Parallel Computing of Kernel Density Estimates with MPI - experimental results

Szymon Łukasik

Department of Automatic Control, Cracow University of Technology,
ul. Warszawska 24, 31-155 Cracow, Poland
Szymon.Lukasik@pk.edu.pl

Abstract. Kernel density estimation is nowadays a very popular tool for nonparametric probabilistic density estimation. One of its most important disadvantages is computational complexity of calculations needed, especially for data-based bandwidth selection and adaptation of bandwidth coefficient. The article presents parallel methods which can significantly improve calculation time. Results of using reference implementation based on *Message Passing Interface* standard in multicomputer environment are included as well as a discussion on effectiveness of parallelization.

Key words: kernel density estimation, plug-in method, least squares cross-validation, adaptive bandwidth, parallel algorithms, MPI

Table 1. Execution times and efficiency of parallel kernel density estimation methods

		no. of CPUs (evaluation level)						*no. of CPUs* (sample level)				
m=p		1	2	3	4	5	6	2	3	4	5	6
1000	$T(c)[s]$	0.438	0.366	0.409	0.876	0.796	0.949	0.358	0.452	1.026	0.897	1.116
	$E(c)[\%]$	-	59.9	35.7	12.5	11.0	7.70	61.4	32.4	10.7	9.79	6.56
2000	$T(c)[s]$	1.771	1.416	0.889	1.440	1.222	1.371	1.526	0.870	1.555	1.439	1.419
	$E(c)[\%]$	-	62.6	66.4	30.8	29.0	21.5	58.1	67.9	28.5	24.6	20.8
3000	$T(c)[s]$	4.094	2.245	1.669	1.833	1.666	1.510	2.211	1.658	1.967	1.719	1.653
	$E(c)[\%]$	-	91.2	81.7	55.8	49.1	45.2	93.4	83.0	52.5	48.0	41.6
4000	$T(c)[s]$	6.918	3.639	2.613	2.578	2.198	2.181	3.875	2.637	2.770	2.498	2.714
	$E(c)[\%]$	-	95.1	88.3	67.1	62.9	52.9	90.9	89.1	63.6	56.4	43.3
5000	$T(c)[s]$	10.84	6.203	4.279	3.983	3.253	3.005	6.279	4.341	3.877	3.238	3.162
	$E(c)[\%]$	-	87.4	84.4	68.0	66.6	60.1	87.6	84.5	70.9	67.9	58.0
7500	$T(c)[s]$	24.49	13.56	9.469	7.638	6.349	5.361	14.01	9.435	7.568	6.180	5.432
	$E(c)[\%]$	-	90.3	86.2	80.1	77.1	76.1	87.4	86.5	80.9	79.2	75.1
10000	$T(c)[s]$	43.09	22.50	15.20	12.08	10.06	8.406	23.89	16.30	13.07	10.61	9.005
	$E(c)[\%]$	-	95.8	94.5	89.2	85.7	85.4	90.2	88.1	82.4	81.3	79.8
15000	$T(c)[s]$	97.51	50.68	34.31	26.63	21.26	18.00	53.32	37.53	28.65	22.91	19.26
	$E(c)[\%]$	-	96.2	94.7	91.6	91.7	90.3	91.4	86.6	85.1	85.1	84.4

Table 2. Execution times and efficiency of parallel plug-in method and parallel least squares cross-validation

		no. of CPUs (plug-in)						*no. of CPUs* (LSCV)					
m		1	2	3	4	5	6	1	2	3	4	5	6
1000	$T(c)[s]$	1.566	0.910	0.815	1.334	1.329	1.193	17.95	9.173	6.296	5.314	4.307	3.921
	$E(c)[\%]$	-	86.0	64.0	29.3	23.6	21.9		97.8	95.0	84.4	83.3	76.3
2000	$T(c)[s]$	6.241	3.251	2.384	2.418	2.162	2.043	71.85	36.26	24.37	19.04	15.30	13.05
	$E(c)[\%]$	-	96.0	87.3	64.5	57.7	50.9		99.1	98.3	94.3	93.9	91.7
3000	$T(c)[s]$	14.14	7.426	5.010	4.390	3.567	3.461	165.5	83.43	55.82	42.57	34.13	28.89
	$E(c)[\%]$	-	95.2	94.1	80.5	79.3	68.1		99.2	98.9	97.2	97.0	95.5
4000	$T(c)[s]$	24.79	12.48	8.436	7.042	5.685	5.179	294.8	148.3	99.54	75.02	60.16	50.37
	$E(c)[\%]$	-	99.3	98.0	88.0	87.2	79.8		99.4	98.7	98.2	98.0	97.6
5000	$T(c)[s]$	38.74	19.56	13.28	10.61	8.811	7.424	460.5	231.7	154.7	116.7	93.44	78.61
	$E(c)[\%]$	-	99.1	97.3	91.3	87.9	87.0		99.4	99.2	98.6	98.6	97.6
10000	$T(c)[s]$	152.6	76.77	51.72	39.19	31.31	26.47	1847	927.6	619.3	464.7	371.6	310.4
	$E(c)[\%]$	-	99.4	98.4	97.4	97.5	96.1		99.6	99.4	99.4	99.4	99.2
15000	$T(c)[s]$	344.8	173.0	116.1	87.25	70.09	58.85	4155	2090	1392	1044	835.7	696.8
	$E(c)[\%]$	-	99.7	99.0	98.8	98.4	97.6		99.4	99.5	99.4	99.4	99.4

Table 3. Execution times and efficiency of parallel bandwidth adaptation

		no. of CPUs (adaptation)					
m=p		1	2	3	4	5	6
1000	$T(c)[s]$	1.450	0.931	0.786	1.198	1.069	0.979
	$E(c)[\%]$	-	77.8	61.5	30.3	27.1	24.7
2000	$T(c)[s]$	5.599	3.301	2.391	2.280	2.011	1.877
	$E(c)[\%]$	-	84.8	78.1	61.4	55.7	49.7
3000	$T(c)[s]$	12.34	7.295	5.045	4.248	3.356	3.421
	$E(c)[\%]$	-	84.6	81.5	72.6	73.5	60.1
4000	$T(c)[s]$	21.29	11.62	8.017	6.442	5.376	4.900
	$E(c)[\%]$	-	91.6	88.5	82.6	79.2	72.4
5000	$T(c)[s]$	36.90	19.58	13.36	10.46	8.503	7.560
	$E(c)[\%]$	-	94.2	92.1	88.2	86.8	81.4
7500	$T(c)[s]$	79.24	40.02	27.57	21.67	17.62	15.03
	$E(c)[\%]$	-	99.0	95.8	91.4	89.9	87.9
10000	$T(c)[s]$	140.7	71.20	47.78	36.29	30.40	25.52
	$E(c)[\%]$	-	98.8	98.1	96.9	92.5	91.8
15000	$T(c)[s]$	313.6	158.0	105.6	79.43	63.61	53.49
	$E(c)[\%]$	-	99.3	99.0	98.7	98.6	97.7