avançado é o choque séptico, no qual o paciente necessita de medicamentos vasopressores devido a um quadro de hipotensão e um nível sérico de lactato elevado, mesmo com reposição volêmica. Na inexistência de um procedimento de teste diagnóstico padrão ouro, identifica-se essa síndrome, em um paciente com suspeita de infecção, através de um conjunto de sinais clínicos e sintomas.

Internacionalmente, a Surviving Sepsis Campaign defende a aplicação de um pacote (*bundle*) de primeira hora, denominado *Hour-1 bundle* [Levy et al. 2018], em caso de consistente constatação de sepse. As ações recomendadas compreendem a medição do nível de lactato, hemocultura antes da aplicação de antibióticos, administração de antibióticos de amplo espectro, ressuscitação volêmica e uso de vasopressores. Identificado

Any opinions, findings, conclusions, or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views, official policy or position of the Itaú Unibanco.

o patógeno pela hemocultura, restringe-se o espectro antimicrobiano, favorecendo o tratamento do paciente e reduzindo o desenvolvimento de infecções multirresistentes e os custos. Logo, um atraso no reconhecimento da sepse prejudica a correta intervenção médica, o que leva à piora do estado do paciente, podendo evoluir para choque séptico, falência de órgãos e óbito. Em caso de choque séptico, o início tardio de antibioticoterapia reduz em aproximadamente 4 a 8% a chance de sobrevivência a cada hora de atraso [Kumar et al. 2006, Seymour et al. 2017].

No Brasil, a sepse é uma das principais causas de mortalidade em hospitais. É a doença mais mortal em unidades de terapia intensiva (UTI), na qual 55,7% dos pacientes internados com sepse evoluem a óbito [Machado et al. 2017]. Outra preocupação é a contrastante letalidade entre hospitais públicos (42,2%) e hospitais privados (17,7%) [ILAS 2018]. Um dos objetivos desse estudo é considerar o cenário de racionalidade e escassez como o Sistema Único de Saúde (SUS) público brasileiro.

Este artigo, portanto, apresenta o estado atual do desenvolvimento de uma solução que permita o reconhecimento precoce de sepse em pacientes internados em UTI.

2. Trabalhos Relacionados

Calvert et al. (2016) avaliam uma solução proprietária, denominada InSight, para prever até 3 horas antes um episódio de 5 horas de duração do critério *Systemic Inflammatory Response Syndrome* (SIRS) [Bone et al. 1992]. Foi utilizado o conjunto de dados *Multi-parameter Intelligent Monitoring in Intensive Care* (MIMIC) II. Obtiveram *recall* de 0,90, especificidade de 0,81 e área sobre a curva *Receiver Operating Characteristic* (AUC) de 0,83. Desautels et al. (2016) realizam mais uma avaliação do InSight, com a base MIMIC-III, adicionando a variável escala de coma Glasgow para avaliação de consciência. Conseguiram 0,58 de acurácia, 0,75 de AUC, 0,80 de *recall* e 0,55 de especificidade para uma predição com 4 horas de antecedência. Barton et al. (2019) tentam realizar a identificação no momento da definição clínica pelo consenso Sepsis-3 e a predição de 24 e 48 horas, utilizando o algoritmo XGBoost. Para 24 horas de antecedência, obtiveram AUC de 0,84, 0,80 de *recall* e 0,72 de especificidade.

O artigo de Ginestra et al. (2019) avalia o impacto de um sistema de alerta para sepse no usuário final, a equipe médica. Argumenta-se que médicos podem considerar algoritmos preditivos complexos não passíveis de confiança e retornar à intuição clínica clássica. As conclusões demonstram o desafio para a computação permear e promover um entrelaçamento técnico-científico capaz de resultar em eficientes novas ferramentas para melhoria da realidade hospitalar. Isto é, abordar a interpretabilidade e explicabilidade dos alertas gerados por sistemas preditivos.

3. Metodologia e Resultados

A competição PhysioNet Challenge 2019 [Reyna et al. 2019] disponibilizou um conjunto de arquivos de 40.336 pacientes internados em UTIs de três hospitais, sendo um desses exclusivo para avaliação de participantes da competição. Os dados foram coletados do hospital Beth Israel Deaconess Medical Center (Boston, Estados Unidos) e Emory University Hospital (Atlanta, Estados Unidos).

