

附录:

附录1: EM聚类结果

```
Scheme:      weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-
6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Relation:    802_ENGLISH_XLS1re
Instances:   802
Attributes:  34
            YEAR_IN
            RECRUITING_WAY
            STYPE
            MAJOR
            NATION
            SEX
            SAGE_IN
            CPP_MEMBER
            INTERVIEW_SCORE_01
            INTERVIEW_SCORE_CLASS
            985_211
            SAGE_OUT
            CONTINUOUS_EDUCATION
            CHANGE_MAJOR
            TAGE_SIN
            TCPP_MEMBER
            TTITLE
            TDUTY
            EXPERT
            CREDIT
            GPA
            GPA_CLASS
            GPA_01
            GPA_MAIN
            GPA_CLASS_MAIN
            GPA_MAIN_01
            PROPOSAL_GRADE
            MID_GRADE
            JUDGING_GRADE
            ANSWERING_GRADE
            TIME_TO_PROPOSAL
            EMPLOYMENT_UNIT_CLASS
            DALEY_YEAR
            DELAY
Test mode:   evaluate on training data
```

=== Clustering model (full training set) ===

EM

==

Number of clusters selected by cross validation: 3

Number of iterations performed: 2

Attribute	Cluster		
	0 (0.2)	1 (0.24)	2 (0.56)
=====			
YEAR_IN			
mean	2011.2296	2010.9417	2011.5848
std. dev.	1.1706	0.9295	1.1193
RECRUITING_WAY			
MASTER_DOCTOR	7.7421	127.7561	7.5018
COMMON_EXAM	1	59	1
UNIFIED_EXAM	97.4092	2.6508	111.94
EXAM_FREE	61.7033	4.5544	330.7423
BACHELOR_DOCTOR	1	2	1
[total]	168.8546	195.9613	452.1841
STYPE			
DOCTOR	7.7421	186.7561	7.5018
MASTER	70.0039	5.5526	382.4435
MASTER_SPECIALIZED	89.1086	1.6526	60.2388
[total]	166.8546	193.9613	450.1841
MAJOR			
CA	36.4836	100.2553	188.261
CAT	26.1362	51.7833	119.0806
CST	15.038	30.2701	73.6918
IS	2.0882	12	10.9118
CT	70.5063	1.079	39.4147
SE	19.6023	1.5736	21.8241
[total]	169.8546	196.9613	453.1841
NATION			
HAN	158.2915	183.9591	429.7494
OTHER	7.5631	9.0022	19.4347
[total]	165.8546	192.9613	449.1841
SEX			
MALE	148.7906	159.6152	343.5941
FEMALE	17.064	33.346	105.59
[total]	165.8546	192.9613	449.1841
SAGE_IN			
mean	22.515	24.6541	22.1784
std. dev.	1.1951	2.0992	0.925
CPP_MEMBER			
NOCPP	104.1616	76.0067	193.8316
CPP	61.693	116.9545	255.3525
[total]	165.8546	192.9613	449.1841
INTERVIEW_SCORE_01			
mean	-0.1102	0.0091	0.0484
std. dev.	1.0135	0.9307	1.0109

