Supplemental material for the article: "Learning-to-Rank *vs* Ranking-to-Learn: Strategies for Test Selection and Prioritization in Continuous Integration"

Anonymous Author(s)

ACM Reference Format:
Anonymous Author(s). 2020. Supplemental material for the article: "Learning-to-Rank *vs* Ranking-to-Learn: Strategies for Test Selection and Prioritization in Continuous Integration". In *ICSE 2020: 42nd International Conference on Software Engineering*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

A ONLINE SUPPLEMENTAL MATERIAL

This appendix is also available online for reviewers' use at the following anonymized GitHub repository: https://github.com/icse20/TSP- CI. Material for replicating the experiments is at the same repository.

The appendix contains: the setting of the prioritization algorithms; results for RQ1, RQ2 and RQ3 not presented in the main article for the sake of space.

A.1 Algorithm parameters

Table 1 reports the parameters setting of the experimented algorithms. In all the cases except two, the default parameters are adopted: for RankNet and LambdaMART, due to overfitting causing NaN occurrences in the ranking, we needed to lower the number of nodes until obtaining a valid result.

A.2 RQ1

Figures 1a-1f and 2a-2f report the *RPA* and *ranking time* boxplots for each subject. Figures 3a-3f report the total training times (ms) per subject, in logarithmic scale.

A.3 RQ2

Figures 4a-4f reports the T^2 Hotelling's statistic used to summarize the trend of the code metrics, computed on a set of principal components able to explain at least the 95% of the original metrics variance.

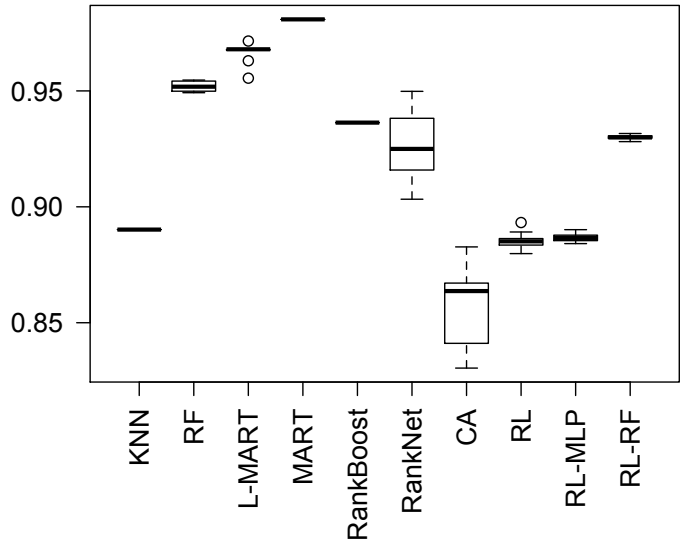
A.4 RQ3

Table 2f reports, for each subject, the difference between the tests execution times taken from the 25%, 50% or 75% of the list of tests sorted according to the *optimal* ranking and those taken from the 25%, 50% or 75% of the list of tests sorted according to the *predicted* ranking. Similarly, the difference between the total number of failing tests considering the 25%, 50% or 75% of optimal and predicted ranking.

