

MANAGEMENT OF MISSING VALUES IN CONTINUOUS BIOMEDICAL DATA

Iapăscuță Victor¹

Scientific adviser: Belii Adrian¹

¹Valeriu Ghereg Department of Anesthesiology and Reanimatology No.1, Nicolae Testemitanu University.

Background. Primary data are an important pillar of practice and, in particular, of biomedical research. They often contain missing values. For „static” data, several recovery methods are proposed. In the case of data representing continuous biomedical signals, the set of methods is limited. **Objective of the study.** Presentation of an algorithm for the recovery of continuous biomedical data for later use for machine learning for clinical purposes. **Material and Methods.** The researched data are publicly available data describing 40,336 patients with sepsis and other pathologies (non-sepsis) provided by the competition „Early Prediction of Sepsis from Clinical Data: the PhysioNet Computing in cardiology Challenge 2019” and contain up to 80.9% of missing data. **Results.** Using the R programming language, an algorithm was created which, unlike other algorithms (e.g., LOCF – last observation carried forward) considers the dynamics (increase or decrease) of a certain parameter of interest. The data restored using the proposed algorithm are finally used to create a system for early prediction (up to 4 hours before onset) of sepsis, which has a predictive performance of 92% by the area under the ROC curve (AUROC). **Conclusion.** The proposed algorithm can be used to restore missing values in continuous biomedical data, describing physiological parameters recorded in intensive care units (heart rate, O₂ blood stasis, blood pressure, etc.).

Keywords: missing data recovery, algorithm, machine learning, artificial intelligence.

GESTIONAREA VALORILOR LIPSĂ ÎN DATE BIOMEDICALE CU CARACTER CONTINUU

Iapăscuță Victor¹

Conducător științific: Belii Adrian¹

¹Catedra de anesteziologie și reanimatologie nr. 1 „Valeriu Ghereg”, USMF „Nicolae Testemitanu”.

Introducere. Datele primare reprezintă un pilon important al practicii și, în deosebi, al cercetării biomedicale. Deseori, ele conțin valori lipsă. Pentru datele „statice” sunt propuse mai multe metode de restabilire. În cazul datelor ce reprezintă semnale biomedicale cu caracter continuu, setul de metode este restrâns. **Scopul lucrării.** Prezentarea unui algoritm de restabilire a datelor biomedicale cu caracter continuu pentru a fi, ulterior, utilizate pentru învățare automată în scopuri clinice. **Material și Metode.** Datele cercetate sunt date cu acces public, care descriu 40336 de pacienți cu sepsis și alte patologii (non-sepsis), furnizate de competiția „Early Prediction of Sepsis from Clinical Data: the PhysioNet Computing in cardiology Challenge 2019” și conțin până la 80,9% de date lipsă. **Rezultate.** Utilizând limbajul de programare R, a fost creat un algoritm care, spre deosebire de alți algoritmi (ex, LOCF – last observation carried forward), ține cont de dinamică (creștere sau descreștere) a unui anumit parametru de interes. Datele restabilite cu ajutorul algoritmului propus sunt, în final, utilizate pentru crearea unui sistem de prezicere timpurie (cu până la 4 ore până la debut) a sepsisului, care are o performanță predictivă de 92% după aria de sub curba ROC (AUROC). **Concluzii.** Algoritmul propus poate fi utilizat pentru restabilirea valorilor lipsă în date biomedicale cu caracter continuu, care descriu parametri fiziologici înregistrați în unitățile de terapie intensivă (frecvența cardiacă, saturația sângelui cu O₂, tensiune arterială etc.).

Cuvinte cheie: restabilirea datelor lipsă, algoritm, învățare automată, inteligență artificială.