Reproducibility Report for ACM SIGMOD 2021 Paper: "Efficient Uncertainty Tracking for Complex Queries with Attribute-level Bounds"

ĐORĐE ŽIVANOVIĆ, University of Oxford, United Kingdom

The main core thesis of this paper was successfully reproduced. The time performance scales and accuracy results are comparable to the ones in the paper.

1 INTRODUCTION

This reproducibility report is for the paper "Efficient Uncertainty Tracking for Complex Queries with Attribute-level Bounds" authored by Su Feng, Boris Glavic from the Illinois Institute of Technology Chicago, United States and Aaron Huber, Oliver A. Kennedy from University at Buffalo Buffalo, United States [1]. The main core thesis of this paper was successfully reproduced on the bellow-mentioned machine different from the one used in the paper by the authors.

2 SUBMISSION

The paper results are easily reproduced on the specified OS (Ubuntu 16.04) by following instructions and evaluating the corresponding scripts. The setup of the paper system alongside the competitors is described in the corresponding repository. The submitted content is listed here:

- GitHub repository with code and scripts at https://github.com/fengsu91/AUDB_Reproducibility. These scripts include installation of necessary libraries, generating the data, evaluating the experiments and generating the results in a suitable format.
- Data generators:
  - synthetic dataset https://github.com/IITDBGroup/pdbench;
  - preprocessed real-world data https://drive.google.com/uc?id=16DfJP6F9aHKs7MzUZtkiiaGWirkJ_RBN;
  - microbenchmarking datasets;
- Data sources at URL:
  - Netflix https://www.kaggle.com/shivamb/netflix-shows;
  - Healthcare https://data.medicare.gov/data/hospital-compare;
  - Chicago Crime https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2.

3 HARDWARE AND SOFTWARE ENVIRONMENT

In Table 1 is listed hardware used in the paper and used in the reproducibility review.

Table 1. Hardware & Software environment

	Paper	Repro Review
CPU	AMD Opteron 4238	Intel Xeon Silver 4214
cores	2x6	24
GHz	3.3	1
RAM	128GiB	188GiB
Storage	HDD	SSD
OS	Ubuntu 16.04	Ubuntu 16.04
Virtualization	Docker	Docker

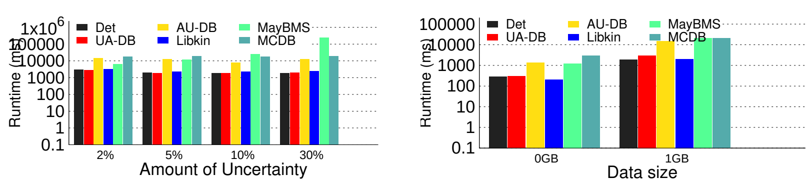
4 REPRODUCIBILITY EVALUATION

4.1 Process

The initial submission had detailed instructions alongside the docker repository that allowed easy evaluation of the experiments. However, the submission did not evaluate all experiments mentioned in the paper. Further, there were no scripts that install libraries on the clean OS image. Furthermore, some of the plots and results did not match the ones in the paper. After several rounds of correspondence, the authors managed to fix all the issues and provide the script that installs all dependencies on a clean image of Ubuntu 16.04.

4.2 Results

The main core thesis of the paper is reproducible on the hardware that differs from the one used in the paper (Figure 1 up to Figure 5 and Table 2, Table 3 and Table 4). The time performance scales and accuracy results are comparable to the ones in the paper. However, the results were observed on different datasets and different query workloads. One issue that remains is that the preprocessed real-world data does not match the ones used in the paper due to the loss of originally preprocessed data. Thus, if possible, the corresponding Reproduced Figure 14 in the paper should be updated with new results.



