DVMMlab
digital video | multimedia laboratory

Coping with Video Domain Change

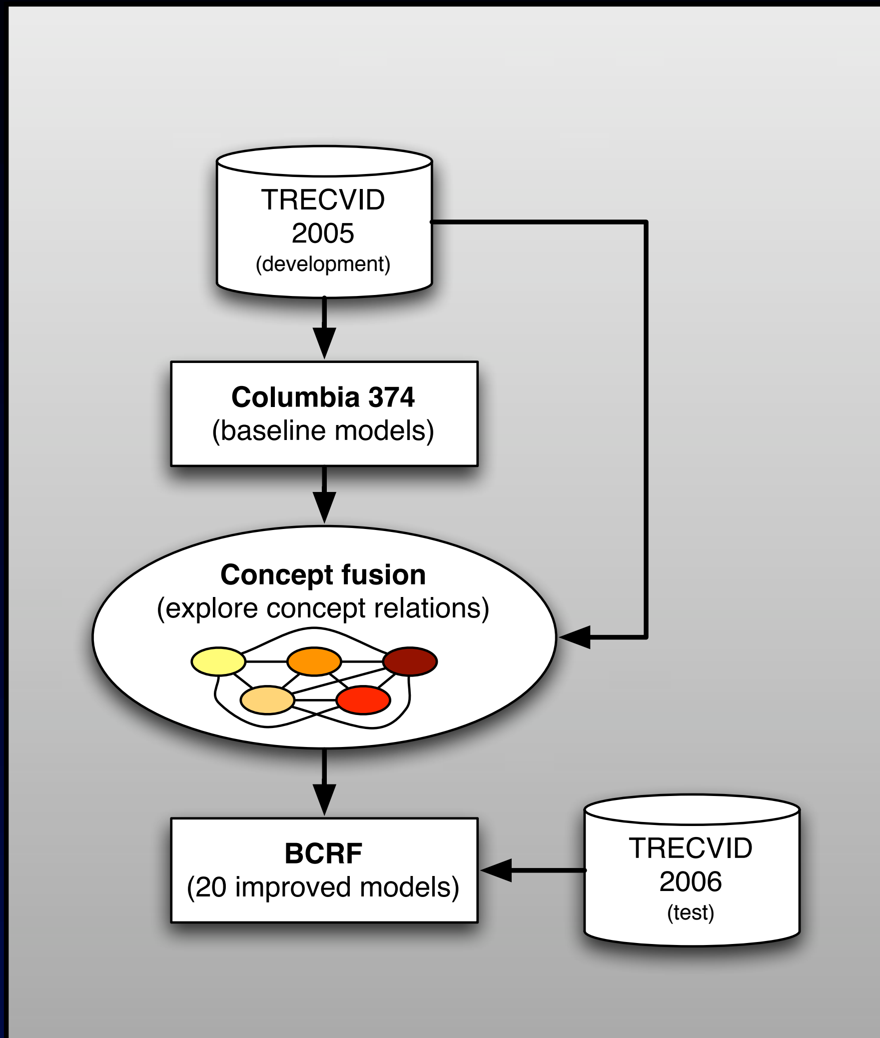
Analysis of Cross-Domain Learning Methods for High-Level Visual Concept Detection

Eric Zavesky, Wei Jiang, Akira Yanagawa, Shih-Fu Chang

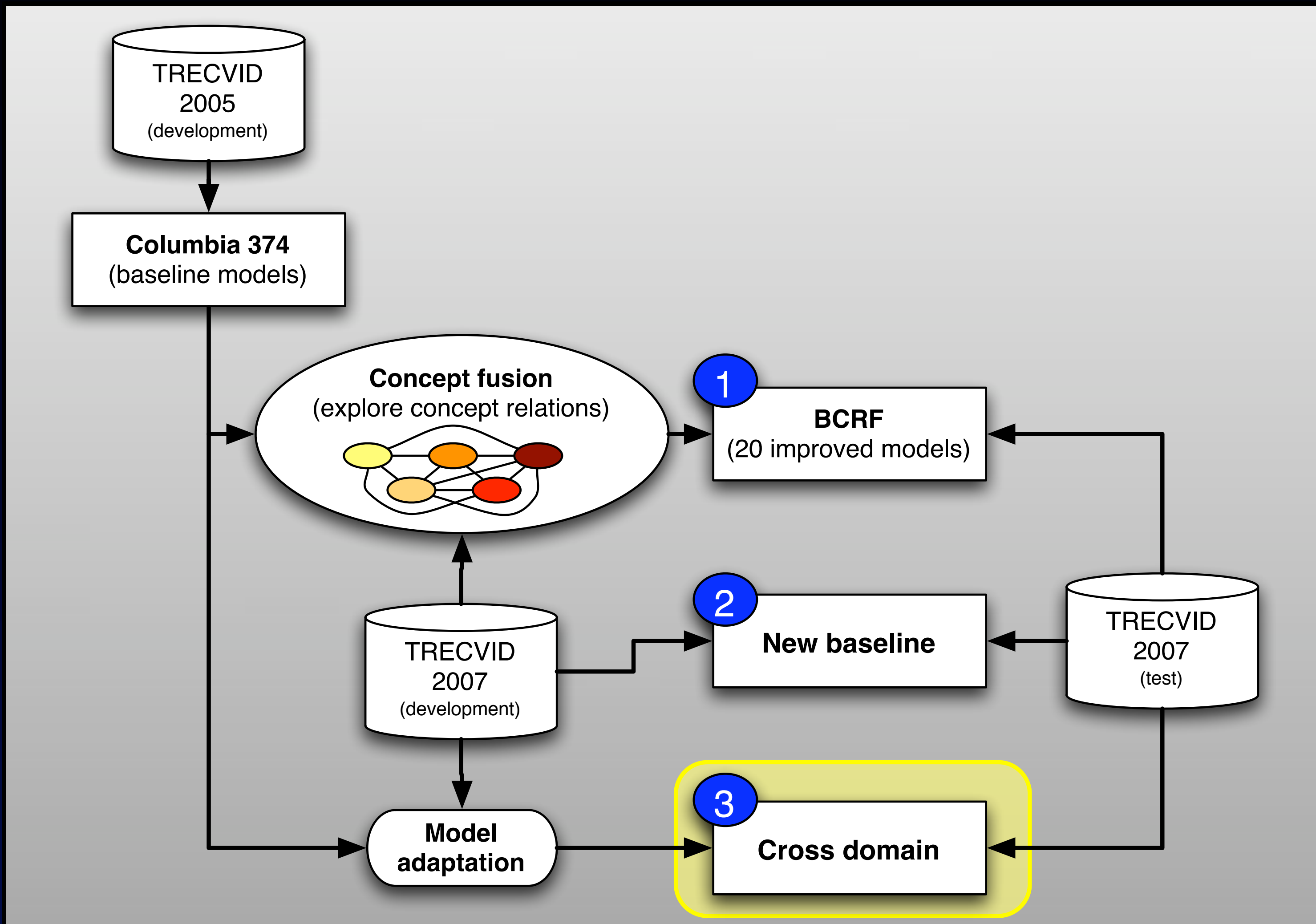
TRECVID HLF 2007

Columbia University
in the City of New York

Columbia HLF: TRECVID2006



Columbia HLF: TRECVID2007



cross-domain learning

Definition:

- <u>Domain</u>: set of content with same production/capture method and content quality

news (old domain)
TRECVID 2005

documentary (new domain)
TRECVID 2007

Problem:

- Not all data sets are created equal; classifiers trained on one domain often do not work well on others

Goal:

- Achieve robust detection in new domain with minimal additional complexity

Cross-Domain Problem: What is it?

