Linear-Complexity Data-Parallel Earth Mover’s Distance Approximations

Kubilay Atasu, Thomas Mittelholzer
Earth/Word Mover’s Distance: Discrete Wasserstein Distance

The Queen to tour Canada

Royal visit to Halifax

embedding space

out-flow constraints

in-flow constraints

<table>
<thead>
<tr>
<th></th>
<th>Search Accuracy</th>
<th>Complexity</th>
<th>GPU friendly</th>
<th>Optimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMD/WMD</td>
<td>Very high</td>
<td>$h^3 \log h$</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Sinkhorn</td>
<td>Very high</td>
<td>$(h^2 \log h) / \epsilon^2$</td>
<td>Yes</td>
<td>Within $\epsilon$</td>
</tr>
<tr>
<td>RWMD</td>
<td>High</td>
<td>$h$</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Our Work</td>
<td>Very high</td>
<td>$hk$</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Our Solution: Iterative Constrained Transfers (ICT) Algorithm

- Relaxed in-flow constraints
- Edge capacity constraints

Sort the edges in the increasing order of costs
Iterative mass transfers under capacity constraints

Approximate ICT (ACT) algorithm: only $k$ iterations
ICT & ACT are tighter lower bounds than RWMD: $\text{RWMD} \leq \text{ACT} \leq \text{ICT} \leq \text{EMD}$
Experiments: Runtime vs Nearest-Neighbors-Search Accuracy

- ACT effective on sparse as well as dense, low- as well as high-dimensional datasets
- 20’000 faster than WMD and matches its search accuracy on 20 Newsgroups
- 10’000 faster and offers a slightly higher search accuracy than Sinkhorn on MNIST

20News: high-dimensional, sparse histograms
MNIST: two-dimensional, dense histograms

WCD: Word centroid distance (Euclidean)
BoW: Bag-of-Words (Cosine similarity)
WMD: Word Mover’s Distance (Kusner et al.)
RWMD: Relaxed Word Mover’s Distance
OMR and ACT-k: the new algorithms
Linear-Complexity Data-Parallel Earth Mover’s Distance Approximations

Thank You!

Check-out our poster #218 in Pacific Ballroom!