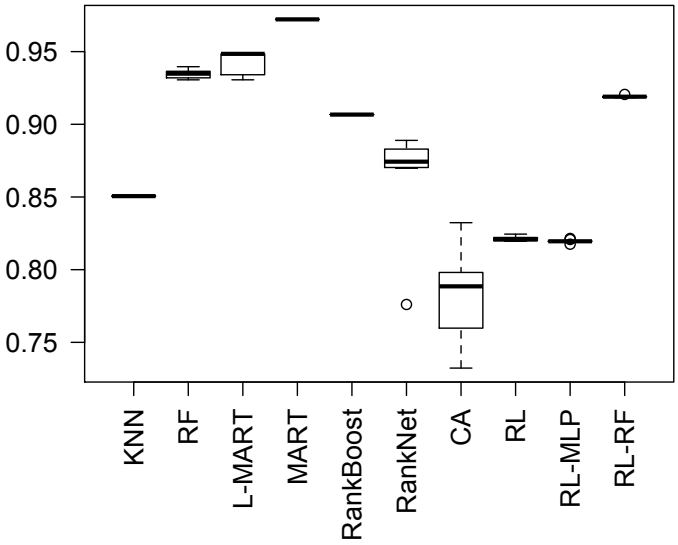
Table 1: Algorithms

Algorithm	Parameters	
KNN	k=4	
RF	Tree=100	SplitCriterion = Information
RF	Tree=100	Gain Ratio
L-MART	Tree=30	Leaf=10
MART	Tree=1000	Leaf=10
RankBoost	round = 300	
RankNet	Epoch=50	Layer=2
CA	randomRestarts=5	iterations=25
RL	Layer=1	Neurons=12
RL-MLP	Layer=4	Neurons=12
RL-RF	Tree=100	SplitCriterion = Information
RL-RF	Tree=100	Gain Ratio

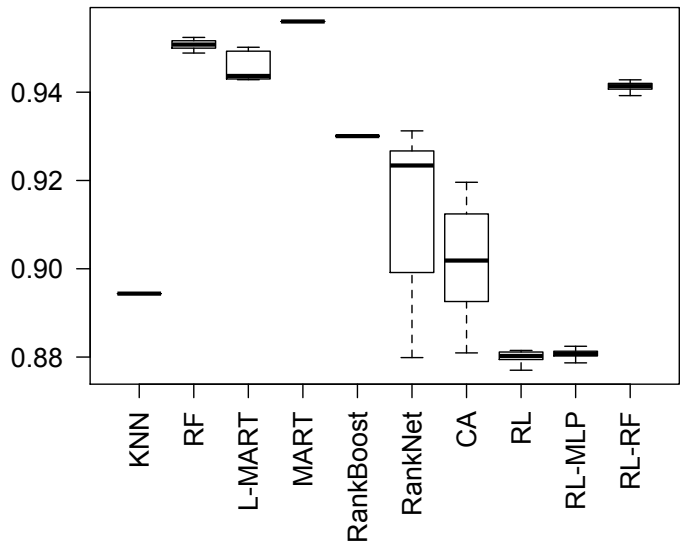
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICSE 2020, May 23–29, 2020, Seoul, South Korea
© 2020 Association for Computing Machinery.
https://doi.org/10.1145/1122445.1122456



