

Experimental Results on Incomplete Operational Transition Complexity of Regular Languages

Eva Maia, Nelma Moreira, Rogério Reis

CMUP & DCC, Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre, 4169-007 Porto, Portugal
e-mail:{emaia,nam,rvr}@dcc.fc.up.pt

1 Regular Languages

We performed some experimental tests in order to analyse how often the upper bounds for state and transition complexity were achieved in practice. Although we fixed the size of the alphabet and consider small values of n and m , the experiments are statistically significant and provide valuable information about the average case behaviour of these operations.

In [1], the authors presented an uniform random generator for complete DFAs. We can use this generator to obtain incomplete DFAs, if we consider the existence of a dead state. However, in this case, the probability that a state has a transition to the dead state is $\frac{1}{n+1}$, where n is the number of useful states of the generated incomplete DFA. Although this corresponds to a uniform distribution, for very large values of n , the referred probability is very low, and thus the generated DFAs are almost always complete. Therefore, in order to generate random incomplete DFAs, we can increase the number of void transitions in the generated DFAs to change the referred probability. For that, the generator accepts a parameter b that defines the multiplicity of dead states. Using b ($0 < b < 1$), we compute the integer part of $m = \frac{b \times n}{1-b}$, which indicates the number of dead states in generated DFA. Note that the generated DFA becomes more incomplete when b tends to 1.

All the tests were performed using the random generator described above. The tests¹ and the generator² were implemented in Python³, and are both publicly available. In the following experiments we consider $b = 0.7$ (Table 2).

As the DFAs were obtained with a uniform random generator, the size of each sample (20000 elements) is sufficient to ensure a 95% confidence level within a 1% error margin. This is calculated with the formula $n = (\frac{z}{2\epsilon})^2$, where z is obtained from the normal distribution table such that $P(-z < Z < z) = \gamma$, ϵ is the error margin, and γ is the desired confidence level.

The following tables show the results of experimental tests with 20000 pairs of incomplete DFAs as operands. We present the results for operands with $m, n \in \{3, 6, 9, 12, 15, 18, 21, 24, 27\}$ states, such that $m+n = 10$ in Table 1, $m+n = 20$

¹ <http://khilas.dcc.fc.up.pt/~eva/>

² <http://fado.dcc.fc.up.pt>

³ <http://www.python.org>

in Table 2, and $m + n = 30$ in Table 3, over an alphabet of $k = 5$ symbols. As union and intersection are symmetric operations, we do not need to present all the results. We considered the following measures for the DFA resulting from the operation: the state and transition complexity, sc and tc , respectively; the upper bounds for this measures, $ubsc$ and $ubtc$, respectively; its density $d = \frac{tc}{k*sc}$; and the ratios $rs = \frac{sc}{ubsc}$ and $rt = \frac{tc}{ubtc}$. Note that the results presented on Table 2 are averages, i.e. we calculate all the referred measures for each pair of operands and then we compute the average of each measure. The measures m_1 , m_2 , m_3 and m_4 are the maximal values of sc , $ubsc$, tc and $ubtc$, respectively. For example, in Table 2, considering $m = 10$ and $n = 10$ we calculate the $ubsc$ for the concatenation of each pair of random incomplete DFAs. Then we do the average of the 20000 obtained values and the result is 8557.90, as we can see in the Table 2. We need to do this because every measure depends of parameters which can be different in each pair of generated DFAs.

