CS466 – Introduction to Bioinformatics

Department of Computer Science

Spring 2021

Grainger College of Engineering, University of Illinois at Urbana–Champaign

---

Mini Project Report

Motif Finding Results

Dr. Saurabh Sinha

Team Members

Liu, Pan *panl3@illinois.edu*

Zhu, Ruike *ruikez2@illinois.edu*

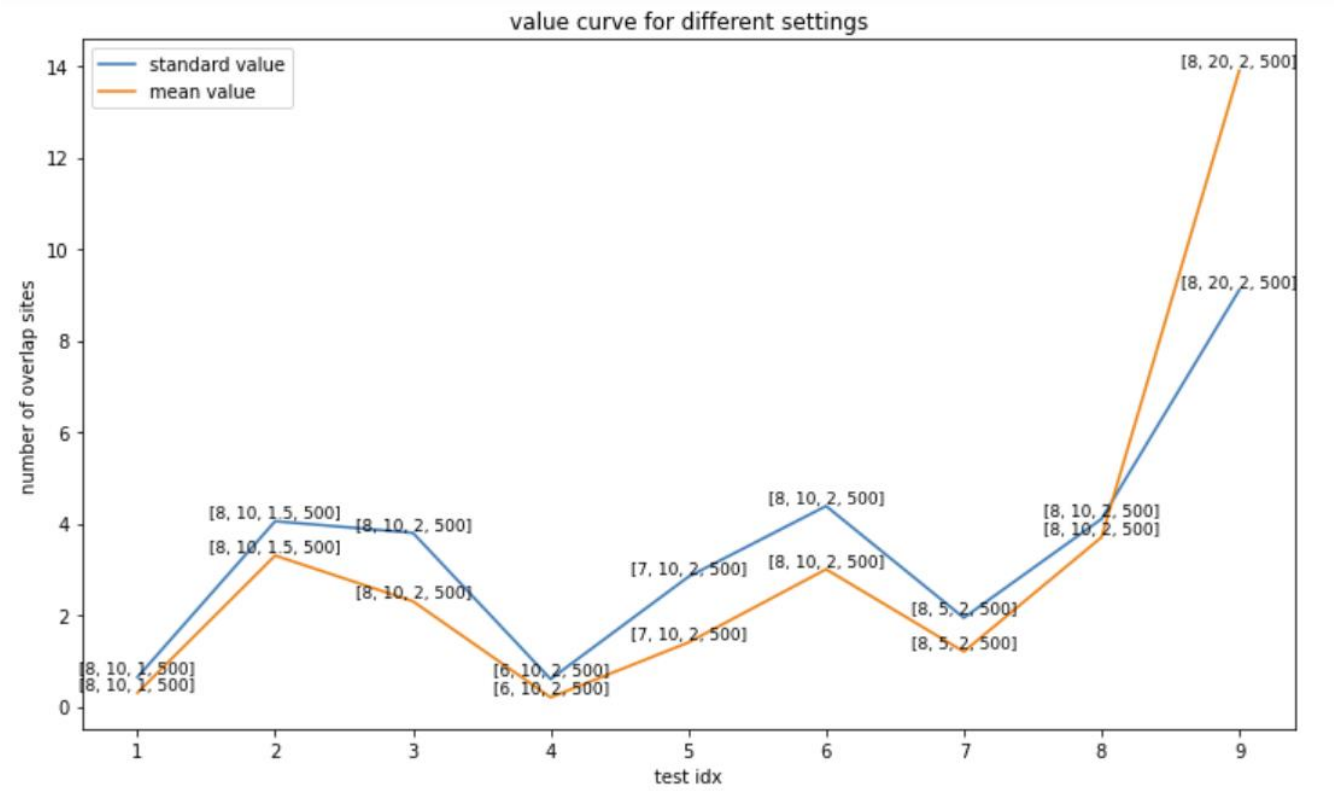
Line charts

Here are the line charts for modifying different parameters.

