Spatio-temporal Range Searching Over Compressed Kinetic Sensor Data

Sorelle A. Friedler and David M. Mount
University of Maryland, College Park

Motivation

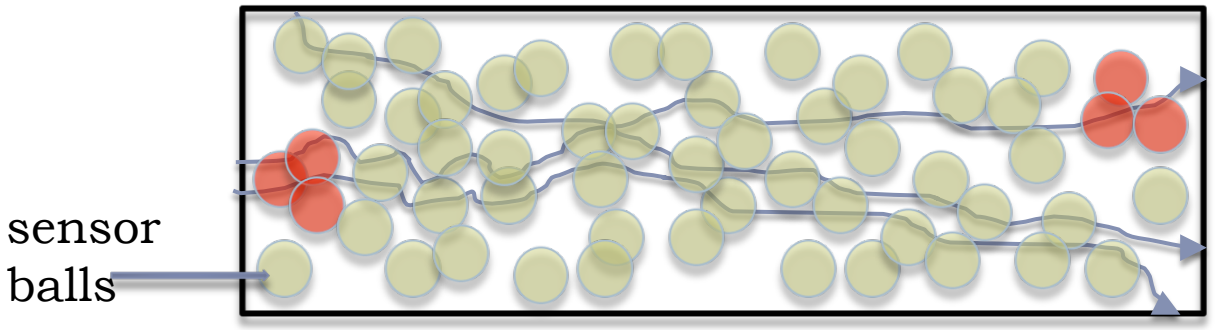
- Kinetic data: data generated by moving objects
- Sensors collect data
- Large amounts of data
- Collect and perform lossless compression
- Goal: retrieve without decompressing
- Next: analyze

Motivation

- Computer Science
  - Graphics: Image and video segmentation, animation
  - Databases: Maintenance over time
  - Sensor Networks: Data analysis
- Physics
  - Simulations
- Biology
  - Mathematical ecology: Migratory paths
  - HIV strain analysis
- Engineering
  - Traffic patterns and identification

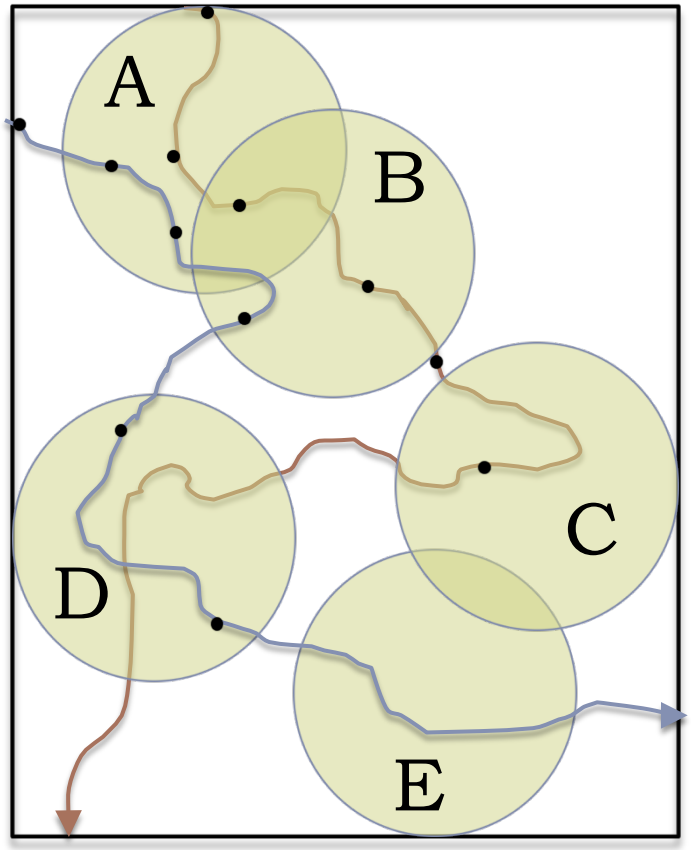
Our Framework [FriedlerMount09]

- Detection region around each sensor (stationary sensors)
- Point motion unrestricted
- No advance knowledge about motion
- Each sensor reports the count of points within its region at each synchronized time step
- *k*-local: Sensor outputs statistically only dependent on *k* nearest neighbors



Data Collection

Data based on underlying geometric motion



Sensor data streams

A	B	C	D	E
1	0	0	0	0
2	0	0	0	0
2	1	0	0	0
0	2	0	0	0
0	0	0	1	0
0	0	1	1	0

time

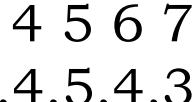
Range Searching: Our Problem

Compress and preprocess the data so as to perform…

- Temporal range query: Given a time interval, return an aggregation of the counts over that time interval.