Approach:

- Leverage pre-trained existing models
- Optimal weighted combination of data from both domains

Data:

- TRECVID2005 (broadcast news @ 100 hours),
- TRECVID2007 (documentaries @ 60 hours)

Cross-Domain Problem: Common approaches

method	training data		applicable condition
	old	new	
use old model	all	-	old domain very similar to new domain

increasing time & complexity

Case 1: old model works best

Studio

top 5 detection results

learn new domain, test new domain

learn old domain, test old domain

learn old domain, test new domain

Cross-Domain Problem: Common approaches

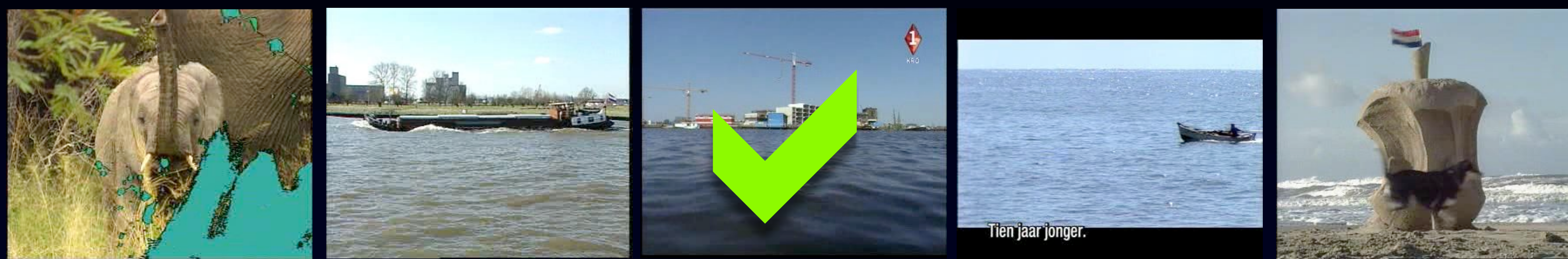
method	training data		applicable condition
	old	new	
use old model	all	-	old domain very similar to new domain
train new domain model	-	all	new and old domains very dissimilar

increasing time & complexity

Case 2: new model works best

Waterscape

top 5 detection results



learn new domain, test new domain

learn old domain, test old domain

learn old domain, test new domain

Cross-Domain Problem: Common approaches

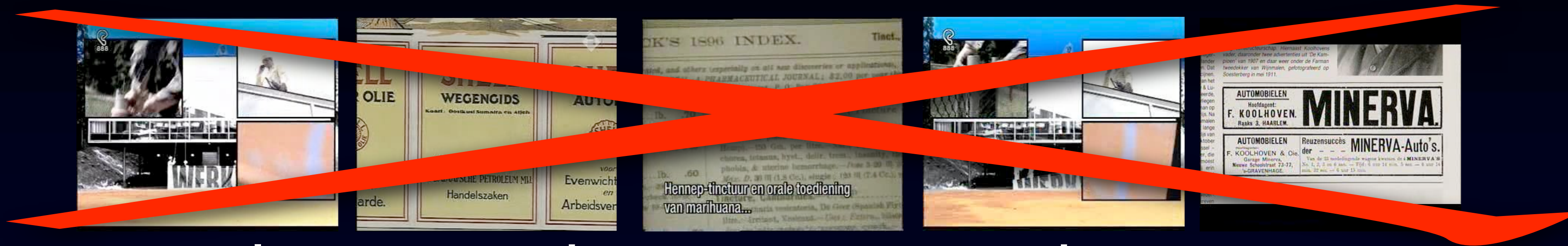
method	training data		applicable condition
	old	new	
use old model	all	-	old domain very similar to new domain
train new model	-	all	new and old domains very dissimilar
adapt old model	small	all	new and old domains slightly dissimilar

increasing time & complexity

Case 3: old model adaptation works best

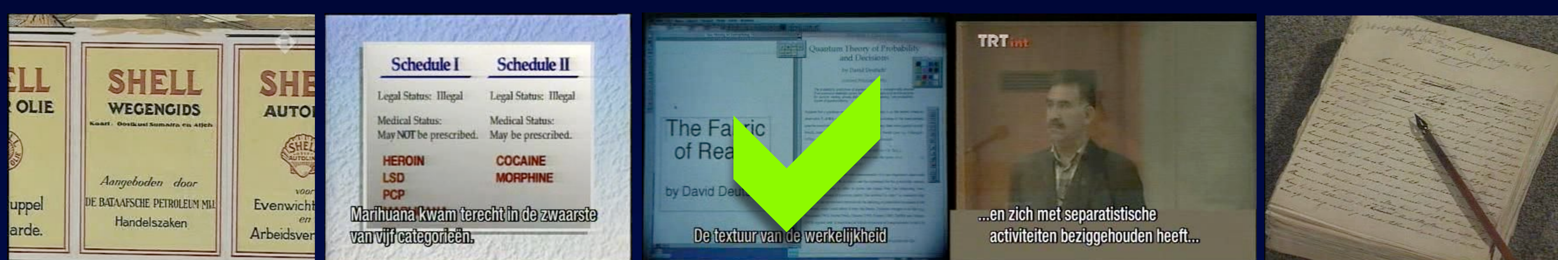
Charts

top 5 detection results



learn new domain, test new domain

learn old domain, test old domain



adapt old domain+new domain, test new domain

Cross-Domain Problem: Common approaches

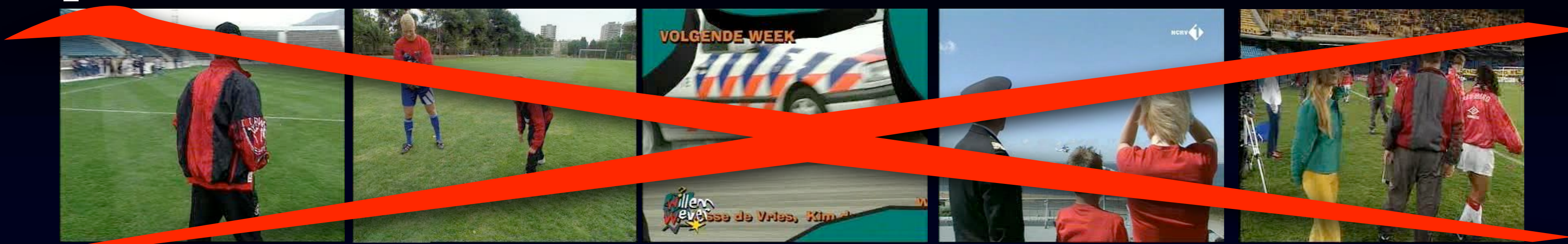
method	training data		applicable condition
	old	new	
use old model	all	-	old domain very similar to new domain
train new model	-	all	new and old domains very dissimilar
adapt old model	small	all	new and old domains slightly dissimilar
train combined new+old model	all	all	old and new domains similar; sparse new domain or strong old model

increasing time & complexity

Case 4: combined model works best

Sports

top 5 detection results



learn new domain, test new domain

learn old domain, test new domain

learn old domain+new domain, test new domain

Cross-Domain Problem: Common approaches

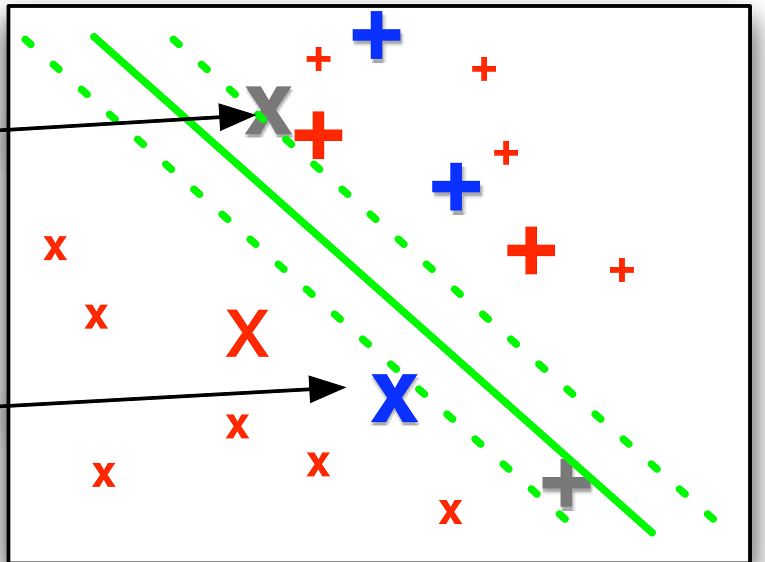
method	training data		applicable condition
	old	new	
use old model	all	-	old domain very similar to new domain
train new model	-	all	new domain and old domains very dissimilar
adapt old model	small	all	new and old domains slightly dissimilar
train combined new+old model	all	all	old and new domains similar; sparse new domain or strong old model

