A Novel Automated Approach for Software Effort Estimation Based on Data Augmentation

LIYAN SONG[1,2], LEANDRO L. MINKU[2], AND XIN YAO[1,2]

[1] *Shenzhen Key Laboratory of Computational Intelligence, Department of Computer Science and Engineering, Shenzhen, China, 518055*
[2] *CERCIA, School of Computer Science, University of Birmingham, UK, B15 2TT*
[*] *Emails: songly@sustc.edu.cn, L.L.Minku@cs.bham.ac.uk, AND xiny@sustc.edu.cn*

1. EFFECT OF SYNTHETIC DATA ON PERFORMANCE IN THE ORIGINAL SPACE

The performance metric reported in our paper [1] is MAE with the effort in the logarithm scale as it is less affected by project size. In this supplementary material, we report the MAE with the effort in the original scale. Specifically, once an SEE method provides an effort estimate, take the exponentiation of that value as the predicted effort and compare that value to the actual effort for calculating MAE. Table S1 lists the performance comparisons in all settings of data set sizes. We can see that the key conclusion that syn.SEEr always performed similarly/better than bsl.SEEr remains the same.

Similar to the MAE in the logarithm effort scale, the performance of LR/ATLM is unstable in some data sets. For instance, ATLM performs extremely poorly in many of the investigated data sets, with infinite MAE (mean MAE of 30 runs) when the training set size is small. Further investigation found that ATLM performs extremely poorly on one or two of the 30 runs. As discussed in section 5.1.1, the unstable performance of LR/ATLM may be due to the scarcity of training data and the consequent ill-conditional problem when doing matrix inversion in the training process. This problem would be even worse in the original effort space.

REFERENCES

1. Liyan Song, Leandro L. Minku, and Xin Yao. 2018. A Novel Automated Approach for Software Effort Estimation based on Data Augmentation. In *ACM Symposium on the Foundations of Software Engineering (FSE)*. Lake Buena Vista, USA.

Table S1. Performance comparisons between each pair of *syn.SEEr* vs *bsl.SEEr* across 14 data sets in terms of MAE in the original effort space for small, medium and large training set sizes respectively. The different training set sizes refer to different *holdout* values of table 1. The reported values are the mean of 30 runs followed by their standard deviations (STDs). The comparison is highlighted in orange (dark grey) and bold font for large, in yellow (light grey) and bold font for medium, and in bold font for small effect size values. The last two rows of each sub table list the results of Wilcoxon tests with Bonferroni correction. The overall comparison between *bsl.SEEr* vs *syn.SEEr* can be seen from *aveRank* (average ranks). The first value 1 (or 0) in *Wilcoxon* row means there is (or not) significant difference detected, and its corresponding *p*-value comes the next. Significant difference is highlighted in orange (dark grey) on this row. 'Inf' indicates *infinite* error in terms of MAE, which only happened to LR or ATLM with small training set sizes. Further investigation found that only one or two out of the total 30 runs committed extremely large error.