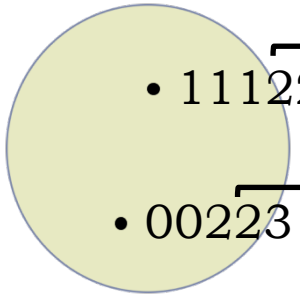
aggregation type: sum

t: 1 2 3 4 5 6 7 8 9 10 11
X: 0,0,4,4,5,4,3,3,1, 1, 0



17

- Spatio-temporal range query: Given a time interval and spherical spatial region, return an aggregation of the counts over that time interval and within that region.



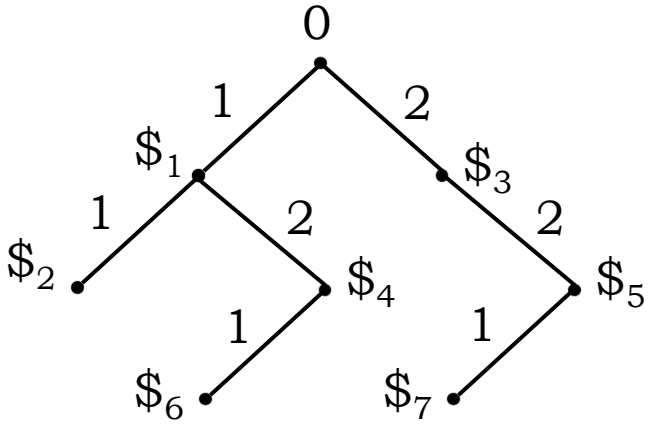
• 11122021…
• 00110123…
• 00223101…

10

aggregation type: sum

Lempel-Ziv Dictionary Compression [LZ78]

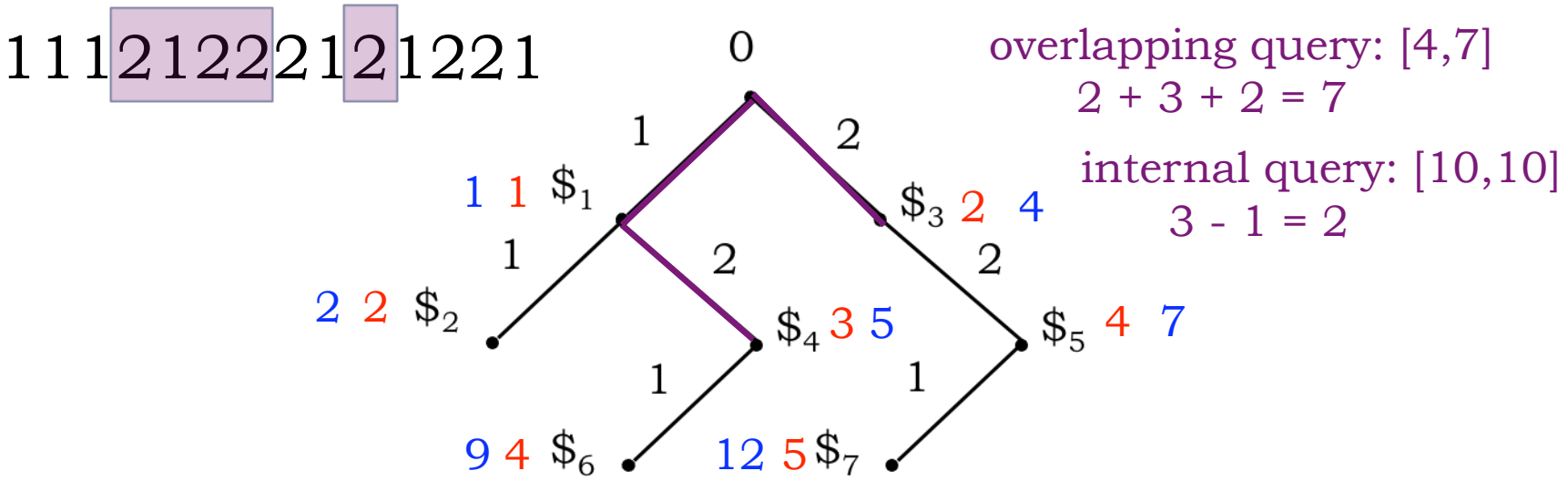
11121222121221



Create a trie while scanning through a string. The compressed string contains pointers to this dictionary.

