1 Analysis of dataset

Analysis about Geolife and T-drive is conducted, below are the histogram graphs of the two datasets based on number of trajectory points.
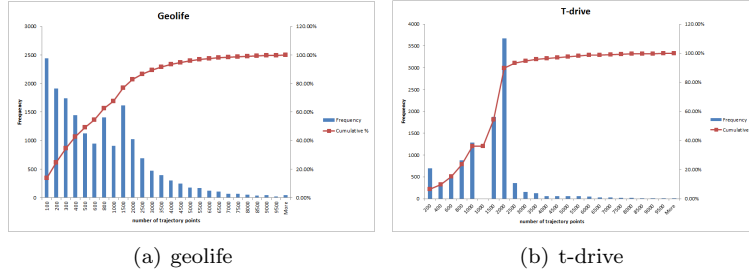

(a) geolife
(b) t-drive

Figure 1: number of trajectory points distribution # different datasets

It seems Geolife has more *short* trajectories which the number of trajectory points are small, while there are a large quantity number of trajectories with a higher number of points in T-drive.

2 running time on different query trajectories

Below are the running time(in log scale) of different similarity measures based on varied query trajectory length.
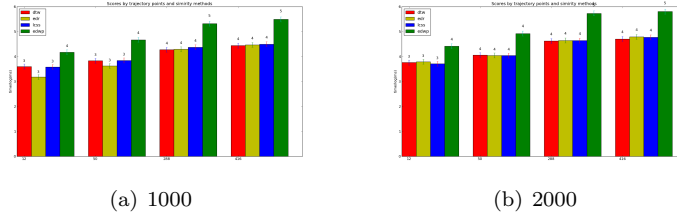

(a) 1000
(b) 2000

Figure 2: running time # different size of dataset

Query trajectories with points of 12, 50, 288 and 416 are chosen as query trajectories, because they can present the major parts of the Geolife dataset according to the histogram graph. And EDwP still runs almost ten times longer than other algorithms.

raw data are listed below:

Table 1: 1000

	dtw	edr	lcss	edwp
12	3996	1507	3835	14934
50	6761	4187	6997	46851
288	18719	19481	23601	208432
416	27584	29180	31221	311053

Table 2: 2000

	dtw	edr	lcss	edwp
12	5780	6168	5138	26087
50	11414	11025	10904	83711
288	41744	43915	43501	537721
416	50788	61113	58896	638533

3 Correlation

The similarities between query trajectory and dataset are recorded in acceding order, to be simple, below are the top 20 results based on DTW and EDwP algorithm.

Table 3: distance results

DTW			EDwP	
trajectory id	distance		trajectory id	distance
16	0		16	0
925	0.260693		146	0.001243
533	0.355299		31	0.002578
510	0.508937		555	0.002686
650	0.760648		239	0.003904
768	0.905939		378	0.004132
537	1.009628		30	0.004135
603	1.021099		222	0.00447
646	1.211956		148	0.004799
725	1.218095		925	0.004913
354	1.278311		898	0.005793
343	1.322424		504	0.005983
491	1.338853		247	0.006372
678	1.35593		975	0.006622
289	1.35958		482	0.006624
103	1.381851		649	0.006679
221	1.384264		484	0.007002
31	1.397892		363	0.007245
166	1.472358		632	0.007635
767	1.501452		732	0.007681

As can be seen from the table, the results differ according to different similarity methods, and the overlap between different methods are quite small so which method is more precise and accurate when computing similarities between trajectories can be also an issue to talk about.