Specialist English: Assignment 10

Rebecca J. Stones
rebecca.stones82@nbjl.nankai.edu.cn

Date due: 27 December 2017

In this tenth assignment (worth 5% of the final mark), we will look at the Experimental Results section.

I'll scale the marks on this assignment according to $m \mapsto \min(m, 10)$ for Master's students and $m \mapsto \lceil m/1.3 \rceil$ for Ph.D. students.

My marking will be affected by (a) your English writing, (b) your LaTeX typesetting, (c) your mathematical presentation, and (d) your understanding of the underlying computer science. Basically, I will "peer review" your assignments.

Problem 1 Looking at Section 4 (entitled Experiment) in the paper Wang et al., *Regularity and Conformity: Location Prediction Using Heterogeneous Mobility Data*, KDD, 2015 (which we looked at in Assignment 7), identify one example of where the authors describe (a) a decision they make, (b) the experimental setup, (c) the baselines, (d) an experimental observation, and (e) some deduction they make from an observation. [5 marks]

Problem 2 Write a paragraph critiquing the following section by Jevring et al. (2008). This paper is available via the third author's webpage `https://www.hesselman.net/`. [3 marks]

5.4 Estimated vs. Real Accuracy

Figure 5 displays the average and standard deviation of the prediction accuracy of different networks that have been optimized by the entropy sort optimizer with different target accuracies. It shows that a higher target accuracy provided to the optimizers results in a higher prediction accuracy of the network when it is actually being used, which suggests that our accuracy metric (Section 4.1) indeed forms a measure of the accuracy of a network.

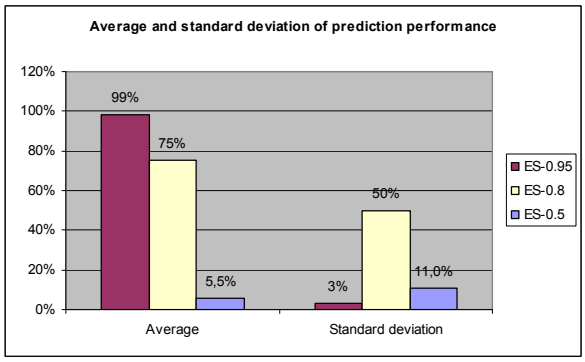


Figure 5. Average and standard deviation of the prediction accuracy (actual accuracy of the localization system) of the entropy sort optimizer using different target accuracies.

— Jevring et al., *Dynamic Optimization of Bluetooth Networks for Indoor Localization*, CSTST, 2008.

(Continued on next page.)

Problem 3 Identify how we can improve the presentation the following table. [1 mark]

	LBR-Meta	LBR	Runtime Ratio of LBR to LBR-Meta
Australian	0.5922	0.4798	0.81
Breast	0.2986	0.1955	0.65
Chess	28.9234	215.403	7.45
Cleve	0.0764	0.0627	0.82
Crx	0.6391	0.5564	0.87
Diabetes	0.2422	0.1235	0.51
German	1.0595	1.1486	1.08
Horse-Colic	0.1124	0.2388	2.12
Hypothyroid	27.9467	36.6264	1.31
Ionosphere	0.250	0.9625	3.85
Mushroom	86.9904	188.848	2.17
Nursery	69.501	71.2436	1.03
Pendigits	62.7794	183.488	2.92
Pima	0.2499	0.1173	0.47
Satimage	46.1528	193.495	4.19
Segment	5.8579	10.8002	1.84
Shuttle-Small	16.386	18.2967	1.11
Sick	24.9702	49.5563	1.98
Solar	0.1062	0.1345	1.27
Soybean-Large	1.4609	15.2015	10.41
Tic-Tac-Toe	0.4281	0.4157	0.97
Vote	0.2048	0.3374	1.65
Waveform-21	17.5639	32.7342	1.86

Table 3: Runtime Comparison (in seconds)

— Xie, *LBR-Meta: An Efficient Algorithm for Lazy Bayesian Rules*, AusDM, 2008.

Problem 4 Identify something suspicious about the following table. [1 mark]

Table 4: Computational cost for different algorithms on the *S. cerevisiae* network (rows indicate different sizes of sub-graph and columns are related to different algorithms), times are in seconds.

	3	4	5	6	7	8	9	10
Kavosh	1.35	34.59	1003.92	20212.99	746385.86	17111178.28	337076691.32	7211199226.13
FANMOD	2.20	41.41	1111.95	24292.05	926745.34	18851135.4	-	-
MAVisto	15784	-	-	-	-	-	-	-
Mfinder	32	306	33548.2	-	-	-	-	-

— Kashani et al., *Kavosh: a new algorithm for finding network motifs*, BMC Bioinformatics, 2009.

(Continued on next page.)

Problem 4 How many data points does Figure 2 in Li et al. (2017) below contain? Based on this, suggest a better way to present this data. [1 mark]

4.2 Coverage of Communities and Recommendation

The coverage of communities is a narrow sense definition. Since, all users registered in system will be allocated to several communities by the demographic information. However, search feature is a more essential factor which reflects user's de-tails. So, in this part, the coverage of communities refers to the coverage of search feature communities. A dozen of new registered users are employed to search in system. All the search behavior will be accorded to analyze the coverage of communities. The statistics information is shown as Figure 2.

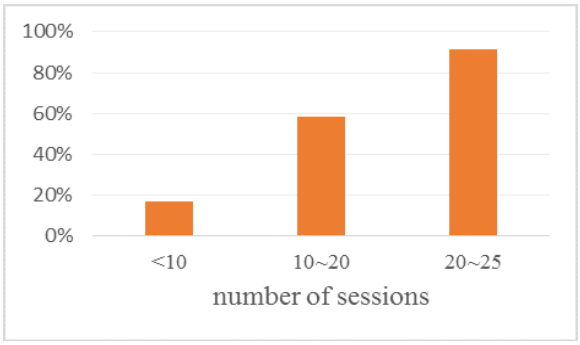


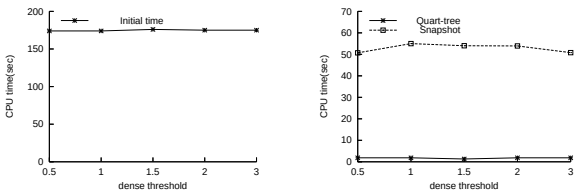
Figure 2. Coverage of communities.

— Li et al., *A Personalized Result Recommendation Method based on Communities*, DMCIT, 2017.

Problem 5 Identify four (or more) ways to improve the following snippet from Hao et al. (2008). [2 marks]

4.2 Results

The first set of experiments studies the impact of ρ on the efficiency of the two algorithms under comparison. Figure 5 shows the CPU time as a function of ρ. Figure 5(a) plots the initialization time for the Quad-tree algorithm, i.e., the Quad-tree building time. Figure 5(b) compares the query answering time for the two algorithms. Both algorithms are not much influenced by ρ, this is because both of them have to search the whole object space, regardless of the value of ρ. Nevertheless, clearly the Quad-tree algorithm is more efficient than the Snapshot algorithm in terms of the query answering time.



(a) initialization time (b) answer time

Figure 5: CPU time vs. density threshold ρ.

— Hao et al., *Continuous Density Queries for Moving Objects*, MobiDE, 2008.

(Continued on next page.)

Problem 6 Include a funny (wholesome) image in your assignment. Last assignment! Yay!
[0 marks; I can give you a smiley face]

Here's one of my favorites ("borrowed" from the Internet):