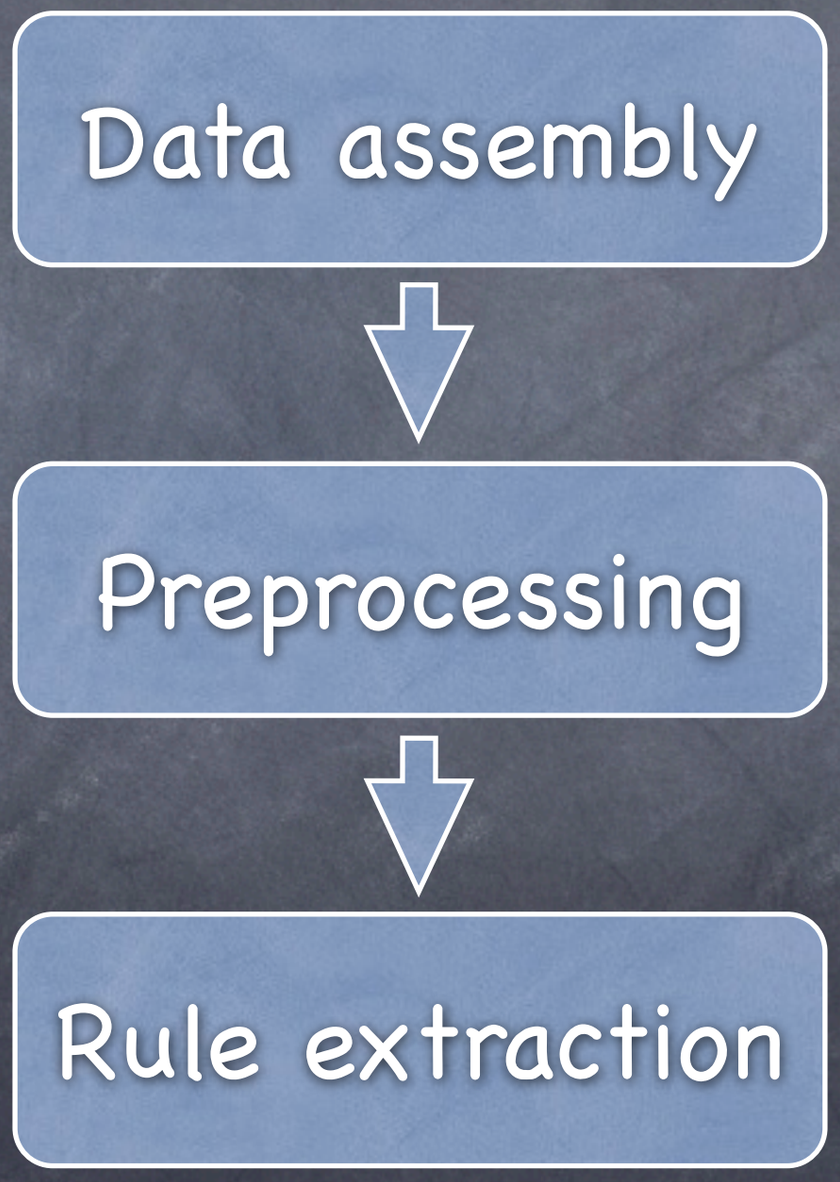
Project Progress Presentation

What are the crucial factors in treatment of diabetes patients?

Yulong Zhang

Oct 12, 2011

Main Idea



Main Idea

Data assembly

Preprocessing

Rule extraction

- Attributes having noting to do with diabetes can be ignored (e.g. payer_id).
- Factors related to patient's status/treatment should be taken into consideration (e.g. diagnosis/medications).

Main Idea

Data assembly

Preprocessing

Rule extraction

- Removing noises
- Fixing missing values
- Relevance analysis
- Data normalization
- Discretization
- .....

Main Idea

Data assembly

Preprocessing

Rule extraction

Construct rule-based classifiers:

If xx AND yy Or zz
THEN “to home”
OTHERWISE “no”

xx,yy,zz are our target

Data assembly

Using encounter_facts_diabetes as the core table, I add several features from other tables:

- Patients' bio-info (3 features)
  - ✳ Race, gender and marital_status might be crucial
- Lab_procedure_facts
  - ✳ Lab_procedure_dim (2437 entries)
  - ✳ Lab_procedure_facts_diabetes (limited to 1508 distinct entries)
  - ✳ Necessary check-ups (further limited to only 66 features)
    - Hemoglobin (39)
    - Creatinine (15)
    - Cholesterol level (12)
    - Blood pressure (0)
    - Feet and lower extremities check (0)
    - Eyes (0)
  - ✳ Average value is chosen if we have different results for one lab_procedure

```
select distinct `lab_procedure_name` from
`lab_procedure_facts_diabetes` left join `lab_procedure_dim` on
`lab_procedure_facts_diabetes`.`detail_lab_procedure_key`=`lab
_procedure_dim`.`lab_procedure_key` where
`lab_procedure_name` like '%creatinine%';
```

Data assembly

Using encounter_facts_diabetes as the core table, I add several features from other tables (continued):

- Procedure_facts (0 features)
  - ✳ procedure_facts_diabetes (2858 distinct procedures)
  - ✳ However, there is no icd9_procedure_dim provided. We can't judge which procedure is crucial given only icd9_procedure_key.

