



Four Horsemen — Assignment VI

Natural Computing

Domain Adaptation

Consider the two algorithms for domain adaptation Ad-REM and CORAL explained in class. Consider the transfer tasks of the Office-caltech dataset available at <https://twanvl.nl/research/domain-adaptation-2017/>.

1. Apply CORAL and Ad-REM

(a) with VGG features and

Answer The recommended Github repository was used (<https://github.com/twanvl/adrem>). For preprocessing all data was normalized by the joint standard deviation of source and target data. The training hyperparameter C has been fixed to 1.0 in order to minimize time for parameter search.

```
fprintf('Using surf features')
result = run_methods({'office-caltech', 'surf'})
```

(b) with SURF features.

Answer With joint-std preprocessing, and C = 1.0

```
fprintf('Using vgg features')
result = run_methods({'office-caltech', 'vgg'})
```

2. Report and discuss the results.

Answer: The tables list individual results for different combinations of source and target datasets as well as the average over all combinations for both ad-REM and Coral for VGG features in subtable 1a and SURF features in subtable 1b. It can be seen that ad-REM performs better than Coral on average for both types of features. Yet the advantage is much more pronounced for the VGG features than for SURF features. On the level of individual comparisons it is noticeable that for the VGG features ad-REM performs better than Coral on each combination of source and target data. Only for *webcam-dslr* both algorithms perform on par with perfect accuracy. For the SURF features the results are much more mixed, with comparable results for both algorithms.

Source	Target	Ad-REM SVM	Coral	Source	Target	Ad-REM SVM	Coral
amazon	Caltech	0.897	0.803	amazon	Caltech10	0.410	0.451
amazon	dslr	0.987	0.873	amazon	dslr	0.369	0.389
amazon	webcam	0.966	0.756	amazon	webcam	0.410	0.420
Caltech	amazon	0.951	0.934	Caltech10	amazon	0.518	0.544
Caltech	dslr	0.911	0.834	Caltech10	dslr	0.439	0.433
Caltech	webcam	0.888	0.763	Caltech10	webcam	0.400	0.298
dslr	amazon	0.953	0.832	dslr	amazon	0.414	0.346
dslr	Caltech	0.869	0.694	dslr	Caltech10	0.325	0.345
dslr	webcam	0.997	0.976	dslr	webcam	0.858	0.814
webcam	amazon	0.927	0.784	webcam	amazon	0.420	0.395
webcam	Caltech	0.874	0.623	webcam	Caltech10	0.343	0.385
webcam	dslr	1.000	1.000	webcam	dslr	0.828	0.841
Average		0.935	0.823	Average		0.478	0.472
(a) VGG features.				(b) SURF features.			

Table 1: Accuracies on a SVM classifier after Ad-REM and CORAL

3. Analyze how the use of labelled target data affects the target accuracy. Modify CORAL and Ad-REM to the setting where a subset of the target data is labelled.

- Select randomly $x = 10\%$ labelled target data for each class.
- Apply your modified CORAL and Ad-REM to the adaptation tasks with partly labelled target data.
- Repeat the experiment 5 times and report average accuracy results and standard deviations.
- Repeat the above three steps for $x = 10, 20, 40, 60\%$
- Report the results in a table and/or in a plot.
- Discuss the results.

You can use the implementations of CORAL and Ad-REM available at <https://github.com/twanvl/adrem>.

The Office data features mentioned can be found at: https://github.com/viggin/domain-adaptation-toolbox/tree/master/ToRelease_GFK/data

Answer

In figure 1 and figure 2 the results have been visualized.

For this experiment partly labeled target data was used, in addition to the experimental setup as described in exercise 1. A percentage of 10%, 20%, 40% or 60% labeled target data was used (data chosen randomly). Furthermore has each experiment with each setting and source-target combination been repeated five times. For sake of readability has the average accuracy been taken over all twelve source-target combinations and five iterations. In figure 1 the errorbars represent the average standard deviation of the source-target combinations. In figure 2 the errorbars represent the average

standard deviation between repetitions.

As can be seen in the figures, the use of labeled target data has a positive effect on the average accuracy. Using a higher percentage of labeled target data gives a consistently higher accuracy, using 60% labeled data compared to 10% labeled data results in an increase in accuracy of approximately 10% for both models with SURF features, and 5% when using VGG features. Aside from that observation, can be seen that both models still perform better with the VGG features compared to the SURF features, and the ad-REM method likewise still has a consistently higher accuracy than the Coral method with both types of features. Also can be seen in Figure 2 that the standard deviation between iterations is very small (less than 2.5%).

