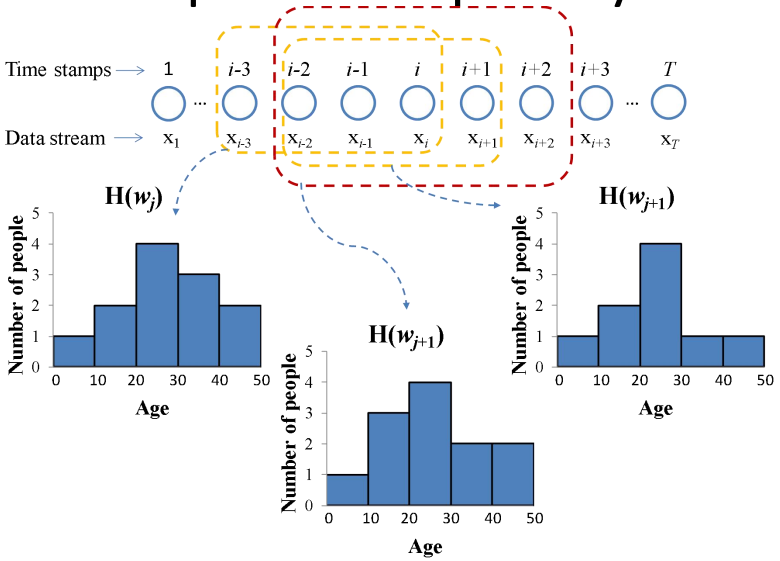
Differential Privacy Histogram Publishing Method based on Dynamic Sliding Window

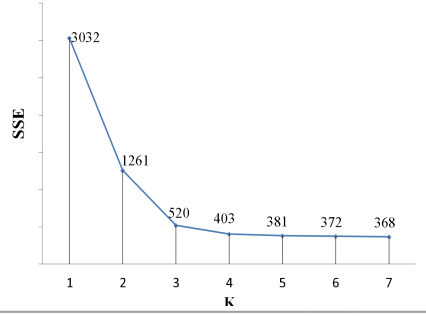
Qian CHEN, Zhiwei NI, Xuhui ZHU, Pingfan XIA

Frontiers of Computer Science, DOI: 10.1007/s11704-022-1651-2

Problems & Ideas

- Problems of conventional dynamic data release approaches:
  - Unreasonable window division leads to privacy leakage when data changes dynamically.
  - Isometric histogram publishing hides the significant data distribution characteristics leading to decreased data availability.
- Ideas: a differential privacy histogram publishing method based on the dynamic sliding window of LSTM (DPHP-DL) is proposed to improve the privacy and availability of dynamic data.



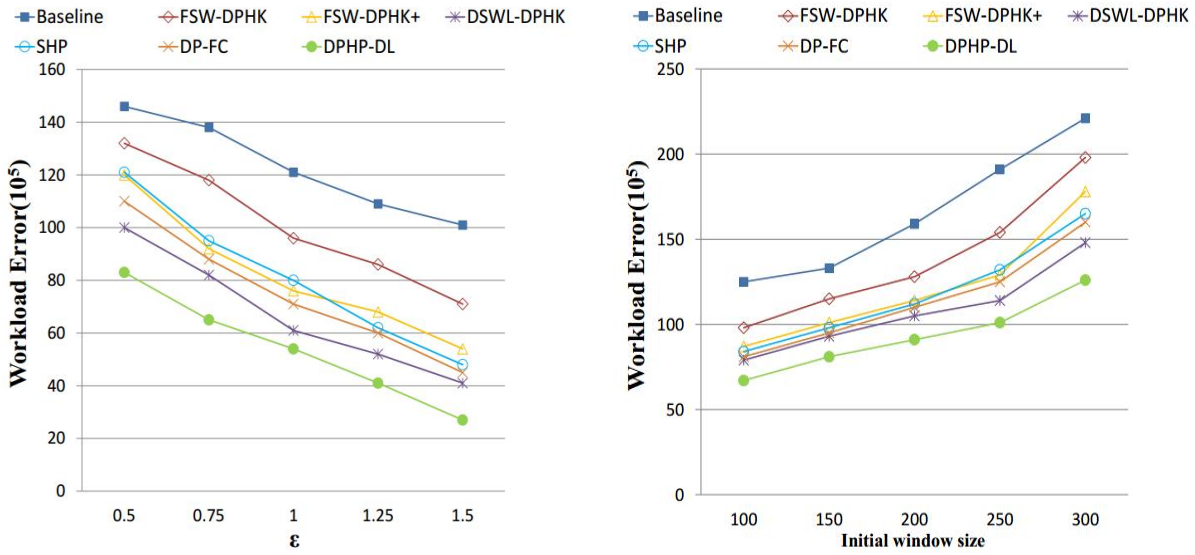


$k_{1,2}$	$k_{2,3}$	$k_{3,4}$	$k_{4,5}$	$k_{5,6}$	$k_{6,7}$
-1771	-741	-117	-22	-9	-4
r_2	r_3	r_4	r_5	r_6	
0.817	1.200	0.236	0.034	0.013	

DPHP-DL is integrated by DSW-LSTM and DPHK+. Left: DSW-LSTM updates the size of sliding windows in a data stream based on data value prediction via Long Short-Term Memory (LSTM). Right: DPHK+ obtains the optimal K value automatically based on k-mean++ to heuristically publish non-isometric histograms.

Main Contributions

- Contributions:
  - The dynamic sliding window based on LSTM (DSW-LSTM) is proposed to update the size of the sliding window adapting to dynamic data changes;
  - Differential privacy histogram publishing based on K-means++ (DPHK+) is proposed to implement heuristic non-isometric histogram publishing, which can accurately reflect the data distribution in each time window;
  - A novel method DPHP-DL is proposed, combining DSW-LSTM and DPHK+. DPHP-DL is experimentally evaluated on real synthetic dynamic datasets to indicate its effectiveness and efficiency in data privacy and availability.



Left: workload error with various privacy budgets on the Adult Dataset; Right: Workload error with various initial window sizes on the Ocular Disease Dataset.