Guest editorial: Special issue on mining for health

Myra Spiliopoulou[1] · Panagiotis Papapetrou[2]

1 Introduction

There is a proliferation of research on health related topics with data mining and machine learning playing a central role in such research, which is sometimes termed as "Medical AI". Medical research and patient caretaking are increasingly benefiting from advances in machine learning. The penetration of smart technologies and the Internet of Things give a further boost to initiatives for patient self-management and empowerment: new forms of health-relevant data become available and require new data acquisition and analytics' workflows. Efficient algorithms are developed and applied on health-related data, clearly motivated by the wish to promote medicine and healthcare, but with limited applicability and by making unrealistic assumptions about the needs of the people they have been designed for. On the other hand, concrete medical investigations or clinical support tasks are performed and demonstrated, but without an elaborate data mining or machine learning methodology.

2 The special issue

This special issue brought together the best of the above-mentioned two threads of research: data mining and machine learning methods for medical tasks and concrete medical and clinical studies and investigations. More concretely, it solicited research of the first type, while demanding evidence of the usefulness of the proposed algorithms and models, and it scrutinized assumptions that seem incompatible to medical research or medical practice. Moreover, it solicited research of the second type, while demanding either innovative algorithms or innovative workflows.

We received a total of 24 submissions, out of which 4 were accepted, 2 are currently under revision, and 18 were rejected. Next, we provide a summary of

✉ Panagiotis Papapetrou
  panagiotis@dsv.su.se

[1] Otto-von-Guericke-University Magdeburg, Magdeburg, Germany

[2] Stockholm University, Stockholm, Sweden

each paper and highlight the main contributions and relevance to the goals of the special issue.

Du et al. (2021) introduce a novel model, referred to as Adversarial Balancing-based representation learning for Causal Effect Inference. The paper focuses on observational data and highlights the selection bias problem by introducing a neural network encoder constrained by a mutual information estimator for minimizing the loss between representation and input covariates. By using the mutual information estimator, the model manages to preserve high predictive performance when employing the latent space. The authors use adversarial learning for balancing treatment and control groups. An extensive experimental evaluation shows that the proposed model outperforms competitors on several benchmark datasets.

Moreover, Varlamis (2021) investigates the prescription behaviour of physicians who treat a chronic disease. The monitoring of chronic diseases through the physician implies regular or irregular visits, whereupon the drug prescription plan may be modified in response to side-effects or emerging (co-)morbidities. Focus of the work is the detection and changes in the physician's drug prescription plan after the detection of a (co-)morbidity, including the prescription of new drugs and the discontinuation of drugs previously prescribed. The proposed approach is based on affinity analysis, a technique that traces co-occurrences of items or events across time stamped data. The approach has been validated on a dataset of more than 4400 patients with hypertension and hyperlipidemia, using the sequences of their visits to the physician(s) and the recorded diagnoses and drug prescriptions. The study is indicatory of the potential of affinity analysis for healthcare applications, and thus of even relatively simple data mining techniques that have been successfully applied in other application domains.

Zhang et al. (2021) focus on electronic health records (EHRs) and stress the importance of identifying and employing appropriate representations of EHR data, varying from simple statistics to temporal abstractions. Moreover, the paper highlights the challenges when dealing with data of unequal length, and proposes a tensor-based time-warping technique called DTW-CP. In contrast to the current state-of-the-art methods DTW-CP is the first one to utilize DTW for extracting patient feature similarities in a given cohort, and then combining them with tensor decomposition for learning the latent feature space. Experiments on the MIMIC-III database demonstrate the effectiveness of the proposed solution.

Finally, Saravanou et al. (2021) perform an extensive comparison of models for infant mortality prediction. The term 'infant mortality' refers to infants that die before their first birthday, and infant mortality prediction is essential for the identification of risk factors and the design of prevention measures. The authors model the prediction task as a classification problem, once with a dichotomous variable (yes/no) and once as a multi-class task by stratifying on the time of death, and they propose a framework that encompasses a training strategy, evaluation metrics, and an importance scoring scheme that assesses the contribution of explanatory variables to classifier quality. classifier quality. The study is indicative of the role that data mining methods can play in the acquisition of insights needed for the design of prevention policies.

References

Du X, Sun L, Duivesteijn W, Nikolaev A, Pechenizkiy M (2021) Adversarial balancing-based representation learning for causal effect inference with observational data. Data Min Knowl Discov. https://doi.org/10.1007/s10618-021-00759-3

Saravanou A, Noelke C, Huntington N et al (2021) Predictive modeling of infant mortality. Data Min Knowl Discov. https://doi.org/10.1007/s10618-020-00728-2

Varlamis I (2021) Affinity analysis for studying physicians' prescription behavior. Data Min Knowl Discov. https://doi.org/10.1007/s10618-021-00758-4

Zhang C, Fanaee-T H, Thoresen M (2021) Feature extraction from unequal length heterogeneous EHR time series via dynamic time warping and tensor decomposition. Data Min Knowl Discov. https://doi.org/10.1007/s10618-020-00724-6