

Clustering Analysis

Experiment details:

In this section we compare performance of our **SparseAgglomerativeClustering** algorithm to Matlab's **linkage** function with 'single' method (with default settings) which is equal to our Minimal linkage criteria.

We will use input data in raw format, i.e. data matrix where columns (rows for Matlab) corresponds to objects, rows (columns for Matlab) to their characteristics. Computation time will include metric calculation. So, we have three main parameters for this method:

- Number of objects (N).
- Number of factors in object characteristic (M).
- Type of the metric to compute.

Performance metrics:

In this test we use the following performance metrics:

1. Total working time – time which was needed to complete specified numbers of computations.
2. FLOPS – approximated number of floating point operations number per second. Typically we use mega flops (MFLOPS which) is $FLOPS \times 10^6$. Denote repeats count as K . So we can write experiment complexity in asymptotic form as $O(N^2MK)$. We will use this expression as approximation of floating point calculations count (FLOPS). So for example if we have 250 objects, 10 factors and we make 10000 runs in 15 seconds we will say that method performance is:

$$\frac{N^2MK}{\text{computation time}} = \frac{250^2 * 10 * 10000}{15} = 416666666 \approx 416 \text{ MFLOPS}$$

3. Time of single computation – time which was needed to solve one problem with selected parameters.

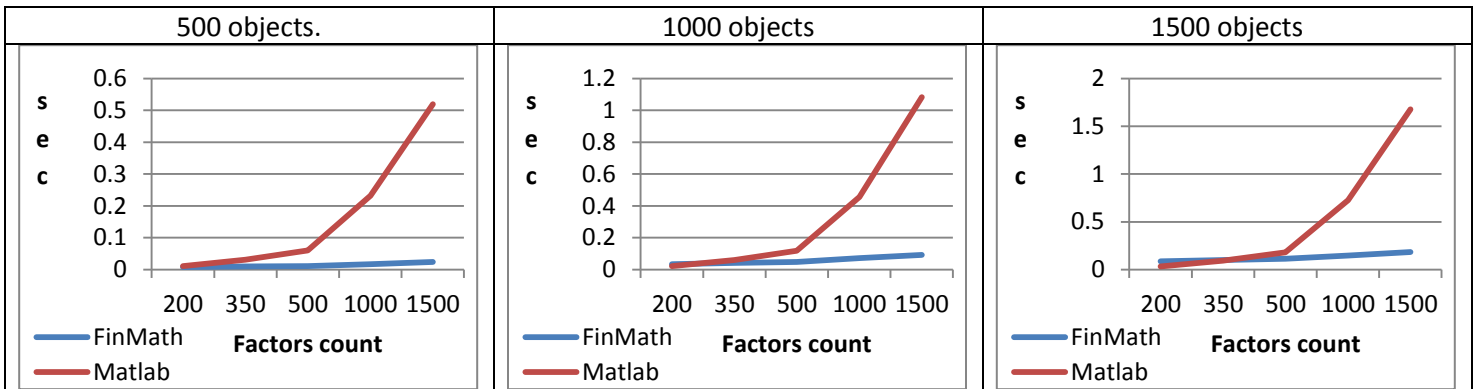
Testing system details:

All tests were performed on typical medium class desktop machine.