INTERVIEW_SCORE_CLASS			
B	21.1854	72.2403	84.5743
D	50.9437	50.0359	107.0205
C	58.176	34.7968	141.0273
A	37.5496	37.8884	118.562
[total]	167.8546	194.9613	451.1841
985_211			
9_8_5	96.6514	99.4826	309.866
OTHER	35.3951	51.525	37.0798
2_1_1	34.808	42.9537	103.2383
[total]	166.8546	193.9613	450.1841
SAGE_OUT			
mean	25.435	28.626	24.9984
std. dev.	1.2062	2.3349	0.9093
CONTINUOUS_EDUCATION			
CONTINUOUS	133.8854	158.3732	428.7413
DISCONTINUOUS	31.9692	34.588	20.4428
[total]	165.8546	192.9613	449.1841
CHANGE_MAJOR			
UNCHANGED	105.4979	109.5441	303.958
CHANGED	60.3567	83.4172	145.2261
[total]	165.8546	192.9613	449.1841
TAGE_SIN			
mean	43.1777	44.4967	40.0483
std. dev.	13.6823	8.3122	5.9897
TCPP_MEMBER			
TCPP	91.6771	150.222	269.1009
NOTCPP	74.1775	42.7393	180.0832
[total]	165.8546	192.9613	449.1841
TTITLE			
SENIOR	102.5425	185.8005	272.6569
SUBSENIOR	50.0146	7.1604	167.825
OTHER	14.2975	1.0004	9.7022
[total]	166.8546	193.9613	450.1841
TDUTY			
DUTYIN	48.7893	79.125	149.0857
DUTYFREE	117.0653	113.8363	300.0984
[total]	165.8546	192.9613	449.1841
EXPERT			
EXPERT	52.0532	116.7195	117.2273
NOEXPERT	113.8014	76.2418	331.9569
[total]	165.8546	192.9613	449.1841
CREDIT			
mean	34.3489	26.545	33.0282
std. dev.	2.2453	15.2697	1.7986
GPA			
mean	77.3482	79.2409	82.8485
std. dev.	3.3973	5.4611	2.6079
GPA_CLASS			
B	20.62	50.6042	135.7758

A	6.7317	52.734	153.5344
D	105.4379	48.6983	46.8639
C	35.0651	42.9248	115.0101
[total]	167.8546	194.9613	451.1841
GPA_01			
mean	-0.7582	0.0611	0.3559
std. dev.	0.8414	0.9836	0.6591
GPA_MAIN			
mean	76.8985	78.6537	82.4033
std. dev.	3.6179	5.2573	2.8111
GPA_CLASS_MAIN			
B	18.0386	46.0255	135.9359
A	9.1783	53.2626	153.5591
D	97.0824	48.2793	47.6384
C	43.5554	47.3939	114.0507
[total]	167.8546	194.9613	451.1841
GPA_MAIN_01			
mean	-0.7402	0.0765	0.3599
std. dev.	0.8656	0.9613	0.7066
PROPOSAL_GRADE			
B	92.0319	64.1692	122.7989
A	70.6796	119.8097	322.5107
C	4.1431	9.9824	4.8745
[total]	166.8546	193.9613	450.1841
MID_GRADE			
B	81.0829	40.124	112.793
A	80.0725	149.8372	336.0902
C	5.6991	4	1.3009
[total]	166.8546	193.9613	450.1841
JUDGING_GRADE			
B	100.5045	122.2799	151.2155
A	65.3501	70.6814	297.9686
[total]	165.8546	192.9613	449.1841
ANSWERING_GRADE			
A	94.7731	106.2012	262.0257
B	71.0815	86.7601	187.1584
[total]	165.8546	192.9613	449.1841
TIME_TO_PROPOSAL			
mean	23.833	26.5267	24.2882
std. dev.	5.4611	9.4187	3.789
EMPLOYMENT_UNIT_CLASS			
NO_EDUCATION_RESEARCH	136.592	95.345	372.063
EDUCATION_RESEARCH	29.2626	97.6163	77.1211
[total]	165.8546	192.9613	449.1841
DALEY_YEAR			
mean	0.063	1.1059	0
std. dev.	0.2211	0.8564	0
DELAY			
DELAY	15.8249	157.1751	1

```

UNDELAY          150.0297    35.7862   448.1841
[total]          165.8546   192.9613   449.1841

```

Time taken to build model (full training data) : 9.09 seconds

=== Model and evaluation on training set ===

Clustered Instances

```

0          93 ( 12%)
1         178 ( 22%)
2         531 ( 66%)

```

Log likelihood: -33.64018

附录2： 贝叶斯分类结果

=== Run information ===

```

Scheme:          weka.classifiers.bayes.BayesNet -D -Q
weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E
weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
Relation:        802_ENGLISH_XLS1re
Instances:       802
Attributes:      34
.....

