Active Nearest Neighbors in Changing Environments

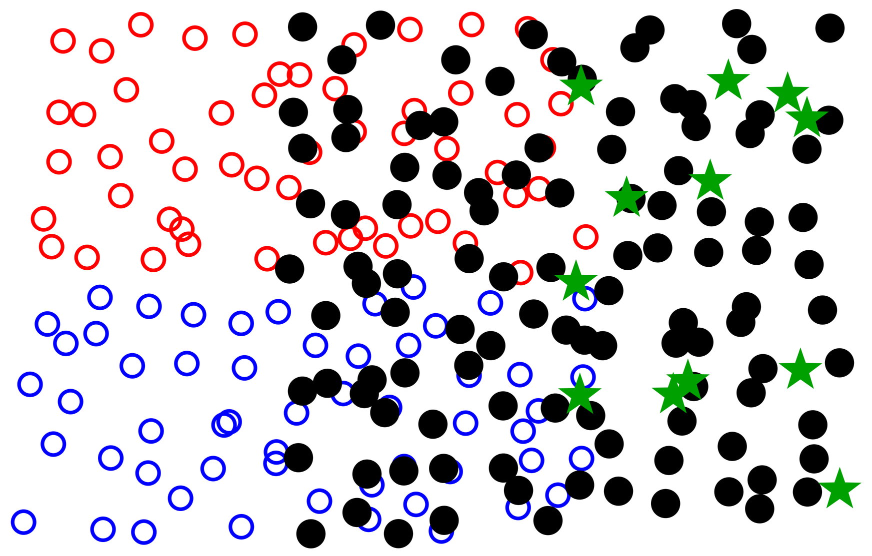
Chris Berlind
Georgia Tech

Ruth Urner
Max Planck Institute

Setting

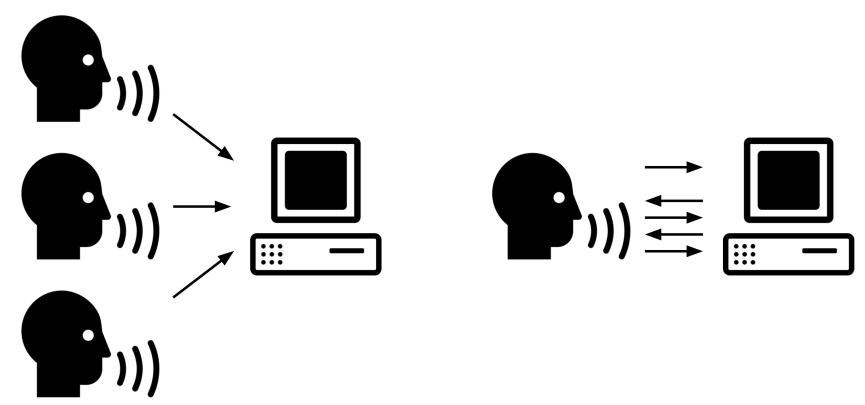
Active Domain Adaptation

- Labeled examples from source distribution
- Unlabeled examples from target distribution
- Active label query ability (target)
- Covariate shift (same labeling function)



Example: Speech recognition software

- Before releasing, train on in-house data set
- Once deployed, needs to learn individual user
- User feedback provides labels for user



Notation and Definitions

- $\eta(x) := \mathbb{P}(Y = 1|x)$ is λ-Lipschitz
- S, T sampled from distributions D_S, D_T
- $\mathcal{X}_S, \mathcal{X}_T \subseteq \mathcal{X}$ are the distribution supports
- $\mathrm{N}_\epsilon(\mathcal{X})$ denotes the ϵ-covering number of $\mathcal{X}$
- $\mathcal{L}_T(h^*)$ is the Bayes error rate of target
- $\beta(A) := D_S(A)/D_T(A)$ is the weight ratio
- $B_{n,A}(x)$ denotes the n-NN ball of x w.r.t. A
- $\mathcal{B}$ is the class of balls in $\mathcal{X}$

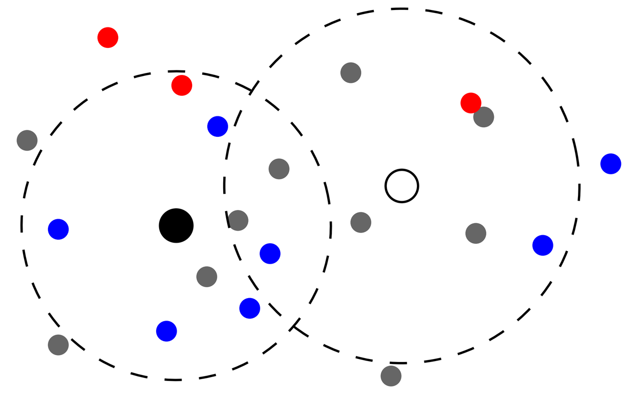
Our Approach

Active adaptive nearest neighbors

- Standard k-nearest-neighbor classification
- Adaptive nearest neighbor query strategy

