– Solving the Longest Common Subsequence Problem Concerning Non-uniform Distributions of Letters in Input Strings –
– Supplementary material –

Bojan Nikolic, Aleksandar Kartelj, Marko Djukanovic, Milana Grbic, Christian Blum, Guenther Raidl

May 15, 2021

1 Remarks and Abbreviations

This is a full report on the computational results of the paper entitled as above. It consists of the full short and long-run result of our algorithm compared to the results of all other competitive literature approaches. The result are given in Tables 1–10. This report follows the notation introduced in the original paper. The most important algorithms are abbreviated as

- BS: Beam search;
- Bs-Ex: a BS configuration which utilizes the expected length calculation heuristic (Ex) [2] to guide search;
- Bs-Pow: a BS configuration which utilizes Pow heuristic [4] from the original paper to guide the search;
- Bs-Gmpsum: a BS configuration which utilizes Gmpsum heuristic to guide search;
- Bs-Blum: a beam search configuration which utilizes rank-sum heuristic [1] to guide the search;
- Bs-Hp: BS configuration which utilizes probability based heuristic [4] to guide the search;
- APS: Anytime pack search algorithm [5];
- A*+ACS: anytime hybrid developed in [3];
- A*+ACS-dist: a configuration of the above A*+ACS framework where Ex is replaced by dist heuristic from [6];
- TRBS-Gmpsum: the time-restricted BS guided by Gmpsum.

2 Complete Results per the Benchmark Sets from the Literature

This section presents full short and long-run results per each benchmark set that are presented in the LCSC literature. The tables reporting detailed short-run results are organized as follows. The first block displays the properties of instances (or instance group). The second block consists of three columns: the first column shows the best (or best average) solution among the compared approaches (including also the new BS guided by novel Gmpsum heuristic) and the second column represents the best (or average best) short-run results from the literature, while the third shows the name of the algorithm which obtains this solution. The third block reports average solutions and average running times of four BS configurations (BS guided by Ex, Pow, Hp, Gmpsum, respectively) for each considered instance (or group of instances). Note that we use the labeling **best** in the case when the best solution from literature is achieved, while in the case that it is outperformed with the proposed algorithm (a new state-of-the-art), the labeling **new** is placed. Otherwise, if certain approach did not achieve either of these two, numerical solution value is shown. Using both **best** and **new** labels in the same table row is therefore impossible.

The tables that report the detailed results of the long-run executions over the considered benchmark sets are organized as follows. The table is split into four blocks. The first block displays the characteristics of the considered instance (or instance group). The second block consists of three columns: the first reports new best (or best average solutions) among all the approaches from literature (including also this new one), the second column report the best (or best average) solution known from the literature so-far, and the third column reports the name of the algorithm from literature that obtains the best (or average best) solution. The third block reports the average solution of the A*+ACS as the best approach from the literature. The fourth block reports average solution and average running time of the TRBS–Gmpsum on the respective instance (or instance group) of the considered benchmark set. Usage of **best** and **new** labeling is same as in the short-run scenario. In the following text, we go set-by-set explaining the obtained results in details.

2.1 Benchmark set Random

Concerning short-run executions, it is quite obvious that BS guided by Ex and Gmpsum achieve together all the best (state-of-the-art) solutions, and performing perform on pair. Other heuristic guidances are much worse. Concerning long-run executions, the TRBS–Gmpsum performs slightly worse in just three out of 20 cases, where it delivers a solution which is just by one letter shorter than the best known solution from the literature. Therefore, we conclude that the best performing algorithms on Random are both, A*+ACS and the TRBS–Gmpsum.

Table 1: Short-run results for benchmark set Random.