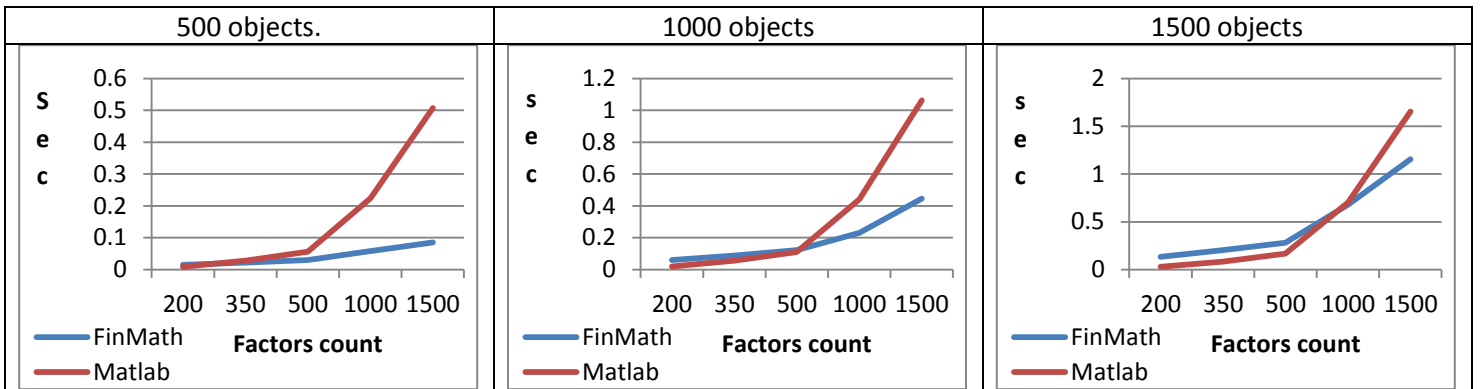
CPU: Core i7 870 (2.93GHz)
Memory: 8GB RAM
OS: Windows 7 x64

All source code of this test are available, one can get performance results on his own system. Changing test parameters is also possible.

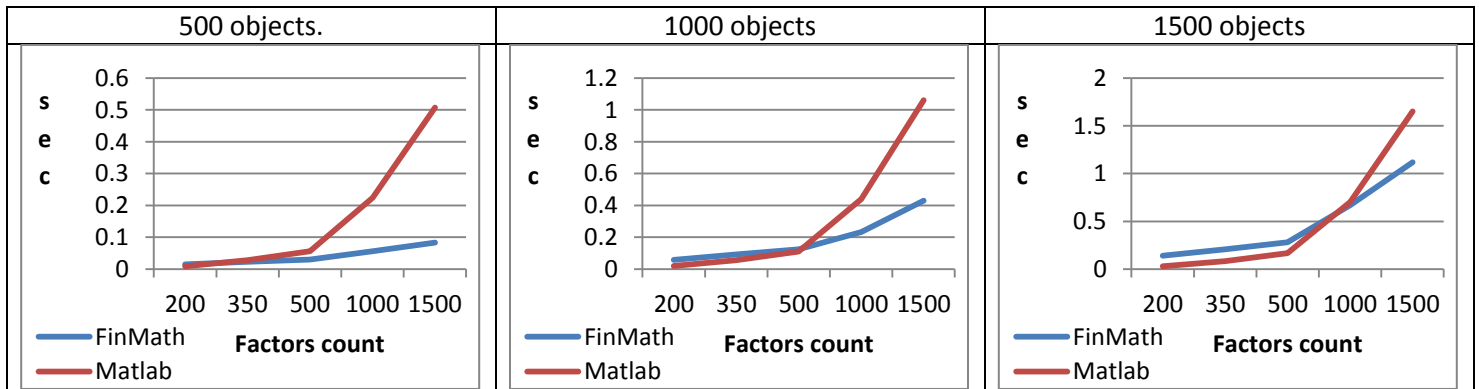
Correlation Distance								
Objects	Factors	Repeats	FinMath			Matlab		
			Total Time	MFLOPS	Single Time	Total Time	MFLOPS	Single Time
500	200	600	4.571	6564	0.008	6.73	4458	0.011
500	350	400	3.806	9196	0.01	12.444	2813	0.031
500	500	300	3.26	11503	0.011	18.072	2076	0.06
500	1000	150	2.605	14396	0.017	34.738	1080	0.232
500	1500	100	2.433	15411	0.024	51.894	723	0.519
1000	200	150	5.132	5846	0.034	3.324	9025	0.022
1000	350	100	4.118	8499	0.041	6.099	5739	0.061
1000	500	70	3.385	10340	0.048	8.337	4199	0.119
1000	1000	40	2.933	13640	0.073	18.279	2189	0.457
1000	1500	20	1.856	16162	0.093	21.654	1386	1.083
1500	200	70	6.021	5232	0.086	2.297	13712	0.033
1500	350	40	3.947	7982	0.099	3.671	8580	0.092
1500	500	30	3.432	9835	0.114	5.374	6281	0.179
1500	1000	12	1.747	15455	0.146	8.713	3099	0.726
1500	1500	7	1.279	18470	0.183	11.738	2013	1.677



