Project Proposal

Analysis of large ICU database for prediction of sepsis

Jinghan LIU (zy18804)

School of Computer Science, University of Nottingham Ningbo China

Motivation and Background

Intensive care and medical records. Electronic health record (EHR) is a systematic and organized record of health information of individual patients that is generated and maintained at a health provider's site. In layperson's words it is a digital version of patient health chart. EHR contains a variety of data components, including, but not limited to demographic data, vitals and daily charts, medications, nursing notes, referrals, present and past complaints, medical history, life style items, physical examination, diagnoses, laboratory test results, medical images, procedures, treatments, hospital admission and discharge history, immunizations, and others [1]. The analysis of thousands of EHR can provide various views of health care, such as analysis of data related to individual patient, patient groups, disease, medication usage, procedures and other dimensions of health care. Various applications are possible, such as summary analytics, trend analytics, or predictive analytics that help decision making in health care and improvement of patient care based on evidence [2]. Critically ill patients admitted to intensive care unit (ICU) generate large amounts of data, because data types may be collected continuously (such as ECG) or hourly (such as vital signs or urine output) [3]. The ICU patients suffer from a serious injury or illness and health care practitioners often need to respond and administer treatment or intervention within in minutes or even seconds, all in condition of high uncertainty [4]. Delay in intervention may result in further injury or even death. Automated system that support the analysis of risk for individual patients and medical decision making can help improve the diagnosis and treatment of ICU patients and ultimately result in better medical outcomes. Clinical analytics uses EHR, disease registries, algorithms, assessment scores, input from monitoring devices, and other items to enable timely and effective response and decision making [5]. This step is computationally intensive; its effectiveness depends on the quality of data and the utility of algorithms that enhance health practitioner's ability to offer optimal response to the patient's needs. Patients with sepsis have higher rate of complications and mortality and bring higher cost of health care along with longer stay in hospital [6]. Prediction methods including statistical and machine learning approaches have been used for prediction of sepsis in ICU patients [6-10], with a variable success. This project will focus on computational aspects of decision support in ICU using clinical analytics with emphasis on early detection of sepsis in patients. We will use machine learning methods combined with the knowledge-based methods.

Computer science/Big Data. MIMIC-III ('Medical Information Mart for Intensive Care') is a large, single-hospital database of ICU patient data. MIMIC-III data include demographics, death data, bedside monitoring (vitals, waveforms, trends and alarms), charts (fluids, medications, and progress notes), laboratory tests, notes and reports (discharge, radiology,

cardiology), orders, and billing (ICD-9 diagnosis codes, drugs, and procedures) [11]. The data are de-identified with random date shifting and format conversions to provide high-level of privacy protection and are, therefore, not considered to be personal data. The MIMIC III data are available in multiple comma-separated values (csv) tables that can be as large as 4 GB. Preliminary analysis indicates that data has inconsistencies and errors that are not previously reported.

Aims and Objectives

This project focusses on applied computer science – I will use computer science techniques and methods to perform clinical analytics tasks. The main aim of this project is to develop and implement a system for online analytics of ICU data that can be used to predict complications in intensive care patients. My will work involve the design and implementation of a system that will perform summary analysis of MIMIC III data and a case study that will focus on prediction of sepsis. The project will involve data cleaning, interface design, summary analytics, and a machine learning case study (prediction of sepsis).

Specific objectives are:

1. Perform the survey of the state-of-the-art solutions for clinical analytics in the ICU environment.
2. Develop and implement a method for cleaning of ICU data.
3. Develop interface for ICU data analytics.
4. Perform a demonstration example of Big Data analytics on ICU data – prediction of sepsis
5. Prepare an article for publication
6. Complete and submit the final year dissertation.

Project Plan

Software development will utilize Waterfall model of software development [13]. The waterfall model is suitable because this work will result in a working prototype. Feature extraction and detecting diseased cells will use machine learning using Python environment, taking advantage of python libraries and the existing code for the in-house single cell database management. The main priority of this project is functional software prototype that will effectively deal with analytics of highly dimensional noisy data. The theoretical technological and engineering aspects of the software development will be considered and deployed as a secondary priority in this project. Specific tasks are:

Preparatory

1.1 Complete and submit supervisor project proposal, detailed project proposal, revised project proposal, and preliminary research ethics checklist

1.2 Review literature, existing ICU data analytics solutions and perform critical utility assessment

1.3 Develop software project plan document, revise

Software development

2.1 Develop and implement user interface data cleaning solution
2.2 Develop and implement user interface for SCT analytics
2.3 Develop and implement solution for summary data analytics (selected representative tasks – TBD)
2.4 Demonstrate the utility of developed software using sepsis as predictive analytics case.

Reporting and publication

3.1 Provide weekly incremental progress reports and short monthly written reports
3.2 Complete and submit interim report (deadline January 7, 2019)
3.3 Develop a plan and schedule for preparing the final dissertation, preliminary and revised.
3.4 Write and submit the final dissertation (deadline May 6, 2019).
3.5 Prepare and submit an article for publication (desired but not compulsory)

