Central Moment Discrepancy
Domain Adaptation via Moment Alignment

W. Zellinger, T. Grubinger, E. Lughofer, T. Natschläger & S. Saminger-Platz

scch software competence center hagenberg

Johannes Kepler University Linz

Problem: Domain Adaptation

Given

- Input space $\mathcal{X}$ (e.g. images)
- Label space $\mathcal{Y}$ (e.g. object categories)
- Source domain D_S (e.g. images captured by webcam) as distribution over $\mathcal{X} \times \mathcal{Y}$
- Target domain D_T (e.g. images from website) as distribution over $\mathcal{X} \times \mathcal{Y}$

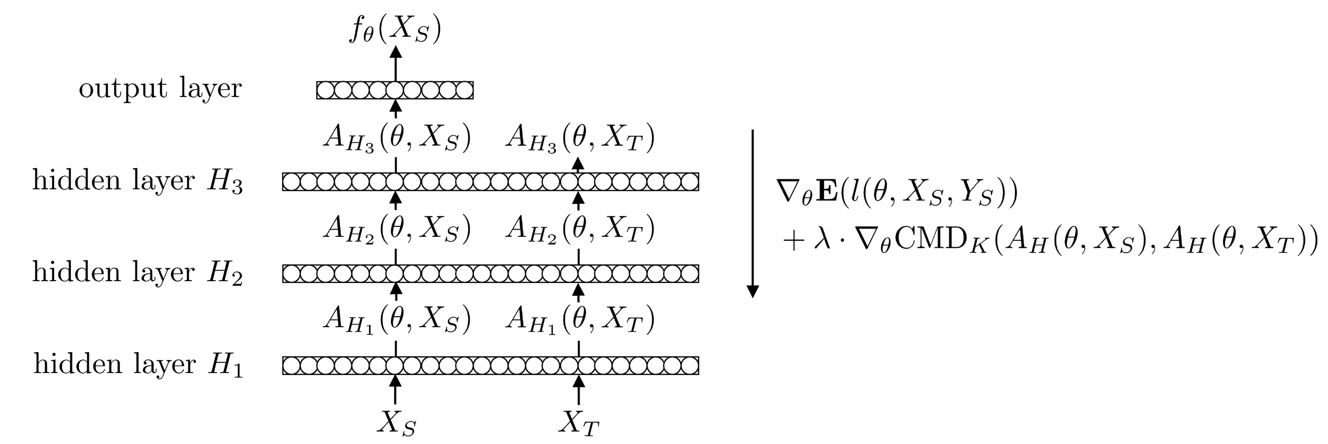
Goal

- Build a classifier $f : \mathcal{X} \to \mathcal{Y}$ with a low target risk $R_T(f) = \Pr_{(x,y)\sim D_T}(f(x) \neq y)$ while no information about the labels in D_T is given.

X. Glorot, A. Bordes, and Y. Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *International Conference on Machine Learning*, pp. 513–520, 2011.

Method: Distribution Matching

Minimization of discrepancy between domain-specific hidden activations $A_{H_3}(\theta, X_S)$ and $A_{H_3}(\theta, X_T)$



M. Long, Y. Cao, J. Wang, and M. Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, pp. 97–105, 2015.

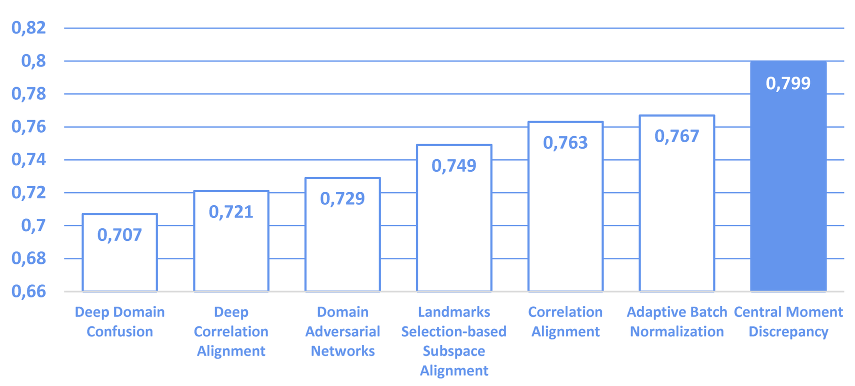
Benchmark Results

Setup

- All results are obtained with default parametrization $K = 5$ (moment order) and $\lambda = 1$ (objective weighting)

Object Recognition

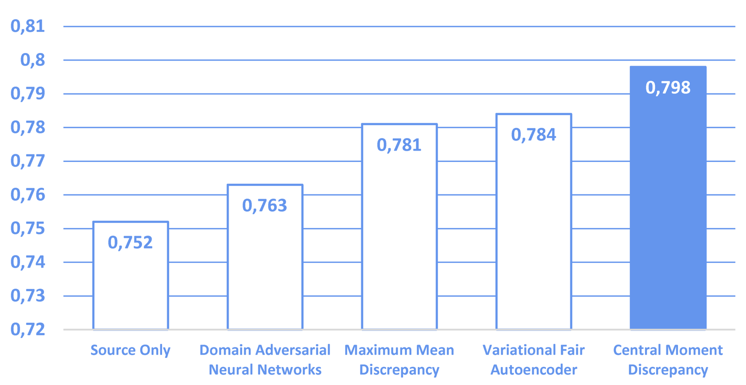
- Office dataset with three domains (amazon, webcam, dslr) and 31 classes (bike, monitor, etc.)
- Average performance improvement of 3.2% compared to previous state-of-the-art (Adaptive Batch Normalization)



K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *European Conference on Computer Vision*, pp. 213–226, 2010.