O arquivo de registros de um paciente contém observações por hora de 8 sinais vitais, 26 resultados laboratoriais e 6 dados demográficos, totalizando 40 atributos. Há uma

coluna de *target* que diz respeito a se o paciente, no instante de determinada observação, é identificado com sepse (1) ou não (0). A definição utilizada de sepse é a mais atual, a Sepsis-3 [Singer et al. 2016]. Isto é, uma deterioração em dois pontos no *score Sequential Organ Failure Assessment* (SOFA) [Vincent et al. 1996] indicando disfunção orgânica e suspeita clínica de infecção, dado pela solicitação de hemocultura ou administração intravenosa de antibióticos.

O conjunto de dados fornecido é esparsos, principalmente os resultados de exames laboratoriais, com mais de 80% dos valores nulos. Isso é compreensível, dado que não se realizam coletas para análise a cada hora. Considerando possíveis fontes de vazamento de informação, a imputação de dados faltantes é realizada por meio de *forward fill*, no qual o último dado disponível é propagado para instantes de tempo posteriores, visto que as observações em sequência formam uma série temporal.

A divisão entre treino e teste é realizada por paciente. Ou seja, todo conjunto de registros médicos de um indivíduo está, exclusivamente, no subconjunto de treinamento ou no subconjunto de teste.

Do total de 40.336 pacientes, apenas 2.932 apresentam quadro de sepse em algum momento de sua internação. Isto é, uma prevalência de sepse de 7,3%. Se for considerada a quantidade total de registros, sem distinção por paciente, são 1.552.210 observações. Destes, apenas 27.916 são de registros cujo atributo alvo “sepsis” é verdadeiro.

A decisão para realizar um balanceamento do *dataset* foi de aplicar *undersampling* dada uma única justificativa: evitar a criação de amostras artificiais em um problema de natureza orgânica, tal como é orquestrada pela fisiologia humana. Em seguida, foi necessário decidir por qual algoritmo de subamostragem. Dentre mais de uma dezena de métodos disponíveis, optou-se pela subamostragem aleatória (*random undersampling*).

A seleção dos atributos foi realizada a partir de duas razões. A primeira resulta da análise de frequência de valores nulos. Adotou-se um limiar máximo para seleção de *features* tal que a quantidade de *missing values* seja inferior a 70%. Assim, todos os resultados laboratoriais são descartados, com exceção do lactato. Um nível de lactato elevado pode indicar má perfusão de oxigênio nos tecidos, um indicativo de disfunção dos órgãos. A segunda razão é uma consideração tática. A monitorização eletrônica de pacientes internados em UTIs não é uma característica heterogênea, quando considerado os diferentes cenários da realidade da saúde pública. É evidente uma discrepante disponibilidade de recursos entre instituições públicas e privadas [Dias et al. 2014]. Ante o exposto, os seguintes sinais vitais escolhidos são: frequência cardíaca (HR), oximetria de pulso (O2Sat), temperatura (Temp), pressão sistólica (SBP), pressão diastólica (DBP), frequência respiratória (Resp), lactato (Lactate), idade (Age), gênero (Gender) e tempo de internação em UTI (ICULOS).

Ao ser constatada a presença de valores extremos inconcebíveis, foi tomada a decisão de limitar os sinais vitais dentro de intervalos considerados aceitáveis. Os intervalos adotados foram: frequência cardíaca entre 30 e 200, oximetria entre 70 e 100, temperatura entre 30 e 41, frequência respiratória entre 10 e 50, pressão sistólica (SBP) entre 50 e 200 e pressão diastólica (DBP) entre 10 e 120.

Ao decidir pelo descarte de 25 variáveis de resultados laboratoriais, surgiu a preocupação com possível perda de informação revelante. Os exames laboratoriais têm

uma frequência inferior aos sinais vitais, eletronicamente adquiridos [Reyna et al. 2019]. Contudo, de acordo com a literatura [Singer et al. 2016], além do lactato, resultados de creatinina, bilirrubina, gasometria e coagulação podem ser indicativos de disfunção orgânica. Decidimos, então, adicionar alguns *scores* médicos relacionados direta e indiretamente à sepse. Optou-se por imputar no conjunto de dados o cálculo dos *scores*: SIRS, SOFA, *quick Sequential Organ Failure Assessment* (qSOFA) [Singer et al. 2016], *Acute physiology and chronic health evaluation* (APACHE II) [Knaus et al. 1985], *Prehospital Early Sepsis Detection* (PRESEP) [Bayer et al. 2015], *National Early Warning Score* (NEWS) 2 [Royal College of Physicians 2017] e *Shock Index* [Allgöwer and Burri 1967].

