Feature

Though unit based feature (number of features = 3+6*38):

• Overall features
  Number of total thought units
  Number of thought units of each annotation
  Percentage of thought units of each annotation
• Wine features
  Number of total thought units
  Number of thought units of each annotation
  Percentage of thought units of each annotation
• Grocery features
  Number of total thought units
  Number of thought units of each annotation
  Percentage of thought units of each annotation

Meta data feature (number of features = 9):

mean_age, negative_experience, education, east.asian.time.in.us, cultural_intelligence, emotional.intelligence, IQ, extroversion, openness

Data conversion

1. Missing data, i.e. 'NA' in meta_fv is replaced with '?' for representing missing value in weka
2. In weka SimpleCLT, the following command would convert .csv file to .arff file:
   java weka.core.converters.CSVLoader C:\Users\pengye\course\trunk\results\baseline_fv.csv > C:\Users\pengye\course\trunk\results\baseline_fv.arff

in weka, you may need to first remove, 'ID' and 'profit.diff' attributes and use the following attributes for regression and choose 'profit.join' to be the regression target

Regression Result

In weka explorer, use weka.classifiers.rules.M5Rules -M 4.0 with 10 fold cross-validation

=== Run information ===

Scheme: weka.classifiers.rules.M5Rules -M 4.0

Relation: baseline_fv-weka.filters.unsupervised.attribute.Remove-R228,238

Instances: 61

Attributes: 242

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

M5 pruned model rules

(using smoothed linear models) :

Number of Rules : 2

Rule: 1

IF

       percentage-of-(OM)-wine > 0.009

       number-of-(IR)-overall > 4.5

       number-of-(IR)-overall > 7.5

THEN

profit.joint =

       695.462 * percentage-of-(QM)-overall

       + 478.7082 * percentage-of-(MIC)-wine

       - 190.037 * percentage-of-(IR)-grocery

       + 131.1652 * percentage-of-(OM)-wine

       + 14.2952 * number-of-(IR)-overall

       + 307.5222 * percentage-of-(PP)-overall

       - 2439.1847 * percentage-of-(SF)-overall

       + 16.0215 * openness

       + 1205.8784 [8/0%]

Rule: 2

profit.joint =

       821.6545 * percentage-of-(QM)-overall

+ 330.9833 * percentage-of-(MIC)-wine

- 32.8251 * number-of-(SF)-overall

+ 18.8568 * number-of-(OM)-grocery

+ 1821.936 * percentage-of-(PO)-overall

+ 1193.0132 [53/63.704%]

Time taken to build model: 0.18 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient   0.5527

Mean absolute error   72.7324

Root mean squared error   91.944

Relative absolute error   81.4745 %

Root relative squared error   87.5159 %

Total Number of Instances   61