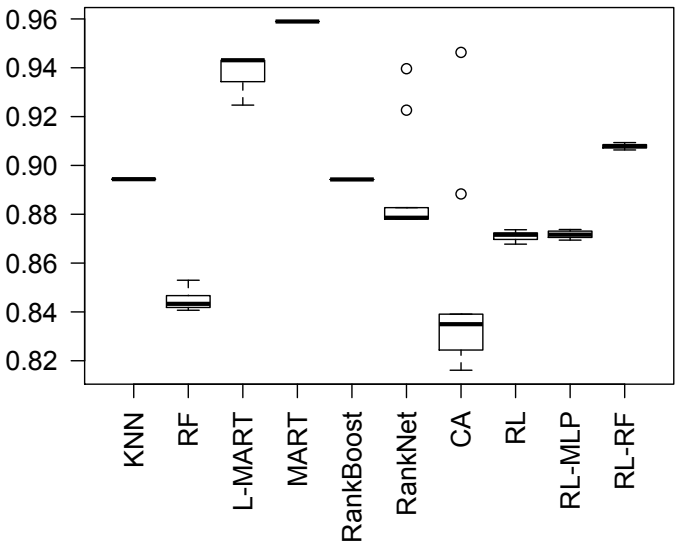
(a) commons-codec



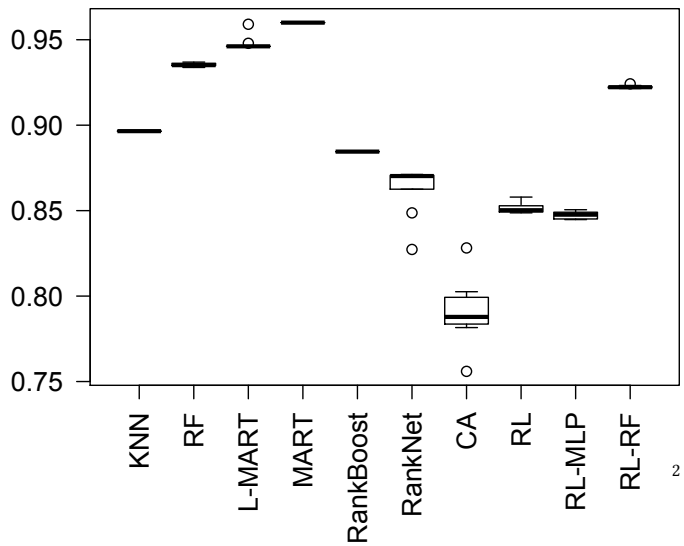
(b) commons-compress



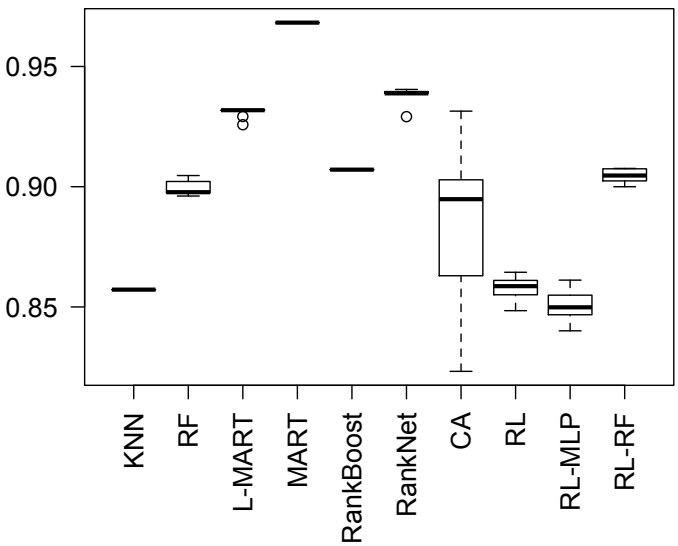
(c) commons-imaging



(d) commons-io

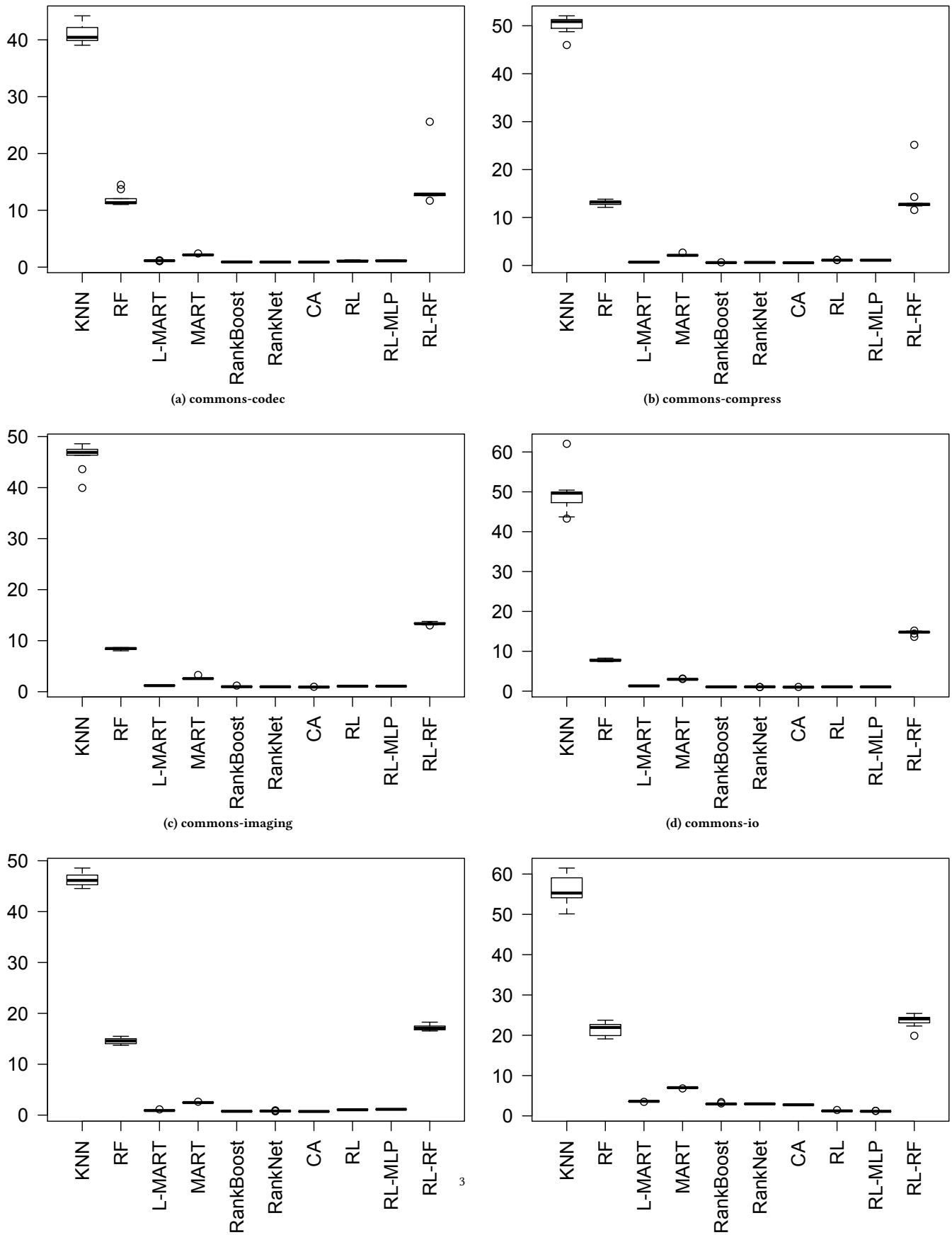


(e) commons-lang



(f) commons-math

Figure 1: RPA



(a) commons-codec

(b) commons-compress

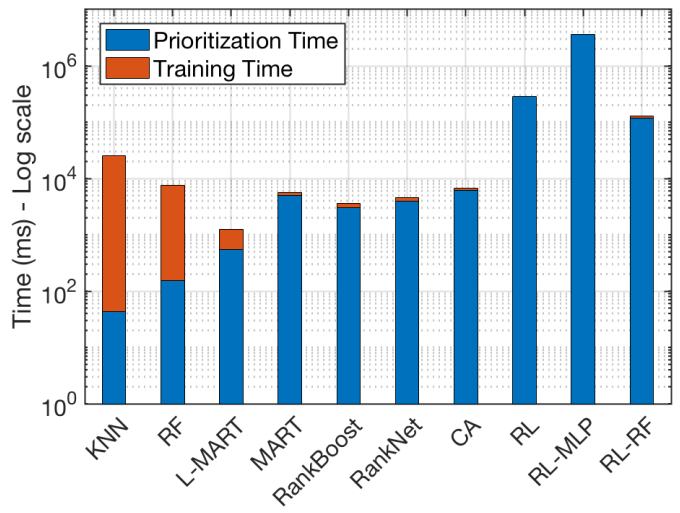
(c) commons-imaging

(d) commons-io

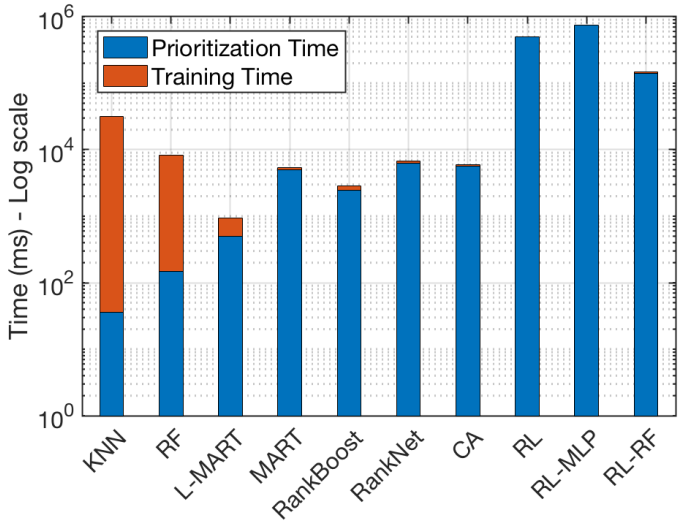
(e) commons-lang

(f) commons-math

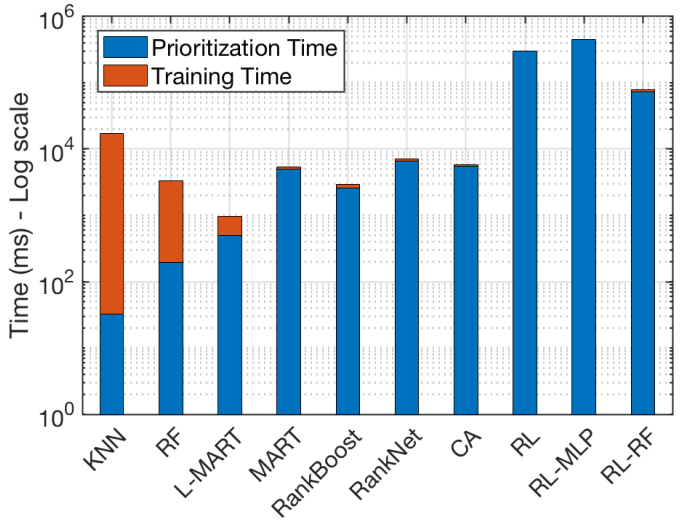