Instance			Best $\|s\|$			BS-Ex		BS-Pow		BS-Hp		BS-Gmpsum	
$\|\Sigma\|$	m	n	New	Lit.	Alg.	$\|s\|$	t	$\|s\|$	t	$\|s\|$	t	$\|s\|$	t
4	10	600	221	221	BS-Ex	**best**	2.8	220	1.3	220	0.8	**best**	2
4	15	600	206	204	BS-Ex	204	2	203	1.3	203	0.8	**best**	2.3
4	20	600	193	193	BS-Ex	**best**	3	191	1.2	192	0.7	**best**	2.8
4	25	600	188	187	BS-Ex	187	3	187	0.9	187	0.6	**best**	3
4	40	600	175	175	BS-Ex	**best**	2.5	173	1.2	173	0.8	**best**	3.8
4	60	600	168	168	BS-Ex	**best**	2.3	166	1.2	166	0.8	167	5.4
4	80	600	163	163	BS-Ex	**best**	2.3	161	1.2	161	0.8	162	6.7
4	100	600	159	159	BS-Ex	**best**	2.2	158	1.3	158	0.7	**best**	7.5
4	150	600	154	153	BS-Ex	153	3	152	1.4	152	1.1	**best**	10.1
4	200	600	151	151	BS-Ex	**best**	3.1	150	1	**best**	1.2	**best**	13.6
20	10	600	63	63	BS-Ex	**best**	4.1	62	2.3	62	2	62	3.3
20	15	600	53	53	BS-Ex	**best**	3.7	**best**	2.3	52	1.3	**best**	3.4
20	20	600	48	48	BS-Ex	**best**	2.6	47	1.7	**best**	1.5	**best**	3.7
20	25	600	45	44	BS-Ex	44	2.5	44	1.8	44	1.4	**best**	4.1
20	40	600	39	39	BS-Ex	**best**	2.6	38	1.6	38	1.2	38	5.1
20	60	600	35	35	BS-Ex	**best**	2.5	**best**	1.2	**best**	1.3	**best**	6.6
20	80	600	33	33	BS-Ex	**best**	2.3	**best**	1.5	**best**	1.2	**best**	8.3
20	100	600	32	32	BS-Ex	**best**	2	**best**	1.1	31	1.2	**best**	10
20	150	600	29	29	BS-Ex	**best**	3	**best**	1.7	**best**	1.3	**best**	14.2
20	200	600	28	28	BS-Ex	**best**	3.3	**best**	1.6	**best**	1.4	**best**	18.8

2.2 Benchmark set Rat

Concerning short-run execution, we emphasize five new state-of-the-art short-run results (out of 20) achieved by BS–Gmpsum. In nine cases, BS–Ex and BS–Gmpsum achieve the same solution. In four cases, the BS–Ex achieves better solution than BS–Gmpsum. The difference is minimal in all cases – a single character. Concerning long-run executions, three new state-of-the-art results were obtained by the TRBS–Gmpsum. Note the instance $|\Sigma| = 4, m = 40$, and $n = 600$ where new state-of-the-art is achieved (156), while the previous best result was 154. In two cases, A*+ACS delivers slightly better solution (by one character), whereas in the remaining cases, the solutions are tied.

Table 2: Long-run results for benchmark set Random.

Instance			Best $\vert s\vert$			A*+ACS	TRBS-Gmpsum 600s	
$\vert\Sigma\vert$	m	n	New	Lit.	Alg.	$\vert s\vert$	$\vert s\vert$	t
4	10	600	223	223	A*+ACS	**best**	222	597.9
4	15	600	206	206	A*+ACS	**best**	**best**	598.5
4	20	600	195	195	A*+ACS	**best**	**best**	595.6
4	25	600	189	189	A*+ACS	**best**	**best**	596.8
4	40	600	177	177	A*+ACS	**best**	**best**	596.5
4	60	600	169	169	A*+ACS	**best**	168	594.7
4	80	600	164	164	A*+ACS	**best**	**best**	593.6
4	100	600	161	161	A*+ACS	**best**	160	592.5
4	150	600	155	155	A*+ACS	**best**	**best**	593
4	200	600	152	152	A*+ACS	**best**	**best**	592.9
20	10	600	63	63	A*+ACS	**best**	**best**	595
20	15	600	53	53	A*+ACS	**best**	**best**	589.3
20	20	600	48	48	A*+ACS	**best**	**best**	583.5
20	25	600	45	45	A*+ACS	**best**	**best**	576.2
20	40	600	39	39	A*+ACS	**best**	**best**	560.7
20	60	600	36	36	A*+ACS	**best**	**best**	561.3
20	80	600	33	33	A*+ACS	**best**	**best**	546.9
20	100	600	32	32	A*+ACS	**best**	**best**	550.4
20	150	600	30	30	A*+ACS	**best**	29	546.7
20	200	600	28	28	A*+ACS	**best**	**best**	545.2

Table 3: Short-run results for benchmark set Rat.