```

Test mode: 10-fold cross-validation

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

```

Bayes Network Classifier
not using ADTree
#attributes=34 #classindex=31
Network structure (nodes followed by parents)
YEAR_IN(1): EMPLOYMENT_UNIT_CLASS
RECRUITING_WAY(5): EMPLOYMENT_UNIT_CLASS
STYPE(3): EMPLOYMENT_UNIT_CLASS
MAJOR(6): EMPLOYMENT_UNIT_CLASS
NATION(2): EMPLOYMENT_UNIT_CLASS
SEX(2): EMPLOYMENT_UNIT_CLASS
SAGE_IN(2): EMPLOYMENT_UNIT_CLASS
CPP_MEMBER(2): EMPLOYMENT_UNIT_CLASS
INTERVIEW_SCORE_01(1): EMPLOYMENT_UNIT_CLASS
INTERVIEW_SCORE_CLASS(4): EMPLOYMENT_UNIT_CLASS
985_211(3): EMPLOYMENT_UNIT_CLASS
SAGE_OUT(2): EMPLOYMENT_UNIT_CLASS
CONTINUOUS_EDUCATION(2): EMPLOYMENT_UNIT_CLASS

```

```

CHANGE_MAJOR(2): EMPLOYMENT_UNIT_CLASS
TAGE_SIN(1): EMPLOYMENT_UNIT_CLASS
TCPP_MEMBER(2): EMPLOYMENT_UNIT_CLASS
TTITLE(3): EMPLOYMENT_UNIT_CLASS
TDUTY(2): EMPLOYMENT_UNIT_CLASS
EXPERT(2): EMPLOYMENT_UNIT_CLASS
CREDIT(3): EMPLOYMENT_UNIT_CLASS
GPA(2): EMPLOYMENT_UNIT_CLASS
GPA_CLASS(4): EMPLOYMENT_UNIT_CLASS
GPA_01(1): EMPLOYMENT_UNIT_CLASS
GPA_MAIN(1): EMPLOYMENT_UNIT_CLASS
GPA_CLASS_MAIN(4): EMPLOYMENT_UNIT_CLASS
GPA_MAIN_01(1): EMPLOYMENT_UNIT_CLASS
PROPOSAL_GRADE(3): EMPLOYMENT_UNIT_CLASS
MID_GRADE(3): EMPLOYMENT_UNIT_CLASS
JUDGING_GRADE(2): EMPLOYMENT_UNIT_CLASS
ANSWERING_GRADE(2): EMPLOYMENT_UNIT_CLASS
TIME_TO_PROPOSAL(1): EMPLOYMENT_UNIT_CLASS
EMPLOYMENT_UNIT_CLASS(2):
DALEY_YEAR(2): EMPLOYMENT_UNIT_CLASS
DELAY(2): EMPLOYMENT_UNIT_CLASS
LogScore Bayes: -16390.47907741949
LogScore BDeu: -16519.331041720074
LogScore MDL: -16528.285518975586
LogScore ENTROPY: -16224.02207731766
LogScore AIC: -16315.02207731766

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	603	75.187 %
Incorrectly Classified Instances	199	24.813 %
Kappa statistic	0.3339	
Mean absolute error	0.2683	
Root mean squared error	0.4803	
Relative absolute error	71.3496 %	
Root relative squared error	110.8278 %	
Total Number of Instances	802	

=== Detailed Accuracy By Class ===

	ROC Area	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
PRC Area		Class					
		0.839	0.507	0.832	0.839	0.835	0.334
0.672	0.822	NO_EDUCATION_RESEARCH					
		0.493	0.161	0.505	0.493	0.499	0.334
0.672	0.439	EDUCATION_RESEARCH					
Weighted Avg.		0.752	0.421	0.750	0.752	0.751	0.334
0.672	0.726						

=== Confusion Matrix ===

