Hyperbolic t-SNE with a Quadtree Splitting in the Cartesian Coordinate System

Author: Yehor Kozyr
Supervisors: Martin Skrodzki, Elmar Eisemann

TU Delft

1 Research questions

- Is it possible to implement a t-SNE algorithm for a Poincaré disk model using a Barnes-Hut Approximation based on a quadtree?
- How does it compare to the implementation with Quadtree with splitting in polar coordinate system?

2 Methodology

- Implemented a quadtree with Cartesian splitting
- Introduced shortcuts
- Proposed a heuristic for approximation of the maximum distance within a cell
- Conducted a series of experiments to compare polar quadtree, Cartesian quadtree and the exact algorithms

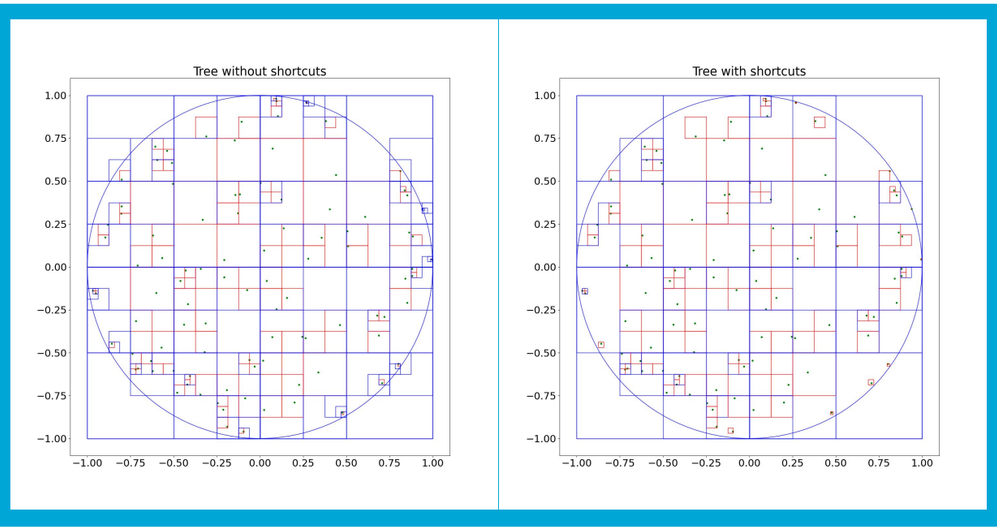

Fig. 1 Cartesian tree before shortcuts were introduced applied to the Poincaré disk model (left) and after they were introduced (right)

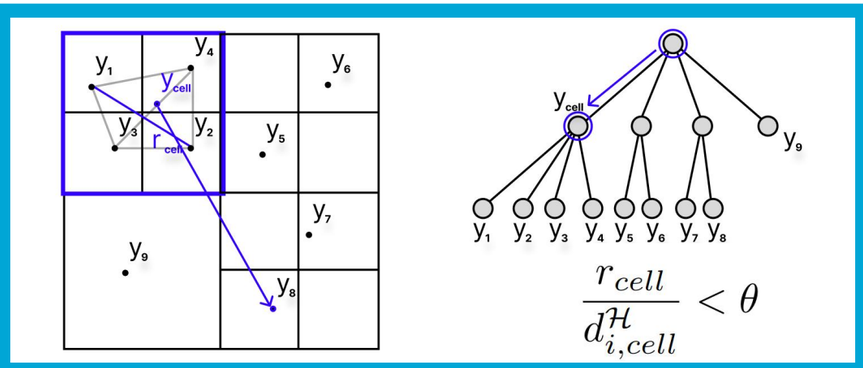

Fig. 2 Schematic drawing of the proposed modification to the Barnes-Hut approximation

