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

Method 1:

=== 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

Method 2:

=== Run information ===

Scheme: weka.classifiers.meta.Bagging -P 100 -S 1 -I 10 -W weka.classifiers.trees.REPTree -- -M 2 -V 0.0010 -N 3 -S 1 -L -1

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

Instances: 61

Attributes: 242

[list of attributes omitted]

Test mode: 10-fold cross-validation

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

All the base classifiers:

REPTree

==========

percentage-of-(IR)-overall < 0.02

| percentage-of-(PP)-overall < 0.01

| | percentage-of-(IP)-grocery < 0.09

| | | percentage-of-(PT)-overall < 0 : 1320 (2/0) [2/0]

| | | percentage-of-(PT)-overall >= 0 : 1240 (2/0) [1/0]

| | percentage-of-(IP)-grocery >= 0.09 : 1170 (7/0) [1/6400]

| percentage-of-(PP)-overall >= 0.01 : 1308.15 (12/766.67) [5/3140]

percentage-of-(IR)-overall >= 0.02

| number-of-(P1)-grocery < 3

| | number-of-(QR)-overall < 1.5 : 1385.95 (8/0) [3/8533.33]

| | number-of-(QR)-overall >= 1.5 : 1462.46 (7/0) [9/6088.89]

| number-of-(P1)-grocery >= 3 : 1240 (2/6400) [0/0]

Size of the tree : 13

REPTree

==========

number-of-(IR)-overall < 4.5

| percentage-of-(PO)-overall < 0.03

| | number-of-(PP)-grocery < 0.5 : 1193.86 (8/1200) [4/6000]

| | number-of-(PP)-grocery >= 0.5 : 1299.66 (9/1580.25) [8/8375.31]

| percentage-of-(PO)-overall >= 0.03 : 1350.41 (10/4736) [2/1744]

number-of-(IR)-overall >= 4.5

| percentage-of-(QR)-overall < 0.01 : 1370 (7/0) [1/57600]

| percentage-of-(QR)-overall >= 0.01 : 1472.8 (6/0) [6/1066.67]

Size of the tree : 9

REPTree

==========

number-of-(IR)-grocery < 1.5

| percentage-of-(PP)-overall < 0.01

| | percentage-of-(QS)-wine < 0.03 : 1269.94 (5/0) [4/8000]

| | percentage-of-(QS)-wine >= 0.03 : 1160 (5/0) [1/0]

| percentage-of-(PP)-overall >= 0.01

| | percentage-of-(PP)-wine < 0.01 : 1430 (3/1422.22) [2/2844.44]

| | percentage-of-(PP)-wine >= 0.01

| | | emotional.intelligence < 3.73 : 1334.92 (8/0) [3/4266.67]

| | | emotional.intelligence >= 3.73

| | | | east.asian.time.in.us < 50.5 : 1290.37 (2/0) [3/2533.33]

| | | | east.asian.time.in.us >= 50.5 : 1240 (4/0) [1/0]

number-of-(IR)-grocery >= 1.5

| percentage-of-(IR)-wine < 0.03 : 1364.38 (6/888.89) [4/3377.78]

| percentage-of-(IR)-wine >= 0.03 : 1462.34 (7/783.67) [3/1654.42]

Size of the tree : 15

REPTree

==========

percentage-of-(OS)-overall < 0.03

| number-of-(IDN)-overall < 1.5 : 1373.2 (8/4075) [7/6739.29]

| number-of-(IDN)-overall >= 1.5 : 1461.25 (7/0) [4/3200]

percentage-of-(OS)-overall >= 0.03 : 1279.96 (25/5647.36) [10/9163.84

Size of the tree : 5

REPTree

==========

number-of-(IR)-overall < 4.5

| number-of-(OM)-grocery < 1.5

| | percentage-of-(P1)-overall < 0.03

| | | percentage-of-(PP)-overall < 0 : 1160 (3/0) [1/0]

| | | percentage-of-(PP)-overall >= 0 : 1236.3 (8/43.75) [2/3556.25]

| | percentage-of-(P1)-overall >= 0.03 : 1308.57 (6/0) [1/6400]

| number-of-(OM)-grocery >= 1.5 : 1332.54 (11/2095.87) [9/4534.44]

number-of-(IR)-overall >= 4.5 : 1445.71 (12/1422.22) [8/1777.78]