(a) **Small training set size.**

Data	syn.LR	bsl.LR	syn.ATLM	bsl.ATLM	syn.*k*-NN	bsl.*k*-NN	syn.RVM	bsl.RVM	syn.RT	bsl.RT	syn.SVR	bsl.SVR
Maxwell	**5207.4±1329.3**	**49636.7±82397.5**	**5654.1±2005.6**	**Inf±NaN**	5244.0±868.6	5285.7±858.9	**4262.7±788.4**	**4728.6±981.5**	**5147.4±693.5**	**5335.0±749.4**	**4230.8±835.1**	**4532.9±850.9**
Cocomo81	**450.8±217.6**	**Inf±Inf**	**648.9±476.4**	**Inf±NaN**	625.5±123.3	625.2±136.5	**390.0±155.6**	**451.3±136.0**	607.4±122.8	627.1±130.1	402.5±128.2	421.2±145.8
Nasa93	**435.6±185.9**	**Inf±Inf**	**552.0±398.5**	**Inf±Inf**	470.7±75.8	478.4±73.0	**317.5±103.8**	**335.8±101.8**	**440.0±77.5**	**467.6±68.8**	339.3±86.7	331.2±69.4
Kitchenham	**2122.3±1141.1**	**3230.0±2871.6**	3850.1±4919.2	5875.8±8794.2	2277.5±317.6	2305.9±350.1	**2171.6±485.4**	**2323.6±364.8**	**2372.9±335.5**	**2413.1±249.5**	**1877.2±616.0**	**2046.8±391.3**
Albrecht	**24.7±48.5**	**Inf±Inf**	**30.1±51.7**	**Inf±NaN**	12.9±4.2	12.5±4.4	**11.0±4.6**	**12.5±5.9**	**15.0±4.8**	**16.3±3.4**	9.7±3.5	10.3±4.0
Kemerer	751.9±2829.2	2272.1±10472.1	**1001.6±3209.8**	**Inf±NaN**	132.2±43.0	140.3±38.1	132.5±50.8	130.2±46.7	151.3±40.5	154.8±32.5	114.8±41.4	118.4±42.3
Deshar	**3851.0±3064.1**	**Inf±Inf**	**4655.2±5003.8**	**Inf±Inf**	2711.9±316.0	2728.7±380.7	2579.1±402.8	2640.7±533.1	**2752.8±346.3**	**2949.0±284.4**	2324.3±264.8	2326.3±310.8
Org1	**5468.8±6301.4**	**Inf±Inf**	**29722.3±88952.6**	**Inf±NaN**	3674.6±804.8	3557.5±378.5	3560.1±365.9	3590.3±407.7	3862.3±678.6	3777.4±440.4	**3476.9±335.2**	**3481.6±384.4**
Org2	273853.8±1480233.9	293784.7±1468380.1	3452.7±2447.7	59061.8±247795.7	2284.0±896.0	2207.2±687.8	2129.6±520.1	2042.0±396.7	2403.3±1054.6	2335.8±825.1	2158.0±720.3	2186.0±874.7
Org3	1291.1±394.8	1434.8±714.0	1310.9±478.7	1494.6±770.8	**1296.9±150.4**	**1312.0±166.5**	**1284.9±255.7**	**1383.7±181.6**	**1365.0±204.5**	**1549.2±118.7**	**1158.3±200.6**	**1217.4±220.8**
Org4	**7623.3±10567.0**	**Inf±Inf**	**6742.3±5388.8**	**Inf±NaN**	4693.1±774.8	4416.2±404.2	4442.2±271.6	4411.3±257.1	4618.0±810.7	4586.8±813.0	4275.5±316.5	4426.0±717.8
Org5	1034100.0±5160676.2	Inf±Inf	**138289.8±418598.6**	**Inf±Inf**	6979.1±1234.0	7129.3±1101.8	7191.9±1707.4	7218.3±967.8	7684.0±1575.6	7509.1±1304.0	**6388.8±2096.3**	**6905.9±1988.9**
Org6	**1309619.1±6963276.4**	**Inf±Inf**	**2683972.5±9401040.3**	**Inf±Inf**	3503.4±919.3	3440.5±839.1	3468.3±817.2	3635.1±805.0	4544.8±1948.0	4631.0±1880.4	**3455.7±1500.7**	**3634.1±1081.8**
Org7	**2803187.7±14862957.4**	**Inf±Inf**	**3333716.6±15060437.3**	**Inf±Inf**	5889.4±1619.0	5529.5±812.5	5577.9±1042.1	5447.3±617.8	5759.9±1314.1	5529.5±812.5	5635.9±1240.5	5396.4±779.7
aveRank	1.00	2.00	1.00	2.00	1.50	1.50	1.29	1.71	1.36	1.64	1.14	1.86
Wilcoxon	1	0.000091	1	0.000091	0	0.669800	0	0.135254	0	0.625732	1	0.003763