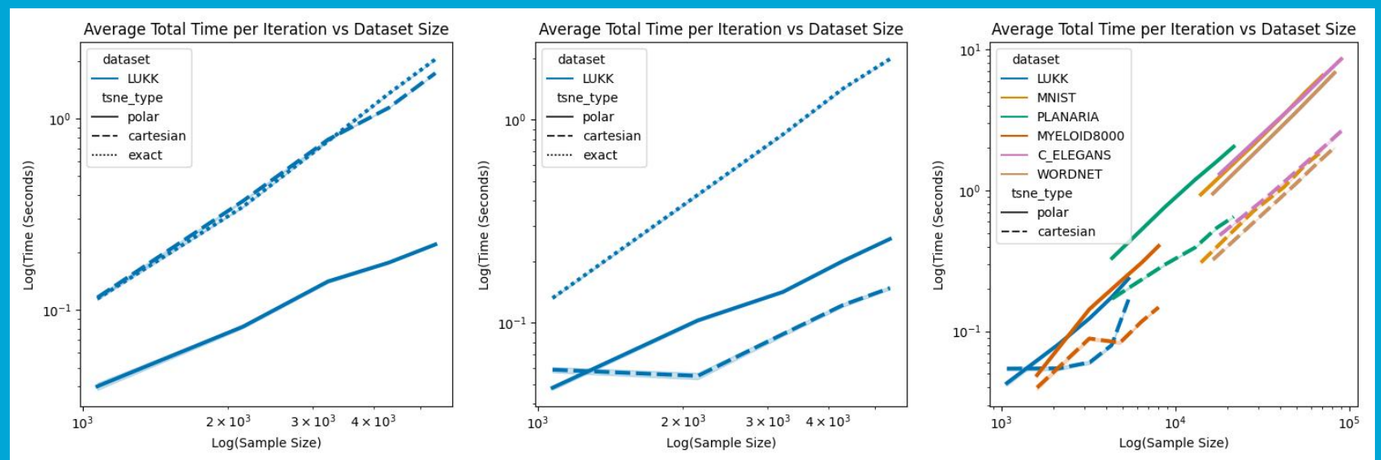

Fig. 3 Graphs of the run time before the change (left) and after (center and right)

Name	Cartesian[s] avg / min / max / std	Polar [s] avg / min / max / std
LUKK	0.17 / 0.06 / 0.70 / 0.08	0.24 / 0.84 / 0.07 / 0.06
MYELOID8000	0.15 / 0.09 / 0.59 / 0.05	0.40 / 0.03 / 1.44 / 0.15
MNIST	1.91 / 1.72 / 2.68 / 0.12	6.53 / 3.45 / 13.9 / 0.42
WORDNET	2.02 / 1.81 / 5.91 / 0.21	6.84 / 2.90 / 14.9 / 0.83
PLANARIA	0.65 / 0.56 / 1.09 / 0.05	2.05 / 1.07 / 4.76 / 0.43
C_ELEGANS	2.64 / 2.35 / 40.9 / 1.83	8.56 / 7.31 / 10.2 / 0.40

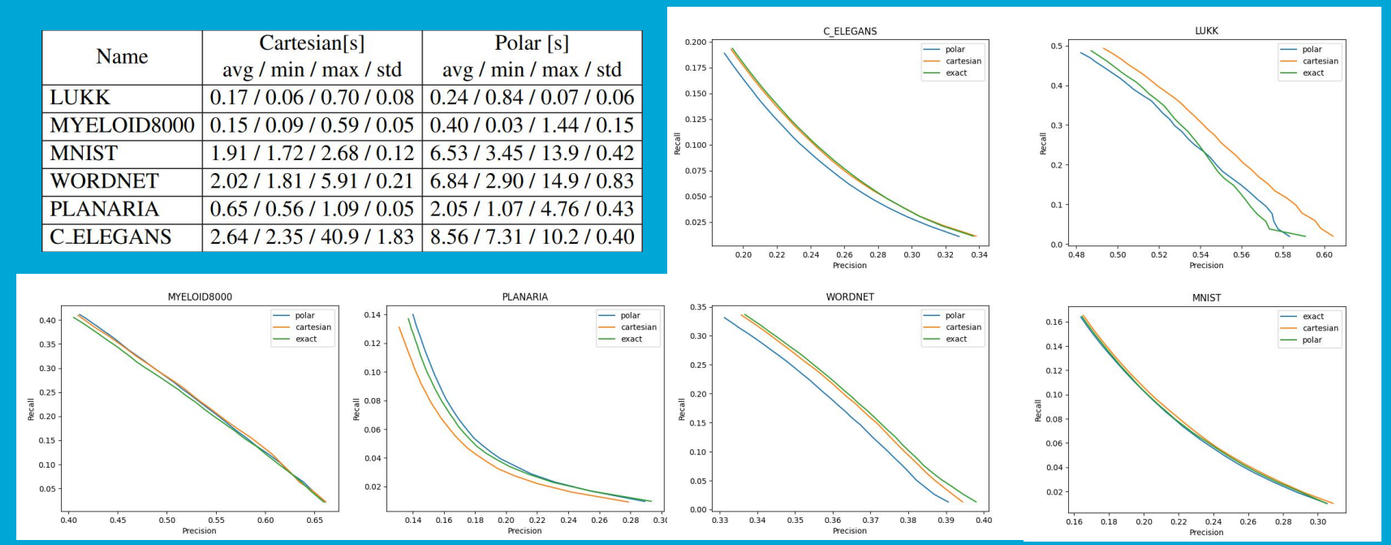

Fig. 4 Precision/recall graphs for each dataset

3 Conslusions

- Shown that the use of the quadtree with splitting in the Cartesian coordinate system for hyperbolic t-SNE is possible.
- Proposed approximation using the maximum distance between the cell's children's barycenters the algorithm was shown to be faster than the exact computation
- Shown that the Cartesian quadtree outperforms the polar quadtree on the metric of time spent per iteration of the algorithm, without a sacrifice in the quality of the embeddings