increasing time & complexity

Topic Review:
Support Vector Machine (SVM)

ignored support vector from old domain

helpful support vector from old domain



new samples

old support vectors

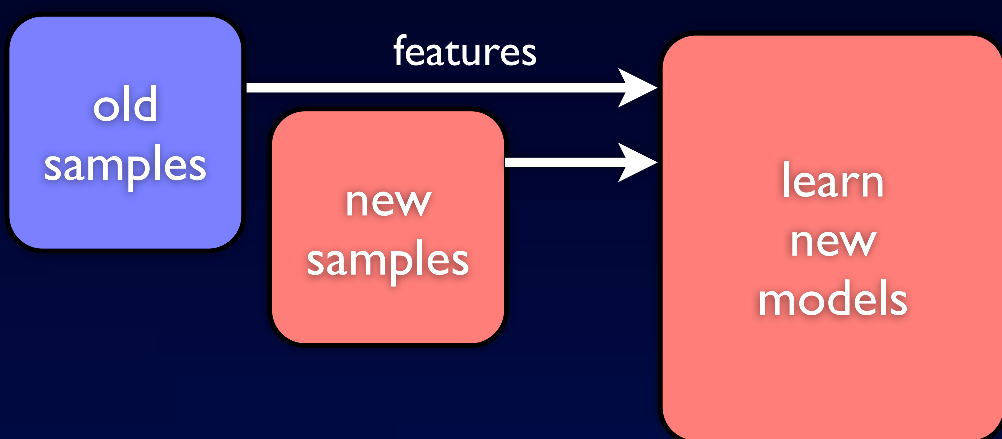
adapted hyperplane

old samples

new samples

Combined model: Uniform sample importance

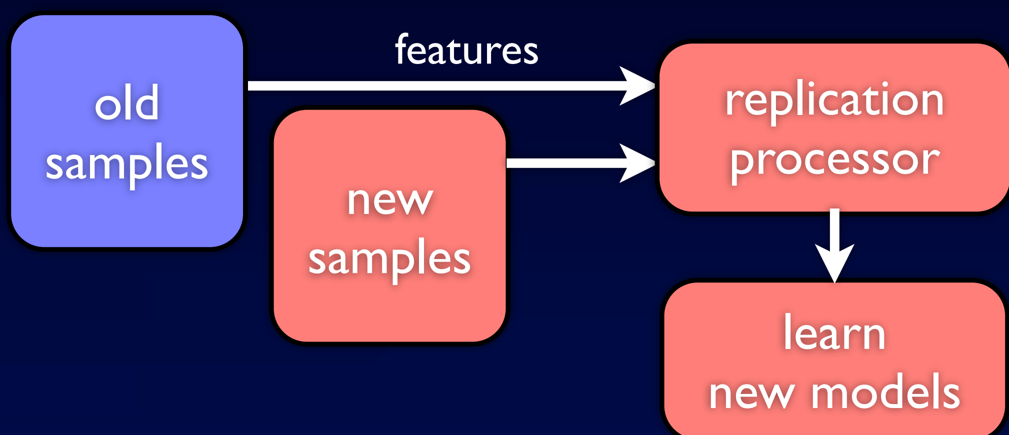
- Idea: includes all data (new and old) in training of new domain models
  - Kernel matrix: equal weights for all samples



	old	new
old	1x	
new		

Replication model:
Kernel matrix replication

- Idea: augment feature vector to learn intra-domain weights across many dimensions
  - Cross-domain training may be quite dissimilar
  - Trust intra-domain similarity more
  - Intelligent method for feature expansion



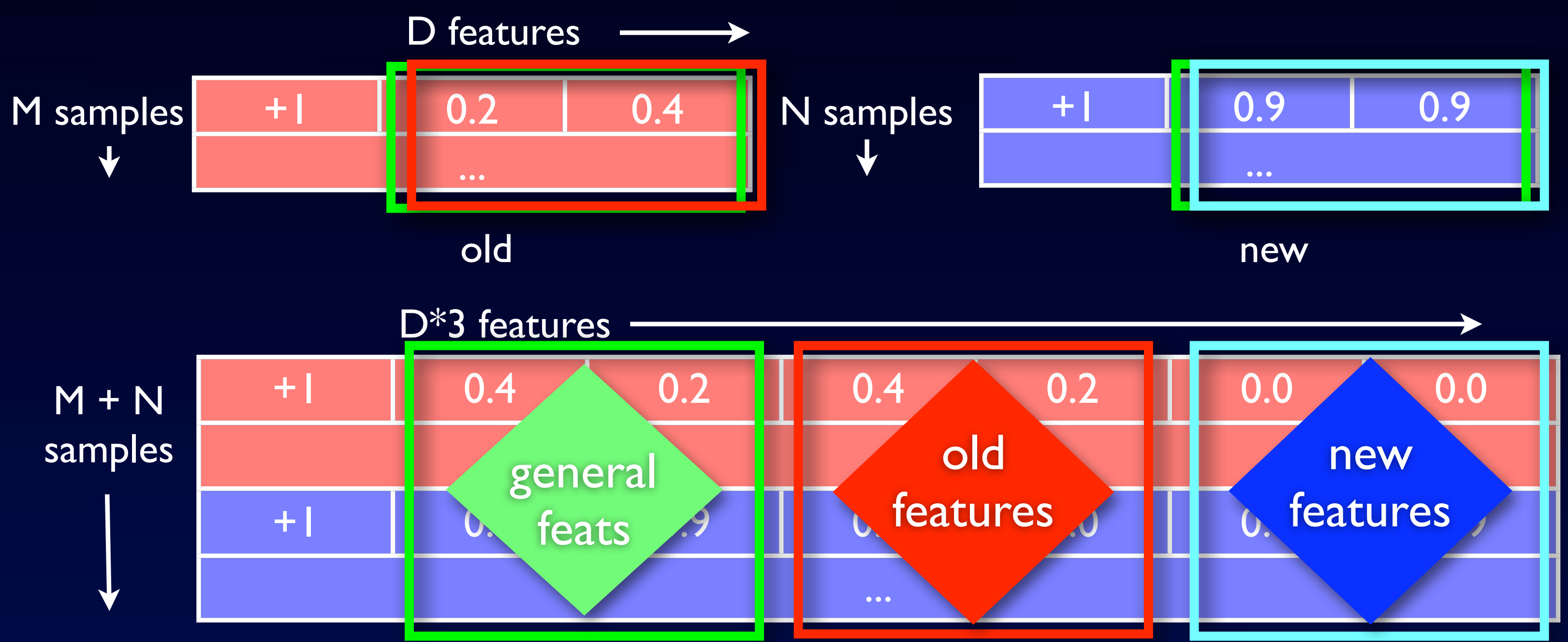
	old	new
old	2x	1x
new	1x	2x

1x

H. Daume III, "Frustratingly easy domain adaptation", Proc. the 45th Annual Meeting of the Association of Computational Linguistics, 2007

Replication model:
Kernel matrix replication

- Idea: augment feature vector to learn intra-domain weights across many dimensions