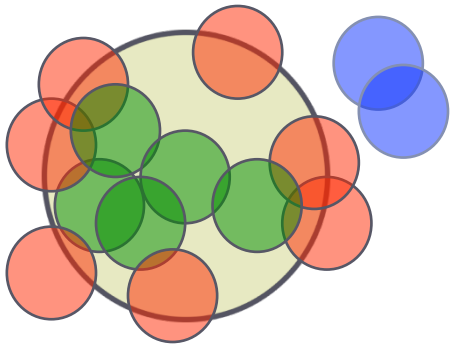
Temporal Range Searching

- Create trie with accompanying pointers
- Annotate trie with aggregate values and word start times
- Given a temporal range $[t_0, t_1]$ find the anchor points $\$^0$ and $\$^1$ such that $\$^0 \leq t_0$ and $\$^1 \geq t_1$
- Use stored prefixes, words, and subtraction of prefixes to find aggregates



Spatio-temporal Range Searching

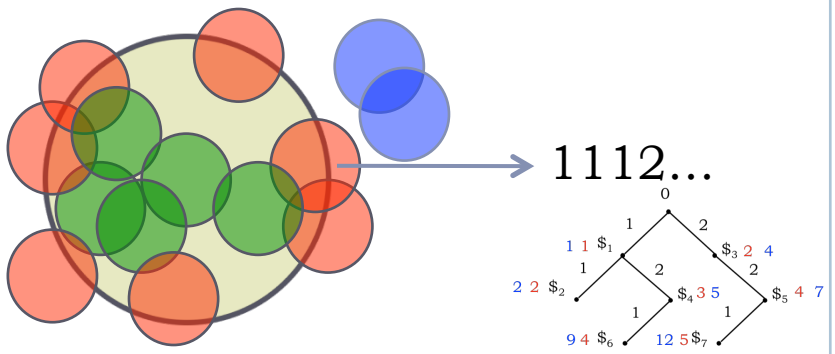
Set of Clumps	Clumps and Ranges
A finite set of balls with the following packing property: Given any ball b with radius r, the number of clumps of radius r' that intersect b is at most $O(1+(r/r')^{\gamma})$ where γ is a constant possibly dependent on dimension.	Given a range Q, a clump may either ▸ be contained in Q (green) ▸ be stabbed by Q (red) ▸ lie outside of Q (blue)



Spatio-temporal Range Searching

- The sensors are clustered and compressed. Each cluster has an associated ball. These balls form a set of clumps.
- Range Searching Among Clumps: Given any query range Q and using a partition-tree, we can report
  - a subset of clump subsets that form a disjoint cover of the clumps within Q
  - the subset of clumps that Q stabs

- Main Theorem: By adding an auxiliary data structure to answer temporal range queries to each node in the range searching among clumps solution we can answer spatio-temporal range queries.



Results

Bounds for Range Searching

	Temporal	Spatio-temporal
Preprocessing time	$O(\mathrm{Enc}(X))$	$O(\mathrm{Enc}(\mathbf{X}))$
Query time	$O(\log T)$	$O(((1/\varepsilon^{d-1}) + \log S) \log T)$
Space	$O(\mathrm{Enc}(X))$	$O(\mathrm{Enc}(\mathbf{X}) \log S)$

- X: The set of sensor system observations
- Enc(X): The encoded size (in bits) of the sensor system
- T: The total time over which data was collected
- S: The total number of sensors
- d: The dimension of the sensor space

First range searching bounds over compressed data

Other Results and Future Work

- Temporal range searching in the semigroup setting ✓
- Extend analysis of compression algorithm to consider empirical entropy (no underlying random process) ✓
- Extend analysis to allow δ-independence instead of strict pure statistical dependence ✓

- Statistical analysis without decompressing the data
- Lossy compression
- Experimental evaluation
- Application in non-sensor contexts

Thank you!

Questions?