```

      a    b    <-- classified as
504  97 |    a = NO_EDUCATION_RESEARCH
102  99 |    b = EDUCATION_RESEARCH

```

但是.....贝叶斯的结果真的要多简洁有多简洁/滑稽

附录3: J48分类结果

```
=== Run information ===
```

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    802_ENGLISH_XLS1re
Instances:   802
Attributes:  34
             YEAR_IN
             RECRUITING_WAY
             STYPE
             MAJOR
             NATION
             SEX
             SAGE_IN
             CPP_MEMBER
             INTERVIEW_SCORE_01
             INTERVIEW_SCORE_CLASS
             985_211
             SAGE_OUT
             CONTINUOUS_EDUCATION
             CHANGE_MAJOR
             TAGE_SIN
             TCPP_MEMBER
             TTITLE
             TDUTY
             EXPERT
             CREDIT
             GPA
             GPA_CLASS
             GPA_01
             GPA_MAIN
             GPA_CLASS_MAIN
             GPA_MAIN_01
             PROPOSAL_GRADE
             MID_GRADE
             JUDGING_GRADE
             ANSWERING_GRADE
             TIME_TO_PROPOSAL
             EMPLOYMENT_UNIT_CLASS
             DALEY_YEAR
             DELAY

```

Test mode: 10-fold cross-validation

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

J48 pruned tree

CREDIT <= 9

```

|   JUDGING_GRADE = B
|   |   CHANGE_MAJOR = UNCHANGED: NO_EDUCATION_RESEARCH (5.0/1.0)
|   |   CHANGE_MAJOR = CHANGED
|   |   |   PROPOSAL_GRADE = B
|   |   |   |   ANSWERING_GRADE = A
|   |   |   |   |   TCPP_MEMBER = TCPP: NO_EDUCATION_RESEARCH (4.0/1.0)
|   |   |   |   |   TCPP_MEMBER = NOTCPP: EDUCATION_RESEARCH (5.0/1.0)
|   |   |   |   |   ANSWERING_GRADE = B: EDUCATION_RESEARCH (4.0)
|   |   |   |   PROPOSAL_GRADE = A
|   |   |   |   |   CPP_MEMBER = NOCPP: NO_EDUCATION_RESEARCH (5.0/2.0)
|   |   |   |   |   CPP_MEMBER = CPP
|   |   |   |   |   TDUTY = DUTYIN
|   |   |   |   |   |   SAGE_OUT <= 32: EDUCATION_RESEARCH (4.0/1.0)
|   |   |   |   |   |   SAGE_OUT > 32: NO_EDUCATION_RESEARCH (2.0)
|   |   |   |   |   TDUTY = DUTYFREE: EDUCATION_RESEARCH (7.0)
|   |   |   |   PROPOSAL_GRADE = C: NO_EDUCATION_RESEARCH (5.0/1.0)
|   |   JUDGING_GRADE = A: EDUCATION_RESEARCH (17.0/1.0)