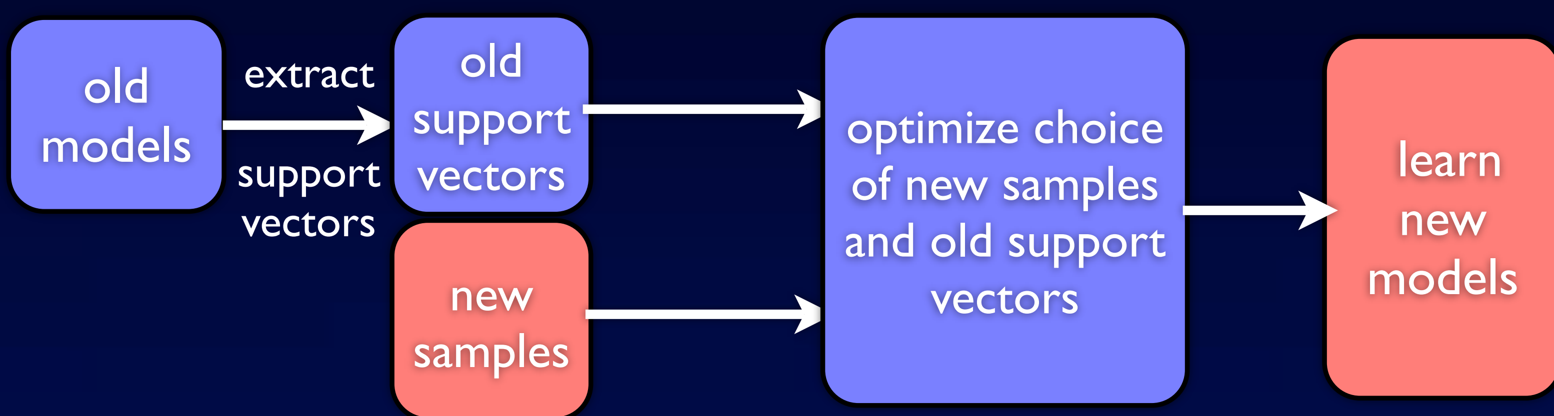


H. Daume III, "Frustratingly easy domain adaptation", Proc. the 45th Annual Meeting of the Association of Computational Linguistics, 2007

Adaptive SVM (A-SVM): Constrained model adaptation

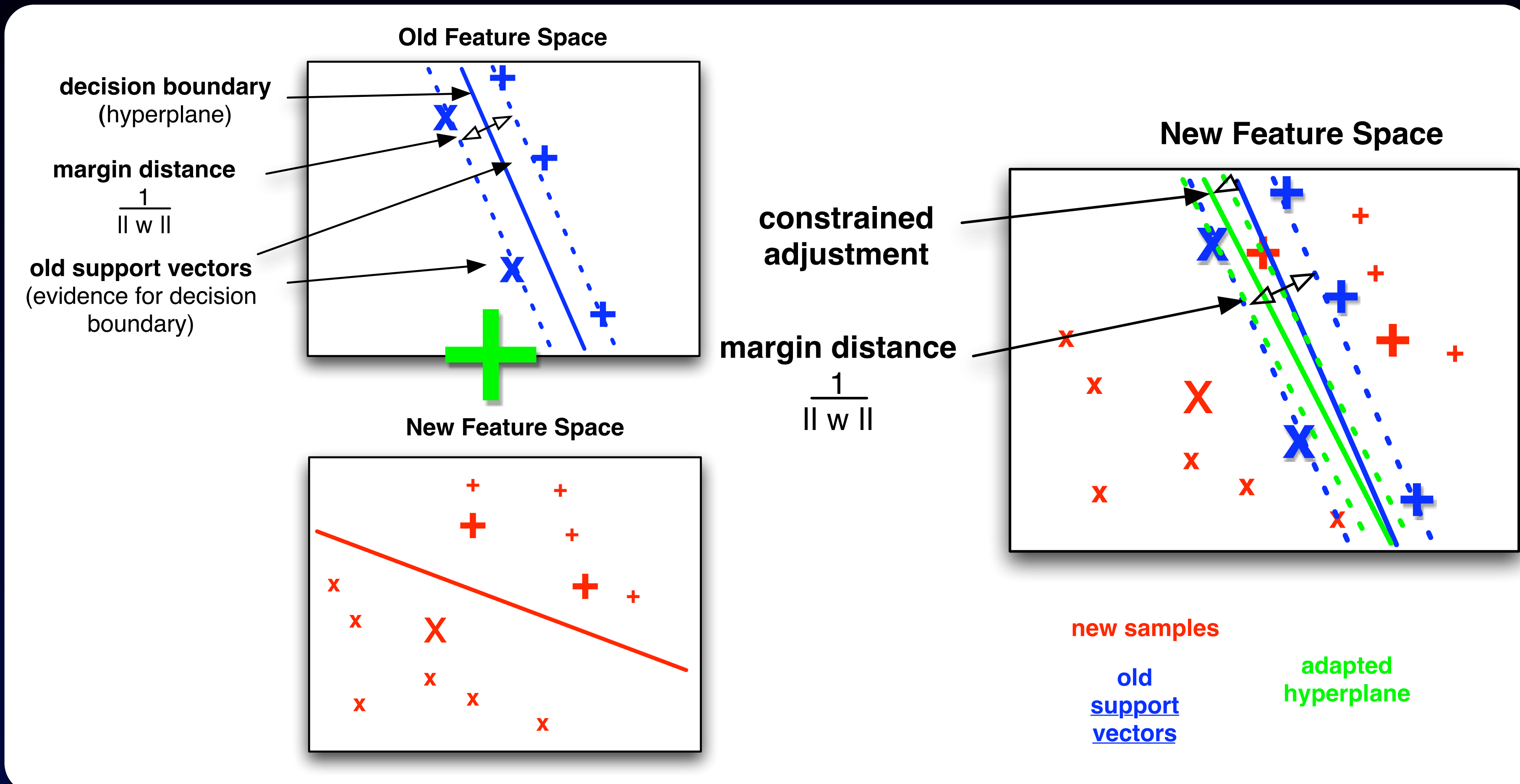
- Idea: trust old domain model more than new domain
  - Perturb old model within some tolerance with weighted new samples and a constant offset

$$f(\mathbf{x}) = f^{old}(\mathbf{x}) + \Delta f(\mathbf{x})$$



J. Yang, et al., "Cross-domain video concept detection using adaptive svms", ACM Multimedia, 2007.

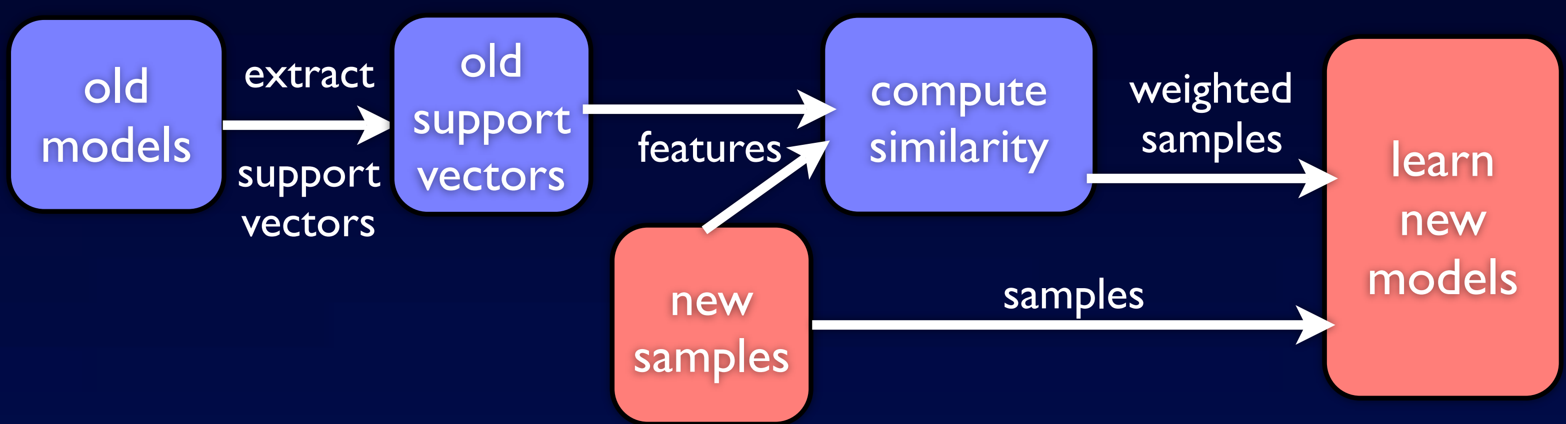
Adaptive SVM (A-SVM): Constrained model adaptation