- Medication_facts

✳ Oral diabetes medication (16 features)
Metformin(723), Repaglinide(15), Nateglinide(11), Chlorpropamide(115), Glimepiride(121), Glipizide(494), Glyburide(616), Tolazamide(105), Tolbutamide, Pioglitazone, Rosiglitazone, Acarbose, Miglitol, Sitagliptin, Saxagliptin, Bromocriptine

✳ Insulin (1 features)

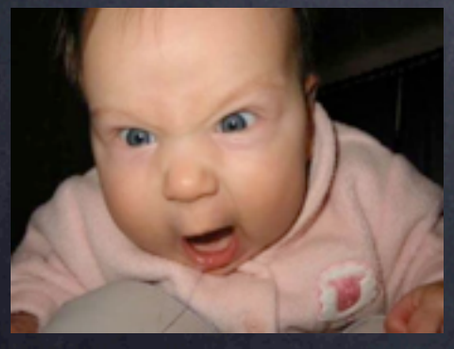
✳ Other injectable diabetes medicine (2 features)
exenatide(4), pramlintide(4),

Data assembly

What about the core table : encounter_facts_diabetes?

discharge_disposition_id
age
weight
weight_unit_id

3+66+16+1+2+4=92 features



Ok! Then let's assemble the data !

Feature list

Discharge_disposition_id
Age
Weight
Weight_unit_id

Race
Gender
Marital_status

Metformin
Repaglinide
Nateglinide
Chlorpropamide
Glimepiride
Glipizide
Glyburide
Tolazamide
Tolbutamide
Pioglitazone
Rosiglitazone
Acarbose
Miglitol
Sitagliptin
Saxagliptin
Bromocriptine
Isulin
exenatide
pramlintide

"Blood Gas Methemoglobin %, Arterial"
"Blood Gas Methemoglobin %, Venous"
"Hemoglobin Total, Urine"
"Hemoglobin, Arterial"
"Hemoglobin, Total"
"Hemoglobin, Venous"
Blood Gas O2 Hemoglobin %
Carboxyhemoglobin
Glyco Hemoglobin
Glyco Hemoglobin Group
Hemoglobin
Hemoglobin 1 Abnormal
Hemoglobin 2 Abnormal
Hemoglobin A
Hemoglobin A1C (Glycosylated Hemoglobin)
Hemoglobin A1C (Glycosylated Hemoglobin) Total
Hemoglobin A2
Hemoglobin Bart's
Hemoglobin C
Hemoglobin D
Hemoglobin E
Hemoglobin Electrophoresis
Hemoglobin F
Hemoglobin F(etal) Quantitative
Hemoglobin Free Plasma/Serum
Hemoglobin Free Urine
Hemoglobin G
Hemoglobin N
Hemoglobin O
Hemoglobin Phenotype
Hemoglobin Plasma/Serum
Hemoglobin S
Hemoglobin S(ickle) Screen

Mean Corpuscular Hemoglobin
Mean Corpuscular Hemoglobin Concentration
Methemoglobin
Methemoglobin by co-oximetry
Oxyhemoglobin by co-oximetry
Sulfhemoglobin
"Creatinine Clearance, Serum"
"Creatinine Clearance, Urine"
"Creatinine, Dialysate"
"Creatinine, Serum"
"Creatinine, Unknown specimen"
"Creatinine, Urine 12 hr"
"Creatinine, Urine 24 hr"
"Creatinine, Urine"
"Creatinine, Urine-Random"
"Creatinine, Urine-Timed"
"Creatinine, Whole Blood ISTAT"
"Creatinine, Whole Blood NOVA"
"Microalbumin/Creatinine Ratio, Urine"
Blood Urea Nitrogen Creatinine Ratio
Creatinine Clearance
"Cholesterol Total, Serum"
Cholesterol / HDL Ratio
Cholesterol / HDL Risk Interp
Cholesterol for Lipid Panel
Cholesterol Fractionation
Cholesterol LDL (Calculated)
Cholesterol Screen
Cholesterol Total & Triglyceride
HDL Cholesterol
LDL Cholesterol
LDL Cholesterol Direct Measure
VLDL Cholesterol

Current Progress

Data assembly — PHP+MYSQL

Preprocessing — WEKA

Rule extraction — WEKA

Thanks!

Yulong Zhang

Oct. 12, 2011