Data Mining with MIMIC-II in BIDMC ICUs

Joseph Paonessa MD, Thomas Brennan PhD, Mengling Feng PhD, Roger Mark MD PhD, Leo Anthony Celi MD, MS, MPH

Beth Israel Deaconess Medical Center – Harvard Medical School

MIT Laboratory of Computational Physiology

Problem

The failure to store and analyze the vast amount of data generated on a daily basis is a key hurdle in advancing the practice of critical care medicine. The intensive care unit (ICU) provides a cogent example of a data rich clinical domain in which an insufficient portion of the data generated has been employed for guiding practice by, for example, supporting the creation of clinical decision support tools, identifying significant patterns in population data, and employing feedback on system outputs for the formulation of systematic process improvements.

Aims

To build a learning system around an open-access ICU database where practice informs research, and research informs practice. Clinicians at the frontline of care should be at the core of this dynamic learning system, fully supported by engineers to collaborate on the daily translation of questions into strategies for database interrogation, modeling and analysis.

The Teams

Teams of clinicians (nurses, doctors, pharmacists) and scientists (database engineers, modelers, epidemiologists) have formed around the Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) database. The database was established in October 2003 from a partnership that combines the resources of a powerful interdisciplinary team from academia (Massachusetts Institute of Technology), industry (Philips Medical Systems and Philips Research North America) and clinical medicine (Beth Israel Deaconess Medical Center). The public-access database now holds over 60,000 ICU stays in the BIDMC ICUs.

The inter-disciplinary teams have been translating day-to-day questions typically asked during rounds that often have no clear answers in the current medical literature into study designs and then perform the modeling and the analysis and publish their findings.

Intervention

The learning system described above has been operating since 2010. The scientists attend ICU rounds, observe the processes surrounding data capture and interact with the rest of the clinical team. The clinicians on the other hand come to MIT and learn the tenets of clinical data analysis. This culture of collaboration is crucial in democratizing research and lowering the barrier for frontline clinicians who are most familiar with the information gaps in practice to participate in knowledge generation, as well as for those not traditionally associated with evidence creation, including patients themselves.

Progress

Predictive Modeling, Prognostication, and Outcomes

The MIMIC database has allowed our group to develop predictive models with actionable outputs that potentially lead to measurable improvements in process and/or outcome. Such models could support appropriate early triage regarding level of care and monitoring, as well as the allotment of costly resources such as specialist-requiring interventions and/or technologies. For example, these tools could assist emergency departments if limitations in ICU resources lead to regionalization of critical care.

Unraveling Complexity and Variability

MIMIC that include detailed clinical information has provided researchers an opportunity to accumulate safety and efficacy evidence, discover patient subpopulations that experience important variances in efficacy or unanticipated delayed adverse effects, and uncover interactions between and among simultaneous treatments as drugs become used in wider, more diverse patient populations than those possible during premarket approval clinical studies.
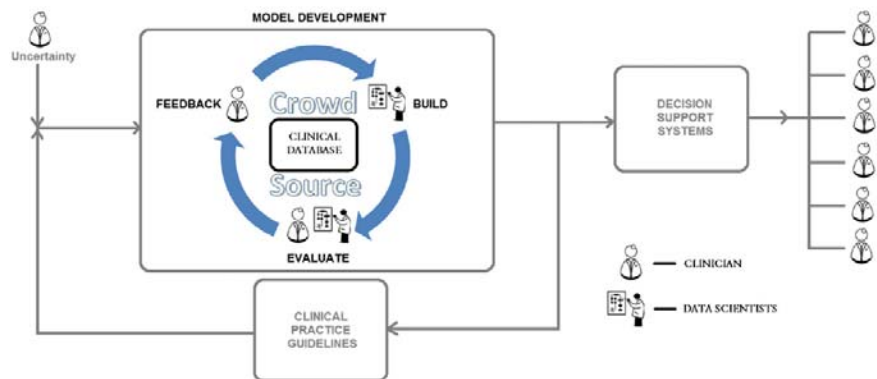
In January 2014, the MIT-BIDMC Critical Data Marathon and Conference was held. The basic premise of the data marathon was to bring together providers and data experts to answer clinically-relevant questions over the course of a weekend. While a truly novel discovery or a fully functioning solution is a rare outcome, these events enable crowdsourcing of valuable, varied points-of-view and new personal connections that will form the basis for longer-term collaborations. More than 80 participants formed 10 teams, including one in London, United Kingdom. The best projects were presented during the Critical Data Conference that followed. The conference was attended by more than 200 participants, and watched by another 400 via a live stream. The talks and discussions revolved around two themes: how to operationalize the vision of a data-driven learning system and how to safeguard big data in healthcare from contributing further to the swaths of unreliable research that plague medicine. The presentations have been made available online and have been viewed by more than 700 unique visitors (http://criticaldata.mit.edu/events/conference/program.html).

Next steps

While the current MIMIC database is limited to one academic hospital in the United States, plans are already in motion to extend the data to other hospitals, including institutions outside the U.S. As the database expands quantitatively and qualitatively across diverse care environments, the power and significance of any individual analysis will only increase over time. Furthermore, such analyses can be easily repeated, modified, and strategically improved based on iterative interpretation of prior findings.

Our vision is for the development of a care system consisting of "clinical informatics without walls", in which the creation of evidence and clinical decision support tools is initiated, updated, honed, and enhanced by crowd sourcing. In this collaborative medical culture, knowledge generation would become *routine and fully integrated* into the clinical workflow. This system would employ individual data to benefit the care of populations and population data to benefit the care of individuals.