In this document, we cover an example of running the KARL using GEO6956.txt as the target dataset and GEO6956\_LTable.csv as the lookup table. These files are prepared in the KARL repository in ‘Demo-Data’ folder.

**Example:** Running KARL on the target dataset, GEO6956.txt by using the lookup table, GEO6956\_LTable.csv. As the discretization method EBD with default lambda parameter, 0.5, has been set as the command arguments:

-lp -tr 2 -lookuptable Demo-data/GEO6956\_LTable.csv -d 9 0.5 n -dp Demo-data/GEO6956.txt

**Output:**

Rule Learner version 2015-10-08

Args: -lp -tr 2 -lookuptable /Users/mahbaneheshaghzadehtorbati/JavaWorkspace/Original-MS-TRL/Demo-data/GEO6956\_LTable.csv -d 9 0.5 n -dp Demo-data/GEO6956.txt

Loading training dataset.....

Lines: 90. Separator is ' '. Parsing........... 12183 attributes, 89 instances.

Loading source dataset(s).....

Starting: 2019-02-12 19:23:44

On machine cl2-wifi-10-215-58-185.wireless.pitt.edu

With max memory 2047m(2147280773)

Loading the functional lookup table.....

Loaded (47) genes-hallmarks associations

Imported (47) genes-hallmarks associations

Inducing new rules with lookup table.....

Discretizing training data....

Discretizing using EBD using lambda of 0.5 (2009) ........... There are 4959 attributes with cut points.

48 rules Induced with lookup table.

Learned 11 rules

=== Data parameters ===

Training data: GEO6956.txt

Test data: None specified, training data used!

=== Learner parameters ===

Discretization method: EBD using lambda of 0.5 (2009)

Learning method: Knowledge Augmented Rule Learning (KARL)

Transfer of prior rules via gene function evidence:

Lookup Table File: /Users/mahbaneheshaghzadehtorbati/JavaWorkspace/Original-MS-TRL/Demo-data/GEO6956\_LTable.csv

No. Gene Function Associations: 47

No. Filtered Gene Function Associations: 47

No. Functional Genes: 29

No. Rules induced from FLTable: 48

Min conjuncts: 1

Max conjuncts: 5

Inference type: Weighted voting (0)

Beam width: 2500

CF function: Laplace Normalized (7)

Min CF: 0.8

Min conjuncts: 1

Max conjuncts: 5

Min coverage: 4.0

Max FP coverage: 0.1

Min TP coverage: 0.05

Inductive strengthening: 1

Specialize satisfactory rules: false

Validation method: training set

=== Rules (11) ===

Pr 1. ((CCNL2 = -inf..8.587) (DVL3 = -inf..7.668)) ==> (@Class = CONTROL)

CF=0.929, PV=4.25978E-10, TP=16, FP=1, Pos=20, Neg=69, TP\_in\_model=16, FP\_in\_model=1

Pr 2. ((DPT = 7.481..inf) (FKBP5 = 6.377..inf)) ==> (@Class = CASE)

CF=0.923, PV=1.52847E-1, TP=11, FP=0, Pos=69, Neg=20, TP\_in\_model=11, FP\_in\_model=0

Pr 3. ((CCNL2 = -inf..8.587)) ==> (@Class = CONTROL)

CF=0.915, PV=2.71599E-9, TP=16, FP=2, Pos=20, Neg=69, TP\_in\_model=16, FP\_in\_model=2

Pr 4. ((DVL3 = -inf..7.668) (SIM2 = -inf..6.414)) ==> (@Class = CONTROL)

CF=0.903, PV=4.74557E-7, TP=11, FP=1, Pos=20, Neg=69, TP\_in\_model=11, FP\_in\_model=1

Pr 5. ((DVL3 = -inf..7.668) (SH3GLB2 = -inf..6.767)) ==> (@Class = CONTROL)

CF=0.899, PV=1.58357E-7, TP=13, FP=2, Pos=20, Neg=69, TP\_in\_model=13, FP\_in\_model=2

Pr 6. ((CCNL2 = 8.587..inf) (PRKCD = 7.905..inf)) ==> (@Class = CASE)

CF=0.88, PV=1.35814E-4, TP=57, FP=2, Pos=69, Neg=20, TP\_in\_model=56, FP\_in\_model=2

Pr 7. ((CSTA = 9.790..inf)) ==> (@Class = CONTROL)

CF=0.875, PV=2.9352E-5, TP=8, FP=1, Pos=20, Neg=69, TP\_in\_model=8, FP\_in\_model=1

Pr 8. ((ERBB3 = -inf..8.351) (TRPM4 = -inf..7.222)) ==> (@Class = CONTROL)

CF=0.874, PV=8.49616E-6, TP=10, FP=2, Pos=20, Neg=69, TP\_in\_model=10, FP\_in\_model=1

Pr 9. ((DVL3 = -inf..7.668)) ==> (@Class = CONTROL)

CF=0.874, PV=1.64516E-7, TP=16, FP=5, Pos=20, Neg=69, TP\_in\_model=16, FP\_in\_model=5

Pr 10. ((CCNL2 = 8.587..inf) (FKBP5 = 6.377..inf)) ==> (@Class = CASE)

CF=0.866, PV=6.09387E-4, TP=50, FP=2, Pos=69, Neg=20, TP\_in\_model=48, FP\_in\_model=2

Pr 11. ((DPT = 7.481..inf) (EHF = -inf..7.202) (FKBP5 = -inf..6.377)) ==> (@Class = CONTROL)

CF=0.826, PV=4.91385E-5, TP=12, FP=6, Pos=20, Neg=69, TP\_in\_model=12, FP\_in\_model=4

Attributes used (11):

CSTA, EHF, DPT, ERBB3, DVL3, CCNL2, TRPM4, SH3GLB2, PRKCD, SIM2, FKBP5

=== Classification performance on training data ===

Accuracy: 87.3563 %

Accuracy including abst.: 85.3933 %

Abstentions: 2 (2.2472 %)

Relative classifier information: 41.6214

Class Sensitivity(%) Specificity(%) Balanced Accuracy(%) AUROC

------------ -------------- -------------- -------------------- --------------

CASE 88.0597 85 86.5299 0

CONTROL 85 88.0597 86.5299 0

Confusion matrix:

c1 c2 Abstentions

59 8 2 | c1 = CASE

3 17 0 | c2 = CONTROL

Total running time: 27 s.