(b) **Medium training set size.**

Data	syn.LR	bsl.LR	syn.ATLM	bsl.ATLM	syn.*k*-NN	bsl.*k*-NN	syn.RVM	bsl.RVM	syn.RT	bsl.RT	syn.SVR	bsl.SVR
Maxwell	**3628.6±1149.4**	**6669.2±7619.6**	**3670.4±1152.7**	**10361.5±22773.4**	4870.7±1984.5	5071.2±1972.5	**3921.5±1019.3**	**4381.8±1340.7**	4656.1±1618.0	4661.2±1786.7	3631.9±1327.4	3766.8±1302.2
Cocomo81	258.9±159.4	261.9±131.6	**295.1±172.4**	**506.3±390.2**	522.6±286.2	548.0±283.9	262.2±164.8	261.3±169.3	**445.6±236.3**	**497.0±231.5**	256.8±151.3	266.1±137.7
Nasa93	327.2±137.1	260.1±157.7	327.2±137.1	260.1±157.7	464.3±152.4	473.6±161.4	246.7±80.3	251.8±79.9	387.2±148.8	393.1±148.9	296.2±113.3	280.4±103.8
Kitchenham	1673.2±282.4	9698.5±42660.7	1723.3±383.6	10005.9±42632.8	2032.3±231.0	2038.2±247.4	**1848.3±282.6**	**2034.8±369.1**	2203.7±302.7	2200.3±290.3	1668.4±364.4	1782.4±478.7
Albrecht	**9.3±4.5**	**18.3±23.0**	**10.3±8.0**	**21.1±24.0**	9.4±4.9	9.4±4.8	8.6±3.9	8.3±3.6	**10.7±4.9**	**14.6±5.9**	8.6±4.3	8.3±3.7
Kemerer	**122.5±83.1**	**1298.4±3830.9**	**118.0±81.8**	**2472.5±7938.6**	111.3±79.3	119.1±79.5	105.5±73.8	112.8±79.4	128.1±79.6	137.9±69.1	97.3±70.7	98.9±74.2
Deshar	**2190.0±355.8**	**2586.6±846.0**	**2188.8±357.8**	**2621.1±850.9**	2436.6±217.8	2425.6±243.5	2227.7±341.5	2289.7±422.4	2570.8±276.4	2590.4±330.7	2018.9±257.9	2035.7±276.1
Org1	3411.0±456.7	3893.2±1931.0	3583.9±703.9	4677.7±4655.4	3463.9±392.2	3488.2±413.2	3552.1±408.3	3567.4±382.3	3692.4±232.8	3632.4±183.2	**3261.0±351.2**	**3451.4±411.2**
Org2	1848.9±310.5	1887.1±421.2	1846.3±318.4	1887.6±429.9	1978.0±318.1	1957.4±282.4	1880.4±361.0	1827.9±357.4	**1954.0±253.0**	**2086.4±299.4**	1817.2±333.8	1833.5±302.0
Org3	1124.0±215.7	1168.6±358.5	1128.1±219.5	1174.2±365.9	1209.1±96.1	1214.3±103.5	**1138.0±130.0**	**1239.2±160.4**	**1251.6±128.0**	**1297.5±77.2**	1081.2±163.6	1068.3±120.1
Org4	**4192.4±449.7**	**8658.1±19680.0**	**4251.6±559.6**	**9879.7±21191.7**	4209.8±435.3	4202.6±379.1	**4052.5±335.1**	**4470.0±452.3**	**4199.7±265.5**	**4374.5±258.5**	**3991.5±287.2**	**4110.5±351.2**
Org5	**7268.3±6623.5**	**23826.5±56644.7**	10483.1±12093.6	27476.3±57745.5	5988.1±1400.1	5944.9±1386.5	**5880.4±1650.5**	**6549.7±2135.5**	**6066.5±1532.7**	**6848.0±1362.1**	4424.9±1586.1	4625.8±1680.3
Org6	**11502.9±20457.8**	**90310.8±394316.3**	**35407.4±97997.7**	**167984.4±527885.4**	2894.5±783.8	2960.8±836.8	**3092.0±825.6**	**3900.1±2564.6**	**3240.8±697.8**	**3830.4±742.2**	**2886.0±1025.2**	**3179.1±1074.6**
Org7	5138.5±825.4	15477.8±58189.8	5294.1±1211.4	15549.2±58178.2	5192.4±856.9	5236.9±747.3	**5011.7±568.7**	**5684.5±1439.5**	**4959.0±993.1**	**5190.6±781.5**	4836.5±779.0	4904.0±818.5
aveRank	1.07	1.93	1.07	1.93	1.29	1.71	1.21	1.79	1.14	1.86	1.21	1.79
Wilcoxon	1	0.000670	1	0.000670	0	0.172607	1	0.016255	1	0.003763	1	0.016255

