说明 1：全集为 49 万条数据，由于时间太长问题，我们分别做 10%、20%、30%的采样，算法的准确率有很大部分受数据量的影响。缺省值构建方案见附录。 **（1）**可以做不同同一采样率下不同缺失度的不同算法准确度比较； **（2）**也可做同一缺失度，不同采样率的不同算法比较； **（3）**也可做同一算法在不同采样率和缺失率下的比较。

说明 2：在缺省值的情况下对抓包提取出来的数据进行分类，因为在缺失值情况下三个算法的分类结果重叠的部分不多，分类效果没有很好地一致性，而且又没有类别标签，不好比较分类的好与坏。
如下为在缺失值情况下的 10 折交叉验证算法详细情况（不含缺失值的在之前的第一个文档）

1、Rough set

1.1 采样率 10% 不同缺失率下的 10 折交叉验证平均时间：80.5 s

1.1.1 每个属性平均缺省率为：8.05%

Results of experiments by cross-validation method: 60_10_1

		Predicted							
		normal	u2r	dos	r2l	probe	No. of obj.	Accuracy	Coverage
Actual	normal	864.5	0.1	0	0.2	0.1	972.6	1	0.889
	u2r	0	0.2	0	0	0	0.6	0.2	0.2
	dos	0.2	0	3,714.3	0	0.1	3,914.7	1	0.949
	r2l	0.1	0	0	9	0	11.7	0.991	0.783
	probe	0.3	0	0.1	0	31.8	40.4	0.988	0.8
	True positive rate	1	0.2	1	0.97	0.99			

更详细的结果信息

TEST RESULTS:

Global coverage=0.9354251012145749

Global accuracy=0.9997405654483783

True positive rates for decision classes:

0.9992985894622555 0.2 0.9999724669603524 0.975 0.9937969924812029

Accuracy for decision classes:

0.9995308338845291 0.2 0.99991977237434 0.990909090909091 0.9883239962651726

Coverage for decision classes:

0.8891623331242628 0.2 0.948885130039718 0.7830088702147525 0.7997333176462793

1.1.2 每个属性平均缺省率为： 15.37%

Results of experiments by cross-validation method: kdd99DC_0.1

		Predicted							
		normal	dos	r2l	probe	u2r	No. of obj.	Accuracy	Coverage
Actual	normal	887.1	0	0.2	0.2	0	988.2	1	0.898
	dos	0.2	3,651.4	0	0	0	3,898.4	1	0.937
	r2l	0	0	9.1	0	0	11.9	1	0.789
	probe	0.1	0	0	32.6	0	40.7	0.997	0.803
	u2r	0.1	0	0	0	0.3	0.8	0.2	0.25
	True positive rate	1	1	0.98	0.99	0.2			

更详细的结果信息

TEST RESULTS:

Global coverage=0.9273886639676114

Global accuracy=0.9998246699650088

True positive rates for decision classes:

0.9995658784933468 1.0 0.9777777777777779 0.9943381180223285 0.2

Accuracy for decision classes:

0.9995531592261935 0.9999446338078364 1.0 0.9973684210526315 0.2

Coverage for decision classes:

0.897862302223604 0.9366514183975297 0.7892183985605039 0.8026031527842508 0.25

1.1.3 每个属性平均缺省率为：22.68%

Results of experiments by cross-validation method: 31

		Predicted							
		normal	dos	r2l	probe	u2r	No. of obj.	Accuracy	Coverage
Actual	normal	869	0	0.4	0	0	971	1	0.895
	dos	0.4	3,665.2	0	0	0	3,911.5	1	0.937
	r2l	0.3	0	9.5	0	0	12.5	0.975	0.781
	probe	0.2	0	0	36.4	0	44.5	0.995	0.82
	u2r	0	0	0	0	0.3	0.5	0.3	0.3
	True positive rate	1	1	0.95	1	0.3			

更详细的结果信息

TEST RESULTS:

Global coverage=0.9274696356275303

Global accuracy=0.9997164328932582

True positive rates for decision classes:

0.9989689045928187 1.0 0.9541666666666666 1.0 0.3

Accuracy for decision classes:

0.999544137015145 0.9998904031217603 0.975 0.9952353942144072 0.3

Coverage for decision classes:

0.8951546429453391 0.9369758685820534 0.7808373326020386 0.8202244530979257 0.3

1.2 采样率 20% 不同缺失率下的 10 折交叉验证平均时间：890 s

1.2.1 每个属性平均缺省率为：8.05%

Results of experiments by cross-validation method: 60_10_2

Actual \ Predicted	normal	dos	r2l	probe	u2r	No. of obj.	Accuracy	Coverage
normal	1,746	0.1	0.4	0.2	0	1,954.6	1	0.894
dos	0.1	7,230.1	0	0.1	0	7,824.9	1	0.924
r2l	0.4	0	16.1	0.1	0	22.1	0.972	0.755
probe	0.1	0.1	0	64.6	0	77.8	0.997	0.832
u2r	0.1	0	0.1	0	0.2	0.6	0.2	0.35
True positive rate	1	1	0.97	0.99	0.2			

更详细的结果信息

TEST RESULTS:

Global coverage=0.9168825910931172

Global accuracy=0.9998006385694452

True positive rates for decision classes:

0.9996130343524079 0.9999728900887632 0.9704995108091085 0.9942702776685366 0.2

Accuracy for decision classes:

0.9995973244881997 0.9999722332806844 0.9721212121212123 0.9974358974358974 0.2

Coverage for decision classes:

0.8936991317450801 0.9239970239983644 0.7546311772456925 0.8322922710637559 0.35

1.2.2 每个属性平均缺省率为： 15.37%

Results of experiments by cross-validation method: 22

Actual \ Predicted	normal	u2r	dos	r2l	probe	No. of obj.	Accuracy	Coverage
normal	1,740.9	0.1	0	0.3	0.1	1,944.1	1	0.896
u2r	0.2	0	0	0	0	0.5	0	0.15
dos	0.3	0	7,189.6	0	0	7,826.7	1	0.919
r2l	0.2	0	0	20.2	0	26.2	0.991	0.78
probe	0.4	0	0	0	68.5	82.5	0.994	0.831
True positive rate	1	0	1	0.99	1			

更详细的结果信息

TEST RESULTS:

Global coverage=0.9130364372469634

Global accuracy=0.9998227632811556

True positive rates for decision classes:

0.9993709557340169 0.0 1.0 0.9884501424501426 0.9986842105263157

Accuracy for decision classes:

0.9997231680907461 0.0 0.9999576739440688 0.9912962962962963 0.9937284144427002

Coverage for decision classes:

0.8955183675992753 0.15 0.9186073996972388 0.7801154364645826 0.8305620258562098

1.2.3 每个属性平均缺省率为：22.68%

Results of experiments by cross-validation method: 32

Actual \ Predicted	normal	dos	r2l	probe	u2r	No. of obj.	Accuracy	Coverage
normal	1,728.9	0.1	0.6	0.1	0.2	1,941.1	0.999	0.891
dos	0.4	7,175.6	0	0.1	0	7,830.4	1	0.916
r2l	0.5	0	16.4	0	0	24	0.971	0.705
probe	0.6	0	0	69.5	0	82.8	0.991	0.845
u2r	0.3	0	0	0	0.3	1.7	0.2	0.3
True positive rate	1	1	0.96	1	0.2			

更详细的结果信息

TEST RESULTS:

Global coverage=0.9102834008097165

Global accuracy=0.9996777809146886

True positive rates for decision classes:

0.998956854383491 0.9999859570285071 0.9639765892861869 0.9968253968253968 0.2

Accuracy for decision classes:

0.9994238860895462 0.9999310067220535 0.9711988115084091 0.9906939825263859 0.2

Coverage for decision classes:

0.8913725027164541 0.9164952461183363 0.705296390405086 0.8450860178015983 0.3

1.3 采样率 30% 不同缺失率下的 10 折交叉验证平均时间：2700 s

1.3.1 每个属性平均缺省率为：8.05%

Results of experiments by cross-validation method: 60_10_3

Actual \ Predicted	normal	u2r	dos	r2l	probe	No. of obj.	Accuracy	Coverage
normal	2,612.6	0	0.1	0.7	0.3	2,913.3	1	0.897
u2r	0.2	0.5	0	0	0	1.9	0.35	0.308
dos	0.2	0	10,908.3	0	0.1	11,748.4	1	0.929
r2l	0.5	0	0	21.6	0	30.7	0.982	0.713
probe	0.4	0	0	0	107	125.7	0.996	0.854
True positive rate	1	0.4	1	0.97	1			

更详细的结果信息

TEST RESULTS:

Global coverage=0.921221322537112

Global accuracy=0.9998169720565755

True positive rates for decision classes:

0.9995037425795065 0.4 0.9999905935471733 0.9725517241379311 0.9959818171454629

Accuracy for decision classes:

0.9995851495271959 0.35 0.9999725728074962 0.9816269284712483 0.9962951999806553
Coverage for decision classes:
0.8970012338100833 0.3083333333333333 0.9285629336712573 0.7129162975877872 0.8539405608031461

1.3.2 每个属性平均缺省率为： 15.37%

Results of experiments by cross-validation method: 23

Actual \ Predicted	normal	u2r	dos	r2l	probe	No. of obj.	Accuracy	Coverage
normal	2,636.4	0	0.1	0.3	0.5	2,938.5	1	0.897
u2r	0	0.7	0	0	0	1.6	0.5	0.353
dos	0.5	0	10,839...	0	0.1	11,722.1	1	0.925
r2l	0.3	0.1	0	26	0	35.2	0.987	0.742
probe	0.4	0	0	0	100.9	122.6	0.996	0.825
True positive rate	1	0.45	1	0.99	1			

更详细的结果信息
TEST RESULTS:
Global coverage=0.9180431848852899
Global accuracy=0.9998304603075198
True positive rates for decision classes:
0.9995484453942669 0.45 0.9999909901792954 0.9884751404202206 0.9951843623866472
Accuracy for decision classes:
0.9996675930292935 0.5 0.9999435201273478 0.987005068102629 0.9960068597712087
Coverage for decision classes:
0.897456512263533 0.35333333333333333 0.9247222525606134 0.7420390725829016 0.8252012174129734

1.3.3 每个属性平均缺省率为：22.68%

Results of experiments by cross-validation method: 33

Actual \ Predicted	normal	dos	r2l	probe	u2r	No. of obj.	Accuracy	Coverage
normal	2,626.7	0	0.6	0.8	0	2,940.2	0.999	0.894
dos	0.6	10,642...	0	0.2	0	11,721.8	1	0.908
r2l	0.4	0	25	0	0	34.8	0.984	0.727
probe	0.4	0	0	102.4	0	120.9	0.996	0.848
u2r	0.5	0	0	0	0.6	2.3	0.242	0.31
True positive rate	1	1	0.97	0.99	0.3			

更详细的结果信息
TEST RESULTS:
Global coverage=0.9042307692307693

Global accuracy=0.9997384034411189

True positive rates for decision classes:

0.9992740495234532 1.0 0.9746904558925019 0.989066376654821 0.3

Accuracy for decision classes:

0.9994755115102244 0.9999251073371687 0.9836521739130436 0.995982696982697 0.24166666666666664

Coverage for decision classes:

0.8939173425447672 0.9079316236990793 0.7265048251042564 0.848482923183256 0.31

2、C45

2.1 采样率 10%

2.1.1 每个属性平均缺省率为：8.05%

```
2017-01-12 18:09:22
RandomDelete
构造缺失记录
每条记录被修改的概率: 60%   每条记录最大被修改的属性个数:10 缺省值替代符:?
每个属性平均缺省率为: 8.05%
预处理耗时: 9.44 s
C4.5分类器 缺省数据集
采样率: 10%
10折交叉验证时间25.131 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    0.004    0.999      0.999    0.999      0.995    1.000     1.000     dos
                 0.111    0.000    0.500      0.111    0.182      0.236    0.859     0.169     u2r
                 0.670    0.000    0.844      0.670    0.747      0.752    0.993     0.847     r2l
                 0.909    0.000    0.974      0.909    0.941      0.941    0.999     0.977     probe
                 0.996    0.002    0.990      0.996    0.993      0.991    1.000     0.998     normal
Weighted Avg.    0.997    0.004    0.997      0.997    0.997      0.993    1.000     0.999

概要
Correctly Classified Instances       49242               99.6761 %
Incorrectly Classified Instances       160                0.3239 %
Kappa statistic                          0.9902
K&B Relative Info Score            4825918.6137 %
K&B Information Score                38926.267  bits      0.7879 bits/instance
Class complexity | order 0           39820.1667 bits      0.806  bits/instance
Class complexity | scheme             3117.5599 bits      0.0631 bits/instance
Complexity improvement     (Sf)      36702.6068 bits      0.7429 bits/instance
Mean absolute error                      0.003
Root mean squared error                  0.0324
Relative absolute error                  2.2621 %
Root relative squared error             12.5787 %
Total Number of Instances            49402

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 39182     0     1     3    32 |     a = dos
     1     1     1     0     6 |     b = u2r
     3     0    65     0    29 |     c = r2l
    13     0     2   411    26 |     d = probe
    26     1     8     8  9583 |     e = normal
```

2.1.2 每个属性平均缺省率为： 15.37%

```
2017-01-12 18:27:18
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:20 缺省值替代符:?
每个属性平均缺省率为：15.37%
预处理耗时：10.347 s
C4.5分类器 缺省数据集
采样率：10%
10折交叉验证时间52.207 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    0.014    0.996      0.999    0.998      0.988    1.000     1.000     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.961     0.048     u2r
                 0.526    0.000    0.927      0.526    0.671      0.698    0.994     0.780     r2l
                 0.794    0.000    0.989      0.794    0.881      0.885    0.997     0.931     probe
                 0.990    0.004    0.986      0.990    0.988      0.985    1.000     0.998     normal
Weighted Avg.    0.994    0.012    0.994      0.994    0.994      0.986    1.000     0.998

概要
Correctly Classified Instances       49109               99.4069 %
Incorrectly Classified Instances       293                0.5931 %
Kappa statistic                          0.982
K&B Relative Info Score            4709262.5245 %
K&B Information Score                37985.3091 bits      0.7689 bits/instance
Class complexity | order 0           39820.1667 bits      0.806  bits/instance
Class complexity | scheme             1873.4327 bits      0.0379 bits/instance
Complexity improvement     (Sf)      37946.7339 bits      0.7681 bits/instance
Mean absolute error                      0.0066
Root mean squared error                  0.044
Relative absolute error                  4.9467 %
Root relative squared error             17.063  %
Total Number of Instances            49402

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 39168     0     1     1    48 |     a = dos
     1     0     0     0     8 |     b = u2r
     6     0    51     0    40 |     c = r2l
    49     0     0   359    44 |     d = probe
    89     0     3     3  9531 |     e = normal
```

2.1.3 每个属性平均缺省率为：22.68%

```
2017-01-12 18:36:17
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:30 缺省值替代符:?
每个属性平均缺省率为：22.68%
预处理耗时：10.123 s
C4.5分类器 缺省数据集
采样率：10%
10折交叉验证时间85.002 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.998    0.026    0.993      0.998    0.996      0.980    0.999     1.000     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.951     0.078     u2r
                 0.629    0.000    0.984      0.629    0.767      0.786    0.995     0.808     r2l
                 0.759    0.000    0.972      0.759    0.852      0.858    0.995     0.900     probe
                 0.980    0.003    0.986      0.980    0.983      0.979    1.000     0.998     normal
Weighted Avg.    0.992    0.021    0.991      0.992    0.991      0.978    0.999     0.998

概要
Correctly Classified Instances       48993               99.1721 %
Incorrectly Classified Instances       409                0.8279 %
Kappa statistic                          0.9748
K&B Relative Info Score            4582121.2965 %
K&B Information Score                36959.7773 bits      0.7481 bits/instance
Class complexity | order 0           39820.1667 bits      0.806  bits/instance
Class complexity | scheme             2901.0852 bits      0.0587 bits/instance
Complexity improvement     (Sf)      36919.0814 bits      0.7473 bits/instance
Mean absolute error                      0.0111
Root mean squared error                  0.0538
Relative absolute error                  8.332  %
Root relative squared error             20.8811 %
Total Number of Instances            49402

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 39151     0     0     1    66 |     a = dos
     0     0     0     1     8 |     b = u2r
     7     0    61     0    29 |     c = r2l
    76     0     0   343    33 |     d = probe
   179     0     1     8  9438 |     e = normal
```