b=0.7												
Concatenation												
m	n	sc	$ubsc$	rs	m_1	m_2	tc	$ubtc$	rt	m_3	m_4	d
2	8	18.14	589.28	0.03	134	639	56.85	2920.17	0.019	605	3528	0.58
4	6	17.91	246.11	0.073	112	287	57.23	1192.79	0.05	496	1396	0.603
6	4	16.17	84.91	0.19	54	103	50.89	400.43	0.13	217	470	0.61
8	2	13.24	26.48	0.50	29	33	39.98	116.28	0.344	120	146	0.59
Union												
2	8	14.00	26	0.54	26	26	36.28	73.21	0.50	102	107	0.50
4	6	14.84	34	0.44	33	34	38.98	100.09	0.39	131	144	0.51
Intersection												
2	8	3.45	16	0.22	16	16	3.67	67.33	0.05	37	200	0.13
4	6	3.60	24	0.15	21	24	3.74	104.34	0.04	43	264	0.13
Star												
2		2.06	3.22	0.64	3	4	5.18	8.72	0.59	15	19	0.50
4		4.65	10.70	0.43	12	16	14.04	40.75	0.34	58	74	0.60
6		8.83	38.21	0.23	31	64	30.71	170.84	0.18	136	305	0.68
8		14.34	141.89	0.10	80	256	53.71	676.75	0.08	350	1228	0.73
Reversal												
2		2.43	3	0.81	3	3	5.26	15	0.35	13	15	0.42
4		6.50	15	0.43	15	15	16.65	75	0.22	65	75	0.49
6		12.19	63	0.19	60	63	34.60	315	0.11	261	315	0.54
8		18.67	255	0.07	99	255	55.24	1275	0.04	368	1275	0.56
Complement												
2		2.99	3	1	3	3	9.57	29.71	0.32	15	3	0.64
4		5.00	5	1	5	5	22.14	50.90	0.43	25	5	0.89
6		7	7	1	7	7	33.93	73.87	0.46	35	7	0.97
8		9	9	1	9	9	44.60	95.34	0.47	45	9	0.99

Table 1. Experimental results for general regular languages with $b = 0.7$ and $m + n = 10$.

2 Finite Languages

Similarly to the previous section, we performed some experimental tests in order to analyse the practical behaviour of the operations over finite languages. All the tests were performed with uniformly random generated acyclic DFAs.

The Tables 4, 5 and 6 show the results of 20000 experimental tests. The number of states of the operands and the measures are the same as used in Section 1.

b=0.7													
Concatenation													
m	n	sc	ubsc	rs	m ₁	m ₂	tc	ubtc	rt	m ₃	m ₄	d	
2	8	12.71	21.11	0.60	31	34	47.26	88.52	0.53	136	165	0.72	
4	6	15.69	59.45	0.26	40	78	60.23	192.96	0.31	186	383	0.75	
6	4	14.56	41.80	0.35	32	43	54.26	149.90	0.36	149	210	0.74	
8	2	11.2046	16	0.7002875	16	16	39.39655	65.738	0.5992964495	69	75	0.70	
Union													
2	8	8.66	14	0.62	12	14	29.67	49.71	0.60	48	65	0.68	
4	6	9.73	22	0.44	16	22	32.30	79.84	0.40	61	103	0.66	
Intersection													
2	8	1.77	2	0.89	2	2	1.30	2.33	0.56	5	5	0.13	
4	6	3.83	10	0.38	9	10	5.99	14.34	0.42	25	38	0.29	
Star													
2		1	0.75	1.34	1	1	2.58	1.93	1.34	5	5	0.52	
4		3.05	4.70	0.65	5	7	11.80	15.21	0.78	25	35	0.78	
6		7.50	18.82	0.40	23	31	33.17	65.20	0.51	111	153	0.88	
8		14.71	71.53	0.21	68	127	68.63	242.52	0.28	332	632	0.93	
Reversal													
2		2	2	1	2	2	2.58	2.58	1	5	5	0.26	
4		5.57	7.99	0.70	8	8	12.87	31.70	0.41	29	35	0.46	
6		11.87	21.00	0.57	20	21	33.98	96.087	0.35	79	100	0.57	
8		21.99	62.00	0.35	40	62	70.73	298.14	0.24	160	305	0.64	
Complement													
2		3	3	1	3	3	7.74	7.74	1	15	15	0.52	
4		5	5	1	5	5	24.13	24.13	1	25	25	0.97	
6		7	7	1	7	7	34.97	34.97	1	35	35	1.00	
8		9	9	1	9	9	45.00	45.00	1	45	45	1.00	

Table 4. Experimental results for finite languages with $m + n = 10$.

References

1. Almeida, M., Moreira, N., Reis, R.: Enumeration and generation with a string automata representation. *Theor. Comput. Sci.* 387(2), 93–102 (2007)