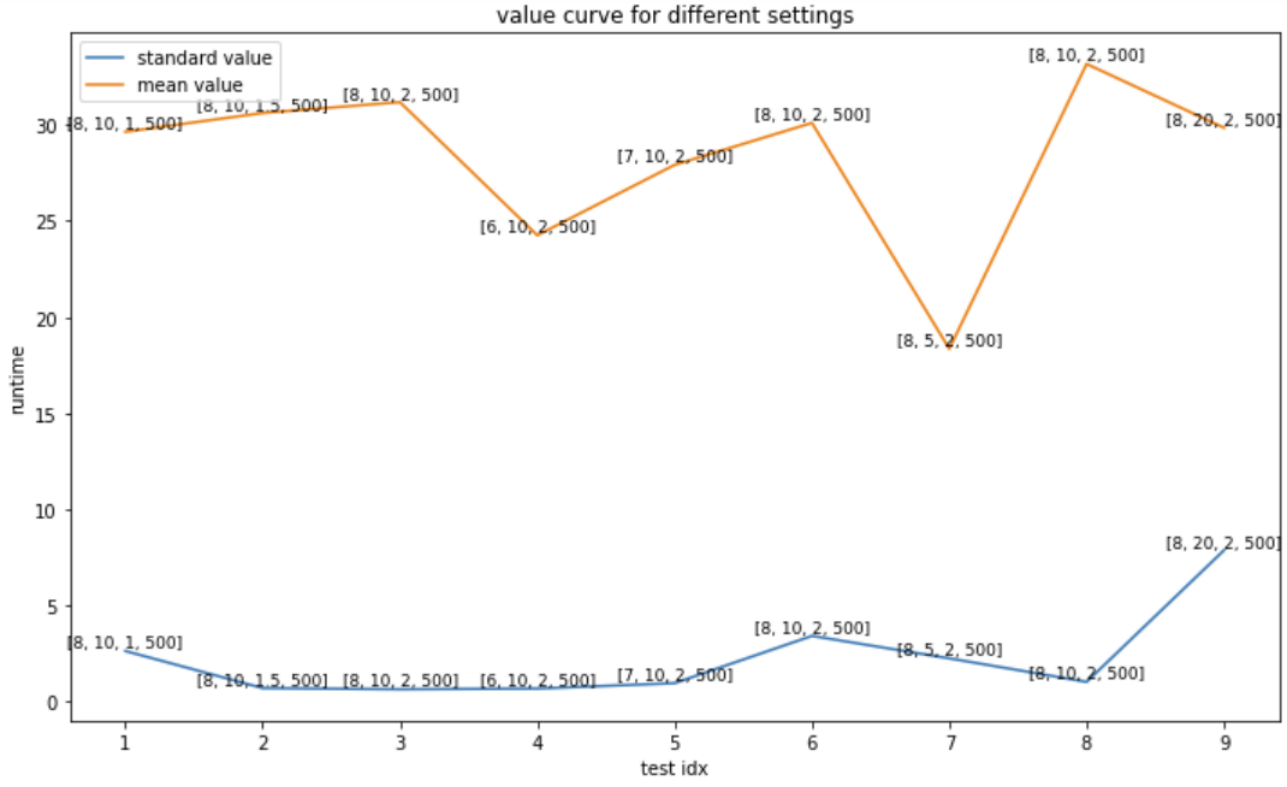
The meaning of the label in the figures represents the value of [ML, SC, ICPC, SL]

So, tests 1-3 represent changing the ICPC (information content per column) value, tests 3-6 mean changing the ML (motif length) value, and tests 6-9 represent changing the SC (sequence count) value.