Size of the tree : 9

REPTree

==========

percentage-of-(IR)-overall < 0.02 : 1278.02 (22/5044.63) [10/9653.55]

percentage-of-(IR)-overall >= 0.02 : 1401.85 (18/6577.78) [11/7729.29]

Size of the tree : 3

REPTree

==========

percentage-of-(OM)-overall < 0.01

| percentage-of-(PP)-overall < 0 : 1167.72 (5/0) [4/1600]

| percentage-of-(PP)-overall >= 0 : 1288.13 (7/1567.35) [3/5746.94]

percentage-of-(OM)-overall >= 0.01

| percentage-of-(SF)-grocery < 0.01

| | percentage-of-(IP)-overall < 0.11 : 1442.04 (11/1480.99) [7/3672.26]

| | percentage-of-(IP)-overall >= 0.11

| | | percentage-of-(OT)-wine < 0.02 : 1341.22 (6/888.89) [2/2311.11]

| | | percentage-of-(OT)-wine >= 0.02 : 1418.75 (7/783.67) [2/2416.33]

| percentage-of-(SF)-grocery >= 0.01 : 1285.44 (4/1600) [3/22933.33]

Size of the tree : 11

REPTree

==========

percentage-of-(IR)-overall < 0.05 : 1329.29 (28/7128.57) [19/13152.63]

percentage-of-(IR)-overall >= 0.05

| percentage-of-(RP)-overall < 0.04 : 1480 (9/0) [2/0]

| percentage-of-(RP)-overall >= 0.04 : 1400 (3/0) [0/0]

Size of the tree : 5

REPTree

==========

percentage-of-(SBR)-wine < 0.07 : 1456.6 (7/1306.12) [4/1208.16]

percentage-of-(SBR)-wine >= 0.07

| percentage-of-(RP)-wine < 0.02

| | percentage-of-(INP)-overall < 0.02

| | | number-of-(P1)-grocery < 0.5 : 1480 (2/0) [1/0]

| | | number-of-(P1)-grocery >= 0.5 : 1352.65 (6/500) [3/10766.67]

| | percentage-of-(INP)-overall >= 0.02 : 1280 (2/900) [1/900]

| percentage-of-(RP)-wine >= 0.02

| | number-of-(IP)-overall < 18.5

| | | number-of-though-unit < 80.5 : 1220 (3/1422.22) [1/711.11]

| | | number-of-though-unit >= 80.5 : 1310.85 (14/424.49) [7/8130.61]

| | number-of-(IP)-overall >= 18.5

| | | number-of-(PP)-grocery < 0.5 : 1160 (4/0) [2/0]

| | | number-of-(PP)-grocery >= 0.5 : 1320 (2/0) [2/28800]

Size of the tree : 15

REPTree

==========

percentage-of-(OM)-overall < 0.01 : 1248.94 (15/5518.22) [6/10055.11]

percentage-of-(OM)-overall >= 0.01

| number-of-(PO)-overall < 4.5

| | percentage-of-(IR)-grocery < 0.02

| | | percentage-of-(QM)-overall < 0.03 : 1240 (2/0) [1/0]

| | | percentage-of-(QM)-overall >= 0.03 : 1318.75 (7/783.67) [2/4244.9]

| | percentage-of-(IR)-grocery >= 0.02 : 1389.46 (8/700) [7/2385.71]

| number-of-(PO)-overall >= 4.5 : 1468.9 (8/700) [5/1060]

Size of the tree : 9

Time taken to build model: 0.1 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.6619
Mean absolute error	67.7539
Root mean squared error	81.4241
Relative absolute error	75.8977 %
Root relative squared error	77.5026 %

Total Number of Instances	61