CSE 291: Domain Adaptation in Vision
Manmohan Chandraker

Lecture 7: Correlation Alignment
Reviewing Papers

• The Feynman approach
  – Read the abstract and problem definition
  – Imagine how you would solve the problem
  – See how the authors approach differs from yours

• The “Newton” approach: stand on the shoulders of giants
  – See which papers are referenced
  – Compare the methods
  – See how papers that cite this one critique it!

• A good approach is somewhere in between the two
  – Rely on your intuition and problem-solving skills
  – Rely on experience gathered by studying several papers in the area
Frustratingly Easy Supervised Adaptation

- Labeled data available in source and (limited) target domain
- Easy adaptation: train classifier in augmented space

Let $\mathcal{X} = \mathbb{R}^F$ be $F$-dimensional feature space.

Define $\Phi^s: x \mapsto \langle x, x, 0 \rangle$,

$\Phi^t: x \mapsto \langle x, 0, x \rangle$
Frustratingly Easy Supervised Adaptation

• Labeled data available in source and (limited) target domain
• Easy adaptation: train classifier in augmented space

$$\mathcal{X} = \mathbb{R}^F$$ be $F$-dimensional feature space.

Define $\Phi^s: x \mapsto \langle x, x, 0 \rangle$
$\Phi^t: x \mapsto \langle x, 0, x \rangle$

• Kernel version of easy adaptation:

$$\Phi^s(x) = \langle \Phi(x), \Phi(x), 0 \rangle$$
$$\Phi^t(x) = \langle \Phi(x), 0, \Phi(x) \rangle$$

• For inputs from the same domain (both $s$ or both $t$):

$$\tilde{K}(x, x') = \langle \Phi(x), \Phi(x') \rangle_\mathcal{X} + \langle \Phi(x), \Phi(x') \rangle_\mathcal{Y} = 2K(x, x')$$

• For inputs from different domains (one $s$, one $t$)

$$\tilde{K}(x, x') = \langle \Phi(x), \Phi(x') \rangle_\mathcal{X} = K(x, x')$$

• Target data has twice as much weight for target domain predictions!
Aligning Domains and Bridging Domains

• A few approaches to align source and target distributions
  – TCA, DDC, DAN, ....

• Another approach:
  – Transform one domain to *match statistics* of other domain
  – Classifier trained on *source labels* likely to work well on target
Matching First Order Statistics

• First attempt: normalize *first order* statistics of source and target
  – Not sufficient to reduce data to zero mean and unit standard deviation
  – Covariance of data distributions can still differ
CORAL: Correlation Alignment

• First attempt: normalize first order statistics
  – Not sufficient to reduce data to zero mean and unit standard deviation
  – Covariance of data distributions can still differ

• CORAL: Match distributions by aligning second-order statistics
  – A frustratingly easy method that does better in some scenarios
  – Does not require re-training or hyperparameters
  – Applicable to arbitrary feature representations
CORAL: Correlation Alignment

- First attempt: normalize first order statistics
  - Not sufficient to reduce data to zero mean and unit standard deviation
  - Covariance of data distributions can still differ

- CORAL: Match distributions by aligning second-order statistics
  - A frustratingly easy method that does better in some scenarios
  - Does not require re-training or hyperparameters
  - Applicable to arbitrary feature representations

- Labeled source examples $D_S = \{x_i\}$, unlabeled target $D_T = \{u_i\}$
- Goal: align covariance matrices of two data distributions

\[
C_S = \frac{1}{n_S - 1} (D_S^T D_S - \frac{1}{n_S} (1^T D_S)^T (1^T D_S))
\]
\[
C_T = \frac{1}{n_T - 1} (D_T^T D_T - \frac{1}{n_T} (1^T D_T)^T (1^T D_T))
\]
Linear Transformation for CORAL

- Labeled source examples $D_S = \{x_i\}$, unlabeled target $D_T = \{u_i\}$
- Goal: align covariance matrices of two data distributions
- Find a linear transformation that maps source covariance to target

$$\min_A \| C_S' - C_T \|_F^2 = \min_A \| A^T C_S A - C_T \|_F^2$$
Linear Transformation for CORAL