J. Yang, et al., "Cross-domain video concept detection using adaptive svms", ACM Multimedia, 2007.

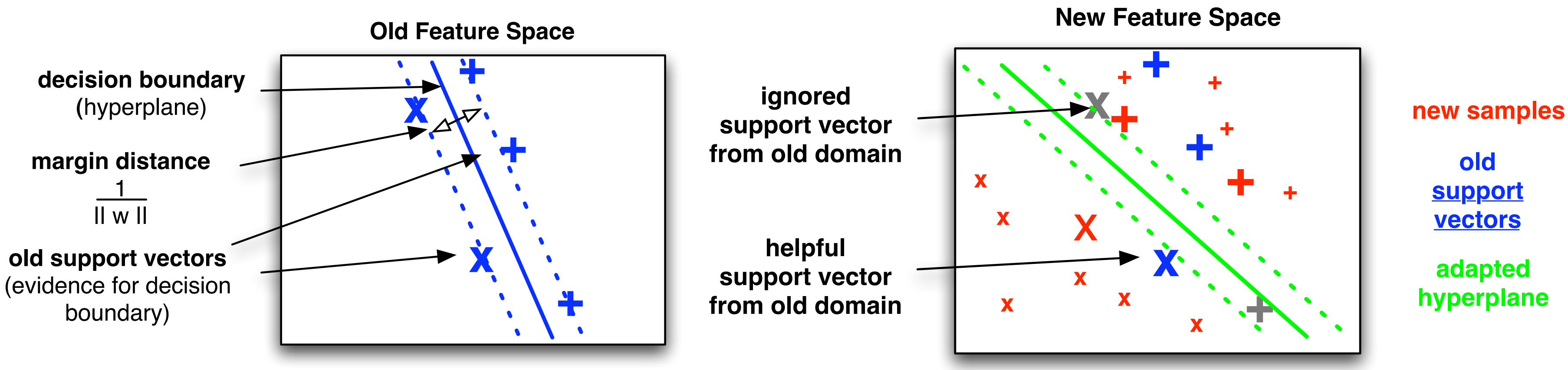
Cross-domain SVM (CD-SVM): Adapting prior models

- Idea: trust support vectors from trained old domain model as best observations in old domain
  - Weigh SVs then combine with new data and retrain



Submitted: W. Jiang, E. Zavesky, S.F. Chang, A. Loui, "Cross-domain learning methods for high-level concept classification," ICASSP 2008.

Cross-domain SVM (CD-SVM): Adapting prior models



$$\min_{w} \frac{1}{2}||\mathbf{w}||_2^2 + C\sum_{i=1}^{|\mathcal{D}^{new}|}\epsilon_i + C\sum_{j=1}^{M}\sigma(\mathbf{v}_j^{old}, \mathcal{D}^{new})\bar{\epsilon}_j$$

$$\sigma(\mathbf{v}_j^{old}, \mathcal{D}^{new}) = \frac{1}{|\mathcal{D}^{new}|}\sum_{(\mathbf{x}_i, y_i)\in\mathcal{D}^{new}} \exp\left\{-\beta||\mathbf{v}_j^{old} - \mathbf{x}_i||_2^2\right\}$$

Submitted: W. Jiang, E. Zavesky, S.F. Chang, A. Loui, "Cross-domain learning methods for high-level concept classification," ICASSP 2008.

Cross-domain methods: Observed speed trends

method	training data		example training cost
	old	new	
old model	all	-	0x
combined	all	all	3x
new model	-	all	1x
replication	all	all	9x
CDSVM	small	all	1.25x

increasing observed performance

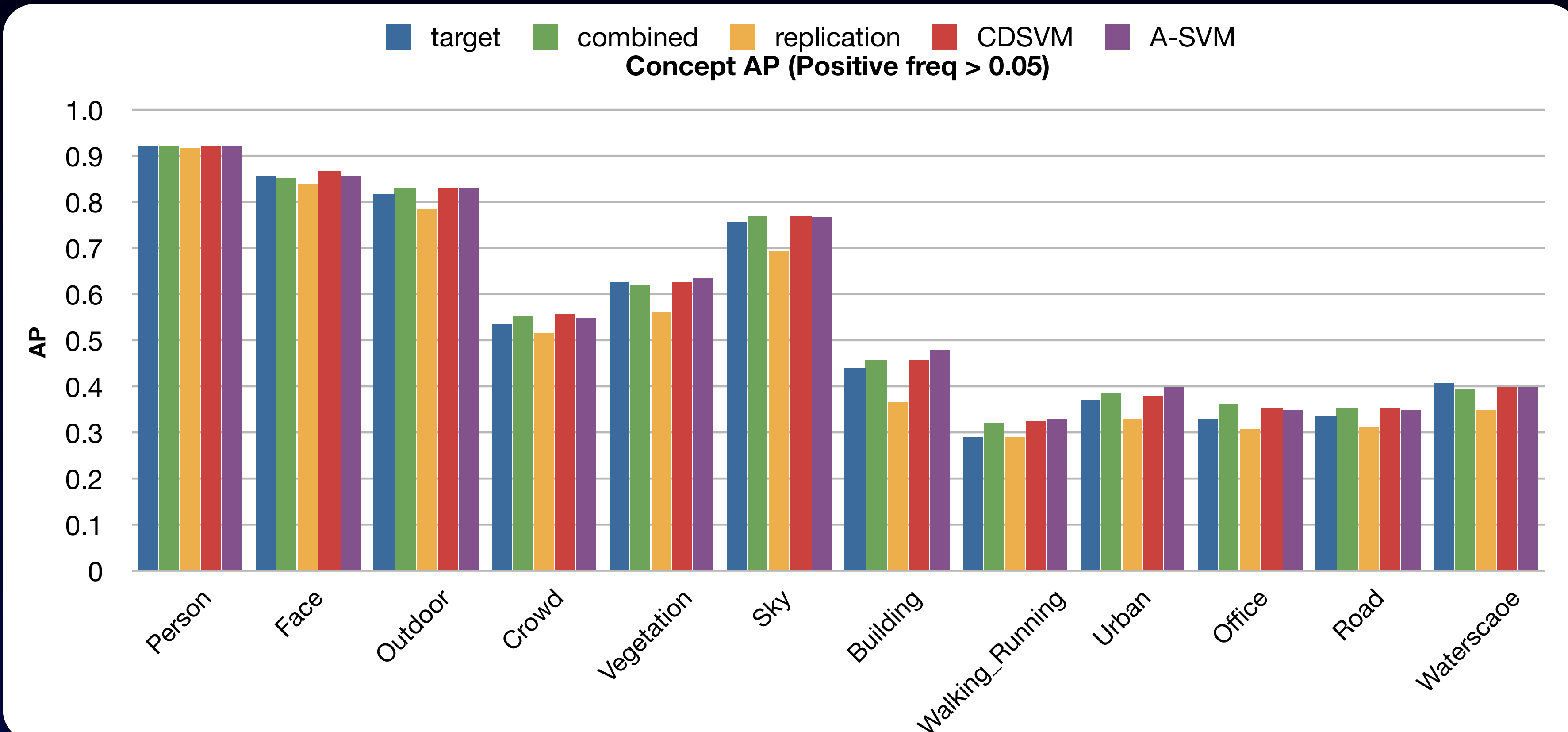
Theoretical training cost with 40k samples in old domain, 20k in new domain (similar to TRECVID problem)

Choosing an approach...

