**On-line Appendix**

This appendix contained tables for the following cases:

1. There are total 4 tables provided in the appendix corresponding to the results of different prediction models.

数据集详细信息描述

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Project | Time Span | Sybsystems | ARB type | # ARBs |
| Linux | Dec 2003 -May 2011 | Network Drivers | Memory-related | 7 |
| Numerical | 1 |
| Other logical resource | 1 |
| SCSI Drivers | Memory-related | 4 |
| EXT3 FileSystem | Memory-related | 3 |
| Storage-related | 2 |
| Networking/IPV4 | Memory-related | 1 |
| Other logical resource | 1 |
| MySQL | Aug 2006 – Feb 2011 | InnoDB Storage engine | Memory-related | 2 |
| Numerical | 1 |
| Other logical resource | 3 |
| Replication | Memory-related |  |
| Other logical resource |  |
| Optimizer | Memory-related |  |
| NetBSD | 2015年10月7版本 | kern | Memory-related | 3 |
| Other logical resource | 1 |
| fs | Memory-related | 0 |
| Other logical resource | 1 |
| usr.bin | Memory-related | 3 |
| Other logical resource | 2 |
| usr.sbin | Memory-related | 5 |
| Other logical resource | 2 |
| crypto | Memory-related | 1 |
| Other logical resource | 0 |

数据特征详细描述

|  |  |  |
| --- | --- | --- |
| Type | Metrics | Description |
| Program size | AltAvgLineBlank, AltAvgLineCode, AltAvgLineComment, AltCountLineBlank, AltCountLineCode, AltCountLineComment, AvgLine, AvgLineBlank, AvgLineCode, AvgLineComment, CountDeclClass, CountDeclFunction, CountLine, CountLineBlank, CountLineCode, CountLineCodeDecl, CountLineCodeExe, CountLineComment, CountLineInactive, CountLinePreprocessor, CountSemicolon, CountStmt, CountStmtDecl, CountStmtEmpty, CountStmtExe, RatioCommentToCode | Metrics related to the amount of lines of code, declarations, statements, and files |
| McCabe’s cyclomatic complexity | AvgCyclomatic, AvgCyclomaticModified, AvgCyclomaticStrict, AvgEssential, MaxCyclomatic, MaxCyclomaticModified, MaxCyclomaticStrict, SumCyclomatic, SumCyclomaticModified, SumCyclomaticStrict, SumEssential | Metrics related to the control flow graph of functions and methods |
| Halstead metrics | Program Volume, Program Length, Program Vocabulary, Program Difficulty, Effort, N1, N2, n1, n2 | Metrics based on operands and operators in the program |
| Aging-Related Metrics (ARMs) | AllocOps, DeallocOps, DerefSet, DerefUse, UniqueDerefSet, UniqueDerefUse | Metrics related to memory usage |

参数选择

|  |  |
| --- | --- |
| Technique | Parameter values |
| K-Nearest Neighbour (KNN) | n\_neighbors = range(1, 21, 2), weights = ['uniform', 'distance'], metric = ['euclidean', 'manhattan', 'minkowski'] |
| Naïve Bayes (NB) | With default param eter in scikitlearn |
| Logistic Regression (LR) | solvers = ['liblinear'],  penalty = ['l2'], c\_values = [100, 10, 1.0, 0.1, 0.01] |
| Decision Tree (DT) | mdepth=np.arange(1,40), grid={"criterion":["gini","entropy"],"max\_depth":mdepth} |
| Random Forest (RF) | 'bootstrap': [True],  'max\_depth': [2,4,5],  'max\_features': ["log2","sqrt"],  'min\_samples\_leaf': [10,30], 'min\_samples\_split': [10,12], 'n\_estimators': [100, 350, 500] |
| Support Vector Machine (SVM) | kernel = ['poly', 'rbf', 'sigmoid'], C parameters= [50, 10, 1.0, 0.1, 0.01] |

|  |  |
| --- | --- |
| Parameter | Values |
| max\_epoch | 10 |
| steps\_per\_epoch | 10 |
| save\_checkpoints | True |
| restore\_session | True |
| batch\_size | 1 |
| z\_dim | 50 |
| noise= | 0.2 |
| l2norm | 0.00001 |
| learning\_rate | 0.001 |
| num\_gen\_rnn | 100 |
| num\_gen\_features | 100 |
| num\_dis\_layers | 2 |
| num\_dis\_hidde | 100 |
| optimizer | 'AdamOptimizer' |

F值的补充实验