2.2 采样率 20%

2.2.1 每个属性平均缺省率为：8.05%

```
2017-01-12 18:11:09
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:10 缺省值替代符:?
每个属性平均缺省率为：8.05%
预处理耗时：12.384 s
C4.5分类器 缺省数据集
采样率：20%
10折交叉验证时间94.454 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.007    0.998      1.000    0.999      0.995    1.000     1.000     dos
                 0.467    0.000    0.875      0.467    0.609      0.639    0.907     0.413     u2r
                 0.708    0.000    0.941      0.708    0.808      0.816    0.992     0.876     r2l
                 0.909    0.000    0.991      0.909    0.948      0.949    0.997     0.977     probe
                 0.994    0.002    0.993      0.994    0.993      0.992    1.000     0.999     normal
Weighted Avg.    0.997    0.006    0.997      0.997    0.997      0.993    1.000     0.999

概要
Correctly Classified Instances       98509               99.7014 %
Incorrectly Classified Instances       295                0.2986 %
Kappa statistic                          0.991
K&B Relative Info Score            9675646.8676 %
K&B Information Score                78051.3136 bits      0.79   bits/instance
Class complexity | order 0           79680.7941 bits      0.8065 bits/instance
Class complexity | scheme             4930.7624 bits      0.0499 bits/instance
Complexity improvement     (Sf)      74750.0317 bits      0.7565 bits/instance
Mean absolute error                      0.0027
Root mean squared error                  0.0305
Relative absolute error                  2.0294 %
Root relative squared error             11.8248 %
Total Number of Instances            98804

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 78268     0     0     2    33 |     a = dos
     1     7     1     0     6 |     b = u2r
     4     0   143     1    54 |     c = r2l
    27     0     1   771    49 |     d = probe
   104     1     7     4 19320 |     e = normal
```

2.2.2 每个属性平均缺省率为： 15.37%

```
2017-01-12 19:01:36
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:20 缺省值替代符:?
每个属性平均缺省率为：15.37%
预处理耗时：12.678 s
C4.5分类器 缺省数据集
采样率：20%
10折交叉验证时间154.104 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.998    0.011    0.997      0.998    0.998      0.989    1.000     1.000     dos
                 0.133    0.000    0.400      0.133    0.200      0.231    0.948     0.227     u2r
                 0.693    0.000    0.940      0.693    0.798      0.807    0.999     0.887     r2l
                 0.817    0.000    0.987      0.817    0.894      0.897    0.997     0.931     probe
                 0.993    0.003    0.987      0.993    0.990      0.987    1.000     0.999     normal
Weighted Avg.    0.995    0.009    0.995      0.995    0.995      0.988    1.000     0.999

概要
Correctly Classified Instances       98307               99.497  %
Incorrectly Classified Instances       497                0.503  %
Kappa statistic                          0.9849
K&B Relative Info Score            9491109.8424 %
K&B Information Score                76562.6941 bits      0.7749 bits/instance
Class complexity | order 0           79680.7941 bits      0.8065 bits/instance
Class complexity | scheme             4221.3887 bits      0.0427 bits/instance
Complexity improvement     (Sf)      75459.4054 bits      0.7637 bits/instance
Mean absolute error                      0.0057
Root mean squared error                  0.0405
Relative absolute error                  4.2701 %
Root relative squared error             15.6754 %
Total Number of Instances            98804

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 78178     0     1     0   124 |     a = dos
     1     2     2     1     9 |     b = u2r
     4     3   140     0    55 |     c = r2l
    82     0     1   693    72 |     d = probe
   129     0     5     8 19294 |     e = normal
```

2.2.3 每个属性平均缺省率为：22.68%

```
2017-01-12 19:39:45
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:30 缺省值替代符:?
每个属性平均缺省率为：22.68%
预处理耗时：12.492 s
C4.5分类器 缺省数据集
采样率：20%
10折交叉验证时间274.798 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    0.013    0.997      0.999    0.998      0.989    1.000     1.000     dos
                 0.067    0.000    0.500      0.067    0.118      0.183    0.940     0.114     u2r
                 0.535    0.000    0.931      0.535    0.679      0.705    0.997     0.741     r2l
                 0.794    0.000    0.967      0.794    0.872      0.875    0.996     0.926     probe
                 0.989    0.004    0.986      0.989    0.987      0.984    1.000     0.999     normal
Weighted Avg.    0.994    0.011    0.994      0.994    0.994      0.986    1.000     0.998

概要
Correctly Classified Instances       98218               99.4069 %
Incorrectly Classified Instances       586                0.5931 %
Kappa statistic                          0.9821
K&B Relative Info Score            9309888.3224 %
K&B Information Score                75100.82   bits      0.7601 bits/instance
Class complexity | order 0           79680.7941 bits      0.8065 bits/instance
Class complexity | scheme             5643.137  bits      0.0571 bits/instance
Complexity improvement     (Sf)      74037.6571 bits      0.7493 bits/instance
Mean absolute error                      0.0089
Root mean squared error                  0.0466
Relative absolute error                  6.7041 %
Root relative squared error             18.0616 %
Total Number of Instances            98804

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 78212     0     0     6    85 |     a = dos
     1     1     0     0    13 |     b = u2r
     4     0   108     0    90 |     c = r2l
    81     0     0   673    94 |     d = probe
   186     1     8    17 19224 |     e = normal
```