(a) (b)

Fig. 1. (a) Reproduced Figure 7 (a) from the paper (b) Reproduced Figure 7 (b) from the paper

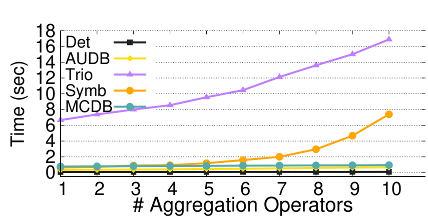
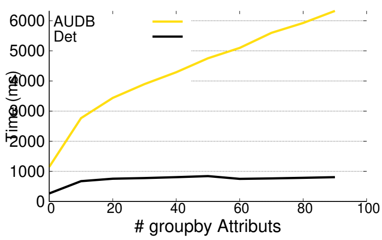


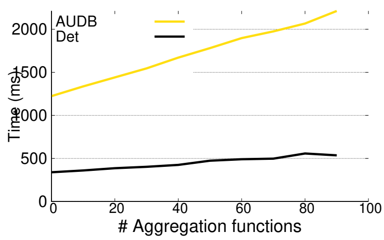
Fig. 2. Reproduced Figure 8 from the paper

Queries	2%/SF0.1	2%/SF0.1	5%/SF1	10%/SF1	30%/SF1
Q1 AU-DB	5131.438	51353.476	52000.801	51344.72	51667.968
Q1 Det	519.419	5357.675	5251.985	5269.145	5236.627
Q1 MCDB	4735.74	46694.57	46744.16	47094.82	46695.89
Q3 AU-DB	1217.677	11488.559	11762.068	12264.907	13435.983
Q3 Det	413.447	2525.695	2591.75	2594.185	2651.765
Q3 MCDB	3993.32	23589.78	23863.92	24167.46	25065.65
Q5 AU-DB	1064.602	13862.823	13885.678	13925.553	14129.551
Q5 Det	198.223	1536.928	1546.541	2517.658	2555.133
Q5 MCDB	1981.71	15305.84	15352.06	25179.01	25317.66
Q7 AU-DB	896.566	8111.511	8128.358	7115.824	12974.659
Q7 Det	163.237	1666.609	1664.348	1946.123	1767.106
Q7 MCDB	1629.81	16625.08	16632.74	19400.22	17707.49
Q10 AU-DB	959.009	11482.409	11472.658	12680.191	13839.897
Q10 Det	164.613	1983.113	1989.965	1974.994	2024.86
Q10 MCDB	1668.77	19811.21	19904.78	19921.54	20032.39

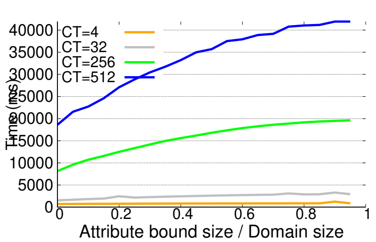
Table 2. Reproduced Figure 9 from the paper



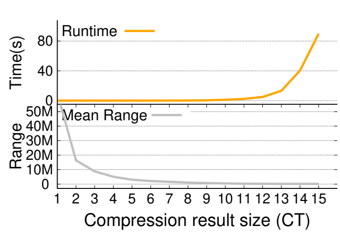
(a)



(b)

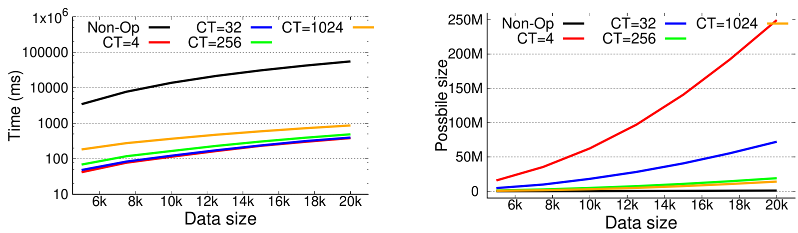


(c)



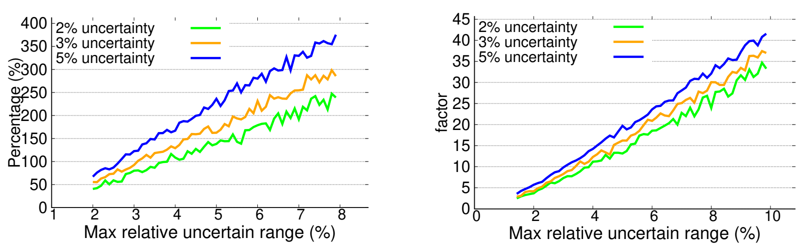
(d)

