Model Transport: Towards Scalable Transfer Learning on Manifolds

Oren Freifeld
MIT

Søren Hauberg
Technical University of Denmark

Michael J. Black
MPI for Intelligent Systems

Introduction

- **Manifold-valued data** are ubiquitous in computer vision: surface normals, shape spaces, histogram-valued features, Symmetric Positive-Definite (SPD) matrices, the Grassmannian, etc.
- **Statistics on manifolds** is often done via tangent-space models; e.g., Gaussians, PCA, regression, classifiers, etc.
- One form of **transfer-learning (TL)** leverages a model learned in one region of $\mathbb{R}^n$ to improve a model in another region.
- **Goal:** Exploit TL ideas in modeling manifold-valued data.
- **Problem 1:** on a manifold, conventional $\mathbb{R}^n$-TL fails.
- **Problem 2:** this is not scalable.

Solution: Transport the model, not the data

Key Contributions

- **Scalability:** We show how models can be transported. $\implies$ # computations is fixed w.r.t. # data points; no need to store the data.
- **Optimality:** We show that for these models, PT and learning commute.

Parallel Transport (PT)

- An established tool to move vectors between tangent spaces.
- A **metric parallel transport (MPT)**: inner-product-preserving PT.
- Both MPT and non-metric PT are widely used in computer vision – but focus has been on expensive data transport.

Covariance/PCA Transport

(In short: keep the std. dev., transport the eig. vecs)

- The transported model can improve the small-sample model.
- Point: only the eigen vectors need to be transported.

Regression/Classification Transport

(In short: transport the coefficient vector)

- Point: only a single vector needs to be transported.
- Similar results hold for linear-regression and SVM (PT the support vectors).

Why Does This Work?

- $M$: an $n$-dimensional manifold; $x \in T_pM$, the (metric) PT of $x \in T_qM$.
- Data: $(x_i)_{i=1}^N \subset T_pM, \hat{X} \triangleq [x_1, \ldots, x_N], \tilde{X} \triangleq [\tilde{x}_1, \ldots, \tilde{x}_N]$: (PT of the data)
- **Proposition 1** (Covariance/PCA Transport):
  - Let $V \in SU_+$, $X \rightarrow VXV^T$ is equivariant under $T_pM \rightarrow T_qM$.
  - Let $\tilde{V} \in SU_+$, $\tilde{X} \rightarrow \tilde{V}X\tilde{V}^T$ is equivariant under $T_pM \rightarrow T_qM$.
  - If $k < n$, then the $k$-dimensional PCA model of $(x_i)_{i=1}^N \subset T_qM$ is given by $\{\tilde{v}_i\}_{i=1}^k$, and $(S/\sqrt{N})\tilde{v}_i^T$ where $[\tilde{v}_1, \ldots, \tilde{v}_k] = V$.

- $(\cdot,\cdot)_p(x,y) \rightarrow x^T A_p y$ and $(\cdot,\cdot)_q(x,y) \rightarrow x^T A_q y$ are inner products on $T_pM$ and $T_qM$, respectively. $A_p, A_q \in \text{SPD}$. Data labels: $(y_i)_{i=1}^N \subset \mathbb{R}$.
  - $L: T_pM \rightarrow T_qM$: the linear map associated with an MPT. A linear regression model $T_pM \rightarrow \mathbb{R}$:
    - $\tilde{x} \rightarrow \tilde{x}^T \alpha + \nu_0 = \langle \tilde{x} A_p^T \alpha + \nu_0 \rangle + \nu = \mathbb{R}$.
    - $\alpha, A_p: \mathbb{R}^n \rightarrow \mathbb{R}^k$.
  - $k: \mathbb{R} \rightarrow \mathbb{R}$: a loss function associated with $\gamma$, e.g., $k: \hat{y} \rightarrow (\hat{y} - y)^2$.

**Proposition 2** (Linear-Regression Transport):

$\gamma, \beta_0 = \arg\min_{\alpha} \sum_{i=1}^N k(\tilde{x}_i^T \alpha + \nu_0) \implies \gamma \triangleq A_p \beta_0 \Delta \beta = \arg\min_{\beta} \sum_{i=1}^N k(L(x_i)^T \beta + \beta_0)$

Acknowledgments

This work is supported in part by NH-NINDS EUREKA (R01-NIN063137). S.H. is supported in part by the Villum Foundation and the Danish Council for Independent Research (Natural Sciences).

Application: Covariance Transport

Woman shapes improve a shape model of men while shapes of people with normal Body-Mass Index (BMI) improve a shape model of high-BMI people.

Application: Classifier Transport

Features were encoded as SPD metrics. PT improves logistic-regression classifier results from 59% to 67%.
- Points: the same performance gain was obtained regardless whether we transported the data (168 vectors) or the model (a single vector).