2.3 采样率 30%

2.3.1 每个属性平均缺省率为：8.05%

```
2017-01-12 18:11:41
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:10 缺省值替代符:?
每个属性平均缺省率为：8.05%
预处理耗时：11.225 s
C4.5分类器 缺省数据集
采样率：30%
10折交叉验证时间192.836 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.003    0.999      1.000    0.999      0.997    1.000     1.000     dos
                 0.250    0.000    0.714      0.250    0.370      0.423    0.911     0.324     u2r
                 0.843    0.000    0.961      0.843    0.898      0.900    0.996     0.949     r2l
                 0.937    0.000    0.983      0.937    0.960      0.960    0.999     0.988     probe
                 0.997    0.001    0.995      0.997    0.996      0.995    1.000     1.000     normal
Weighted Avg.    0.998    0.003    0.998      0.998    0.998      0.996    1.000     1.000

概要
Correctly Classified Instances      147939               99.8198 %
Incorrectly Classified Instances       267                0.1802 %
Kappa statistic                          0.9946
K&B Relative Info Score            14629218.1621 %
K&B Information Score                117702.4689 bits      0.7942 bits/instance
Class complexity | order 0           119210.753  bits      0.8044 bits/instance
Class complexity | scheme             6887.4152 bits      0.0465 bits/instance
Complexity improvement     (Sf)      112323.3378 bits      0.7579 bits/instance
Mean absolute error                      0.0018
Root mean squared error                  0.0236
Relative absolute error                  1.3335 %
Root relative squared error              9.14   %
Total Number of Instances           148206

分类的混合矩阵
      a      b      c      d      e   <-- classified as
 117545      1      0      6     37 |      a = dos
      1      5      1      2     11 |      b = u2r
      3      0    269      0     47 |      c = r2l
     39      1      1   1168     37 |      d = probe
     59      0      9     12  28952 |      e = normal
```

2.3.2 每个属性平均缺省率为： 15.37%

```
2017-01-12 19:01:42
RandomDelete
构造缺失记录
每条记录被修改的概率：60%    每条记录最大被修改的属性个数:20 缺省值替代符:?
每个属性平均缺省率为：15.37%
预处理耗时：11.351 s
C4.5分类器 缺省数据集
采样率：30%
10折交叉验证时间338.829 s

类别准确率的详细情况
                 TP Rate   FP Rate   Precision   Recall   F-Measure   MCC      ROC Area  PRC Area  Class
                 0.999     0.007     0.998       0.999    0.999       0.993    1.000     1.000     dos
                 0.250     0.000     0.714       0.250    0.370       0.423    0.969     0.342     u2r
                 0.774     0.000     0.969       0.774    0.861       0.866    0.999     0.927     r2l
                 0.864     0.000     0.981       0.864    0.919       0.920    0.999     0.971     probe
                 0.995     0.002     0.990       0.995    0.993       0.991    1.000     0.999     normal
Weighted Avg.    0.997     0.006     0.996       0.997    0.996       0.992    1.000     0.999

概要
Correctly Classified Instances      147689               99.6512 %
Incorrectly Classified Instances       517                0.3488 %
Kappa statistic                          0.9895
K&B Relative Info Score             14382958.9968 %
K&B Information Score                 115721.1387 bits      0.7808 bits/instance
Class complexity | order 0            119210.753  bits      0.8044 bits/instance
Class complexity | scheme               3497.3691 bits      0.0236 bits/instance
Complexity improvement     (Sf)       115713.3839 bits      0.7808 bits/instance
Mean absolute error                      0.0045
Root mean squared error                  0.034
Relative absolute error                  3.4047 %
Root relative squared error             13.2087 %
Total Number of Instances           148206

分类的混合矩阵
      a      b      c      d      e   <-- classified as
 117480      1      0      4    104 |      a = dos
      0      5      1      1     13 |      b = u2r
      7      1    247      2     62 |      c = r2l
     69      0      1   1077     99 |      d = probe
    132      0      6     14  28880 |      e = normal
```

2.3.3 每个属性平均缺省率为：22.68%

```
2017-01-12 20:25:19
RandomDelete
构造缺失记录
每条记录被修改的概率：60%    每条记录最大被修改的属性个数:30 缺省值替代符:?
每个属性平均缺省率为：22.68%
预处理耗时：10.992 s
C4.5分类器 缺省数据集
采样率：30%
10折交叉验证时间574.638 s

类别准确率的详细情况
                 TP Rate   FP Rate   Precision   Recall   F-Measure   MCC      ROC Area  PRC Area  Class
                 0.999     0.012     0.997       0.999    0.998       0.990    1.000     1.000     dos
                 0.250     0.000     0.625       0.250    0.357       0.395    0.985     0.292     u2r
                 0.608     0.000     0.951       0.608    0.742       0.760    0.997     0.780     r2l
                 0.823     0.000     0.978       0.823    0.894       0.896    0.998     0.941     probe
                 0.990     0.003     0.988       0.990    0.989       0.986    1.000     0.999     normal
Weighted Avg.    0.995     0.010     0.995       0.995    0.995       0.988    1.000     0.999

概要
Correctly Classified Instances      147440               99.4832 %
Incorrectly Classified Instances       766                0.5168 %
Kappa statistic                          0.9844
K&B Relative Info Score             14055021.0154 %
K&B Information Score                 113082.6443 bits      0.763  bits/instance
Class complexity | order 0            119210.753  bits      0.8044 bits/instance
Class complexity | scheme               6138.59   bits      0.0414 bits/instance
Complexity improvement     (Sf)       113072.163  bits      0.7629 bits/instance
Mean absolute error                      0.0081
Root mean squared error                  0.0437
Relative absolute error                  6.0968 %
Root relative squared error             16.9535 %
Total Number of Instances           148206

分类的混合矩阵
      a      b      c      d      e   <-- classified as
 117464      0      0      6    119 |      a = dos
      2      5      3      2      8 |      b = u2r
     11      0    194      0    114 |      c = r2l
    111      0      0   1025    110 |      d = probe
    255      3      7     15  28752 |      e = normal
```

