CSAIL

OUR APPROACH

OPTIMIZATION

- The problem is non-convex (even after entropic regularization!)

- Naive solution involves operating on 4-th order tensor

- Yet, regularized version **solved efficiently**! [Peyre et al. 2016]

- Iterative: projected gradient descent

- Projections involve solving a classic entropy-regularized OT problem (Sinkhorn)

- For very large problems, we propose a **two-step approach**:

1. Solve GW problem for a subset of the points

2. Use predicted matches to learn orthogonal mapping across spaces

$$\mathrm{GW}(\mathbf{C}, \mathbf{C}', \mathbf{p}, \mathbf{q}) = \min_{\Gamma \in \Pi(\mathbf{p}, \mathbf{q})} \sum_{i,j,k,l} \mathcal{L}(\mathbf{C}_{ik}, \mathbf{C}'_{jl}) \Gamma_{ij} \Gamma_{kl}$$

OPTIMIZATION

$$\mathrm{GW}(\mathbf{C}, \mathbf{C}', \mathbf{p}, \mathbf{q}) = \min_{\Gamma \in \Pi(\mathbf{p}, \mathbf{q})} \sum_{i,j,k,l} \mathcal{L}(\mathbf{C}_{ik}, \mathbf{C}'_{jl}) \Gamma_{ij} \Gamma_{kl}$$

- The problem is non-convex (even after entropic regularization!)
- Naive solution involves operating on 4-th order tensor
- Yet, regularized version **solved efficiently**! [Peyre et al. 2016]
  - Iterative: projected gradient descent
  - Projections involve solving a classic entropy-regularized OT problem (Sinkhorn)
- For very large problems, we propose a **two-step approach**:
  1. Solve GW problem for a subset of the points
  2. Use predicted matches to learn orthogonal mapping across spaces

ALGORITHM: GW ACROSS WORD EMBEDDING SPACES

Algorithm 1 Gromov-Wasserstein Computation for Word Embedding Alignment

Input: Source and target embeddings $\mathbf{X}$, $\mathbf{Y}$. Regularization λ. Probability vectors $\mathbf{p}, \mathbf{q}$.
// Compute intra-language similarities
$\mathbf{C}_s \leftarrow \cos(\mathbf{X}, \mathbf{X}), \quad \mathbf{C}_t \leftarrow \cos(\mathbf{Y}, \mathbf{Y})$
$\mathbf{C}_{st} \leftarrow \mathbf{C}_s^2 \mathbf{p} \mathbb{1}_m^\top + \mathbb{1}_n \mathbf{q} (\mathbf{C}_t^2)^\top$
while not converged **do**
  // Compute pseudo-cost matrix (Eq. (9))
  $\hat{\mathbf{C}}_\Gamma \leftarrow \mathbf{C}_{st} - 2\mathbf{C}_s \Gamma \mathbf{C}_t^\top$
  // Sinkhorn iterations (Eq. (7))
  $\mathbf{a} \leftarrow \mathbb{1}, \quad \mathbf{K} \leftarrow \exp\{-\hat{\mathbf{C}}_\Gamma / \lambda\}$
  while not converged **do**
    $\mathbf{a} \leftarrow \mathbf{p} \oslash \mathbf{K}\mathbf{b}, \ \mathbf{b} \leftarrow \mathbf{q} \oslash \mathbf{K}^\top \mathbf{a}$
  end while
  $\Gamma \leftarrow \operatorname{diag}(\mathbf{a}) \, \mathbf{K} \operatorname{diag}(\mathbf{b})$
end while
// Optional step: Learn explicit projection
$\mathbf{U}, \Sigma, \mathbf{V}^\top \leftarrow \mathrm{SVD}(\mathbf{X} \Gamma \mathbf{Y}^\top)$
$\mathbf{P} = \mathbf{U}\mathbf{V}^\top$
return $\Gamma, \mathbf{P}$

CSAIL