A manifestação fisiológica da sepse dá-se pela evolução de métricas do organismo no tempo [Singer et al. 2016]. Sendo assim, condiciona-se indispensável considerar não somente os dados de apenas uma observação, mas a evolução dentro de uma janela temporal. Portanto, define-se uma janela deslizante de 6 horas. Isto é, dada uma observação no instante de tempo t , são concatenados a este os dados de t_{-1} até t_{-5} .

Os algoritmos de aprendizado de máquina avaliados neste trabalho foram: regressão logística (*logistic regression*), árvore de decisão (*decision tree*), floresta aleatória (*random forest*) e *naïve Bayes*. Em seguida, técnicas baseadas em *boosting* foram adotadas, como o AdaBoost, XGBoost e LightGBM.

Para avaliação final da performance dos modelos gerados no conjunto de teste, dá-se maior enfoque nas métricas de sensibilidade (*recall* ou *true positive rate* (TPR)), precisão e *F1 score*. Dado que o *dataset* é desbalanceado, não é adequado uma simples avaliação de acurácia e AUC. A técnica de validação cruzada com k -pastas (k -fold) é utilizada com divisão $k = 5$ pastas. Nessa técnica $k - 1$ conjuntos são utilizados para treino e o restante para teste [James et al. 2013].

Inicialmente, com o intuito de estabelecer um ponto de partida, foram executados os métodos *logistic regression*, *decision tree*, *random forest* e *naïve Bayes*. Devido à baixa performance, prosseguiu-se aos modelos AdaBoost, XGBoost e LightGBM sendo este último o escolhido devido à combinação de performance preditiva e velocidade.

A avaliação do modelo gerado foi feita por meio de três cenários distintos. Pode-se analisar a performance do modelo calculando as métricas sobre a quantidade total de observações disponibilizadas para teste. A segunda opção é realizar um *broadcast* de uma predição positiva de sepse para os instantes posteriores. A última alternativa é a análise por paciente. Isto é, deseja-se saber, dentre todos os paciente que desenvolvem sepse em algum momento de sua internação na UTI, quantos deles foram identificados pelo modelo. A intenção foi a de avaliar três faces do modelo proposto, a cobertura dos pacientes e a correta identificação precoce da sepse. Assim sendo, após a aplicação de validação cruzada com *k-Folds* com cinco divisões, obtivemos métricas para o conjunto total de observações de teste e para a visão por paciente como um todo, tal como na Tab. 1.

4. Conclusão e Trabalhos Futuros

A predição de casos de sepse em pacientes internados em UTI foi o tema do presente trabalho. Com um *dataset* de sinais vitais e resultados laboratoriais de milhares de pacientes, propôs-se o emprego de *machine learning* para construção de um modelo capaz de identificar com antecedência o risco de evolução de um quadro infeccioso para sepse.

Tabela 1. Resultados com LightGBM

Métrica	Individual	Broadcast	Paciente
Acurácia	0,812 ± 0,010	0,626 ± 0,013	0,639 ± 0,011
AUROC	0,760 ± 0,003	0,730 ± 0,003	N/A
Precisão	0,068 ± 0,004	0,041 ± 0,001	0,142 ± 0,004
Recall	0,707 ± 0,013	0,837 ± 0,016	0,844 ± 0,019
F1 score	0,123 ± 0,006	0,077 ± 0,003	0,243 ± 0,006

A predição de sepse é uma ferramenta auxiliar para a equipe médica, pois possibilita que haja intervenção em tempo hábil. Quanto mais cedo a sepse for identificada e o *bundle* for aplicado, maiores as chances de recuperação. Para tanto, uma série de atividades de pré-processamento e *feature engineering* foram empregadas, face aos desafios identificados de dados faltantes e alto desbalanceamento entre as classes alvo. Tomou-se a decisão de uso do algoritmo LightGBM para modelagem do problema. Os resultados do modelo proposto foram tais que tenta-se realizar a melhor cobertura possível dos casos positivos de sepse, de modo a não descartar qualquer paciente com a condição.

