## Less is more: Temporal fault predictive performance over multiple Hadoop releases

June 6, 2014

This document contains additional data and results related to the paper M. Harman, S. Islam, Y. Jia, L. Minku, F. Sarro and K. Srivisut, Less is more: Temporal fault predictive performance over multiple Hadoop releases, which is currently to appaer in the proceedings of the 5th International Symposium on Search-Sased Software Enginering 2014 .

In the paper, we address the challenge of understanding the way in which information about different versions of the software system impacts upon our ability to define search based fault prediction systems over time. We extracted and curated data from 8 versions of the Hadoop system, using it to train a search based fault prediction system. Our prediction system uses a Genetic Algorithm to train a Support Vector Machine, which predicts whether a class is faulty. This is the first time that results have been reported on temporal fault predictive performance, over multiple releases.

Our results reveal that, as expected, overall predictive performance (measured using G-mean) is statistically significantly better when augmented with the data from the entire version history. However, perhaps more surprisingly, we also found that, for half the versions considered, Recall is statistically significantly better using *solely* the previous version. Therefore, this study calls for a fundamental change in the way we view software defect prediction, in order to take chronology into account.

 $<sup>^1</sup>$ All data is available at http://www0.cs.ucl.ac.uk/staff/F.Sarro/projects/hadoop/.