```

CREDIT > 9

```

|   DELAY = DELAY
|   |   985_211 = 9_8_5
|   |   |   TIME_TO_PROPOSAL <= 31
|   |   |   |   MAJOR = CA
|   |   |   |   |   GPA_CLASS = B: EDUCATION_RESEARCH (9.0/1.0)
|   |   |   |   |   GPA_CLASS = A
|   |   |   |   |   |   JUDGING_GRADE = B
|   |   |   |   |   |   |   INTERVIEW_SCORE_01 <= -0.089143:
|   |   |   |   |   |   |   EDUCATION_RESEARCH (2.0)
|   |   |   |   |   |   |   |   INTERVIEW_SCORE_01 > -0.089143:
|   |   |   |   |   |   |   |   NO_EDUCATION_RESEARCH (4.0)
|   |   |   |   |   |   |   |   |   JUDGING_GRADE = A: EDUCATION_RESEARCH (4.0)
|   |   |   |   |   |   |   |   GPA_CLASS = D
|   |   |   |   |   |   |   |   SAGE_IN <= 24: EDUCATION_RESEARCH (3.0/1.0)
|   |   |   |   |   |   |   |   SAGE_IN > 24: NO_EDUCATION_RESEARCH (3.0)
|   |   |   |   |   |   |   GPA_CLASS = C
|   |   |   |   |   |   |   |   YEAR_IN <= 2010: NO_EDUCATION_RESEARCH (2.0)
|   |   |   |   |   |   |   |   YEAR_IN > 2010: EDUCATION_RESEARCH (4.0/1.0)
|   |   |   |   |   |   MAJOR = CAT
|   |   |   |   |   |   |   CPP_MEMBER = NOCPP: NO_EDUCATION_RESEARCH (5.0)
|   |   |   |   |   |   |   CPP_MEMBER = CPP: EDUCATION_RESEARCH (13.0/3.0)
|   |   |   |   |   |   MAJOR = CST: EDUCATION_RESEARCH (5.0/2.0)
|   |   |   |   |   |   MAJOR = IS: NO_EDUCATION_RESEARCH (3.0/1.0)
|   |   |   |   |   |   MAJOR = CT: NO_EDUCATION_RESEARCH (1.0)
|   |   |   |   |   |   MAJOR = SE: EDUCATION_RESEARCH (0.0)
|   |   |   |   |   TIME_TO_PROPOSAL > 31: NO_EDUCATION_RESEARCH (16.0/2.0)
|   |   985_211 = OTHER
|   |   |   CHANGE_MAJOR = UNCHANGED: NO_EDUCATION_RESEARCH (12.0)

```



```

| | | CHANGE_MAJOR = CHANGED
| | | | TAGE_SIN <= 39: EDUCATION_RESEARCH (2.0)
| | | | TAGE_SIN > 39: NO_EDUCATION_RESEARCH (3.0)
| | 985_211 = 2_1_1
| | | INTERVIEW_SCORE_CLASS = B: EDUCATION_RESEARCH (11.0/4.0)
| | | INTERVIEW_SCORE_CLASS = D: NO_EDUCATION_RESEARCH (12.0/1.0)
| | | INTERVIEW_SCORE_CLASS = C: NO_EDUCATION_RESEARCH (3.0)
| | | INTERVIEW_SCORE_CLASS = A: EDUCATION_RESEARCH (2.0)
| DELAY = UNDELAY
| | STYPE = DOCTOR
| | | MID_GRADE = B: NO_EDUCATION_RESEARCH (4.0)
| | | MID_GRADE = A
| | | | TTITLE = SENIOR
| | | | GPA_CLASS = B
| | | | SAGE_OUT <= 27
| | | | SEX = MALE: EDUCATION_RESEARCH (5.0)
| | | | SEX = FEMALE: NO_EDUCATION_RESEARCH
(3.0/1.0)
| | | | SAGE_OUT > 27: NO_EDUCATION_RESEARCH (3.0)
| | | | GPA_CLASS = A
| | | | TDUTY = DUTYIN
| | | | INTERVIEW_SCORE_01 <= -0.572356:
EDUCATION_RESEARCH (3.0)
| | | | INTERVIEW_SCORE_01 > -0.572356:
NO_EDUCATION_RESEARCH (2.0)
| | | | TDUTY = DUTYFREE: EDUCATION_RESEARCH (5.0)
| | | | GPA_CLASS = D
| | | | CREDIT <= 35.5: EDUCATION_RESEARCH (2.0)
| | | | CREDIT > 35.5: NO_EDUCATION_RESEARCH (4.0)
| | | | GPA_CLASS = C: NO_EDUCATION_RESEARCH (8.0/1.0)
| | | | TTITLE = SUBSENIOR: EDUCATION_RESEARCH (3.0)
| | | | TTITLE = OTHER: EDUCATION_RESEARCH (0.0)
| | | | MID_GRADE = C: NO_EDUCATION_RESEARCH (0.0)
| | | STYPE = MASTER: NO_EDUCATION_RESEARCH (443.0/66.0)
| | | STYPE = MASTER_SPECIALIZED
| | | | JUDGING_GRADE = B
| | | | TIME_TO_PROPOSAL <= 24
| | | | RECRUITING_WAY = MASTER_DOCTOR:
NO_EDUCATION_RESEARCH (0.0)
| | | | RECRUITING_WAY = COMMON_EXAM: NO_EDUCATION_RESEARCH
(0.0)
| | | | RECRUITING_WAY = UNIFIED_EXAM
| | | | | NATION = HAN
| | | | | TTITLE = SENIOR
| | | | | GPA <= 75.46: EDUCATION_RESEARCH (4.0)
| | | | | GPA > 75.46: NO_EDUCATION_RESEARCH
(11.0/1.0)
| | | | TTITLE = SUBSENIOR
| | | | | CHANGE_MAJOR = UNCHANGED:
EDUCATION_RESEARCH (4.0)
| | | | | CHANGE_MAJOR = CHANGED:
NO_EDUCATION_RESEARCH (2.0)
| | | | TTITLE = OTHER: NO_EDUCATION_RESEARCH
(8.0/1.0)