Figure 2: Ranking time



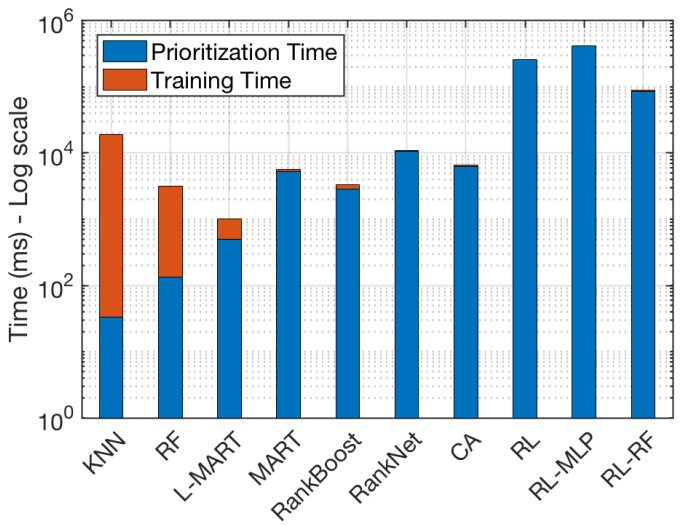
(a) commons-codec



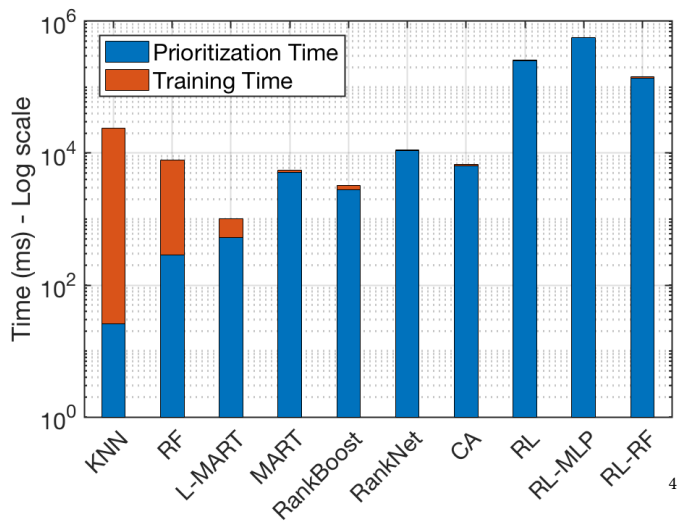
(b) commons-compress



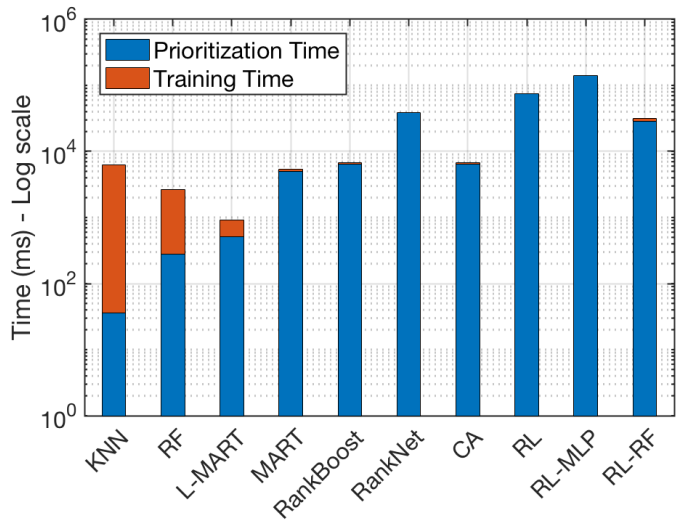
(c) commons-imaging



(d) commons-io

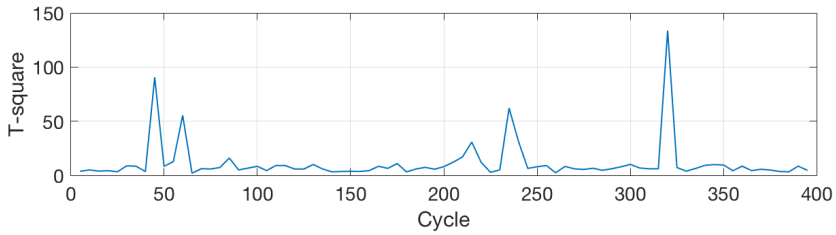


(e) commons-lang

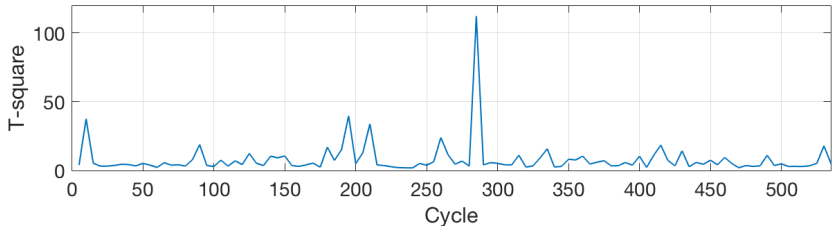


(f) commons-math

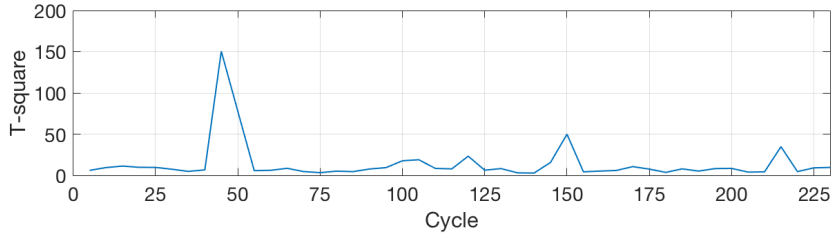
Figure 3: Training time



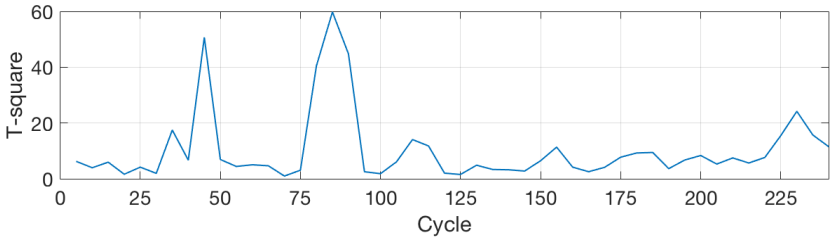
(a) commons-codec



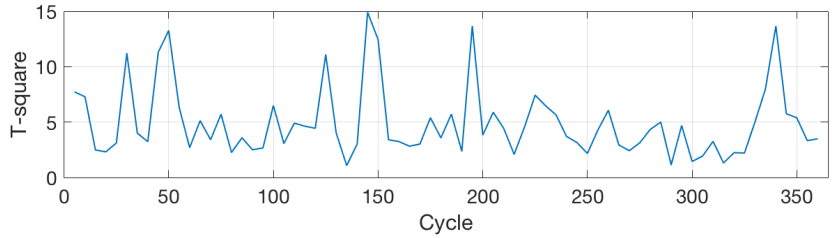