- No single approach is always optimal, but predictions can be found in a piece-wise manner
  - Based on available statistics
  - Positive new domain samples strongly relates to ideal training conditions for each approach...

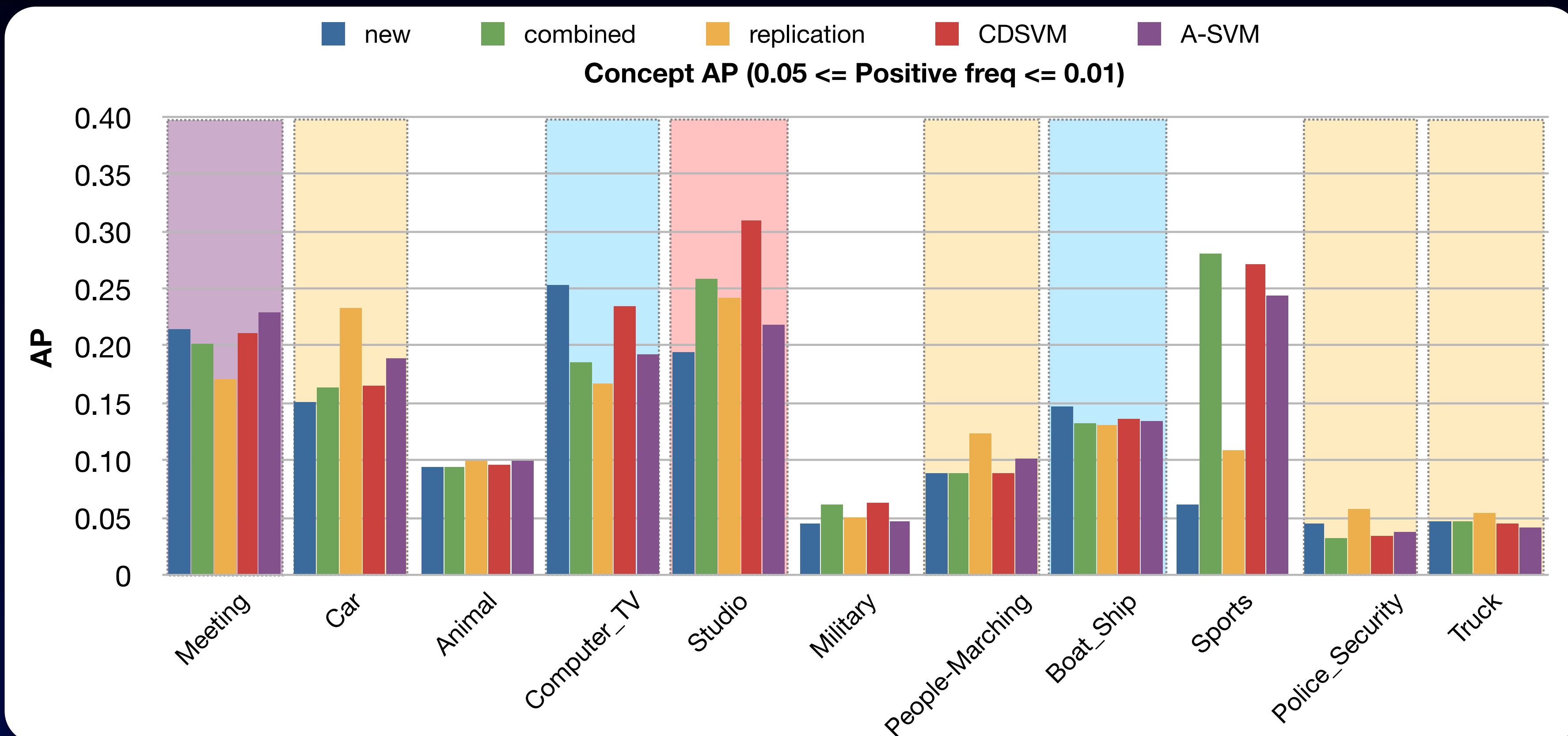
Performance comparison: High positive frequency

- No clear winners



Performance comparison:
Mid positive frequency

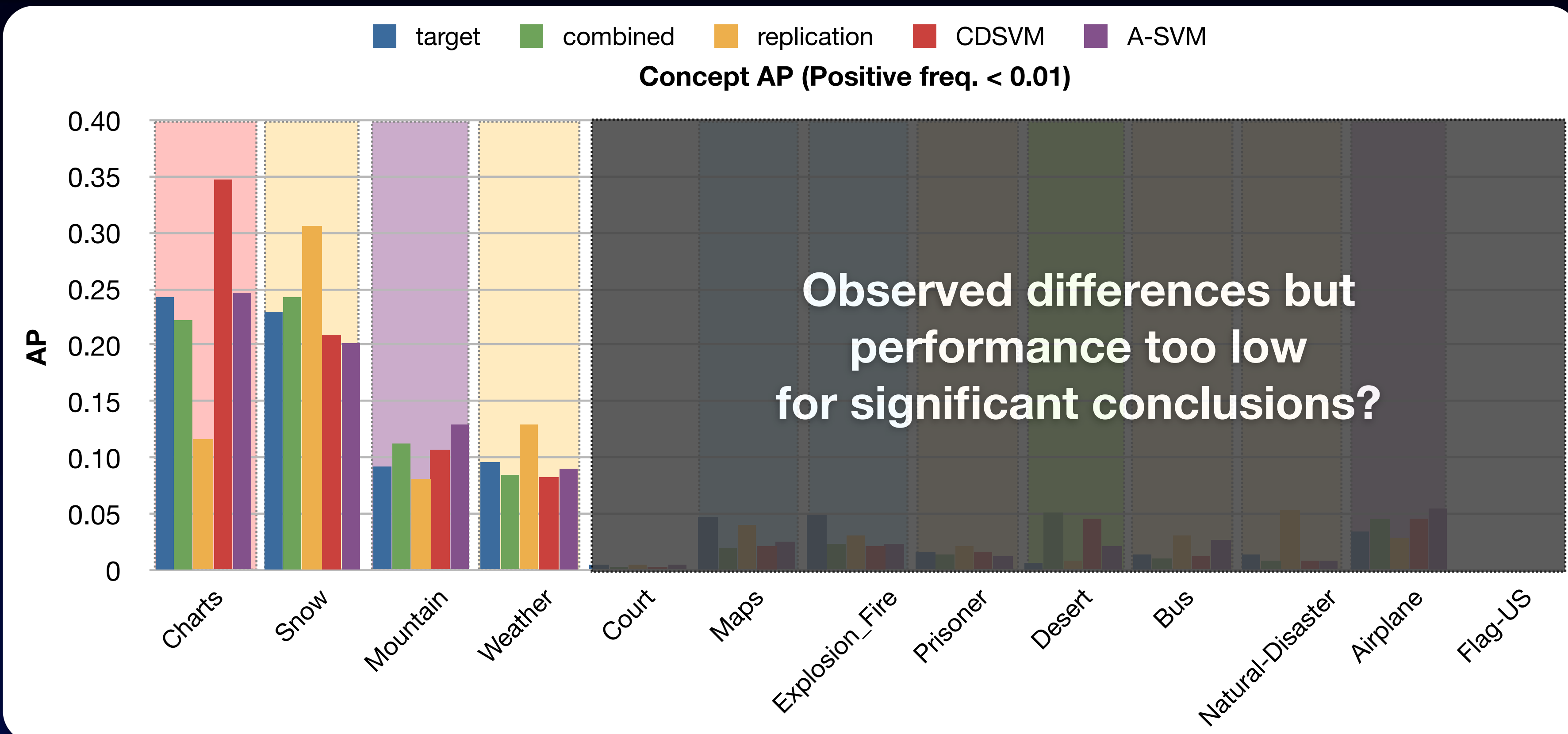
- Clear differentiation seen



shaded regions best per concept (5% relative improvement over all others)

Performance comparison:
Low positive frequency

- More differentiation, but less reliable in for low performance



shaded regions best per concept (5% relative improvement over all others)

Which approach to choose to obtain good new domain performance

- Decision based on frequency of positive samples and performance of old model...
- High frequency (old or new more than 5%)
  select CDSVM (adapts old to well-defined new domain)
  - person, sky, road, ...