Instance			Best $\vert s\vert$			BS-Ex		BS-Pow		BS-Hp		BS-Gmpsum	
$\vert\Sigma\vert$	m	n	New	Lit.	Alg.	$\vert s\vert$	t	$\vert s\vert$	t	$\vert s\vert$	t	$\vert s\vert$	t
4	10	600	205	205	BS-Ex	**best**	3.1	204	1.2	200	0.6	**best**	1.9
4	15	600	185	185	BS-Ex	**best**	2.7	183	1.1	184	0.6	**best**	2.3
4	20	600	173	172	BS-Ex	172	2.3	170	0.9	168	0.5	**new**	2.5
4	25	600	170	170	BS-Ex	**best**	2.7	168	1	166	0.6	169	2.9
4	40	600	154	152	BS-Ex	152	1.8	150	1	145	0.5	**new**	3.4
4	60	600	152	152	BS-Ex	**best**	2.3	151	1.2	150	0.7	**best**	4.7
4	80	600	142	142	BS-Ex	**best**	2.5	139	1.1	139	0.7	141	5.7
4	100	600	137	137	BS-Ex	**best**	2.5	131	1	135	0.5	**best**	6.5
4	150	600	129	129	BS-Ex	**best**	2	126	0.9	125	0.6	128	7.8
4	200	600	124	123	BS-Ex	123	2.7	123	0.7	122	0.8	**new**	9.9
20	10	600	71	71	BS-Ex	**best**	3.4	**best**	2.5	**best**	1.9	**best**	3.6
20	15	600	63	63	BS-Ex	**best**	2.6	62	1.6	62	1.4	62	3.5
20	20	600	55	54	BS-Ex	54	2.5	54	1.7	54	1.2	**new**	3.5
20	25	600	52	52	BS-Ex	**best**	2.9	51	1.4	51	1.1	**best**	3.8
20	40	600	50	49	BS-Ex	49	3	49	1.1	49	1.2	**new**	4.6
20	60	600	47	47	BS-Pow	46	2.4	**best**	1.5	46	1.2	**best**	7.1
20	80	600	44	44	BS-Hp	43	2.6	43	1.5	**best**	1.1	43	7.7
20	100	600	40	40	BS-Ex	**best**	2.5	39	1.2	39	1	**best**	8.5
20	150	600	37	37	BS-Ex	**best**	2	**best**	0.9	36	0.6	**best**	8.7
20	200	600	34	34	BS-Ex	**best**	2.7	**best**	1.4	33	1	33	11.1

Table 4: Long-run results for benchmark set Rat.

Instance			Best $\vert s\vert$			A*+ACS	TRBS-Gmpsum 600s	
$\vert\Sigma\vert$	m	n	New	Lit.	$\vert s\vert$	$\vert s\vert$	t	
4	10	600	206	206	A*+ACS	**best**	**best**	599.2
4	15	600	189	189	A*+ACS	**best**	188	597.3
4	20	600	175	174	A*+ACS	174	**new**	597.3
4	25	600	173	173	A*+ACS	**best**	**best**	595.4
4	40	600	156	154	A*+ACS	154	**new**	592.1
4	60	600	154	154	A*+ACS	**best**	**best**	593.8
4	80	600	144	144	A*+ACS	**best**	**best**	596.2
4	100	600	139	139	A*+ACS	**best**	**best**	594
4	150	600	131	131	A*+ACS	**best**	**best**	595
4	200	600	126	126	A*+ACS	**best**	**best**	593.4
20	10	600	72	72	A*+ACS	**best**	71	586.3
20	15	600	63	63	A*+ACS	**best**	**best**	574.8
20	20	600	55	55	A*+ACS	**best**	**best**	575.7
20	25	600	53	52	A*+ACS	52	**new**	575.9
20	40	600	50	50	A*+ACS	**best**	**best**	558.8
20	60	600	47	47	A*+ACS	**best**	**best**	542.6
20	80	600	44	44	A*+ACS	**best**	**best**	524.2
20	100	600	40	40	A*+ACS	**best**	**best**	531.2
20	150	600	38	38	A*+ACS	**best**	**best**	547.1
20	200	600	35	35	A*+ACS	**best**	**best**	545.2

2.3 Benchmark set Virus

Concerning the short-run execution, in seven (out of 20) cases, new-state-of-the-art solutions are obtained by BS–Gmpsum. In just two cases, BS–Ex outperforms BS–Gmpsum. In all other cases, the results of BS guided by the aforementioned two heuristic are tied. Other two guiding heuristics perform significantly worse. Concerning long-run executions, in six out of 20 cases new state-of-the-art results are obtained by the TRBS–Gmpsum. In just one case, A*+ACS delivers a better solution. In other cases, the results of the two algorithms are tied.