Cityblock Metric								
Objects	Factors	Repeats	FinMath			Matlab		
			Total Time	MFLOPS	Single Time	Total Time	MFLOPS	Single Time
500	200	600	8.924	3362	0.015	5.637	5322	0.009
500	350	400	8.924	3922	0.022	11.243	3114	0.028
500	500	300	8.877	4225	0.03	16.82	2230	0.056
500	1000	150	8.737	4293	0.058	33.63	1116	0.224
500	1500	100	8.55	4387	0.085	50.74	740	0.507
1000	200	150	8.799	3410	0.059	2.802	10708	0.019
1000	350	100	8.877	3943	0.089	5.494	6371	0.055
1000	500	70	8.628	4057	0.123	7.798	4489	0.111
1000	1000	40	9.205	4346	0.23	17.634	2269	0.441
1000	1500	20	8.893	3374	0.445	21.254	1412	1.063
1500	200	70	9.283	3394	0.133	1.94	16237	0.028
1500	350	40	8.128	3876	0.203	3.309	9520	0.083
1500	500	30	8.44	3999	0.281	4.991	6763	0.166
1500	1000	12	8.144	3316	0.679	8.453	3195	0.704
1500	1500	7	8.082	2924	1.155	11.573	2042	1.653



Euclidean Distance								
Objects	Factors	Repeats	FinMath			Matlab		
			Total Time	MFLOPS	Single Time	Total Time	MFLOPS	Single Time
500	200	600	9.002	3333	0.015	5.553	5403	0.009
500	350	400	9.08	3855	0.023	11.187	3129	0.028
500	500	300	9.142	4102	0.03	16.761	2238	0.056
500	1000	150	8.456	4435	0.056	33.57	1118	0.224
500	1500	100	8.316	4510	0.083	50.719	740	0.507
1000	200	150	8.768	3422	0.058	2.792	10744	0.019
1000	350	100	9.158	3822	0.092	5.467	6402	0.055
1000	500	70	8.658	4043	0.124	7.753	4515	0.111
1000	1000	40	9.282	4310	0.232	17.598	2273	0.44
1000	1500	20	8.596	3491	0.43	21.246	1413	1.062
1500	200	70	9.813	3211	0.14	1.931	16311	0.028
1500	350	40	8.222	3832	0.206	3.308	9523	0.083
1500	500	30	8.409	4014	0.28	4.973	6787	0.166
1500	1000	12	8.034	3361	0.67	8.426	3205	0.702
1500	1500	7	7.832	3017	1.119	11.556	2045	1.651



Chebyshev Distance								
Objects	Factors	Repeats	FinMath			Matlab		
			Total Time	MFLOPS	Single Time	Total Time	MFLOPS	Single Time
500	200	600	8.596	3491	0.014	5.913	5074	0.01
500	350	400	8.799	3978	0.022	11.943	2931	0.03
500	500	300	9.173	4089	0.031	17.818	2105	0.059
500	1000	150	8.378	4477	0.056	35.636	1053	0.238
500	1500	100	8.222	4562	0.082	53.768	698	0.538
1000	200	150	8.44	3555	0.056	2.704	11097	0.018
1000	350	100	8.939	3916	0.089	5.242	6677	0.052
1000	500	70	8.58	4080	0.123	7.519	4656	0.107
1000	1000	40	9.189	4354	0.23	16.756	2388	0.419
1000	1500	20	8.627	3478	0.431	20.493	1464	1.025
1500	200	70	8.736	3606	0.125	1.808	17422	0.026
1500	350	40	8.019	3929	0.2	3.17	9938	0.079
1500	500	30	8.128	4153	0.271	4.613	7317	0.154
1500	1000	12	8.065	3348	0.672	7.916	3411	0.66
1500	1500	7	7.769	3041	1.11	10.94	2160	1.563