Which approach to choose to obtain good new domain performance

- Decision based on frequency of positive samples and performance of old model...
- High frequency (old or new more than 5%)
- Mid-frequency (new < 5%, new > 1%)
  - If performance (AP) of old model was high, select replication (learn combined trends)
    - truck, car, people-marching
  - If AP was too low, select new domain only (not enough evidence)

Which approach to choose to obtain good new domain performance

- Decision based on frequency of positive samples and performance of old model...
- High frequency (old or new more than 5%)
- Mid-frequency (new < 5%, new > 1%)
- Low-frequency (new < 1%)
  - If sparse old (old < %1) select new (sparsity risk too high)
    - boats, computer-tv, map, explosion-fire

Which approach to choose to obtain good new domain performance

- Decision based on frequency of positive samples and performance of old model...
- High frequency (old or new more than 5%)
- Mid-frequency (new < 5%, new > 1%)
- Low-frequency (new < 1%)
- Otherwise, choose default model...

Approach selection: Empirical rule set

- Aggregating these intuitions, we can create a ruleset to choose an approach that optimizes new domain performance

if $(freq(\mathcal{D}^t_+) > T^t_1) \cup (freq(\mathcal{D}^s_+) > T^s)$ **then**
  Selected model = CDSVM
else if $AP(\mathcal{D}^s) > MAP(\mathcal{D}^s)$ **then**
  Selected model = Feature Replication
else if $(freq(\mathcal{D}^t_+) < T^t_2)$ & $(freq(\mathcal{D}^s_+) < T^s)$ **then**
  Selected model = SVM over Target Labeled Set $\mathcal{D}^t_l$
else
  Selected model = CDSVM
end if

← high frequency

← strong old model

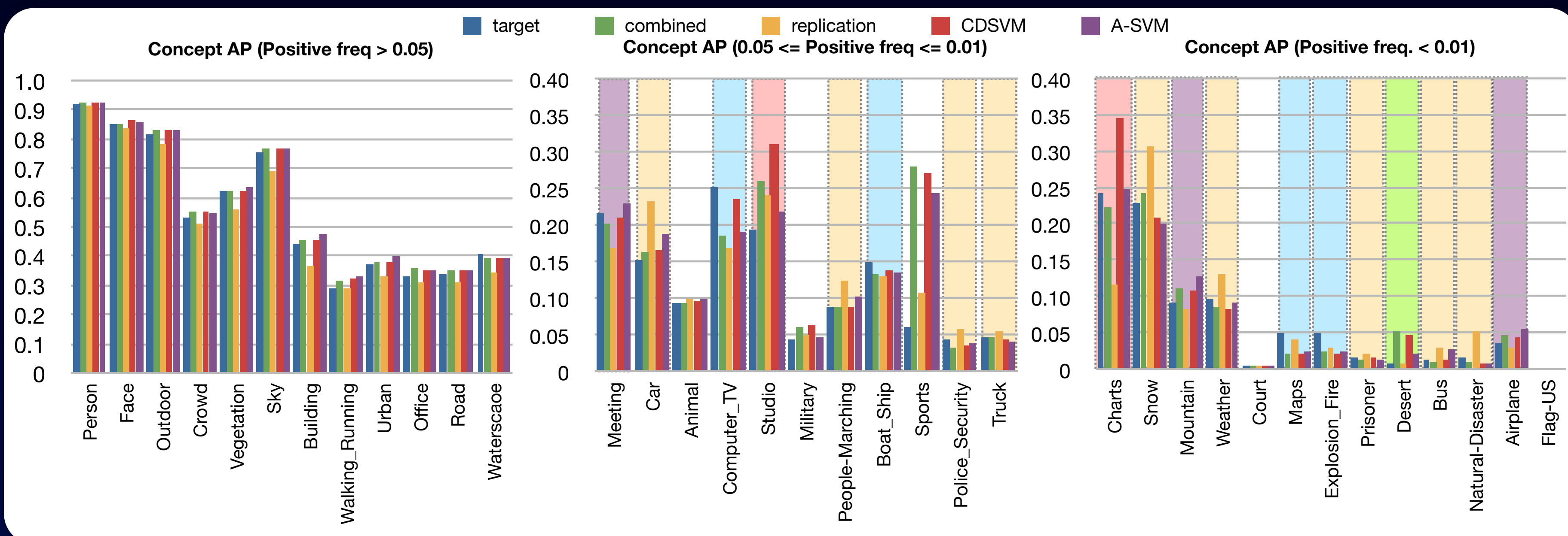
← new and old too sparse

← default choice

Approach selection: Rule-based benefits

high frequency	mid frequency	low frequency
8.7%	29.8%	24.6%

Observed MAP improvement over new model alone



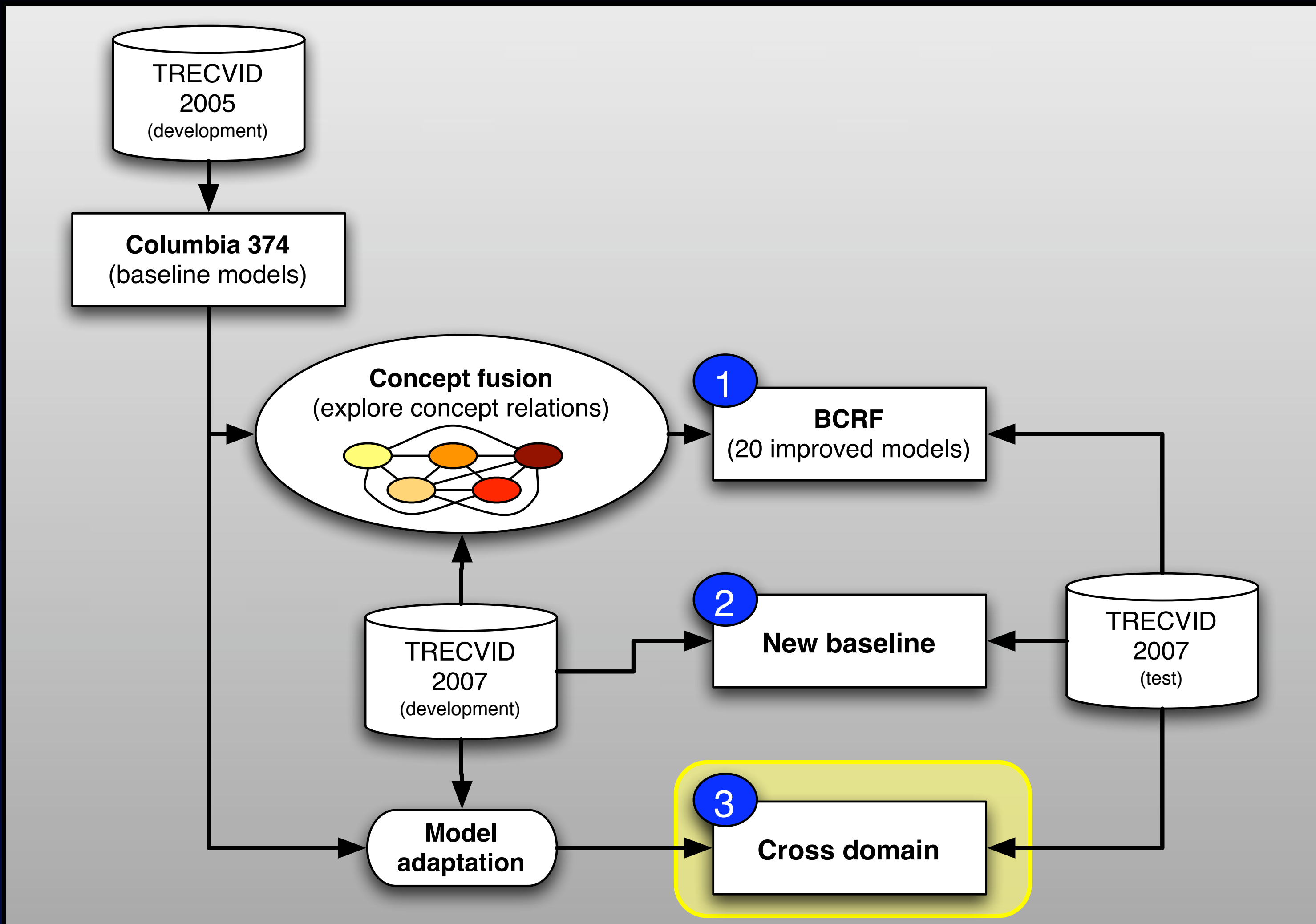
shaded regions best per concept (5% relative improvement over all others)

