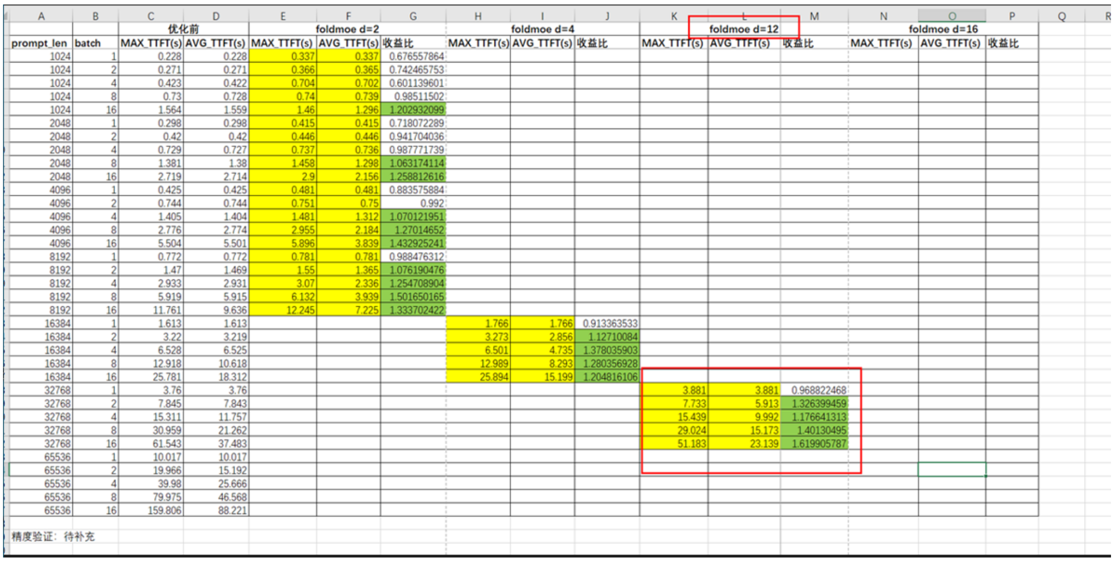

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
		优化前		foldmoe d=2			foldmoe d=4			foldmoe d=12			foldmoe d=16		
prompt_len	batch	MAX_TTFT(s)	AVG_TTFT(s)	MAX_TTFT(s)	AVG_TTFT(s)	收益比	MAX_TTFT(s)	AVG_TTFT(s)	收益比	MAX_TTFT(s)	AVG_TTFT(s)	收益比	MAX_TTFT(s)	AVG_TTFT(s)	收益比
1024	1	0.228	0.228	0.337	0.337	0.676557864									
1024	2	0.271	0.271	0.366	0.365	0.742465753									
1024	4	0.423	0.422	0.704	0.702	0.601139601									
1024	8	0.73	0.728	0.74	0.739	0.98511502									
1024	16	1.564	1.559	1.46	1.296	1.202932099									
2048	1	0.298	0.298	0.415	0.415	0.718072289									
2048	2	0.42	0.42	0.446	0.446	0.941704036									
2048	4	0.729	0.727	0.737	0.736	0.987771739									
2048	8	1.381	1.38	1.458	1.298	1.063174114									
2048	16	2.719	2.714	2.9	2.156	1.258812616									
4096	1	0.425	0.425	0.481	0.481	0.883575884									
4096	2	0.744	0.744	0.751	0.75	0.992									
4096	4	1.405	1.404	1.481	1.312	1.070121951									
4096	8	2.776	2.774	2.955	2.184	1.27014652									
4096	16	5.504	5.501	5.896	3.839	1.432925241									
8192	1	0.772	0.772	0.781	0.781	0.988476312									
8192	2	1.47	1.469	1.55	1.365	1.076190476									
8192	4	2.933	2.931	3.07	2.336	1.254708904									
8192	8	5.919	5.915	6.132	3.939	1.501650165									
8192	16	11.761	9.636	12.245	7.225	1.333702422									
16384	1	1.613	1.613				1.766	1.766	0.913363533						
16384	2	3.22	3.219				3.273	2.856	1.12710084						
16384	4	6.528	6.525				6.501	4.735	1.378035903						
16384	8	12.918	10.618				12.989	8.293	1.280356928						
16384	16	25.781	18.312				25.894	15.199	1.204816106						
32768	1	3.76	3.76							3.881	3.881	0.968822468			
32768	2	7.845	7.843							7.733	5.913	1.326399459			
32768	4	15.311	11.757							15.439	9.992	1.176641313			
32768	8	30.959	21.262							29.024	15.173	1.40130495			
32768	16	61.543	37.483							51.183	23.139	1.619905787			
65536	1	10.017	10.017												
65536	2	19.966	15.192												
65536	4	39.98	25.666												
65536	8	79.975	46.568												
65536	16	159.806	88.221												

精度验证：待补充