```

```

| | | | | | NATION = OTHER: NO_EDUCATION_RESEARCH (2.0)
| | | | | | RECRUITING_WAY = EXAM_FREE: NO_EDUCATION_RESEARCH
(8.0)
| | | | | | RECRUITING_WAY = BACHELOR_DOCTOR:
NO_EDUCATION_RESEARCH (0.0)
| | | | | | TIME_TO_PROPOSAL > 24: NO_EDUCATION_RESEARCH (38.0/1.0)
| | | | | | JUDGING_GRADE = A
| | | | | | ANSWERING_GRADE = A: NO_EDUCATION_RESEARCH (37.0/6.0)
| | | | | | ANSWERING_GRADE = B
| | | | | | PROPOSAL_GRADE = B: NO_EDUCATION_RESEARCH (7.0/1.0)
| | | | | | PROPOSAL_GRADE = A
| | | | | | SAGE_OUT <= 26
| | | | | | MAJOR = CA: EDUCATION_RESEARCH (0.0)
| | | | | | MAJOR = CAT: EDUCATION_RESEARCH (0.0)
| | | | | | MAJOR = CST: EDUCATION_RESEARCH (0.0)
| | | | | | MAJOR = IS: EDUCATION_RESEARCH (0.0)
| | | | | | MAJOR = CT: EDUCATION_RESEARCH (11.0/2.0)
| | | | | | MAJOR = SE
| | | | | | TTITLE = SENIOR: NO_EDUCATION_RESEARCH
(3.0)
| | | | | | TTITLE = SUBSENIOR: EDUCATION_RESEARCH
(2.0)
| | | | | | TTITLE = OTHER: NO_EDUCATION_RESEARCH
(0.0)
| | | | | | SAGE_OUT > 26: NO_EDUCATION_RESEARCH (2.0)
| | | | | | PROPOSAL_GRADE = C: NO_EDUCATION_RESEARCH (1.0)

```

Number of Leaves : 69

Size of the tree : 112

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	596	74.3142 %
Incorrectly Classified Instances	206	25.6858 %
Kappa statistic	0.2236	
K&B Relative Info Score	9426.9747 %	
K&B Information Score	76.6847 bits	0.0956
bits/instance		
Class complexity order 0	651.4445 bits	0.8123
bits/instance		
Class complexity scheme	59601.1647 bits	74.3157
bits/instance		
Complexity improvement (Sf)	-58949.7202 bits	-73.5034
bits/instance		
Mean absolute error	0.3209	
Root mean squared error	0.4644	
Relative absolute error	85.3476 %	
Root relative squared error	107.1615 %	
Total Number of Instances	802	

