Can we learn in the ER?

William Elazmeh
Ottawa-Carleton Institute for Computer Science
University of Ottawa
Ottawa, Ontario K1N 6N5 Canada
welazmeh@site.uottawa.ca

Abstract

This paper argues that machine learning research can benefit from assessment and decision making experience and methodologies developed by physicians in departments of medical emergencies (ER).

Physicians in the ER have standard methodologies to follow in diagnosing diseases, predicting prognosis, selecting treatments, identifying subjects at risk for diseases, and so forth. These are all tasks that present challenges similar to many faced by machine learning researchers. I propose that adapting such methods to develop, validate and test learning systems can potentially lead support the reliability, the sensibility, and the deployment of confident learning systems.

Through clinical experience, physicians develop an intuitive sense of which findings on history, physical examination, and investigation are critical for making accurate assessment of patients' conditions. Such knowledge have been formalized in various forms, such are CDRs. A clinical decision rule (CDR) is a tool that quantifies individual contributions that various components (history, physical examination, and laboratory results) make towards the diagnosis, prognosis, or likely response to treatment in patients[1]. Developing and testing CDRs involves (1) creating the rule, (2) testing and validating the rule, and (3) assessing the impact of the rule on clinical behavior (impact analysis) [2]. Rule validation studies assess the accuracy of the rule in several clinical sites and in particular settings. Before a physician can apply such rules, they must assess the evidence supporting the use of these rules in their particular practice. This assessment is guided by a well-thought hierarchy of evidence of 4 levels. These levels range from: **level 1** the rule can be used in a wide variety of settings with confidence that it has an impact on clinical behavior and improves patient outcome, **level 2** the rule can be used in various settings with confidence in its accuracy, **level 3** the rule can be considered by clinicians to be used with caution and only if patients in the study (from which the rule was derived) are similar to those in the clinical settings, and **level 4** the rule needs further evaluation before it can be applied clinically.

We, machine learning researchers, develop learning systems and validate their performance using selected methods and metrics common to machine learning. In many cases, we fail to perform any significant analysis of the relationship between sample and population to help understand how well data samples represent the population. In most cases, we assume a distribution in the sample and/or in the population. This is clearly an issue that hinders the progress of applying machine learning techniques in real-life domains, particularly in medical domains. When creating CDRs, physicians require clear definition and description of: (1) the **relationship** between predictive variables (attributes) and outcome (class or decision) along with their clinical **importance** [1, 2], (2) the **characteristics** of patients that are likely to affect the performance of the decision rule, (3) the **study site** (clinic, office, ER, or hospital), (4) the level of **supporting evidence** (based on the hierarchy mentioned above), (5) the **effects** of clinical use (impact analysis), (6) the mathematical **technique** used to develop the rule, (7) the **results** of using the rule (sensitivity, specificity, likelihood ratio, etc.) (9) the **reproducibility** of predictors and outcome, and finally, (10) the **sensibility** of the rule in itself.

Unfortunately, such knowledge remains, in its strongest form, absent or implicit in most machine learning studies. I propose that we can learn form the experience of others to increase the reliability and sensibility of learning machines. Data and study analysis before building a learning system is important to the assessment of the quality of learning. The comprehensibility of the learning model can also contribute to the evaluation or assessment of the learning machine. Performance testing, validation, and long term effectiveness measures can demonstrate reliability and sensibility. In my opinion, machine learning research has reached a level of maturity capable of producing reliable and sensible learning machines. Our task is to demonstrate strong supporting evidence of due credibility.

References

[1] A. Laupacis, N. Sekar, and I. Stiell. Clinical prediction rules: a review and suggested modifications of methodological standards. *JAMA*, 277:488–494, 1997.

[2] T.G. McGinn, G.H. Guyatt, P.C. Wayer, C.D. Naylor, I.G. Stiell, and W.S. Richardson. User's guide to the medical literature: Xxii: how to use articles about clinical decision rules. *JAMA*, 284(1):79–84, 2000.