More interesting is the accuracy of the ad-REM method with VGG features, which seems to improve only slightly with the use of labeled target data, in comparison to the Coral method which has a more significant improvement in accuracy. Both methods seem to converge to an average accuracy over the source-target tasks of approximately 95%, reaching perfect scores for some tasks, and scores of 90% on others. The cause of this might be that a subset of the data contain label errors, meaning a perfect score is not possible for all source-target tasks. An alternative explanation is that the correlation between source and target data is too complex or far-fetched for the model to capture, resulting in an erroneous prediction. Finally, a third option could be that the tasks differ too much for the models to capture.

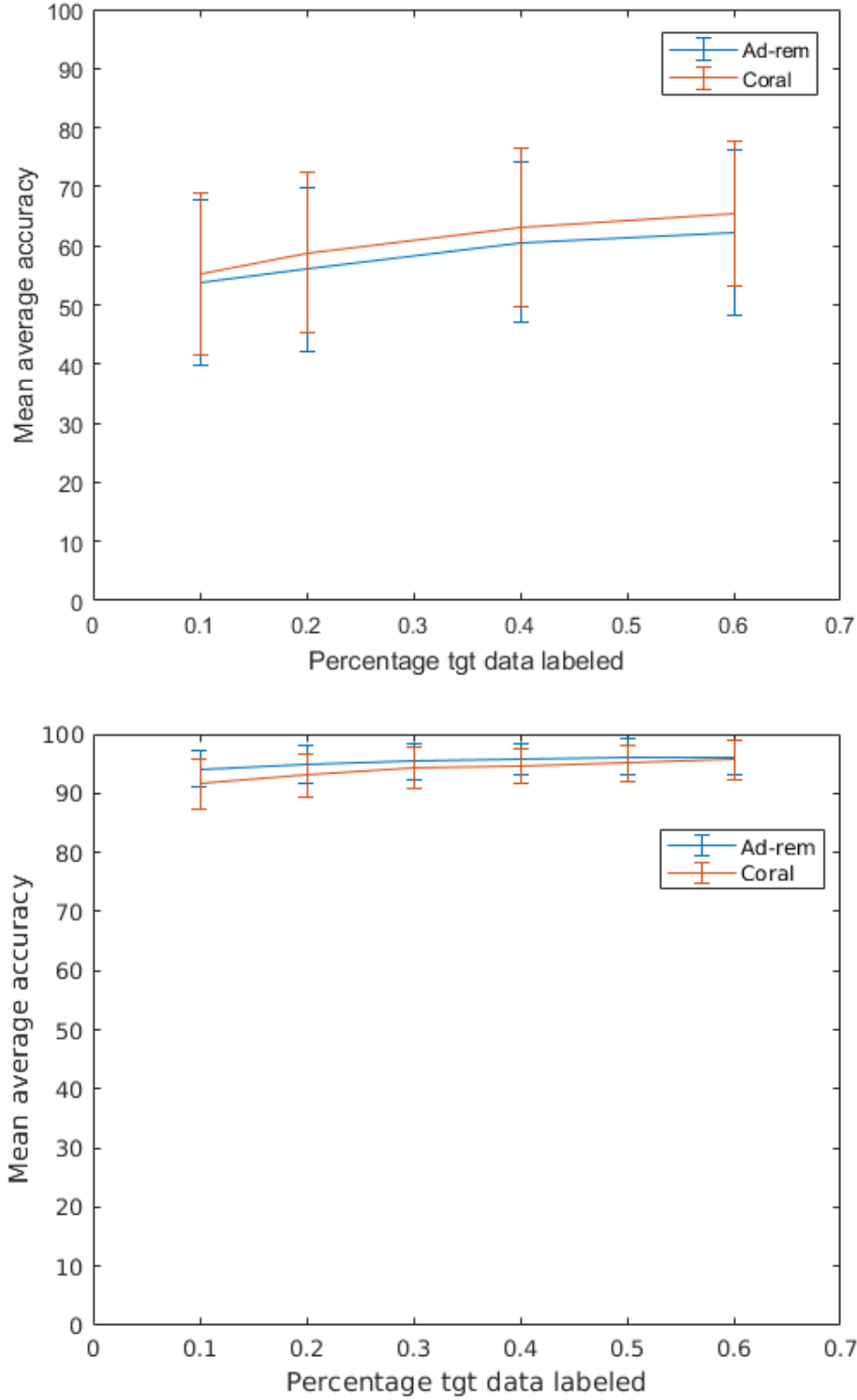


Figure 1: Coral and ad-REM applied to the Office-caltech dataset, with different percentages of labeled target data (x-axis). In the top figure SURF features were used, in the bottom figure VGG features. Five repetitions have been done for each source-target dataset combination. Accuracy is calculated as the mean over the average of the 5 repetitions and 12 source-target combinations, the errorbars represent standard deviation of the accuracies of the 12 source-target combinations.