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

TRECVID 2007 high level features

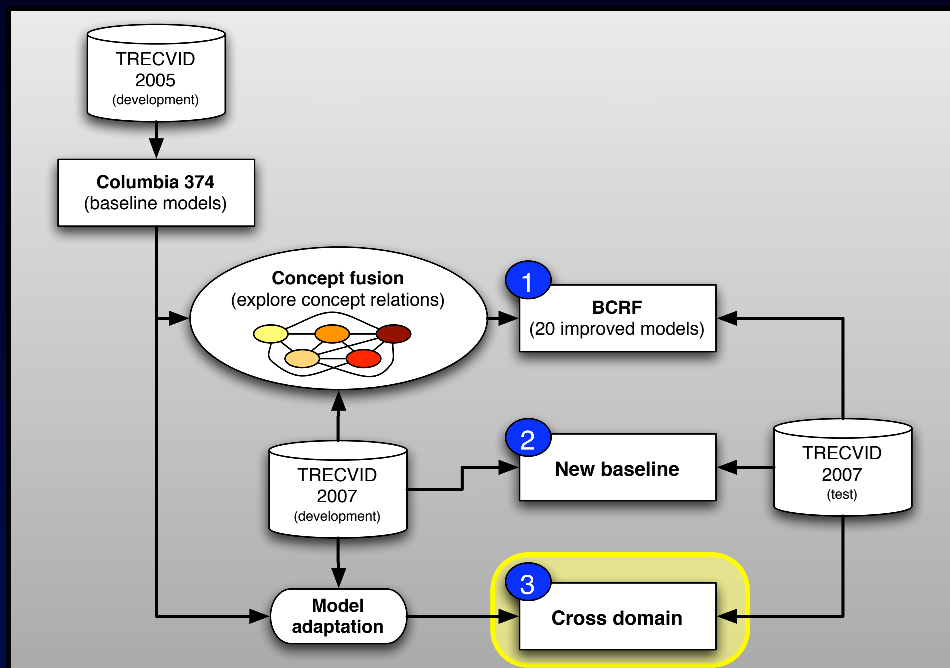
Columbia University in the City of New York

Columbia HLF: TRECVID2007

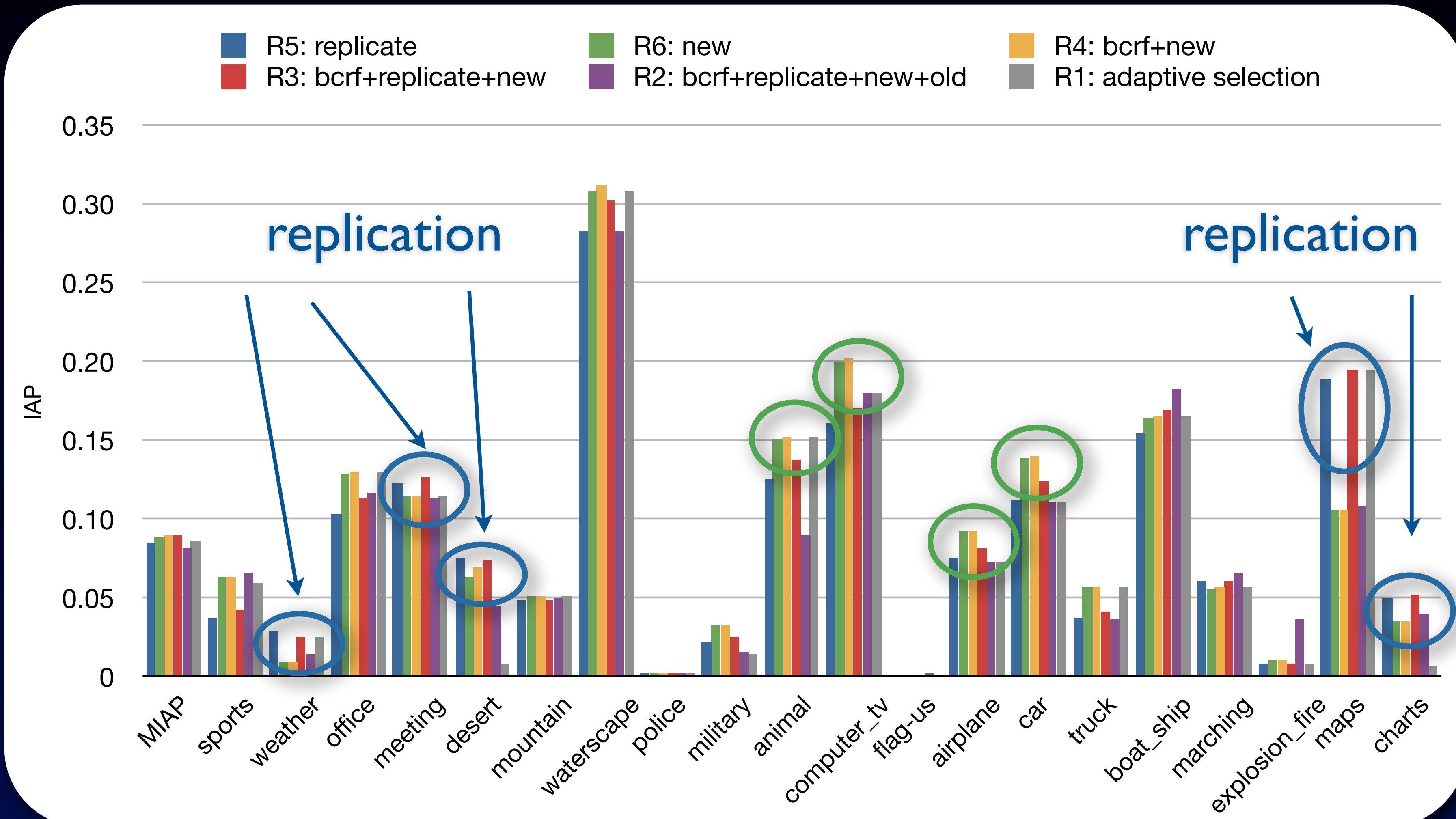


Empirical Results: TRECVID2007

- 4 of 6 runs in top 20
- Less than 0.005 MIAP difference between new models and replicated models
  - Only replication model was submitted
  - Cross-domain fusion improved performance for most concepts
  - Color moment, edge direction histogram Gabor texture



Empirical performance:
Method comparisons; new vs. replication



conclusions
&
next steps

Conclusions

- Cross-domain helps to cope with domain change
  - When new domain model is weak, good to use old domain data and models
  - Move models into new domain with minimal complexity increase and maintain performance
- Explore different different model approaches
  - No universally superior approach
    - Performance predictors: frequency of new and old domain and domain similarity
  - Prediction using domain properties works reasonably well

Next Steps:
Technical questions for adaptation

- When to adapt vs. training new model
  - Rules are first step, but deeper data distribution analysis is underway
- Next problem: few or no labels on new domain
- Leveraging large concept ontology (LSCOM)
  - Adaptation needed for concept-based approaches on new data

Thanks for your time.