Computer Science
Defense

Predicting Rare Clinical Events in Complex and Dynamic Environments

Azade Tabaie
Emory University

Abstract: Traditional machine learning classification algorithms assume a balanced proportion of classes in the data. However, class-imbalanced data is a challenge for training predictive models in many fields such as the medical domain. Although patient adverse outcomes occur rarely, they are worthy of prediction to improve the quality of care that patients have received; therefore, monitoring systems are needed in the hospital setting to capture the adverse rare events and improve patient health outcomes. ¡br¿¡br¿

To that end, machine learning and natural language processing (NLP) techniques were used along with clinical expert knowledge to address the issue of rare event classification in a complex environment such as a hospital setting. In particular, two different patient cohort with distinct characteristics and objectives were investigated.¡br¿¡br¿

First, strategies were proposed to predict a rare type of infection among hospitalized children with central venous lines (CVLs). This cohort of pediatric patients are at high risk of morbidity and mortality from hospital acquired infections. Many serious infections in hospitalized children are likely preventable through interventions that prevent the infection or identify them early to initiate antimicrobial therapy. Besides being considered as a rare clinical event, the definitions that have been proposed for bloodstream infection commonly have inadequate sensitivity for clinically important infections and may be difficult to generalize across electronic health records (EHR) platforms. To infer the onset of the infection from EHR and eliminate the need for extensive chart reviews, a surrogate definition for bloodstream infection was proposed and validated. Then, two study designs were tested to improve the prediction accuracy of the onset of the infection during hospitalization. Finally, a data fusion approach was undertaken to integrate structured and unstructured information from EHR to boost the prediction performance. Incremental but meaningful improvements in the predictions were observed after each step.¡br¿¡br¿

Second, an algorithm was proposed to monitor the visits to an emergency department to detect intimate partner violence (IPV). IPV is a pervasive social challenge with severe health and demographic consequences. People experiencing IPV may seek care in emergency settings. Despite the urgency of this critical public health issue, IPV continues to be profoundly underdiagnosed and is considered a persistent hidden epidemic. IPV is frequently undercoded, undetected without appropriate screening tools, and underreported, rendering it a rare encounter in EHRs. The early and appropriate detection of and response to such cases is critical in disrupting the cycle of abuse including IPV related morbidity and mortality. Our proposed algorithm benefits from NLP techniques and domain expert knowledge. It can identify victims of IPV with a high sensitivity by analyzing the recorded provider notes and patient narratives.¡br¿¡br¿

We argue that all the techniques incorporated in this thesis are transferable to identify other rare clinical events with the ultimate goal of improving the level of care.

Tuesday, November 16, 2021, 3:00 pm

https://us02web.zoom.us/j/3978183382?pwd=anFkN2V3d2VBWEErd2VCOEdWL2xiU