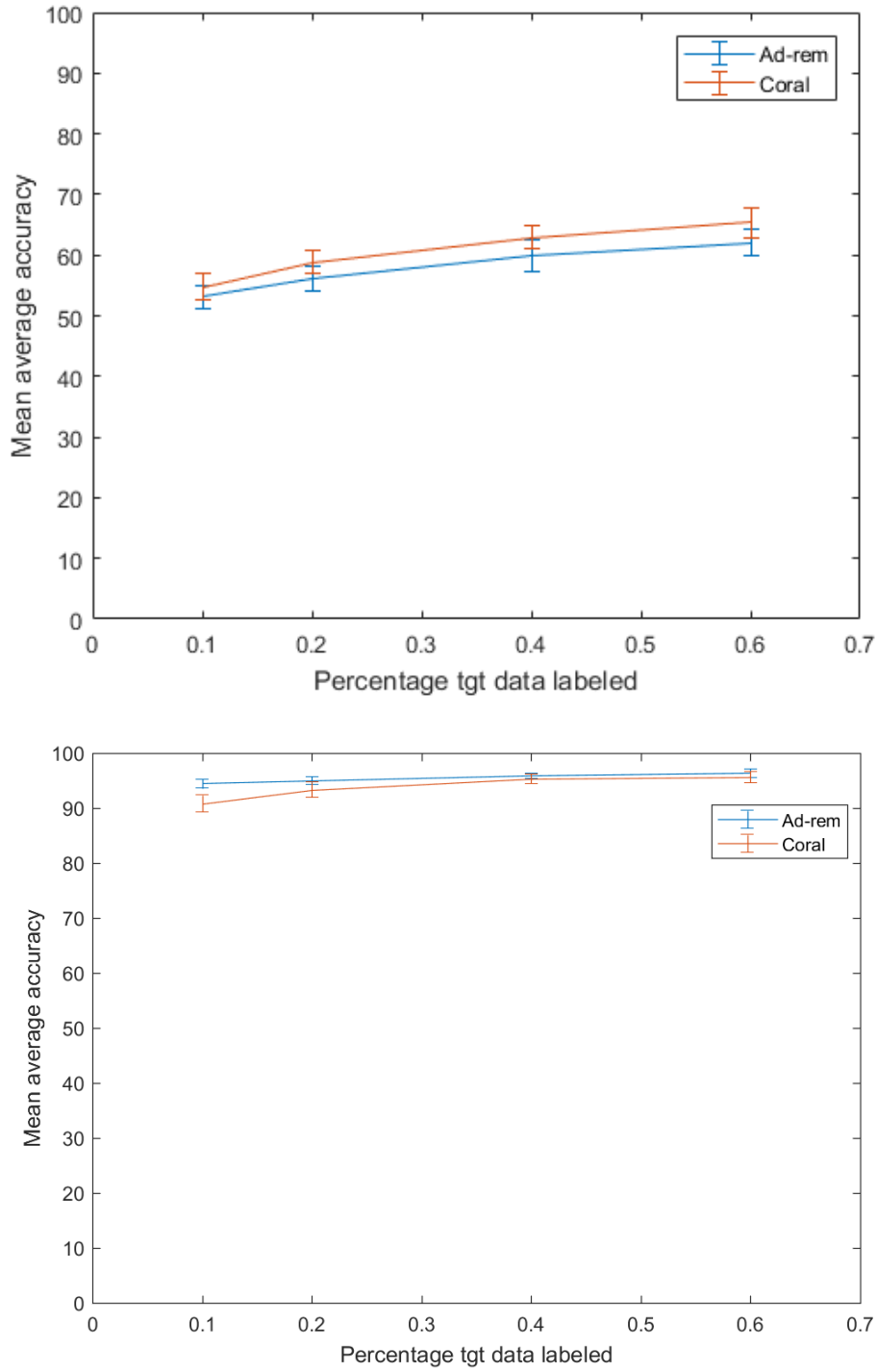


Figure 2: Identical experimental setup to Figure 1, with in the left figure SURF features and in the right figure VGG features being used. Only difference being the errorbars which represent the standard deviation of the average accuracy between iterations.