(c) **Large training set size.**

Data	syn.LR	bsl.LR	syn.ATLM	bsl.ATLM	syn.*k*-NN	bsl.*k*-NN	syn.RVM	bsl.RVM	syn.RT	bsl.RT	syn.SVR	bsl.SVR
Maxwell	4314.8±4878.2	4089.5±5153.2	4326.1±4891.4	4180.9±5174.5	6561.9±10529.7	6606.7±10506.1	3983.2±4636.6	4999.8±6782.3	**4450.5±7220.8**	**4806.7±7286.5**	3859.9±3985.7	3562.1±3844.0
Cocomo81	232.1±401.6	247.8±531.5	248.6±449.0	423.7±1360.0	326.5±454.4	387.8±485.6	136.2±220.2	172.4±279.4	**242.5±421.3**	**359.9±520.5**	183.8±340.5	206.0±375.4
Nasa93	174.8±250.6	130.5±190.6	174.8±250.6	130.5±190.6	282.6±458.5	302.3±463.6	135.5±171.4	184.6±214.6	177.9±363.7	166.1±263.9	117.2±156.3	147.2±206.8
Kitchenham	1552.2±849.6	1562.6±861.0	1483.3±771.7	1499.2±767.0	1802.0±1042.0	1817.2±1089.8	1513.3±847.2	1551.3±902.0	**1893.0±1080.8**	**1977.6±1045.2**	1412.9±705.4	1410.8±684.8
Albrecht	6.3±6.8	11.5±31.6	6.3±6.8	11.5±31.6	6.7±10.2	8.5±14.6	5.5±5.6	7.0±7.8	**6.9±6.9**	**11.7±12.0**	5.8±9.4	6.3±11.4
Kemerer	110.3±194.3	115.6±193.3	106.8±194.5	112.8±194.0	119.2±207.2	120.4±208.0	85.5±148.0	103.1±227.7	125.1±220.6	145.1±222.0	99.7±200.7	101.5±197.6
Deshar	2030.8±420.4	2114.8±442.7	2031.5±419.5	2117.0±443.7	2325.5±487.4	2377.4±456.9	2011.1±474.4	2089.2±483.2	2239.2±572.5	2321.3±707.8	2016.4±425.3	2041.7±422.8
Org1	2779.6±2475.8	2764.4±2417.1	2914.8±2447.3	3077.2±2420.1	3448.6±2740.6	3431.4±2938.9	3153.3±3000.7	3176.2±3043.7	**3378.9±2829.1**	**3933.4±2645.8**	2849.2±2521.8	2881.8±2446.5
Org2	1575.7±442.9	1582.9±486.4	1565.9±447.4	1567.1±485.5	1772.5±501.5	1835.2±489.7	1631.7±537.7	1636.4±587.9	1738.2±465.8	1684.7±510.9	1524.7±396.8	1520.6±439.7
Org3	944.6±199.9	942.4±206.3	944.7±197.2	942.4±206.3	1065.4±230.5	1078.7±216.8	950.9±206.1	964.5±220.7	998.3±222.3	976.0±235.2	931.1±193.7	929.4±200.4
Org4	3909.8±879.0	3963.8±859.0	3944.6±878.5	4046.7±917.9	4126.1±1009.8	4239.5±939.1	3753.7±880.5	3815.0±892.5	3997.0±899.0	4074.2±946.9	3838.6±837.2	3896.8±794.7
Org5	3806.3±1886.2	4671.7±3252.0	3976.3±1817.3	4733.0±3233.7	5628.3±2677.5	5783.1±3051.8	**4478.5±2463.8**	**5715.3±3091.7**	5515.4±3079.3	5719.0±3411.1	3767.8±1656.0	3364.6±1482.1
Org6	4269.9±5608.5	10318.5±12906.8	4580.1±5872.7	10343.3±12890.4	2909.3±1396.8	2838.7±1415.9	2135.7±1039.4	2006.8±781.1	2864.0±1368.8	2845.3±1401.7	2085.0±891.3	2165.6±743.8
Org7	4648.4±1676.9	4656.7±1571.9	4660.3±1696.1	4663.1±1595.4	5065.9±1561.5	4915.4±1565.3	4649.9±1709.4	4675.9±1643.7	4458.3±1427.5	4394.8±1541.3	4470.1±1448.1	4366.9±1328.7
aveRank	1.29	1.71	1.21	1.79	1.21	1.79	1.07	1.93	1.36	1.64	1.43	1.57
Wilcoxon	0	0.153076	1	0.016255	0	0.153076	1	0.000670	0	0.067627	0	0.903198