3、SVM

3.1 采样率 10%

3.1.1 每个属性平均缺省率为：8.05%

```
2017-01-12 18:10:03
RandomDelete
构造缺失记录
每条记录被修改的概率：60%    每条记录最大被修改的属性个数:10 缺省值替代符:?
每个属性平均缺省率为：8.05%
预处理耗时：7.791 s
缺省值填充、归一化耗时：0.328 s
SVM分类器 缺省数据
采样率：10%
10折交叉验证时间171.715 s

类别准确率的详细情况
                 TP Rate   FP Rate   Precision   Recall   F-Measure   MCC      ROC Area  PRC Area  Class
                 0.994     0.007     0.998       0.994    0.996       0.981    0.994     0.997     dos
                 0.000     0.000     0.000       0.000    0.000       0.000    0.500     0.000     u2r
                 0.000     0.000     0.000       0.000    0.000       0.000    0.500     0.002     r2l
                 0.834     0.000     0.943       0.834    0.885       0.886    0.917     0.788     probe
                 0.994     0.010     0.961       0.994    0.977       0.972    0.992     0.957     normal
Weighted Avg.    0.990     0.007     0.988       0.990    0.989       0.976    0.991     0.985

概要
Correctly Classified Instances       48923               99.0304 %
Incorrectly Classified Instances       479                0.9696 %
Kappa statistic                          0.9709
K&B Relative Info Score            4655096.0764 %
K&B Information Score                37548.398  bits      0.7601 bits/instance
Class complexity | order 0           39820.1667 bits      0.806  bits/instance
Class complexity | scheme           514446      bits     10.4135 bits/instance
Complexity improvement     (Sf)    -474625.8333 bits     -9.6074 bits/instance
Mean absolute error                      0.0039
Root mean squared error                  0.0623
Relative absolute error                  2.9223 %
Root relative squared error             24.1775 %
Total Number of Instances            49402

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 38979     0     0     0   239 |     a = dos
     0     0     0     0     9 |     b = u2r
     1     0     0     0    96 |     c = r2l
    33     0     0   377    42 |     d = probe
    36     0     0    23  9567 |     e = normal
```

3.1.2 每个属性平均缺省率为： 15.37%

```
2017-01-12 18:28:26
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:20 缺省值替代符:?
每个属性平均缺省率为：15.37%
预处理耗时：7.236 s
缺省值填充、归一化耗时：0.323 s
SVM分类器 缺省数据
采样率：10%
10折交叉验证时间252.554 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.993    0.017    0.996      0.993    0.994      0.973    0.988     0.994     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.000     u2r
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.002     r2l
                 0.768    0.000    0.948      0.768    0.848      0.852    0.884     0.730     probe
                 0.987    0.010    0.959      0.987    0.972      0.966    0.988     0.948     normal
Weighted Avg.    0.988    0.015    0.986      0.988    0.987      0.969    0.986     0.981

概要
Correctly Classified Instances       48802               98.7855 %
Incorrectly Classified Instances       600                1.2145 %
Kappa statistic                          0.9634
K&B Relative Info Score            4599897.3364 %
K&B Information Score                37103.1603 bits      0.751  bits/instance
Class complexity | order 0           39820.1667 bits      0.806  bits/instance
Class complexity | scheme           644400      bits     13.044  bits/instance
Complexity improvement     (Sf)    -604579.8333 bits    -12.238  bits/instance
Mean absolute error                      0.0049
Root mean squared error                  0.0697
Relative absolute error                  3.6605 %
Root relative squared error             27.0594 %
Total Number of Instances            49402

分类的混合矩阵
    a     b     c     d     e   <-- classified as
 38956     0     0     0   262 |     a = dos
     1     0     0     0     8 |     b = u2r
     1     0     0     0    96 |     c = r2l
    60     0     0   347    45 |     d = probe
   108     0     0    19  9499 |     e = normal
```

3.1.3 每个属性平均缺省率为：22.68%

```
2017-01-12 18:37:58
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:30 缺省值替代符:?
每个属性平均缺省率为：22.68%
预处理耗时：8.408 s
缺省值填充、归一化耗时：0.359 s
SVM分类器 缺省数据
采样率：10%
10折交叉验证时间359.053 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.993    0.038    0.990      0.993    0.991      0.958    0.977     0.989     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.000     u2r
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.002     r2l
                 0.673    0.000    0.935      0.673    0.782      0.792    0.836     0.632     probe
                 0.967    0.011    0.954      0.967    0.960      0.951    0.978     0.929     normal
Weighted Avg.    0.983    0.032    0.980      0.983    0.981      0.953    0.975     0.972

概要
Correctly Classified Instances       48544               98.2632 %
Incorrectly Classified Instances       858                1.7368 %
Kappa statistic                          0.9473
K&B Relative Info Score            4492721.5136 %
K&B Information Score                36238.6711 bits      0.7335 bits/instance
Class complexity | order 0           39820.1667 bits      0.806  bits/instance
Class complexity | scheme           921492      bits     18.6529 bits/instance
Complexity improvement     (Sf)    -881671.8333 bits    -17.8469 bits/instance
Mean absolute error                      0.0069
Root mean squared error                  0.0833
Relative absolute error                  5.2345 %
Root relative squared error             32.3584 %
Total Number of Instances            49402

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 38929     0     0     1   288 |     a = dos
     0     0     0     0     9 |     b = u2r
     3     0     0     0    94 |     c = r2l
    86     0     0   304    62 |     d = probe
   295     0     0    20  9311 |     e = normal
```

3.2 采样率 20%

3.2.1 每个属性平均缺省率为：8.05%

```
2017-01-12 18:13:02
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:10 缺省值替代符:?
每个属性平均缺省率为：8.05%
预处理耗时：8.73 s
缺省值填充、归一化耗时：0.432 s
SVM分类器 缺省数据
采样率：20%
10折交叉验证时间1534.933 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.994    0.004    0.999      0.994    0.996      0.983    0.995     0.998     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.000     u2r
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.002     r2l
                 0.866    0.000    0.965      0.866    0.912      0.913    0.933     0.836     probe
                 0.997    0.010    0.962      0.997    0.979      0.974    0.994     0.959     normal
Weighted Avg.    0.991    0.005    0.989      0.991    0.990      0.978    0.993     0.987

概要
Correctly Classified Instances       97931               99.1164 %
Incorrectly Classified Instances       873                0.8836 %
Kappa statistic                          0.9736
K&B Relative Info Score            9358036.02   %
K&B Information Score                75489.2168 bits      0.764  bits/instance
Class complexity | order 0           79680.7941 bits      0.8065 bits/instance
Class complexity | scheme           937602      bits      9.4895 bits/instance
Complexity improvement     (Sf)    -857921.2059 bits     -8.6831 bits/instance
Mean absolute error                      0.0035
Root mean squared error                  0.0594
Relative absolute error                  2.6519 %
Root relative squared error             23.0309 %
Total Number of Instances            98804

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 77824     0     0     0   479 |     a = dos
     0     0     0     0    15 |     b = u2r
     1     0     0     0   201 |     c = r2l
    42     0     0   734    72 |     d = probe
    36     0     0    27 19373 |     e = normal
```