Fig. 3. (a) Reproduced Figure 10 (a) from the paper (b) Reproduced Figure 10 (b) from the paper (c) Reproduced Figure 10 (c) from the paper (d) Reproduced Figure 10 (d) from the paper



(a) (b)

Fig. 4. (a) Reproduced Figure 11 (a) from the paper (b) Reproduced Figure 11 (b) from the paper



(a) (b)

Fig. 5. (a) Reproduced Figure 12 (a) from the paper (b) Reproduced Figure 12 (b) from the paper

Comp.Size	1 join	2 joins	3 joins	4 joins
4(3%)	0.002597	0.0027	0.003405	0.004578
4(10%)	0.002065	0.002161	0.003321	0.003321
16(3%)	0.002496	0.005529	0.005463	0.007476
16(10%)	0.003347	0.005025	0.005433	0.007375
64(3%)	0.006737	0.018196	0.030347	0.043732
64(10%)	0.006619	0.018759	0.032512	0.048082
256(3%)	0.026172	0.07803	0.120723	0.186269
256(10%)	0.026321	0.088994	0.149281	0.214891
No comp.(3%)	0.012324	0.075853	0.559922	2.285259
No comp.(10%)	0.009872	0.105315	2.34798	59.49469

Table 3. Reproduced Figure 13 from the paper

5 SUMMARY

I want to thank the authors, for all the effort that they put into addressing all the suggestions.

Dataset & Queries	Time (sec)	cert tup.	attr. min	attr. max	pos. tup by id	pos. tup. by val
Netflix SPJ AU-DB	0.0141351	100.00%	1	1	1	1
Netflix SPJ Trio	1.2002151	100.00%	1	1	1	1
Netflix SPJ MCDB	0.5976732	N.A.	1	1	0.9944	0.986
Netflix SPJ UA-DB	0.00690898	100.00%	N.A.	N.A.	0.9835	0.9791
Netflix GB AU-DB	0.0717098	100.00%	1	4	1	1
Netflix GB Trio	1.800192378	100.00%	1	1	0.9883	0.9769
Netflix GB MCDB	0.10238722	N.A.	1	1	0.9984	0.9795
Netflix GB UA-DB	0.006121	0.00%	N.A.	N.A.	0.9912	0.9636
Crimes SPJ AU-DB	1.2221122	100.00%	1	1	1	1
Crimes SPJ Trio	48.021395	100.00%	1	1	1	1
Crimes SPJ MCDB	6.0121108	N.A.	0.6	1	0.999	0.9198
Crimes SPJ UA-DB	0.54749201	100.00%	N.A.	N.A.	0.999	0.8707
Crimes GB AU-DB	2.46722	100.00%	1	1.01	1	1
Crimes GB Trio	100.753	100.00%	1	1	1	1
Crimes GB MCDB	5.0125221	N.A.	0.99	0	1	0.0001
Crimes GB UA-DB	0.591063	0.00%	N.A.	N.A.	1	0
Health SPJ AU-DB	0.293522	100.00%	1	1	1	1
Health SPJ Trio	23.12132	100.00%	1	1	1	1
Health SPJ MCDB	0.617443	N.A.	0.4	1	0.999	0.8759
Health SPJ UA-DB	0.0590882	98.21%	N.A.	N.A.	0.9921	0.6551
Health GB AU-DB	0.8472451	100.00%	1	45	1	1
Health GB Trio	40.472956	100.00%	1	1	1	1
Health GB MCDB	3.42879362	N.A.	0.78	1	1	0
Health GB UA-DB	0.2982354	0.00%	N.A.	N.A.	1	0.0001

Table 4. Reproduced Figure 14 from the paper

REFERENCES

[1] Su Feng, Boris Glavic, Aaron Huber, and Oliver A Kennedy. 2021. Efficient Uncertainty Tracking for Complex Queries with Attribute-level Bounds. In *Proceedings of the 2021 International Conference on Management of Data*. 528–540.