Table 5: Short-run results for benchmark set Virus.

Instance			Best $\lvert s\rvert$			BS-Ex		BS-Pow		BS-Hp		BS-Gmpsum	
$\lvert\Sigma\rvert$	m	n	New	Lit.	Alg.	$\lvert s\rvert$	t	$\lvert s\rvert$	t	$\lvert s\rvert$	t	$\lvert s\rvert$	t
4	10	600	227	227	BS-Ex	**best**	2.9	225	1	226	0.7	**best**	2.2
4	15	600	205	205	BS-Ex	**best**	2.2	203	1.2	204	0.7	**best**	2.4
4	20	600	192	192	BS-Ex	**best**	2.7	190	1.1	190	0.7	**best**	2.8
4	25	600	195	194	BS-Ex	194	2.2	192	1.2	194	0.7	**new**	3.1
4	40	600	172	170	BS-Ex	170	2.2	170	1.2	169	0.9	**new**	3.8
4	60	600	168	166	BS-Ex	166	2.4	165	0.8	166	0.7	**new**	5.1
4	80	600	163	163	BS-Ex	**best**	2.7	157	1	159	0.7	161	6.6
4	100	600	160	158	BS-Ex	158	2.3	155	1.2	158	0.9	**new**	7.8
4	150	600	157	156	BS-Ex	156	2.4	147	1.2	156	0.7	**new**	11
4	200	600	156	155	BS-Hp	154	2.6	148	1.4	155	1.2	**new**	14.8
20	10	600	77	77	BS-Pow	76	2.9	**best**	2.4	75	1.7	76	3.4
20	15	600	64	64	BS-Ex	**best**	2.9	**best**	2.1	**best**	1.6	63	3.6
20	20	600	60	60	BS-Ex	**best**	2.7	**best**	2.1	**best**	1.6	**best**	4.2
20	25	600	55	55	BS-Ex	**best**	2.7	**best**	1.8	**best**	1.1	**best**	4
20	40	600	51	50	BS-Ex	50	2.9	49	1.9	50	0.9	**new**	5.5
20	60	600	48	48	BS-Ex	**best**	3.3	47	1.2	47	1.1	**best**	7.4
20	80	600	46	46	BS-Ex	**best**	2.6	**best**	1.5	**best**	1.4	**best**	9.3
20	100	600	45	45	BS-Ex	**best**	2.3	44	1.5	44	1.4	**best**	10.8
20	150	600	45	45	BS-Ex	**best**	2.8	**best**	2.1	**best**	1.3	**best**	17.4
20	200	600	44	44	BS-Hp	43	3.2	43	2.1	**best**	1.7	**best**	23.3

Table 6: Long-run results for benchmark set Virus.

Instance			Best $\lvert s\rvert$			A*+ACS	TRBS-Gmpsum 600s	
$\lvert\Sigma\rvert$	m	n	New	Lit.	Alg.	$\lvert s\rvert$	$\lvert s\rvert$	t
4	10	600	229	228	A*+ACS	228	**new**	598.3
4	15	600	207	206	A*+ACS	206	**new**	598.6
4	20	600	194	194	A*+ACS	**best**	193	595
4	25	600	196	196	A*+ACS	**best**	**best**	597
4	40	600	174	174	A*+ACS	**best**	**best**	597.1
4	60	600	169	168	A*+ACS	168	**new**	596.5
4	80	600	164	163	A*+ACS	163	**new**	595.8
4	100	600	162	160	A*+ACS	160	**new**	596
4	150	600	158	157	A*+ACS	157	**new**	596
4	200	600	156	156	A*+ACS	**best**	**best**	594.7
20	10	600	77	77	A*+ACS	**best**	**best**	590.4
20	15	600	64	64	A*+ACS	**best**	**best**	573
20	20	600	61	61	A*+ACS	**best**	**best**	570.4
20	25	600	56	56	A*+ACS	**best**	**best**	569
20	40	600	51	51	A*+ACS	**best**	**best**	554.2
20	60	600	48	48	A*+ACS	**best**	**best**	541.6
20	80	600	46	46	A*+ACS	**best**	**best**	543.2
20	100	600	45	45	A*+ACS	**best**	**best**	534.9
20	150	600	46	46	A*+ACS	**best**	**best**	549.9
20	200	600	44	44	A*+ACS	**best**	**best**	538.5

