

CS 3750 final project: Overview of MIMIC III and extensions

The final projects for the class uses MIMIC III database (<https://www.nature.com/articles/sdata201635>). MIMIC-III is a database of Electronic Health Records (EHRs) for patients admitted to ICUs of Beth Israel Deaconess Medical Center in Boston, Ma, which is one of the teaching hospitals in Boston area.

MIMIC data consists of different kinds of events, observations, readings and records that are recorded in multiple database tables. For complete list and explanation of the tables please see:

<https://mimic.physionet.org/mimictables/>

What follows is a brief description of tables you are most likely to use and some recommendations of how to select and use the data.

Admissions. The ADMISSIONS table is the basic table that list times of admissions and discharges of a patient from ICU. Field HADM_ID is the primary identifier of an admission (or encounter). It can be used to access and combine the different kinds of information about the patient from different table. The ADMISSION table also includes information about the type of discharge (patient may die or may be transferred to other units of the hospital). Please note that admission and discharge times (as well as all other times in the DB) are shifted randomly in time (a fixed shift for each individual patients) to avoid possible identification and matching of the patient based on time of the admission. Overall, there are 55,765 rows in the admission table. **Note:** In your project you do not have to use all admissions and actually I do not recommend to use all admission when your models rely on medications. The reason is that the different admissions in the table were covered by the different recording and mapping systems these used different (disjoint!!) medication vocabularies. These differences are recorded in the DB_SOURCE column of the admissions. The options here are metavision, carevue, both, and null. I recommend using the metavision admission subset which has more compact medication vocabulary and hence it is easier to work with.

Labevents. The LABEVENTS table represent the readings of specific test values for a patient in time. The lab test done is defined by the ITEMID code. The time when the results was reported is CHARTTIME. The value and units are in VALUE, VALUENUM and VALUEOM fields. In some cases, when the value is outside of the normal range it is labeled as abnormal in the FLAG field. The ITEMID codes used in the LABEVENTS table are defined in D_LABITEMS table. **Note:** Lab events with the same HADM_ID and ITEMID define a time series of observations for one patient and a specific lab. So if you want to analyze a numerical time series and model them you can pull the same time series for the different patients and learn the model. Also please note we used LABEVENTS table to define events in the midterm project database. However, in that case we ignored any value specifically assigned to them via results. Well, sometimes what tests are done (ordered by clinicians) tells you a lot about the patient and clinician's thoughts.

Medications represent administrations of the different medications in time and their dosages. These are recorded in two tables INPUTEVENTS_CV and INPUTEVENTS_MV. The difference is that they were recorded by the different systems with two different medication vocabularies. As mentioned earlier I recommend to use patients who were recorded by the metavision system and their medications are recorded in INPUTEVENTS_MV table. STARTTIME and ENDTIME define time interval during which the medication was given which is used often for continuous infusions. ITEMID defines a medication and the

ITEMID codes are defined in the D_ITEMS table. **Note:** Please note medications have dosages or infusion rates so it is possible to use these to come up with some numerical time series models.

Chart events stored in the CHARTEVENTS table record vital signs. These are readings of the most critical values reflecting body functions such as blood pressures, heart rate, respiratory rates, or temperatures and other records. The ITEMID column defines the type of the readings, and the code is defined in the D_ITEMS dictionary table. The time of the measurement is the CHARTTIME. **Note:** Please note that a chart-event may have a value and for a specific (fixed) code and HADM_ID these define a time series of the same observations. So this table, similarly to LABEVENTS table may be a good source of numerical time series if you want to model those.

Procedures. PROCEDURES_ICD is a table that records procedures done on patient during the admission. The procedures in this table are not logged in time and are associated with the admission only. The procedures for the different HADM_IDs are defined by the ICD-9 codes in the ICD9_CODE column. The dictionary of these codes is in the D_ICD_PROCEDURES table. **Note:** In addition to ICD-9 codes we mapped procedures also to procedure categories that are listed in the PROCEDURE_CATEGORY column. These codes are listed in D_ITEMS dictionary table. If you plan to use procedures in your models I recommend to try the PROCEDURE_CATEGORY first. ICD-9 codes may have very low counts.

Diagnoses. DIAGNOSIS_ICD is a table that lists all major diagnoses the patient had during the admission. The diagnoses in this table are not logged in time and are associated with the admission only. The diagnoses for the different HADM_IDs are defined by the ICD-9 codes in the ICD9_CODE column of the table. The ICD-9 codes for the diagnosis are defined in the D_ICD_DIAGNOSES table.

Microbiology. MICROBIOLOGYEVENTS table list the tests done on the patient in order to identify the presence of a foreign organism in the sample (blood, or urine, etc) causing infection. Please note that just the entry in the table signals there was a suspicion of infection to be present in the body. Some of the entries in the table lead to identification of the organism. When the organism is present the labs often test whether the strain of the organism is sensitive or resistant to a certain set of drugs that can be used to treat the infection. These are recorded in other columns of the MICROBIOLOGYEVENTS table.

Text records. NOTEEVENTS table records text written by a clinician. These consist of different types of text reports such as: progress notes or discharge notes. Progress notes are recorded daily for a patient and should reflect the current status of the patient.

ADDITIONAL TABLES

In addition to tables in MIMIC distribution we are building two tables that reflect adverse events in time for the metavision patients.

The first table SEPSIS_TABLE estimates and records the occurrence of sepsis in time.

The second table AKI_TABLE records scores reflecting different stages of Acute Kidney Injury (AKI) in time. The level 0 is related to 'no-AKI' state. Values 1, 2, and 3 reflect AKI-scores related to the different stages of AKI with stage 3 being the worst. These values are estimated from creatinine lab values and the creatinine baselines. Please note that the AKI_TABLE records only the changes in the scores. Also please note that we calculate AKI scores for all patients. However, the scores are clinically relevant only for patients who do not suffer from chronic kidney disease.

These two tables can be used for building event detection model where the goal is to detect sepsis and AKI as early as possible before they occur. Other events you can define and predict include:

- Death: the time of death is recorded in the ADMISSION table
- Discharge from ICU (other than death) which again is recorded in the ADMISSION table.