Key Structure: (k, k')-NN-cover for T

- Definition: every example in T is either in the cover R or has k neighbors in R among the k' nearest neighbors in $T \cup R$
- Meaning: every target example is either labeled or has many labeled examples nearby



Algorithm

ANDA: Active NN for Domain Adaptation

- **Input:** labeled S, unlabeled T, params k, k'
- Find $Q \subseteq T$: $S \cup Q$ is (k, k')-NN-cover of T
- Query labels of the examples in Q
- **Output:** k-NN classifier on $S \cup Q$

Algorithm Variants

ANDA-Safe

- For each target example, query label if k'-NN ball has fewer than k labels
- Safe queries: *only* points not covered by source

ANDA-Safe-EMMA

- Efficient Multiset Multicover Approximation
- Finds approximate minimum (k, k')-NN-cover
- Potentially makes many fewer queries
- Retains query safety guarantee

Error Bound

Theorem 1. *For all ϵ, if η is λ-Lipschitz, the expected target error of* ANDA(S, T, k, k') *is*

$$\leq (1 + \sqrt{8/k})\mathcal{L}_T(h^*) + 9\lambda\epsilon + \frac{2\,\mathrm{N}_\epsilon(\mathcal{X}_T)\,k'}{|T|}.$$

Proof sketch:

- Consider target test point $x \sim D_T$
- k'-th nearest neighbor is not too far away
- (k, k')-NN-cover: k-th nearest label not far
- η cannot change much over short distance
- k nearest labels provide good approx. at x

Query Bound

Theorem 2. *If $|S| \geq \tilde{\Omega}(\frac{\mathrm{VC}(\mathcal{B})\ln(1/\delta)|T|}{C\,k\,w})$ and $|S| \geq \frac{9|T|}{C\,w}$ with $k \geq \Omega(\mathrm{VC}(\mathcal{B})\ln(|T|/\delta))$ and $|T| > k' = (C+1)k$, then, w.p. $\geq 1 - \delta$,*

ANDA-Safe- will not query any $x \in T$ with $\beta(B_{Ck,T}(x)) > w$.*

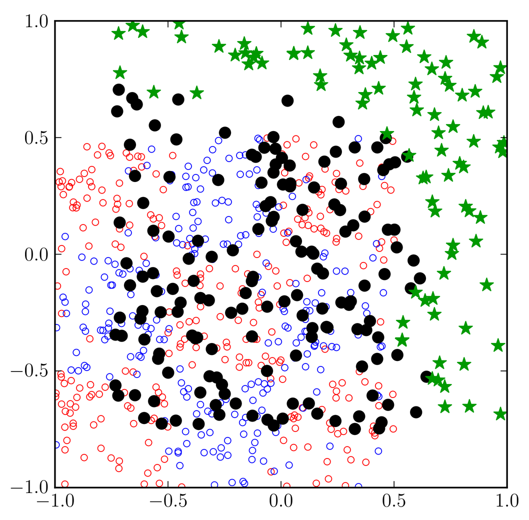
Proof sketch:

- Relative VC bounds: relate empirical weights to true probability weights of balls in $\mathcal{X}$
- Weight ratio: Source has significant weight in Ck-NN-ball $B_{Ck,T}(x)$ around target point x
- Source hits $B_{Ck,T}(x)$ at least k times

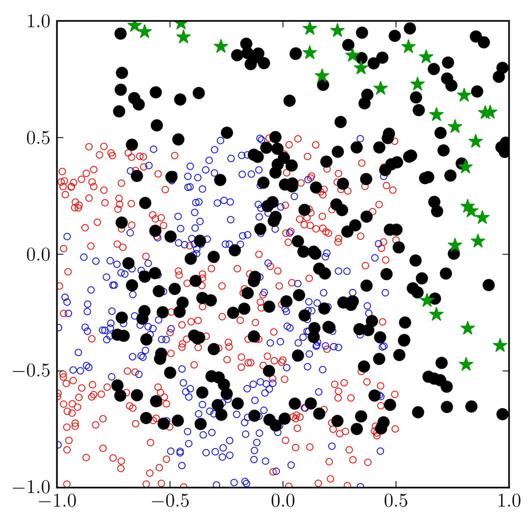
Discussion

- First formal demonstration of benefits from active learning for domain adaptation
- First algorithm with finite sample bounds when target is not fully supported by source
- Query complexity automatically adjusts to similarity between source and target
- Both error and query consistency