Noc	Version	Metric	Min	Max	Mean	Median	St.Dev
CBO         1         93         16.82         10         17.72           RFC         0         307         49.96         25         62.42           LCOM         0         1492         75.07         4         212.73           NPM         0         55         11.96         7         13.01           LOC         1         2867         391.32         136         595.74           hadoop-0.2         DIT         0         10         1.38         1         1.46           NOC         0         7         0.34         0         1.17           CBO         0         135         11.12         6         15.59           RFC         0         353         32.78         15         47.74           LCOM         0         1899         47.24         2         202.98           NPM         0         76         7.97         4         10.93           LOC         1         15765         293.69         80         618.33           hadoop-0.3         DIT         0         11         1.64         1         1.81           LOC         1         160         14.92	hadoop-0.1	DIT	0	15	2.01	1	2.38
RFC         0         307         49.96         25         62.42           LCOM         0         1492         75.07         4         212.73           NPM         0         55         11.96         7         13.01           LOC         1         2867         391.32         136         595.74           hadoop-0.2         DIT         0         10         1.38         1         1.46           NOC         0         7         0.34         0         1.17         6         15.59           RFC         0         353         32.78         15         47.74         10.99           LCOM         0         1899         47.24         2         202.98         NPM         0         76         7.97         4         10.99         10.13         16.14         1         1.81         1.00         11         1.64         1         1.81         1.00         1.09         1.11         1.64         1         1.81         1.09         1.09         1.01         1         1.14         1         1.65         1.09         1.14         1         1.65         1.00         1         1.00         1.18         1.00 <td< td=""><td></td><td>NOC</td><td>0</td><td>7</td><td>0.32</td><td>0</td><td>1.03</td></td<>		NOC	0	7	0.32	0	1.03
LCOM					16.82	10	
NPM						25	
NOC							
hadoop-0.2   DIT							
NOC							
CBO         0         135         11.12         6         15.59           RFC         0         353         32.78         15         47.74           LCOM         0         1899         47.24         2         202.98           NPM         0         76         7.97         4         10.99           LOC         1         5765         293.69         80         618.33           hadoop-0.3         DIT         0         11         1.64         1         1.81           NOC         0         8         0.26         0         1         1.81           NOC         0         8         0.26         0         1         6.64         4         232.69         16.55         4         68.45         4         232.69         16.55         16.84         4         232.69         16.54         4         232.69         16.54         4         232.69         16.54         4         232.69         16.54         4         232.69         16.54         4         232.69         18.41         1         1.65         16.44         4         232.69         18.41         1         1.65         16.44         4         232.69	hadoop-0.2						
RFC							
LCOM							
NPM							
LOC							
hadoop-0.3   DIT							
NOC	hadoon 0.3						
CBO         1         100         14.92         9         16.55           RFC         0         386         49.15         24         68.45           LCOM         0         2194         65.44         4         232.69           NPM         0         70         10.32         4         13.38           LOC         1         6180         434.06         126         819.42           hadoop-0.4         DTT         0         9         1.41         1         1.65           NOC         0         8         0.27         0         1.01           CBO         1         93         13.53         8         15.08           RFC         0         402         42.39         24         56.38           LCOM         0         2270         64.14         4         250.35           NPM         0         75         9.12         4         12.35           LOC         1         6404         367.82         130         692.37           hadoop-0.5         DIT         0         14         1.46         1         1.72           NOC         0         9         0.32 <t< td=""><td>nadoop-0.5</td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	nadoop-0.5						
RFC         0         386         49.15         24         68.45           LCOM         0         2194         65.44         4         232.69           NPM         0         70         10.32         4         13.38           LOC         1         6180         434.06         126         819.42           hadoop-0.4         DIT         0         9         1.41         1         1.65           NOC         0         8         0.27         0         1.01           CBO         1         93         13.53         8         15.08           RFC         0         402         42.39         24         56.38           LCOM         0         2270         64.14         4         250.35           NPM         0         75         9.12         4         12.35           LOC         1         6404         367.82         130         692.37           hadoop-0.5         DIT         0         14         1.46         1         1.72           NOC         0         9         0.32         0         1.18           RFC         0         343         45.28 <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>							
LCOM							
NPM							
LOC							
hadoop-0.4   DIT							
NOC	hadoop-0.4						
CBO         1         93         13.53         8         15.08           RFC         0         402         42.39         24         56.38           LCOM         0         2270         64.14         4         250.35           NPM         0         75         9.12         4         12.35           LOC         1         6404         367.82         130         692.37           hadoop-0.5         DIT         0         14         1.46         1         1.72           NOC         0         9         0.32         0         1.18           CBO         1         88         14.74         9         15.38           RFC         0         343         45.28         32         50.97           LCOM         0         2544         65.93         5         253.01           NPM         0         70         9.49         5         11.64           LOC         1         5059         394.41         188         625.37           hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35 <td< td=""><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></td<>	1						
LCOM		CBO					
NPM		RFC	0	402	42.39	24	56.38
LOC         1         6404         367.82         130         692.37           hadoop-0.5         DIT         0         14         1.46         1         1.72           NOC         0         9         0.32         0         1.18           CBO         1         88         14.74         9         15.38           RFC         0         343         45.28         32         50.97           LCOM         0         2544         65.93         5         253.01           NPM         0         70         9.49         5         11.64           LOC         1         5059         394.41         188         625.37           hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35         0         1.17           CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05 <t< td=""><td></td><td>LCOM</td><td>0</td><td>2270</td><td>64.14</td><td>4</td><td>250.35</td></t<>		LCOM	0	2270	64.14	4	250.35
hadoop-0.5         DIT         0         14         1.46         1         1.72           NOC         0         9         0.32         0         1.18           CBO         1         88         14.74         9         15.38           RFC         0         343         45.28         32         50.97           LCOM         0         2544         65.93         5         253.01           NPM         0         70         9.49         5         11.64           LOC         1         5059         394.41         188         625.37           hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35         0         1.17           CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         1		NPM	0	75	9.12	4	12.35
NOC		LOC	1	6404	367.82	130	692.37