3.2.2 每个属性平均缺省率为： 15.37%

```
2017-01-12 19:04:29
RandomDelete
构造缺失记录
每条记录被修改的概率：60%    每条记录最大被修改的属性个数:20 缺省值替代符:?
每个属性平均缺省率为：15.37%
预处理耗时：9.309 s
缺省值填充、归一化耗时：0.435 s
SVM分类器 缺省数据
采样率：20%
10折交叉验证时间1147.33 s

类别准确率的详细情况
                 TP Rate   FP Rate   Precision   Recall   F-Measure   MCC      ROC Area  PRC Area  Class
                 0.994     0.013     0.997       0.994    0.995       0.977    0.991     0.995     dos
                 0.000     0.000     0.000       0.000    0.000       0.000    0.500     0.000     u2r
                 0.000     0.000     0.000       0.000    0.000       0.000    0.500     0.002     r2l
                 0.775     0.000     0.943       0.775    0.850       0.853    0.887     0.732     probe
                 0.989     0.010     0.960       0.989    0.974       0.968    0.989     0.951     normal
Weighted Avg.    0.989     0.012     0.987       0.989    0.988       0.972    0.988     0.982

概要
Correctly Classified Instances       97700               98.8826 %
Incorrectly Classified Instances      1104                1.1174 %
Kappa statistic                          0.9665
K&B Relative Info Score            9241626.3434 %
K&B Information Score                74550.1656 bits      0.7545 bits/instance
Class complexity | order 0           79680.7941 bits      0.8065 bits/instance
Class complexity | scheme          1185696      bits     12.0005 bits/instance
Complexity improvement     (Sf)    -1106015.2059 bits    -11.194  bits/instance
Mean absolute error                      0.0045
Root mean squared error                  0.0669
Relative absolute error                  3.3536 %
Root relative squared error             25.8993 %
Total Number of Instances            98804

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 77820     0     0     0   483 |     a = dos
     0     0     0     0    15 |     b = u2r
     6     0     0     0   196 |     c = r2l
    81     0     0   657   110 |     d = probe
   173     0     0    40 19223 |     e = normal
```

3.2.3 每个属性平均缺省率为：22.68%

```
2017-01-12 19:44:38
RandomDelete
构造缺失记录
每条记录被修改的概率：60%    每条记录最大被修改的属性个数:30 缺省值替代符:?
每个属性平均缺省率为：22.68%
预处理耗时：6.447 s
缺省值填充、归一化耗时：0.418 s
SVM分类器 缺省数据
采样率：20%
10折交叉验证时间5474.04 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.993    0.029    0.992      0.993    0.993      0.966    0.982     0.991     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.000     u2r
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.002     r2l
                 0.752    0.001    0.922      0.752    0.829      0.832    0.876     0.696     probe
                 0.973    0.010    0.959      0.973    0.966      0.957    0.981     0.938     normal
Weighted Avg.    0.985    0.025    0.983      0.985    0.984      0.961    0.980     0.976

概要
Correctly Classified Instances       97337               98.5152 %
Incorrectly Classified Instances      1467                1.4848 %
Kappa statistic                          0.9553
K&B Relative Info Score            9112133.9424 %
K&B Information Score                73505.5789 bits      0.744  bits/instance
Class complexity | order 0           79680.7941 bits      0.8065 bits/instance
Class complexity | scheme          1575558      bits     15.9463 bits/instance
Complexity improvement     (Sf)    -1495877.2059 bits    -15.1398 bits/instance
Mean absolute error                      0.0059
Root mean squared error                  0.0771
Relative absolute error                  4.4563 %
Root relative squared error             29.8551 %
Total Number of Instances            98804

分类的混合矩阵
     a     b     c     d     e   <-- classified as
 77790     0     0     1   512 |     a = dos
     1     0     0     0    14 |     b = u2r
     7     0     0     0   195 |     c = r2l
   113     0     0   638    97 |     d = probe
   474     0     0    53 18909 |     e = normal
```

3.3 采样率 30%

3.3.1 每个属性平均缺省率为：8.05%

```
2017-01-12 18:15:12
RandomDelete
构造缺失记录
每条记录被修改的概率：60%    每条记录最大被修改的属性个数:10 缺省值替代符:?
每个属性平均缺省率为：8.05%
预处理耗时：8.2 s
缺省值填充、归一化耗时：0.564 s
SVM分类器 缺省数据
采样率：30%
10折交叉验证时间1738.613 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.994    0.003    0.999      0.994    0.997      0.984    0.995     0.998     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.000     u2r
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.002     r2l
                 0.880    0.000    0.987      0.880    0.930      0.931    0.940     0.870     probe
                 0.998    0.010    0.962      0.998    0.979      0.975    0.994     0.960     normal
Weighted Avg.    0.992    0.004    0.989      0.992    0.990      0.979    0.994     0.987

概要
Correctly Classified Instances      146947               99.1505 %
Incorrectly Classified Instances      1259                0.8495 %
Kappa statistic                          0.9746
K&B Relative Info Score            14055343.6351 %
K&B Information Score                113085.24   bits      0.763  bits/instance
Class complexity | order 0           119210.753  bits      0.8044 bits/instance
Class complexity | scheme          1352166      bits      9.1236 bits/instance
Complexity improvement     (Sf)    -1232955.247 bits     -8.3192 bits/instance
Mean absolute error                      0.0034
Root mean squared error                  0.0583
Relative absolute error                  2.5582 %
Root relative squared error             22.6203 %
Total Number of Instances           148206

分类的混合矩阵
      a      b      c      d      e   <-- classified as
 116886      0      0      1    702 |      a = dos
      0      0      0      0     20 |      b = u2r
      3      0      0      0    316 |      c = r2l
     39      0      0   1097    110 |      d = probe
     54      0      0     14  28964 |      e = normal
```