• Labeled source examples $D_S = \{x_i\}$, unlabeled target $D_T = \{u_i\}$
• Goal: align covariance matrices of two data distributions
• Find a linear transformation that maps source covariance to target

$$\min_A \| C_\hat{S} - C_T \|_F^2 = \min_A \| A^\top C_S A - C_T \|_F^2$$

• Closest rank $r$ matrix $X$ to a given matrix $Y$ with $r < \text{rank}(Y)$

$$\arg \min_X \| X - Y \|_F^2 = U_{Y[1:r]} \Sigma_{Y[1:r]} V_{Y[1:r]}^\top$$
Linear Transformation for CORAL

- Labeled source examples $D_S = \{x_i\}$, unlabeled target $D_T = \{u_i\}$
- Goal: align covariance matrices of two data distributions
- Find a linear transformation that maps source covariance to target

$$\min_A \|C_S - C_T\|_F^2 = \min_A \|A^\top C_S A - C_T\|_F^2$$

- Closest rank $r$ matrix $X$ to a given matrix $Y$ with $r < \text{rank}(Y)$

$$\arg \min_X \|X - Y\|_F^2 = U_{Y[1:r]} \Sigma_{Y[1:r]} V_{Y[1:r]}^\top$$

- Solution:
  - Let $C_S = U_S \Sigma_S U_S^\top$ and $C_T = U_T \Sigma_T U_T^\top$
  - Let $r$ be the minimum of ranks of $C_S$ and $C_T$
  - Optimal solution: $A^* = U_S \Sigma_S^\frac{1}{2} U_S^\top U_{T[1:r]} \Sigma_{T[1:r]}^\frac{1}{2} U_{T[1:r]}^\top$
Linear Transformation for CORAL

• Find a linear transformation that maps source covariance to target

\[ A^* = U_S \Sigma_S^{\frac{1}{2}} U_S^\top U_T[1:r] \Sigma_T[1:r]^{\frac{1}{2}} U_T[1:r]^\top \]

• Intuition for CORAL transformation:
  – First part “whitens” source data, second “recolors” with target covariance
Relationship to Data Whitening

- Data whitening
  - Given data $X$, with mean 0 and non-singular covariance $\Sigma$
  - Find transformation $Y = WX$ to obtain unit diagonal covariance
  - That is, solve: $W^T W = \Sigma^{-1}$
  - Many solutions, a common choice is $W = \Sigma^{-1/2}$

- Just whitening both source and target is not sufficient
  - Data might lie on different subspaces
Relationship to Data Whitening

• In practice, must account for singular covariance matrices
  – Data often lies on low-dimensional manifold
  – CORAL transform accounts for low-rank covariance

• In practice, just do traditional whitening and recoloring
  – Add a regularization to diagonal to make covariance matrix full-rank
  – More efficient and stable SVD computation

\[
\begin{align*}
\text{Input:} & \quad \text{Source Data } D_S, \text{ Target Data } D_T \\
\text{Output:} & \quad \text{Adjusted Source Data } D_S^* \\
C_S &= \text{cov}(D_S) + \text{eye(size}(D_S, 2)) \ast \lambda \\
C_T &= \text{cov}(D_T) + \text{eye(size}(D_T, 2)) \ast \lambda \\
D_S &= D_S \ast C_S^{-\frac{1}{2}} \quad \% \text{ whitening source} \\
D_S^* &= D_S \ast C_T^{\frac{1}{2}} \quad \% \text{ re-coloring with target covariance}
\end{align*}
\]
Relationship to Data Whitening

• In practice, must account for singular covariance matrices
  – Data often lies on low-dimensional manifold
  – CORAL transform accounts for low-rank covariance

• In practice, just do traditional whitening and recoloring
  – Add a regularization to diagonal to make covariance matrix full-rank
  – More efficient and stable SVD computation
  – Adaptation results quite stable with varying regularization term
Efficiency Considerations

- Match distributions by aligning second-order statistics (covariance)
- CORAL whitens source data and recolors with target covariance