CBO         1         88         14.74         9         15.38           RFC         0         343         45.28         32         50.97           LCOM         0         2544         65.93         5         253.01           NPM         0         70         9.49         5         11.64           LOC         1         5059         394.41         188         625.37           hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35         0         1.17         CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01         1         100         17         10.05         5         14.07           LCOM         0         2701         68.94         5         267.72         14.07         100         1         10.05         5         14.07         14.07         100         1         10.05         5         14.07         10.06         1         2.35         10         1.08         10.07         10.08         10.07         1.08         10	hadoop-0.5		0	14		1	1.72
RFC         0         343         45.28         32         50.97           LCOM         0         2544         65.93         5         253.01           NPM         0         70         9.49         5         11.64           LOC         1         5059         394.41         188         625.37           hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35         0         1.17           CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         <							
LCOM   NPM   0   2544   65.93   5   253.01     NPM   DOC   1   5059   394.41   188   625.37     hadoop-0.6   DIT   DOC   DIT   OCBO   DIT   DOC   DIT     CBO   DIT   DOC   DIT   DOC   DIT     CBO   DIT   DOC   DIT   DOC   DOC   DOC     RFC   DOC   DOC   DOC   DOC     RFC   DOC   DOC   DOC   DOC     RFC   DOC   DOC   DOC     LCOM   DOC   DOC   DOC     LCOM   DOC   DOC   DOC     DIT   DOC   DIT   DOC   DOC     LOC   DIT   DOC   DOC     NOC   DOC   DOC   DOC     NOC   DOC   DOC   DOC     NOC   DOC   DOC     DIT   DOC   DOC   DOC     DOC   DOC     DOC   DOC     DOC   DOC   DOC     DOC   DOC     DOC   DOC   DOC     DOC							
NPM         0         70         9.49         5         11.64           LOC         1         5059         394.41         188         625.37           hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35         0         1.17           CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2752         70.5 <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>							
LOC         1         5059         394.41         188         625.37           hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35         0         1.17           CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2752         70.5         4         288.88           NPM         0         76         9.88 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>							
hadoop-0.6         DIT         0         20         1.59         1         2.38           NOC         0         9         0.35         0         1.17           CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2752         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         1							
NOC 0 9 0.35 0 1.17 CBO 1 101 17.29 9 21.37 RFC 0 420 48.15 28 63.01 LCOM 0 2701 68.94 5 267.72 NPM 0 71 10.05 5 14.07 LOC 1 5646 413.02 167 725.04  hadoop-0.7 DIT 0 21 1.66 1 2.35 NOC 0 9 0.35 0 1.08 CBO 1 132 17.96 9 23.85 RFC 1 483 49.69 28 68.33 LCOM 0 2752 70.5 4 288.88 NPM 0 76 9.88 4 13.31 LOC 1 6294 418.04 172 719.82  hadoop-0.8 DIT 0 20 1.6 1 2.36 NOC 0 9 0.31 0 1.06 CBO 1 125 16.16 9 20.11 RFC 0 533 46.24 26 63.96 LCOM 0 2850 65.18 4 293.09 NPM 0 73 9.1 4 12.52	1 1 0 0						
CBO         1         101         17.29         9         21.37           RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2352         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31 <t< td=""><td>hadoop-0.6</td><td></td><td></td><td></td><td></td><td></td><td></td></t<>	hadoop-0.6						
RFC         0         420         48.15         28         63.01           LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2352         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16 <td< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>							
LCOM         0         2701         68.94         5         267.72           NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2752         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24							
NPM         0         71         10.05         5         14.07           LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2752         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24         26         63.96           LCOM         0         2850         65.18							
LOC         1         5646         413.02         167         725.04           hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2752         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24         26         63.96           LCOM         0         2850         65.18         4         293.09           NPM         0         73         9.1         4							
hadoop-0.7         DIT         0         21         1.66         1         2.35           NOC         0         9         0.35         0         1.08           CBO         1         132         17.96         9         23.85           RFC         1         483         49.69         28         68.33           LCOM         0         2752         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24         26         63.96           LCOM         0         2850         65.18         4         293.09           NPM         0         73         9.1         4         12.52							
NOC 0 9 0.35 0 1.08 CBO 1 132 17.96 9 23.85 RFC 1 483 49.69 28 68.33 LCOM 0 2752 70.5 4 288.88 NPM 0 76 9.88 4 13.31 LOC 1 6294 418.04 172 719.82  hadoop-0.8 DIT 0 20 1.6 1 2.36 NOC 0 9 0.31 0 1.06 CBO 1 125 16.16 9 20.11 RFC 0 533 46.24 26 63.96 LCOM 0 2850 65.18 4 293.09 NPM 0 73 9.1 4 12.52	hadoop-0.7						
CBO       1       132       17.96       9       23.85         RFC       1       483       49.69       28       68.33         LCOM       0       2352       70.5       4       288.88         NPM       0       76       9.88       4       13.31         LOC       1       6294       418.04       172       719.82         hadoop-0.8       DIT       0       20       1.6       1       2.36         NOC       0       9       0.31       0       1.06         CBO       1       125       16.16       9       20.11         RFC       0       533       46.24       26       63.96         LCOM       0       2850       65.18       4       293.09         NPM       0       73       9.1       4       12.52	nadoop o.,						
RFC       1       483       49.69       28       68.33         LCOM       0       2752       70.5       4       288.88         NPM       0       76       9.88       4       13.31         LOC       1       6294       418.04       172       719.82         hadoop-0.8       DIT       0       20       1.6       1       2.36         NOC       0       9       0.31       0       1.06         CBO       1       125       16.16       9       20.11         RFC       0       533       46.24       26       63.96         LCOM       0       2850       65.18       4       293.09         NPM       0       73       9.1       4       12.52							
LCOM         0         2752         70.5         4         288.88           NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24         26         63.96           LCOM         0         2850         65.18         4         293.09           NPM         0         73         9.1         4         12.52							
NPM         0         76         9.88         4         13.31           LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24         26         63.96           LCOM         0         2850         65.18         4         293.09           NPM         0         73         9.1         4         12.52							
LOC         1         6294         418.04         172         719.82           hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24         26         63.96           LCOM         0         2850         65.18         4         293.09           NPM         0         73         9.1         4         12.52							
hadoop-0.8         DIT         0         20         1.6         1         2.36           NOC         0         9         0.31         0         1.06           CBO         1         125         16.16         9         20.11           RFC         0         533         46.24         26         63.96           LCOM         0         2850         65.18         4         293.09           NPM         0         73         9.1         4         12.52							
NOC     0     9     0.31     0     1.06       CBO     1     125     16.16     9     20.11       RFC     0     533     46.24     26     63.96       LCOM     0     2850     65.18     4     293.09       NPM     0     73     9.1     4     12.52	hadoop-0.8						
CBO 1 125 16.16 9 20.11 RFC 0 533 46.24 26 63.96 LCOM 0 2850 65.18 4 293.09 NPM 0 73 9.1 4 12.52	-						
LCOM 0 2850 65.18 4 293.09 NPM 0 73 9.1 4 12.52		CBO	1	125	16.16	9	20.11
NPM 0 73 9.1 4 12.52		RFC	0	533	46.24	26	63.96
			0			4	293.09
LOC 1 6931 399.92 159 701.74			0				
		LOC	1	6931	399.92	159	701.74