(b) commons-compress



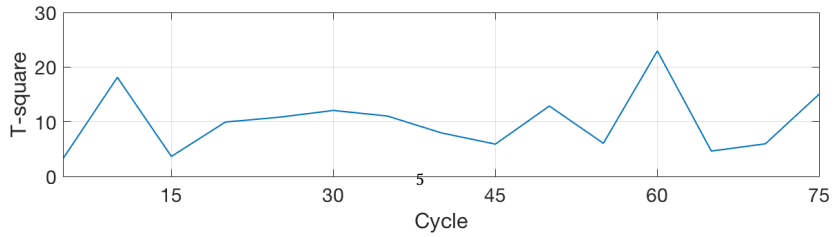
(c) commons-imaging



(d) commons-io



(e) commons-lang



(f) commons-math

Table 2: Optimal-predicted difference of tests execution times (ms), averaged over all the commits and subjects, and of total number of failing tests, averaged over subjects

(a) commons-codec

Algorithms	Time-constrained scenarios					
	25%		*50%*		*75%*	
	Time	Failures	Time	Failures	Time	Failures
KNN	0.2846	0.0000	0.5775	0.0000	0.9354	0.0000
RF	0.0196	0.0000	0.1166	0.0000	0.3651	0.0000
L-MART	0.0298	0.0000	0.3240	0.0000	0.6993	0.0000
MART	0.0103	0.0000	0.0680	0.0000	0.0205	0.0000
RankBoost	0.2944	0.0000	0.5288	0.0000	0.5511	0.0000
RankNet	0.2633	0.0000	0.4158	0.0000	0.5592	0.0000
CA	0.6300	-5.4000	0.8352	0.0000	0.8903	0.0000
RL	0.2967	-1.0000	0.5357	0.0000	0.6005	0.0000
RL-MLP	0.2707	-1.0000	0.5257	0.0000	0.5641	0.0000
RL-RF	0.0510	-1.0000	0.2363	-1.0000	0.5359	-1.0000

(b) commons-compress

Algorithms	Time-constrained scenarios					
	25%		*50%*		*75%*	
	Time	Failures	Time	Failures	Time	Failures
KNN	0.2951	0.0000	0.7672	0.0000	1.7559	0.0000
RF	0.0386	-5.2000	0.4788	-1.2000	1.3699	0.0000
L-MART	0.0366	-6.8000	0.2798	-4.8000	0.7783	-0.8000
MART	0.0120	0.0000	0.0136	0.0000	0.0819	0.0000
RankBoost	0.2848	0.0000	0.9204	0.0000	1.1516	0.0000
RankNet	0.3164	-4.0000	0.8582	-1.6000	1.2750	-0.4000
CA	0.8784	0.0000	1.4143	-4.2000	1.7709	-3.0000
RL	0.6250	-15.5000	1.1784	-12.2000	1.6203	-8.4000
RL-MLP	0.6521	-17.4000	1.2036	-13.3000	1.6207	-6.2000
RL-RF	0.0311	-22.0000	0.6720	-10.8000	1.8496	-5.8000