Synthetic Data

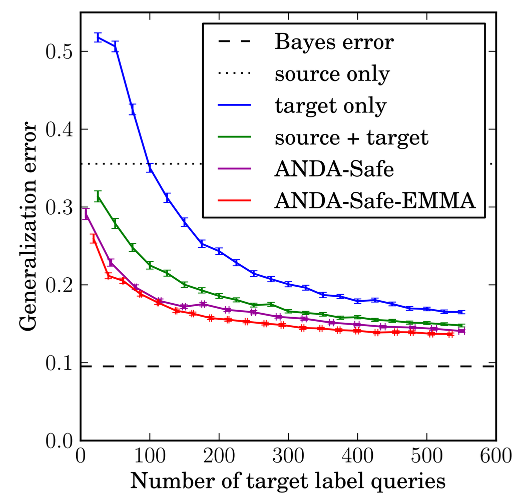


ANDA-Safe



ANDA-Safe-EMMA

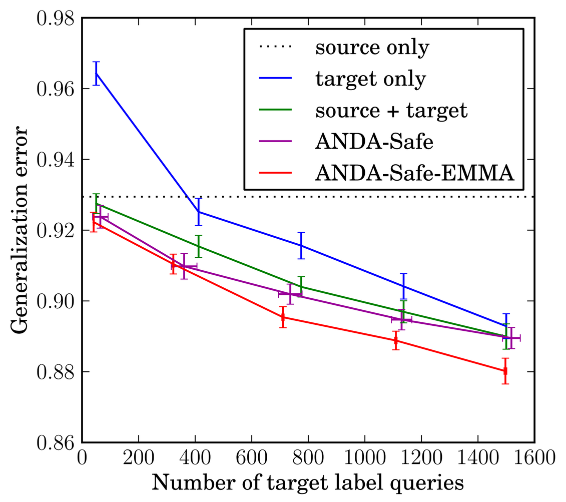
- Negative source example
- Positive source example
- Unlabeled target example
- Active label query



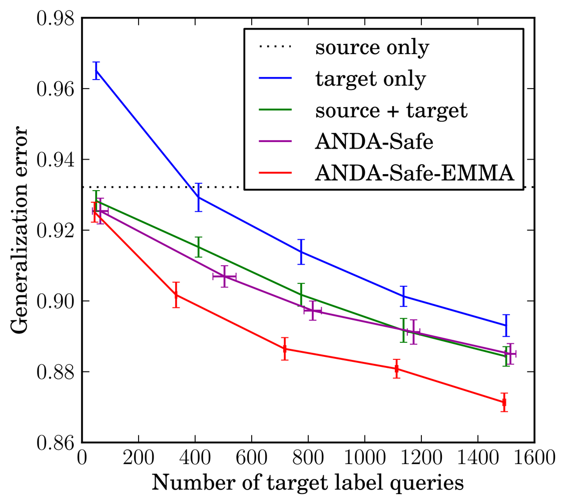
- $|S| = 3200$
- $|T|$ varies
- $k = 7$
- $k' = 21$
- 100 trials

Image Classification

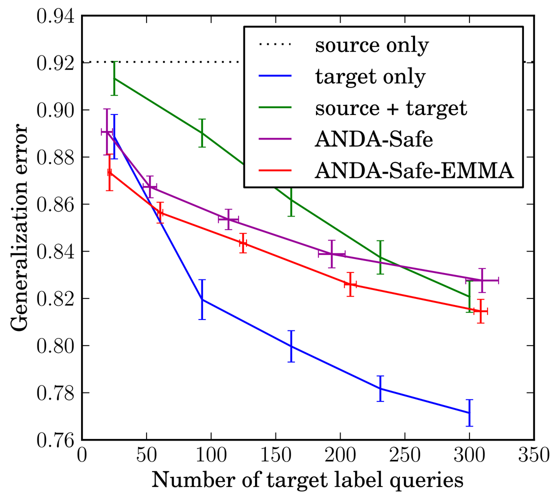
- Dataset bias: 4 datasets from Tommasi et al.
- 40 object classes, 1000 SIFT BOW features



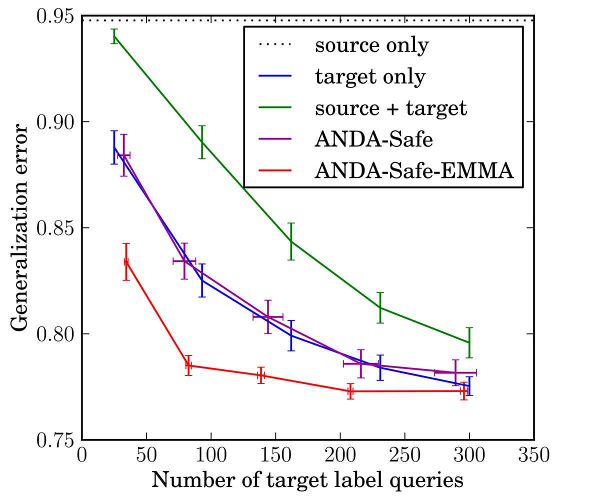
Caltech256 → Imagenet



Bing → Imagenet



Imagenet → Caltech256



Bing → Caltech256