Table 1: Summary Statistics.

Version	Faulty	Non-Faulty
hadoop-0.1	35.46	64.54
hadoop-0.2	21.99	78.01
hadoop-0.3	25.12	74.88
hadoop-0.4	20.90	79.10
hadoop-0.5	17.05	82.95
hadoop-0.6	13.25	86.75
hadoop-0.7	19.20	80.80
hadoop-0.8	6.67	93.33

Table 2: Percentage of Faulty and Non-Faulty components in each release.

$M_v < MP_v$	G-mean	Recall	Specificity
V = 1	1	1	< 0.001
V = 2	< 0.001	0.002	0.370
V = 3	< 0.001	0.072	0.002
V = 4	< 0.001	1	< 0.001
V = 5	0.001	1	< 0.001
V = 6	< 0.001	< 0.001	1

Table 3: Wilcoxon Test beetween  ${\cal M}_v$  and  ${\cal M} P_v$ 

$M_v > RG_v$	G-Mean	Recall
V=1	0.999(0.13)	1 (0.11)
V = 2	<0.001(1)	< 0.001 (1)
V = 3	< 0.001 (1)	< 0.001(1)
V = 4	< 0.001 (0.97)	< 0.001 (1)
V = 5	0.398 (0.55)	0.147(0.41)
V = 6	< 0.001(1)	< 0.001 (1)

Table 4: Wilcoxon Test between  $M_v$  and RG,  $\hat{A}12$  effect size between brackets

$MP_v > RG_v$	G-Mean	Recall
V = 0	< 0.001 (0.98)	< 0.001(0.95)
V = 1	< 0.001 (0.99)	< 0.001 (1)
V = 2	< 0.001 (0.81)	< 0.001 (0.80)
V = 3	< 0.001 (0.94)	< 0.001 (1)
V = 4	$0.761\ (0.40)$	< 0.001 (1)
V = 5	0.999(0.26)	< 0.001 (0.93)
V = 6	0.967(0.39)	$0.998 \ (0.29)$

Table 5: Wilcoxon Test between  $MP_v$  and RG,  $\hat{A}12$  effect size between brackets