2.4 Benchmark set BB

In the short-run scenario, BS–Pow heuristic and BS–Gmpsum are performing well, since they are able to deliver best average solution in six versus five (out of eight) instance groups. However, slightly better performing approach in terms of solution quality is BS–Pow. BS–Gmpsum is able to deliver a new state-of-the-art short-run solution for one instance group. In the remaining group, other BS configuration from the literature (BS–Blum, see [1]) delivers the best known solution. Concerning the long-run executions, in six out of eight cases, the best average solutions is obtained by the TRBS–Gmpsum, from which tow new state-of-the-art results are achieved. In three cases the best average solution produced by A*+ACS is also matched by TRBS–Gmpsum. In the remaining two cases, the best performing algorithms are those from the literature: APS [3, 5] and BS–Pow.

Table 7: Short-run results for BB instances.

Instance			Best $|s|$			BS-Ex		BS-Pow		BS-Hp		BS-Gmpsum	
$|\Sigma|$	m	n	New	Lit.	Alg.	$|s|$	t	$|s|$	t	$|s|$	t	$|s|$	t
2	10	1000	676.5	676.5	BS-Hp	673.5	5.5	**best**	4	**best**	1.2	**best**	6.2
2	100	1000	560.8	560.7	BS-Pow	536.6	6.1	560.7	5.7	558.9	1.9	**new**	23.7
4	10	1000	545.4	545.4	BS-Hp	545.2	6.2	**best**	9.5	**best**	1.7	**best**	7.1
4	100	1000	388.8	388.8	BS-Pow	329.5	5.9	**best**	2.9	368	2.6	379.4	22.3
8	10	1000	462.7	462.7	BS-Ex	**best**	7.9	**best**	12.5	**best**	2.6	**best**	9.9
8	100	1000	272.1	272.1	BS-Blum	210.6	8	271.8	5.3	247.7	3.5	253	26.4
24	10	1000	385.6	385.6	BS-Ex	**best**	16.2	**best**	7	**best**	8.2	**best**	31.1
24	100	1000	149.5	149.5	BS-Pow	113.3	12.5	**best**	3.6	138.7	7.2	135.5	88.6

Table 8: Long-run results for BB instances.

Instance			Best $|s|$			A*+ACS	TRBS-Gmpsum 600s	
$|\Sigma|$	m	n	New	Lit.	Alg.	$|s|$	$|s|$	t
2	10	1000	676.7	676.7	A*+ACS	676.6	**best**	598.1
2	100	1000	571.1	563.6	APS	547.1	**new**	598
4	10	1000	545.5	545.5	A*+ACS	**best**	**best**	599.7
4	100	1000	391.8	390.2	APS	344.3	**new**	598.8
8	10	1000	462.7	462.7	A*+ACS	**best**	**best**	599.9
8	100	1000	273.4	273.4	APS	223.7	265.6	598.9
24	10	1000	385.6	385.6	A*+ACS	**best**	**best**	600.5
24	100	1000	149.5	149.5	BS-Pow	117	143.2	598.9

2.5 Benchmark set ES

These uniform-at-random instances perfectly fit the Ex guidance, as is already shown in the literature. So, it is not surprising that for eight out of twelve groups BS–Ex reaches the best average solution. In four out of twelve groups, BS–Gmpsum is able to reach the best average solution, all representing the new state-of-the-art results. Concerning long-run executions, in eight out of twelve groups, A*+ACS beats the result of the TRBS–Gmpsum. In three out of twelve groups, new state-of-the-art results are obtained by the TRBS–Gmpsum, and for one group, the result was tied.

Table 9: Short-run results for ES instances.