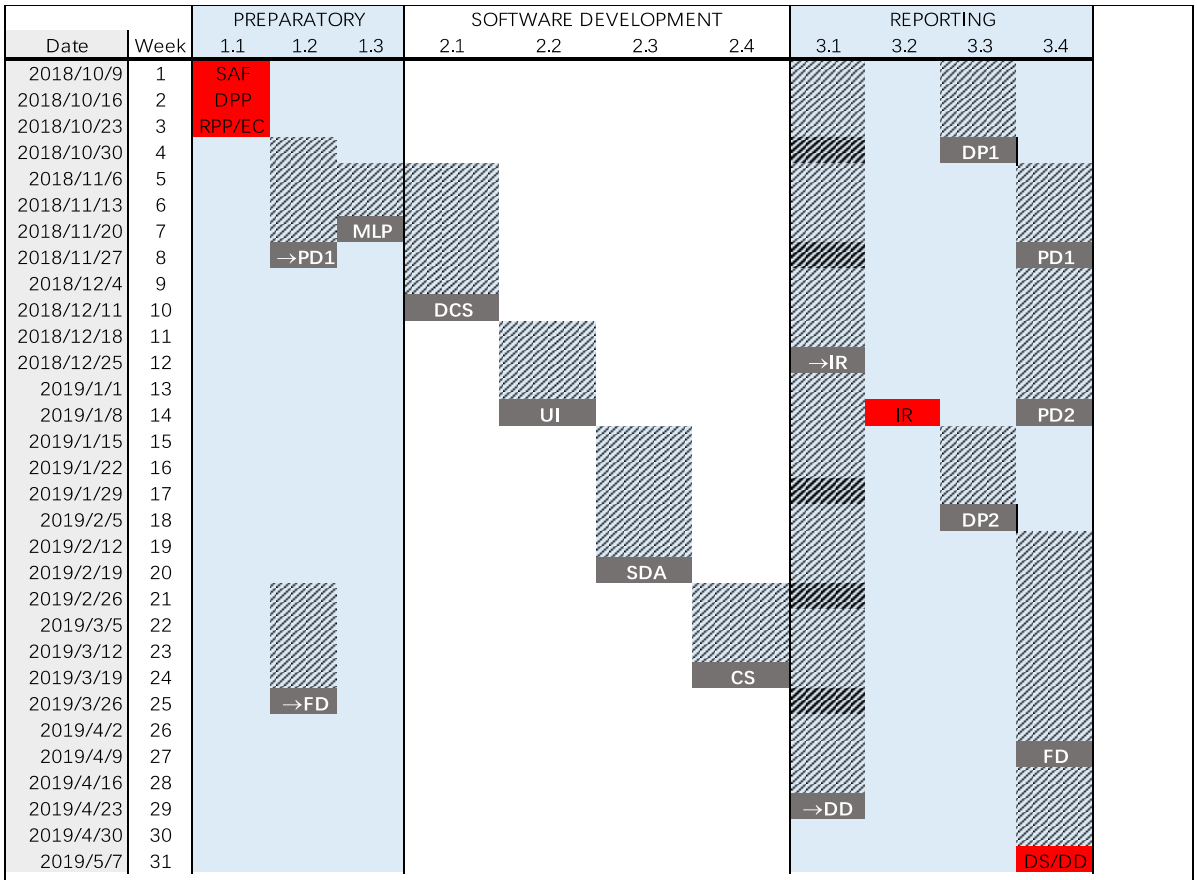
References

1. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. International journal of medical informatics. 2008 May 1;77(5):291-304.
2. Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care. Chest. 2018 May 9.
3. Zhang Z. Big data and clinical research: focusing on the area of critical care medicine in mainland China. Quantitative imaging in medicine and surgery. 2014 Oct;4(5):426.
4. Ghassemi M, Celi LA, Stone DJ. State of the art review: the data revolution in critical care. Critical Care. 2015 Dec;19(1):118.
5. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs. 2014 Jul 1;33(7):1123-31.
6. Sheetrit E, Nissim N, Klimov D, Fuchs L, Elovici Y, Shahar Y. Temporal Pattern Discovery for Accurate Sepsis Diagnosis in ICU Patients. arXiv preprint arXiv:1709.01720. 2017 Sep 6.
7. Siddiqui S, Chua M, Kumaresh V, Choo R. A comparison of pre ICU admission SIRS, EWS and q SOFA scores for predicting mortality and length of stay in ICU. Journal of critical care. 2017 Oct 1;41:191-3.
8. Schlapbach LJ, MacLaren G, Festa M, Alexander J, Erickson S, Beca J, Slater A, Schibler A, Pilcher D, Millar J, Straney L. Prediction of pediatric sepsis mortality within 1 h of intensive care admission. Intensive care medicine. 2017 Aug 1;43(8):1085-96.
9. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman MD, Barton C, Wales DJ. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR medical informatics. 2016 Jul;4(3).
10. Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. Journal of biomedical informatics. 2017 Feb 1;66:19-31.

11. Johnson AE, Pollard TJ, Shen L, Li-wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24;3:160035.
12. Petersen K, Wohlin C, Baca D. The waterfall model in large-scale development. In International Conference on Product-Focused Software Process Improvement 2009 Jun 15 (pp. 386-400). Springer, Berlin, Heidelberg.

Project schedule and deliverables

This chart is based on the activities detailed in the Project Plan section.



ADMINISTRATIVE DELIVERABLES

2018/10/12	SAF	Supervisor agreement form
2018/10/15	DPP	Detailed project proposal
2018/10/22	RPP	Revised project proposal
2018/10/22	EC	Preliminary research ethics checklist
2019/1/7	IR	Interim report submission
2019/5/6	DS	Dissertation submission
2019/5/7	DD	Demonstration day

INTERIM DELIVERABLES

DP1/2	Dissertation plan and schedule
PD1/2	Preliminary draft (structure)
FD	First draft
SPP	Software project plan
DEMO	Demonstration
MLP	Machine learning plan
DCS	Data cleaning solution
UI	User interface
SDA	Summary data analytics
CS-S	Case study - sepsis