3.3.2 每个属性平均缺省率为： 15.37%

```
2017-01-12 19:07:39
RandomDelete
构造缺失记录
每条记录被修改的概率：60%    每条记录最大被修改的属性个数:20 缺省值替代符:?
每个属性平均缺省率为：15.37%
预处理耗时：9.388 s
缺省值填充、归一化耗时：0.535 s
SVM分类器 缺省数据
采样率：30%
10折交叉验证时间2462.939 s

类别准确率的详细情况
                 TP Rate   FP Rate   Precision   Recall   F-Measure   MCC      ROC Area  PRC Area  Class
                 0.994     0.011     0.997       0.994    0.996       0.979    0.992     0.996     dos
                 0.000     0.000     0.000       0.000    0.000       0.000    0.500     0.000     u2r
                 0.000     0.000     0.000       0.000    0.000       0.000    0.500     0.002     r2l
                 0.821     0.000     0.962       0.821    0.886       0.888    0.910     0.792     probe
                 0.991     0.010     0.961       0.991    0.976       0.970    0.991     0.954     normal
Weighted Avg.    0.990     0.010     0.988       0.990    0.988       0.974    0.990     0.984

概要
Correctly Classified Instances      146669               98.9629 %
Incorrectly Classified Instances      1537                1.0371 %
Kappa statistic                          0.9688
K&B Relative Info Score             13924291.4781 %
K&B Information Score                 112030.8322 bits      0.7559 bits/instance
Class complexity | order 0            119210.753  bits      0.8044 bits/instance
Class complexity | scheme           1650738      bits     11.1381 bits/instance
Complexity improvement     (Sf)    -1531527.247 bits    -10.3338 bits/instance
Mean absolute error                      0.0041
Root mean squared error                  0.0644
Relative absolute error                  3.1231 %
Root relative squared error             24.9932 %
Total Number of Instances           148206

分类的混合矩阵
      a      b      c      d      e   <-- classified as
 116877      0      0      0    712 |      a = dos
      0      0      0      0     20 |      b = u2r
      6      0      0      0    313 |      c = r2l
    102      0      0   1023    121 |      d = probe
    223      0      0     40  28769 |      e = normal
```

3.3.3 每个属性平均缺省率为：22.68%

```
2017-01-12 20:35:12
RandomDelete
构造缺失记录
每条记录被修改的概率：60%   每条记录最大被修改的属性个数:30 缺省值替代符:?
每个属性平均缺省率为：22.68%
预处理耗时：9.565 s
缺省值填充、归一化耗时：0.604 s
SVM分类器 缺省数据
采样率：30%
10折交叉验证时间15700.87 s

类别准确率的详细情况
                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.994    0.027    0.993      0.994    0.993      0.968    0.984     0.992     dos
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.000     u2r
                 0.000    0.000    0.000      0.000    0.000      0.000    0.500     0.002     r2l
                 0.765    0.000    0.965      0.765    0.853      0.858    0.882     0.740     probe
                 0.976    0.010    0.959      0.976    0.967      0.959    0.983     0.940     normal
Weighted Avg.    0.986    0.023    0.984      0.986    0.985      0.963    0.981     0.977

概要
Correctly Classified Instances      146133               98.6013 %
Incorrectly Classified Instances      2073                1.3987 %
Kappa statistic                          0.9577
K&B Relative Info Score             13709964.7925 %
K&B Information Score                 110306.4215 bits      0.7443 bits/instance
Class complexity | order 0            119210.753  bits      0.8044 bits/instance
Class complexity | scheme            2226402      bits     15.0223 bits/instance
Complexity improvement     (Sf)     -2107191.247 bits    -14.218  bits/instance
Mean absolute error                      0.0056
Root mean squared error                  0.0748
Relative absolute error                  4.2123 %
Root relative squared error             29.0258 %
Total Number of Instances           148206

分类的混合矩阵
     a      b      c      d      e   <-- classified as
 116851      0      0      0    738 |      a = dos
      0      0      0      0     20 |      b = u2r
      9      0      0      0    310 |      c = r2l
    140      0      0    953    153 |      d = probe
    668      0      0     35  28329 |      e = normal
```

附录：人为构建随机缺省的数据集

构建方案

A: 每一行是否含有缺省值的概率；

B: 含有缺省值的那一行最多含有缺省属性的最大个数，个数的分布服从 1-b 的均匀分布

C：由于训练集的最后一个属性是类型属性，所以不对此属性进行随机缺省，所以参与缺省的属性个数为所有属性数量减 1

说明：每一次随机挑选将要进行缺失的属性，不重复挑选（无放回抽样）

每一个属性的平均缺省率计算公式为：

$$MissingRatio = A \times \sum_{i=1}^{B} i \div B \div C$$

例如：

A 设置为 60%

B 设置为 10

C 因为一共有 42 个属性，所以 C = 42-1

计算得出 $MissingRatio = ¿$ 8.05%