Instance			Best $|s|$			BS-Ex		BS-Pow		BS-Hp		BS-Gmpsum	
$|\Sigma|$	m	n	New	Lit.	Alg.	$|s|$	t	$|s|$	t	$|s|$	t	$|s|$	t
2	10	1000	615.1	615.06	BS-Ex	615.06	4.4	614.2	1.4	612.5	0.9	**new**	5.1
2	50	1000	538.24	538.24	BS-Ex	**best**	4.4	535.56	1.6	536.46	1.1	536.28	12.3
2	100	1000	519.84	519.84	BS-Ex	**best**	4.8	516.24	1.9	518.56	1.3	516.42	22
10	10	1000	203.1	203.1	BS-Ex	**best**	5.6	202.72	2.5	201.42	1.6	203.08	4.9
10	50	1000	136.34	136.32	BS-Ex	136.32	3.9	135.52	2.1	135.22	1.4	**new**	9.9
10	100	1000	123.32	123.32	BS-Ex	**best**	4.3	122.18	2.2	122.4	1.5	122.96	17.1
25	10	2500	235.58	235.22	BS-Pow	231.12	19.1	235.22	10.5	233.34	8	**new**	29
25	50	2500	139.5	139.5	BS-Ex	**best**	14.5	138.56	7.2	137.76	5.5	139.44	53.7
25	100	2500	122.88	122.88	BS-Ex	**best**	16	121.62	7.3	121.6	5.9	122.7	96.7
100	10	5000	145.1	144.9	BS-Pow	144.18	91.9	144.9	75.9	143.62	71.6	**new**	185.4
100	50	5000	71.94	71.94	BS-Ex	**best**	53.5	71.32	39.1	70.86	35	71.8	365.4
100	100	5000	60.66	60.66	BS-Ex	**best**	53.7	60.06	36	59.96	32.4	60.6	624.3

Table 10: Long-run results for ES instances.

Instance			Best $|s|$			A*+ACS	TRBS-Gmpsum 600s	
$|\Sigma|$	m	n	New	Lit.	Alg.	$|s|$	$|s|$	t
2	10	1000	619.1	618.9	A*+ACS	618.9	**new**	599
2	50	1000	540.9	540.9	A*+ACS	**best**	540.3	598.1
2	100	1000	522.1	522.1	A*+ACS	**best**	520.4	597.1
10	10	1000	205	205	A*+ACS	**best**	204.9	599.3
10	50	1000	137.6	137.5	A*+ACS	137.5	**new**	594.5
10	100	1000	124.1	124.1	A*+ACS	**best**	**best**	591.7
25	10	2500	238	236.6	A*+ACS-dist	235	**new**	599.2
25	50	2500	140.4	140.4	A*+ACS	**best**	140.1	593.3
25	100	2500	123.4	123.4	A*+ACS	**best**	123.2	588.5
100	10	5000	145.7	145.7	A*+ACS	**best**	145.1	597.8
100	50	5000	72	72	A*+ACS	**best**	71.8	580.5
100	100	5000	60.8	60.8	A*+ACS	**best**	60.1	619.3

References

[1] C. Blum, M. J. Blesa, and M. López-Ibáñez. Beam search for the longest common subsequence problem. *Computers & Operations Research*, 36(12):3178–3186, 2009.

[2] M. Djukanovic, G. Raidl, and C. Blum. A beam search for the longest common subsequence problem guided by a novel approximate expected length calculation. In *Proceedings of LOD 2019 – The 5th International Conference on Machine Learning, Optimization, and Data Science*, LNCS. Springer, 2019. to appear.

[3] M. Djukanovic, G. R. Raidl, and C. Blum. Finding longest common subsequences: New anytime a search results. *Applied Soft Computing*, 95:106499, 2020.

[4] F. S. Tabataba and S. R. Mousavi. A hyper-heuristic for the longest common subsequence problem. *Computational Biology and Chemistry*, 36:42–54, 2012.

[5] S. G. Vadlamudi, S. Aine, and P. P. Chakrabarti. Anytime pack search. *Natural Computing*, 15(3):395–414, 2016.

[6] J. Yang, Y. Xu, G. Sun, and Y. Shang. A new progressive algorithm for a multiple longest common subsequences problem and its efficient parallelization. *IEEE Transactions on Parallel and Distributed Systems*, 24(5):862–870, 2012.