1. Number of overlapping sites



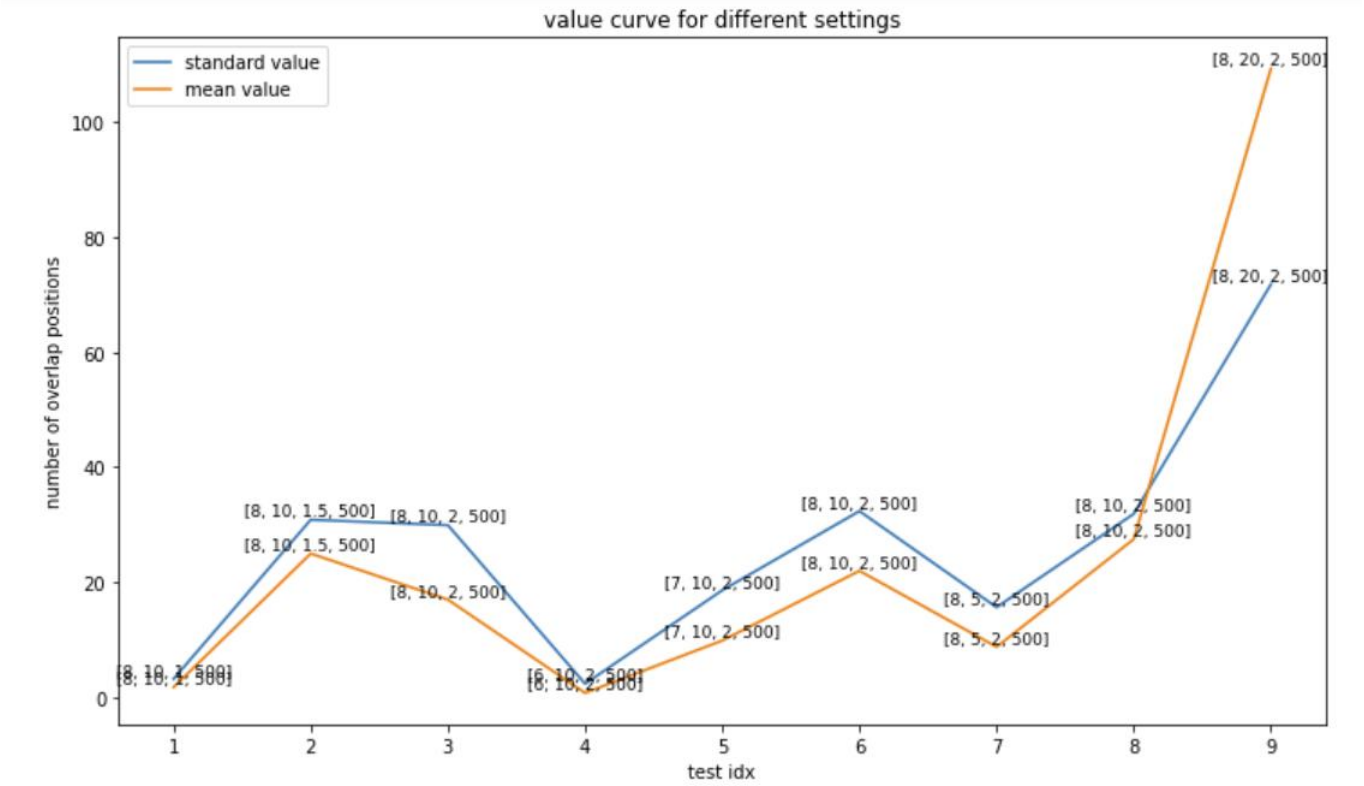
2. Runtime



3. Relative entropy

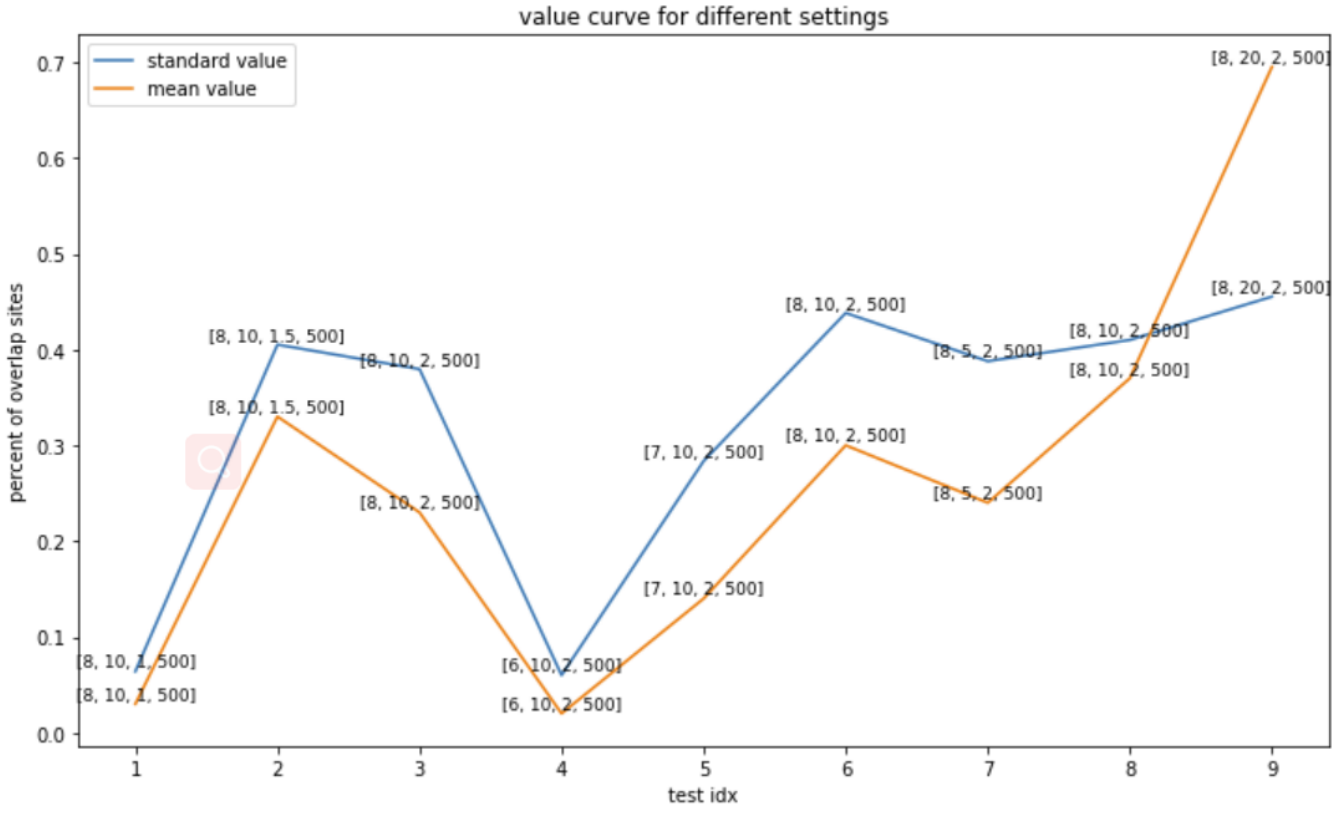
   This is the criterion we use to compare the discovered PWM with the planted PWM. Sine the greedy algorithm cannot always find the optimal solution, so the entropy of many of our tests is INF. Instead of drawing figures, we record the value of each test in table 4.

4. Number of overlapping positions



5. Percent of overlap sites

   We use this as an additional criterion to evaluate the accuracy of our motif finder.



Tables

Below are the tables recording the detailed value of each test.