(c) commons-imaging

Algorithms	Time-constrained scenarios					
	25%		*50%*		*75%*	
	Time	Failures	Time	Failures	Time	Failures
KNN	0.9200	0.0000	0.9884	0.0000	4.0614	0.0000
RF	0.0814	0.0000	0.6912	0.0000	3.8510	0.0000
L-MART	0.4546	0.0000	1.3764	0.0000	1.9815	0.0000
MART	0.0664	-2.0000	0.0188	0.0000	0.1173	0.0000
RankBoost	1.0048	0.0000	1.5274	0.0000	2.1156	0.0000
RankNet	1.2028	0.0000	2.1866	0.0000	3.3407	0.0000
CA	1.0650	-1.2000	2.6108	0.0000	3.6653	0.0000
RL	1.7687	-1.4000	3.3307	-0.6000	4.6458	-0.1000
RL-MLP	1.7952	-1.5000	3.4279	-1.0000	4.7758	-0.7000
RL-RF	0.0529	-1.2000	1.7756	0.0000	5.0547	0.0000

(d) commons-io

Algorithms	Time-constrained scenarios					
	25%		*50%*		*75%*	
	Time	Failures	Time	Failures	Time	Failures
KNN	1.5958	-2.0000	4.9703	-2.0000	15.9193	-2.0000
RF	5.3489	-1.0000	12.7569	-0.8000	16.5187	-0.3000
L-MART	0.6970	-2.0000	2.4218	-2.0000	8.9455	-1.9000
MART	0.0170	-4.0000	0.2378	-2.0000	0.1483	-2.0000
RankBoost	8.0699	-2.0000	9.7164	-1.0000	10.3045	-1.0000
RankNet	4.2097	-0.4000	8.0319	-0.4000	9.1499	-0.3000
CA	10.3991	-2.2000	13.5966	-0.7000	14.8289	-0.5000
RL	5.4701	-3.0000	9.6149	-2.8000	13.5486	-1.3000
RL-MLP	5.0147	-3.2000	9.6170	-3.0000	13.4622	-0.6000
RL-RF	3.1774	-6.0000	9.3008	-6.0000	16.6132	-3.0000

(e) commons-lang

Algorithms	Time-constrained scenarios					
	25%		*50%*		*75%*	
	Time	Failures	Time	Failures	Time	Failures
KNN	0.0564	-4.0000	0.5346	-4.0000	2.1290	0.0000
RF	0.0109	-3.6000	0.2227	-3.6000	1.4451	0.0000
L-MART	0.1703	-4.0000	1.1242	-4.0000	1.7714	-2.0000
MART	0.0115	-6.0000	0.0244	-2.0000	0.0605	-2.0000
RankBoost	0.4790	0.0000	1.6340	0.0000	2.0571	0.0000
RankNet	0.1823	0.0000	0.8609	0.0000	0.9708	0.0000
CA	0.9939	-6.0000	1.5273	-1.4000	1.6984	-1.0000
RL	0.5868	-3.0000	1.1955	-1.5000	1.6581	-0.7000
RL-MLP	0.6391	-2.8000	1.1720	-2.0000	1.6530	-0.9000
RL-RF	0.0281	-4.0000	0.3530	-4.0000	1.8433	0.0000

(f) commons-math

Algorithms	Time-constrained scenarios					
	25%		*50%*		*75%*	
	Time	Failures	Time	Failures	Time	Failures
KNN	0.1643	-6.0000	2.1354	-6.0000	4.4231	-6.0000
RF	0.1120	-6.0000	1.3305	-6.0000	2.9984	-6.0000
L-MART	0.0826	-6.0000	0.4000	-6.0000	0.5274	-6.0000
MART	0.0150		0.0177	-6.0000	0.0250	-6.0000
RankBoost	0.3496	-6.0000	0.8658	-6.0000	3.2967	-6.0000
RankNet	0.1544	-6.0000	2.5655	-6.0000	2.4697	-6.0000
CA	1.0478		1.4488	-6.0000	2.0388	-6.0000
RL	2.5887	-7.4000	5.1019	-4.5000	6.9543	-2.8000
RL-MLP	2.2968	-6.8000	5.0432	-4.6000	6.9328	-3.4000
RL-RF	0.1372	-9.0000	1.4215	-5.4000	4.8902	-4.3000