=== Detailed Accuracy By Class ===

ROC Area	PRC Area	TP Rate Class	FP Rate	Precision	Recall	F-Measure	MCC
0.563	0.758	0.889 NO_EDUCATION_RESEARCH	0.692	0.793	0.889	0.838	0.232
0.563	0.349	0.308 EDUCATION_RESEARCH	0.111	0.481	0.308	0.376	0.232
Weighted Avg.		0.743	0.546	0.715	0.743	0.722	0.232
0.563	0.656						

=== Confusion Matrix ===

```

a   b   <-- classified as
534 67 |   a = NO_EDUCATION_RESEARCH
139 62 |   b = EDUCATION_RESEARCH

```

附录4：决策树桩分类结果

=== Run information ===

```

Scheme:      weka.classifiers.trees.DecisionStump
Relation:     802_ENGLISH_XLS1re
Instances:    802
Attributes:   34

```

```
Test mode:    10-fold cross-validation
```

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

```
Decision Stump
```

```
Classifications
```

```

STYPE = DOCTOR : EDUCATION_RESEARCH
STYPE != DOCTOR : NO_EDUCATION_RESEARCH
STYPE is missing : NO_EDUCATION_RESEARCH

```

```
Class distributions
```

```

STYPE = DOCTOR
NO_EDUCATION_RESEARCH   EDUCATION_RESEARCH
0.49246231155778897     0.507537688442211
STYPE != DOCTOR
NO_EDUCATION_RESEARCH   EDUCATION_RESEARCH
0.8341625207296849     0.16583747927031509
STYPE is missing

```

```
NO_EDUCATION_RESEARCH      EDUCATION_RESEARCH
0.7493765586034913         0.2506234413965087

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      595           74.1895 %
Incorrectly Classified Instances    207           25.8105 %
Kappa statistic                    0.2522
Mean absolute error                 0.3325
Root mean squared error             0.4081
Relative absolute error             88.4299 %
Root relative squared error         94.1648 %
Total Number of Instances          802
```

```
=== Detailed Accuracy By Class ===

ROC Area   PRC Area   TP Rate   FP Rate   Precision   Recall   F-Measure   MCC
          Class
0.639      0.809      0.869     0.637     0.803       0.869     0.835       0.256
          NO_EDUCATION_RESEARCH
0.639      0.375      0.363     0.131     0.480       0.363     0.414       0.256
          EDUCATION_RESEARCH
Weighted Avg. 0.742     0.510     0.722       0.742     0.729       0.256
0.639      0.700
```

```
=== Confusion Matrix ===

  a    b  <-- classified as
522  79 |  a = NO_EDUCATION_RESEARCH
128  73 |  b = EDUCATION_RESEARCH
```

附录5: meta算法分类结果

```
weka.classifiers.meta.Vote -S 1 -B "weka.classifiers.bayes.BayesNet -D -
Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E
weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5" -B
"weka.classifiers.bayes.NaiveBayes " -B
"weka.classifiers.trees.DecisionStump " -B "weka.classifiers.trees.J48 -
C 0.25 -M 2" -R AVG
```

```
Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
```

=== Summary ===

Correctly Classified Instances	608	75.8105 %
Incorrectly Classified Instances	194	24.1895 %
Kappa statistic	0.3294	
Mean absolute error	0.2948	
Root mean squared error	0.4258	
Relative absolute error	78.4196 %	
Root relative squared error	98.2609 %	
Total Number of Instances	802	

=== Detailed Accuracy By Class ===

		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC
ROC Area	PRC Area	Class					
		0.859	0.542	0.826	0.859	0.842	0.331
0.668	0.820	NO_EDUCATION_RESEARCH					
		0.458	0.141	0.520	0.458	0.487	0.331
0.668	0.457	EDUCATION_RESEARCH					
Weighted Avg.		0.758	0.442	0.749	0.758	0.753	0.331
0.668	0.729						

=== Confusion Matrix ===

```

  a    b    <-- classified as
516  85 |   a = NO_EDUCATION_RESEARCH
109  92 |   b = EDUCATION_RESEARCH

```