**Input:** Source Data $D_S$, Target Data $D_T$

**Output:** Adjusted Source Data $D^*_S$

$$C_S = \text{cov}(D_S) + \text{eye}(\text{size}(D_S, 2))$$

$$C_T = \text{cov}(D_T) + \text{eye}(\text{size}(D_T, 2))$$

$$D_S = D_S \times C_S^{-1}$$  \hspace{1cm} \% whitening source

$$D^*_S = D_S \times C_T^{1/2}$$  \hspace{1cm} \% re-coloring with target covariance

- Train classifier on adjusted source data $D^*_S$ and apply directly to $D_T$
  - For linear classifier $w^T u$, can apply transformation directly to $w$
  - More efficient when number and dimensionality of target samples is high
Efficiency Considerations

• Match distributions by aligning second-order statistics (covariance)
• CORAL whitens source data and recolors with target covariance

\[
\begin{align*} 
\text{Input: } & \text{Source Data } D_S, \text{ Target Data } D_T \\
\text{Output: } & \text{Adjusted Source Data } D_S^* \\
C_S &= \text{cov}(D_S) + \text{eye(size}(D_S, 2)) \\
C_T &= \text{cov}(D_T) + \text{eye(size}(D_T, 2)) \\
D_S &= D_S \times C_S^{-1} \\
D_S^* &= D_S \times C_T^{1/2} \\
\end{align*}
\]

• Train classifier on adjusted source data \( D_S^* \) and apply directly to \( D_T \)
  – For linear classifier \( w^T u \), can apply transformation directly to \( w \)
  – More efficient when number and dimensionality of target samples is high

• Since CORAL only transforms data, can be used with any classifier
  – Efficient, advantageous when target domain changes rapidly (such as videos)
CORAL: Properties

• Relationship to MMD
  – CORAL equivalent to minimizing MMD with a particular polynomial kernel
  – Choice of kernel: $k(x, y) = (1 + x^Ty)^2$
  – MMD can match arbitrary moments, CORAL only covariances
  – CORAL a lot simpler to implement
CORAL: Properties

• Relationship to MMD
  – CORAL equivalent to minimizing MMD with a particular polynomial kernel
  – Choice of kernel: \(k(x, y) = (1 + x^\top y)^2\)
  – MMD can match arbitrary moments, CORAL only covariances
  – CORAL a lot simpler to implement

• CORAL is asymmetric (whiten source data, recolor with target data)
  – Symmetric transforms must find an “invariant” space to which both source and target can be mapped
  – Asymmetric transforms must find a “bridge” between the two domains
  – Can argue that asymmetric methods have an easier task
CORAL: Properties

• Relationship to MMD
  – CORAL equivalent to minimizing MMD with a particular polynomial kernel
  – Choice of kernel: \( k(x, y) = (1 + x^T y)^2 \)
  – MMD can match arbitrary moments, CORAL only covariances
  – CORAL a lot simpler to implement

• CORAL is asymmetric (whiten source data, recolor with target data)
  – Symmetric transforms must find an “invariant” space to which both source and target can be mapped
  – Asymmetric transforms must find a “bridge” between the two domains
  – Can argue that asymmetric methods have an easier task

• CORAL less effective for transforming target covariance to source
  – Whiten source, recolor with target: classifier imbibes information from both labeled source and unlabeled target data
  – Whiten target, recolor with source: classifier only sees source information
Deep CORAL

• Train feature by backpropagating the CORAL loss
  – Discriminative on source and adaptable to target

\[
\ell = \ell_{\text{CLASS.}} + \sum_{i=1}^{t} \lambda_i \ell_{\text{CORAL}}
\]
Deep CORAL

• Trade-off between classification and adaptation losses
• Only classification loss
  – Overfits to source domain
• Only CORAL loss
  – Degenerate low-rank features
  – Trivial solution: projecting both source and data to single
Deep CORAL

- Trade-off between classification and adaptation losses
- Only classification loss
  - Overfits to source domain
- Only CORAL loss
  - Degenerate low-rank features
  - Trivial solution: projecting both source and data to single

- More scalable
  - CORAL requires computation of SVD, which is not scalable
  - Deep CORAL objective minimized by batchwise SGD, which is more scalable