Table 1: Number of overlapping positions:

Test	ICPC=1.0	ICPC=1.5	ICPC=2.0	ML=6	ML=7	ML=8	SC=5	SC=10	SC=20
1	1	72	0	0	0	0	40	0	140
2	0	0	0	0	1	0	0	0	160
3	1	0	0	0	36	0	0	56	160
4	0	58	80	0	3	7	0	0	160
5	0	0	16	0	56	80	0	80	0
6	6	64	0	0	3	0	4	35	152
7	0	0	0	0	0	0	4	24	0
8	10	0	2	0	0	0	0	80	160
9	0	56	0	8	0	63	40	0	160
10	0	0	72	0	0	70	0	0	0
Mean value	1.8	25	17	0.8	9.9	22	8.8	27.5	109.2
Standard value	3.245	30.87	29.92	2.40	18.63	32.37	15.6	31.9	71.7

Table 2: Number of overlapping sites:

Test	ICPC=1.0	ICPC=1.5	ICPC=2.0	ML=6	ML=7	ML=8	SC=5	SC=10	SC=20
1	0	9	0	0	0	0	5	0	20
2	0	0	0	0	0	0	0	0	20
3	0	0	0	0	6	0	0	8	20
4	0	8	10	0	0	1	0	0	20
5	0	0	4	0	8	10	0	10	0
6	1	8	0	0	0	0	1	5	19
7	0	0	0	0	0	0	1	4	0
8	2	0	0	0	0	0	0	10	20
9	0	8	0	2	0	9	5	0	20
10	0	0	9	0	0	10	0	0	0
Mean value	0.3	3.3	2.3	0.2	1.4	3.0	1.2	3.7	13.9
Standard value	0.640	4.05	3.79	0.60	2.83	4.38	1.93	4.1	9.10

Table 3: Runtime

Test	ICPC=1.0	ICPC=1.5	ICPC=2.0	ML=6	ML=7	ML=8	SC=5	SC=10	SC=20
1	22.046	31.203	31.875	23.734	28.719	31.766	16.157	33.328	34.734
2	30.813	29.64	31.015	23.734	26.703	32.172	17.609	34.391	35.36
3	29.984	30.969	31.344	24.0	28.516	26.172	16.969	32.938	34.234
4	29.5	30.031	31.125	25.125	28.609	21.25	16.281	32.937	35.656
5	31.219	29.406	29.859	24.469	27.172	31.61	17.469	32.281	33.735
6	31.391	31.094	31.422	23.422	28.5	32.015	21.562	32.844	35.468
7	30.797	30.406	30.797	24.562	28.688	31.094	17.203	35.094	35.641
8	30.656	31.578	31.0	23.719	28.078	31.078	17.172	33.953	19.156
9	30.766,	31.031	31.125	23.953	25.875	31.812	23.0	31.437	17.312
10	29.297	30.891	32.344	25.656	28.39	32.125	20.047	32.594	17.25
Mean value	29.65	30.62	31.19	0.66	0.94	3.40	2.23	1.01	7.86
standard value	2.62	0.684	0.623	24.26	27.93	30.11	18.35	33.18	29.85

Table 4: Relative entropy:

Test	ICPC=1.0	ICPC=1.5	ICPC=2.0	ML=6	ML=7	ML=8	SC=5	SC=10	SC=20
1	inf	inf	inf	inf	inf	inf	inf	inf	inf
2	inf	inf	inf	inf	inf	inf	inf	inf	0
3	inf	inf	inf	inf	inf	inf	inf	inf	0
4	inf	inf	0	inf	inf	inf	inf	inf	0
5	inf	inf	inf	inf	0	0	inf	0	inf
6	inf	inf	inf	inf	inf	inf	inf	inf	0
7	inf	inf	inf	inf	inf	inf	inf	inf	inf
8	inf	inf	inf	inf	inf	inf	inf	0	0
9	inf	inf	inf	inf	inf	inf	0	inf	0
10	inf	inf	0	inf	inf	inf	inf	inf	inf
Mean value	inf	inf	inf	inf	inf	inf	inf	inf	inf
Standard value	inf	inf	inf	inf	inf	inf	inf	inf	inf