Como trabalho futuro, quer-se avaliar modelos de aprendizado profundo. Entre eles: *multilayer perceptron* (MLP), rede neural convolucional (CNN), rede neural recorrente (RNN), *long short-term memory* (LSTM) e *autoencoder* com uma abordagem de detecção de anomalias. Ainda, serão realizados experimentos com *ensembles* de diferentes modelos de aprendizado de máquina com o objetivo de melhorar a performance da solução. Considerando um subconjunto sequencial de sinais vitais, pode-se modelar cada uma delas como uma série temporal e realizar predições. A projeção da evolução desses indicadores pode auxiliar na predição de sepse. Pode-se, ainda, descartar pacientes com menos de 18 anos de idade, visto que metabolismo e fisiologia diferem de um adulto, assim como algumas métricas de avaliação [Le et al. 2019]. Por fim, é relevante a obtenção de dados de mais pacientes, de diferentes unidades de saúde, de diversas populações. Para isso, é objetivo a incorporação do MIMIC-III, curado pelo MIT Lab for Computational Physiology [Johnson et al. 2016] com mais de 60.000 registros de admissões na unidade de terapia intensiva do Beth Israel Deaconess Medical Center.

Referências

- Allgöwer, M. and Burri, C. (1967). Schockindex. *DMW - Deutsche Medizinische Wochenschrift*, 92(43):1947–1950.
- Bayer, O., Schwarzkopf, D., Stumme, C., Stacke, A., Hartog, C. S., Hohenstein, C., Kabisch, B., Reichel, J., Reinhart, K., and Winning, J. (2015). An early warning scoring system to identify septic patients in the prehospital setting: the PRESEP score. *Academic emergency medicine*, 22(7):868–871.
- Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A., Schein, R. M., and Sibbald, W. J. (1992). Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest*, 101(6):1644–1655.
- Dias, F. S., Rezende, E. A. d. C., Mendes, C. L., Silva Jr, J. M., and Sanches, J. L. (2014). Hemodynamic monitoring in the intensive care unit: a Brazilian perspective. *Revista Brasileira de terapia intensiva*, 26(4):360–366.

- ILAS (2018). Morte por infecção é duas vezes maior na rede pública do que na privada. <https://ilas.org.br/interacao/?p=357>. Accessed: 2020-03-07.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer New York.
- Johnson, A. E., Pollard, T. J., Shen, L., Li-wei, H. L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., and Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, 3:160035.
- Knaus, W. A., Draper, E. A., Wagner, D. P., and Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. *Critical care medicine*, 13(10):818–829.
- Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., Taiberg, L., et al. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical care medicine*, 34(6):1589–1596.
- Le, S., Hoffman, J., Barton, C., Fitzgerald, J. C., Allen, A., Pellegrini, E., Calvert, J., and Das, R. (2019). Pediatric severe sepsis prediction using machine learning. *Frontiers in Pediatrics*, 7:413.
- Levy, M. M., Evans, L. E., and Rhodes, A. (2018). The Surviving Sepsis Campaign Bundle: 2018 Update. *Intensive care medicine*, 44(6):925–928.
- Machado, F. R., Cavalcanti, A. B., Bozza, F. A., Ferreira, E. M., Carrara, F. S. A., Sousa, J. L., Caixeta, N., Salomao, R., Angus, D. C., Azevedo, L. C. P., et al. (2017). The epidemiology of sepsis in Brazilian intensive care units (the Sepsis PREvalence Assessment Database, SPREAD): an observational study. *The Lancet Infectious Diseases*, 17(11):1180–1189.
- Reyna, M., Josef, C., Jeter, R., Shashikumar, S., Westover, M., Nemati, S., Clifford, G., and Sharma, A. (2019). Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*.
- Royal College of Physicians (2017). National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS.
- Seymour, C. W., Gesten, F., Prescott, H. C., Friedrich, M. E., Iwashyna, T. J., Phillips, G. S., Lemeshow, S., Osborn, T., Terry, K. M., and Levy, M. M. (2017). Time to treatment and mortality during mandated emergency care for sepsis. *New England Journal of Medicine*, 376(23):2235–2244.
- Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., et al. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA: The Journal of the American Medical Association*, 315(8):801–810.
- Vincent, J.-L., Moreno, R., Takala, J., Willatts, S., De Mendonça, A., Bruining, H., Reinhart, C., Suter, P., and Thijs, L. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive care medicine*, 22(7):707–710.