Sentiment Analysis of Product Reviews

- Amazon Reviews binary classification dataset with four domains (books, electronics, kitchen, dvds) and bag of words text representations



M. Chen, Z. E. Xu, K. Q. Weinberger, and F. Sha. Marginalized denoising autoencoders for domain adaptation. In *International Conference on Machine Learning*, pp. 767–774, 2012.

Approach: The Probability Metric

We define the Central Moment Discrepancy (CMD) between two probability distributions μ and ν, of two random vectors X and Y on $[0,1]^N$, by

$$\mathrm{CMD}_K(\mu,\nu) = \|\mathbb{E}(X) - \mathbb{E}(Y)\|_2 + \sum_{k=2}^{K} \|c_k(X) - c_k(Y)\|_2$$

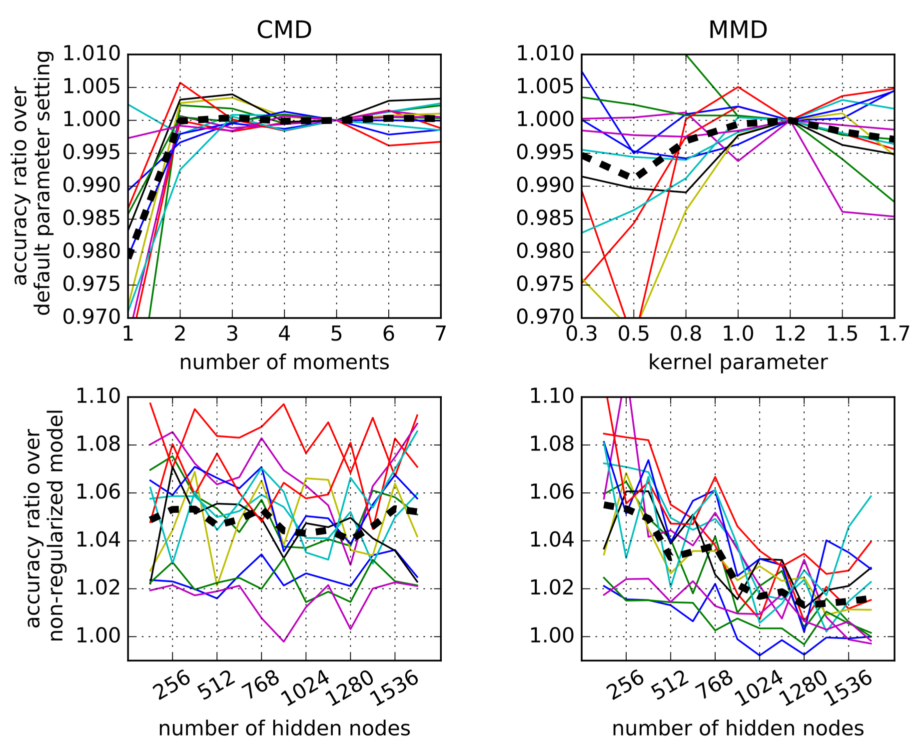
with the expectation $\mathbb{E}(X)$ and the central moment vector $c_k(X)$ of order k,

$$c_k(X) = \left(\mathbb{E}\Big(\prod_{i=1}^{N} (X_i - \mathbb{E}(X_i))^{r_i} \Big) \right)_{\substack{r_1+\ldots+r_N=k \\ r_1,\ldots,r_n\geq 0}}$$

W. Zellinger, T. Grubinger, E. Lughofer, T. Natschläger, and S. Saminger-Platz. Central moment discrepancy for domain-invariant representation learning. In *International Conference on Learning Representations*, 2017.

Sensitivity Analysis

Accuracy sensitivity w.r.t. parameter changes compared to Maximum Mean Discrepancy (MMD) with Gaussian kernel



Properties of CMD

Empirical Estimation

- Consistent estimator by means of empirical expectation
- Computation time $\mathcal{O}(n)$ when cross-moments are neglected, i.e.

$$c_k(X) := \frac{1}{|X|} \sum_{x\in X} \Big(x - \frac{1}{|X|}\sum_{x\in X} x\Big)^k$$

- Strictly decreasing upper bound for marginal moment terms

$$\|c_k(X) - c_k(Y)\|_2 \leq 2\sqrt{N} \left(\frac{1}{k+1} \left(\frac{k}{k+1}\right)^k + \frac{1}{2^{1+k}} \right)$$

Probability Metric

- Metric for random variables supported on $[0,1]^N$

$$\mathrm{CMD}_\infty(\mu,\nu) = 0 \;\Leftrightarrow\; X = Y$$

- Minimization implies convergence in distribution

$$\mathrm{CMD}_\infty(\mu_n,\mu) \to 0 \;\Rightarrow\; X_n \xrightarrow{d} X$$

Contact

Werner Zellinger
Department of Knowledge-Based Mathematical Systems, JKU Linz
Web http://flll.jku.at/staff/wernerz
Email werner.zellinger@jku.at

The research reported on this poster has been supported by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry of Science, Research and Economy, and the Province of Upper Austria in the frame of the COMET center SCCH. We would like to thank Bernhard Moser and Florian Sobieczky for fruitful discussions on metric spaces.